
Unicamp - Universidade Estadual de Campinas
Instituto de Química - Departamento de Físico-Química



TESE DE DOUTORADO

Estudos teóricos (modelagem molecular e QSAR) de
inibidores de HIV-1 integrase.

Autor

Eduardo Borges de Melo

Orientadora

Profª Drª Márcia Miguel Castro Ferreira

Campinas, setembro de 2009.

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE
QUÍMICA DA UNICAMP**

M491e	<p>Melo, Eduardo Borges de. Estudos teóricos (modelagem molecular e QSAR) de inibidores de HIV-1 integrase / Eduardo Borges de Melo. – Campinas, SP: [s.n], 2009.</p> <p style="text-align: center;">Orientadora: Márcia Miguel Castro Ferreira.</p> <p style="text-align: center;">Tese - Universidade Estadual de Campinas, Instituto de Química.</p> <p style="text-align: center;">1. QSAR. 2. HIV-1 integrase. 3. Modelagem molecular. 4. Química medicinal. I. Ferreira, Marcia Miguel Castro. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.</p>
-------	--

Título em inglês: Theoretical studies (molecular modeling and QSAR) of HIV-1 integrase inhibitors

Palavras-chaves em inglês: QSAR, HIV-1 integrase, Molecular modeling, Medicinal chemistry

Área de concentração: Físico-Química

Titulação: Doutor em Ciências

Banca examinadora: Profa. Dra. Márcia Miguel Castro Ferreira (orientadora), Prof. Dr. Rogério Custódio (IQ-UNICAMP), Prof. Dr. Nelson Henrique Morgon (IQ-UNICAMP), Profa. Dra. Nelilma Correia Romeiro (IQ-UFRJ); Profa. Dra. Lílian Sibelle Campos Bernardes (DCF-UFSC)

Data de defesa: 18/09/2009

À minha esposa, Tânia.

AGRADECIMENTOS

A minha orientadora, Prof^a Dr^a Márcia Miguel Castro Ferreira, pela orientação no desenvolvimento deste trabalho.

A minha esposa, Tânia Cristina Amaral de Melo, a quem dedico esta tese, por toda ajuda, apoio, compreensão e sacrifício pessoal.

À todos os membros do Laboratório de Quimiometria Teórica e Aplicada (LQTA), pela companhia e amizade prestados nos últimos anos, essenciais para que eu conseguisse chegar ao final deste trabalho.

Agradecimentos especiais aos membros:

- ♦ João Paulo Athaíde Martins, Kerly Fernanda Mesquita Pasqualoto e Euzébio Guimarães Barbosa, pelo desenvolvimento do programa LQTA-QSAR, sem o qual este trabalho não teria sido possível;
- ♦ Raimundo Clécio Dantas Muniz Filho, por toda a ajuda prestada durante a disciplina Introdução a Química Quântica, e por todas as demais ajudar prestadas no decorrer do curso;
- ♦ Flávia da Silva Pereira, pela ajuda prestada em diversos momentos e pela companhia como companheira de república.

Aos Profs. Drs. Yuji Takahata, Nelson Henrique Morgon, Pedro Antônio Muniz Vazquez e Rogério Custódio, pelas disciplinas que cursei com os mesmos durante o curso, cujo aprendizado foi-me muito útil para o desenvolvimento desta tese, e que com certeza ainda será muito mais útil para os trabalhos futuros.

A todos os funcionários do Instituto de Química da Unicamp, pois sem o apoio de cada um deles nem o meu trabalho, nem o de nenhum outro poderia ser desenvolvido.

Ao curso de Farmácia da Universidade Estadual do Oeste do Paraná (Unioeste) e a instituição como um todo, pelo apoio e afastamento total remunerado.

A meus colegas docentes do curso de Farmácia da Unioeste, muitos dos quais dispuseram de muitas horas de seu tempo para ministrar as disciplinas pelas quais sou responsável e por me ajudar a resolver problemas burocráticos, quando necessário.

A toda minha família, especialmente meus pais, irmã, cunhado e sogros, por todo o apoio em muitos momentos.

E a Deus, sem o qual não teria conseguido chegar ao final deste trabalho.

CURRICULUM VITAE

1. Dados Pessoais

Eduardo Borges de Melo.

Brasileiro, natural de Riolândia-SP.

ebmelo@unioeste.br; ebormelo@yahoo.com.br

2. Formação Acadêmica

Mestrado em Ciências Farmacêuticas – área Fármaco e Medicamentos

Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade Estadual de São Paulo (FCFRP/Usp)

Período: 08/1999 a 08/2001

Orientadora: Prof^a Dr^a Ivone Carvalho

Graduação de Farmácia

Universidade Paulista (Unip) – campus de São José do Rio Preto-SP

Período: 1995 a 1998

3. Experiência profissional

Universidade Estadual do Oeste do Paraná (Unioeste) – Curso de Farmácia, campus de Cascavel-PR

Período: 08/2002 até o momento

Função: docente universitário.

Atividades: ensino e pesquisa.

Universidade Paulista (Unip) – Curso de Farmácia, campus de Araçatuba-SP

Período: 03/2002 a 06/2002

Função: docente universitário.

Atividades: ensino.

4. Atividades acadêmicas

4.1 Artigos publicados

Melo, E.B.; Ferreira, M.M.C . Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors. *European Journal of Medicinal Chemistry*, v. 44, p. 3577-3583, 2009.

Pugens, A.M.; Donaduzzi, C.M.; Melo, E.B. Controle de qualidade total e equivalência farmacêutica de três apresentações de captopril. *Revista Eletrônica da Faculdade de Farmácia*, v. 5, p. 32-45, 2008.

Melo, E.B.; Teixeira, J.J.V.; Mânica, G.C.M. Histórico das tentativas de liberação da venda de medicamentos em estabelecimentos leigos no Brasil a partir da implantação do Plano Real. *Ciência & Saúde Coletiva*, v. 12, p. 1333-1339, 2007.

Dal Curtivo, C. P. ; Melo, E.B; Donaduzzi, C.M.; Bedor, D.C.G. ; Mota, T.G. ; Santana, D.P. A Bioequivalence Study Analyzed in Face of the Brazilian Generic Drugs Policy. *Latin American Journal of Pharmacy*, v. 26, p. 859-865, 2007.

Melo, E.B.; Bruni, A.T.; Ferreira, M.M.C. Inibidores da HIV-integrase: potencial abordagem farmacológica para o tratamento da AIDS. *Química Nova*, v. 29, n. 3, p. 555-562, 2006.

Melo, E.B.; Carvalho, I. Alfa e Beta-glucosidases como Alvos Moleculares para Desenvolvimento de Fármacos. *Química Nova*, v. 29, n. 4, p. 840-843, 2006.

Melo, E.B.; Minneto, A.; Violin, I.; Braga, R.S.; Andrade, E. Mariot, S.; Schneider, D.S.L.; Godoi, M.F.S.; Lucca, P.S.R. Medicamentos Similares e Saúde Pública: Controle de Qualidade Físico-Químico de Comprimidos de Similar de Ácido Acetilsalicílico do Estoque da Farmácia Básica do Município de Cascavel, PR, Brasil.. *Acta Farmaceutica Bonaerense*, v. 25, n. 3, p. 344-350, 2006.

Lima Neto, Q.A.; Reis, R.R.; Nogueira, M.A.; Melo, E.B. Estudo da relação estrutural-atividade quantitativa de dibenzoilmetanos alfa-substituídos quanto a atividade anticâncer de mama (linhagem celular MCF7). *Revista Brasileira de Cancerologia*, v. 52, p. 345-357, 2006.

Melo, E.B.; da Silveira, A.G.; Carvalho, I. α - and β -Glucosidase inhibitors: chemical structure and biological activity. *Tetrahedron*, v. 62, n. 44, p. 10277-10302, 2006.

Carvalho, I; Melo, E.B. Synthesis of (+)-(2R,3S,4R)-2,3,4-trihydroxycyclohexanone from d-glucose. *Carbohydrate Research*, v. 339, n. 2, p. 361-365, 2004.

4.2 Artigos aceitos para publicação

Melo, E.B.; Ferreira, M.M.C. Nonequivalent effects of diverse Log P algorithms in three QSAR studies. *QSAR & Combinatorial Science*, 2009.

Melo, E.B.; Knappmann, A.L. Qualidade de medicamentos isentos de prescrição: um estudo com marcas de dipirona comercializadas em uma drogaria de Cascavel - PR. *Ciência & Saúde Coletiva*, 2009.

4.3 Capítulo de livro publicado.

Melo, E.B.; Silva, C.H.T.P; Taft, C.; Carvalho, I . Alfa and Beta-Glucosidase Inhibitors: Importance and Perspectives for the Development of New Pharmaceuticals. In: Taft, C.. (Org.). *Current Trends in Biotechnology. : Research Signpost (Trivandrum)*, 2005, v. 1º, p. 42- 65.

4.4 Resumos simples de congressos publicados em anais

Melo, E.B.; Ferreira, M.M.C. A receptor independent 4D-QSAR study by means LQTA-QSAR package. In: 5th International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (CMTPI 2009), 2009, Istambul, Turquia. Abstract Book of CMTPI 2009, 2009. v. único.

Melo, E.B.; Ferreira, M.M.C. Several Log P in a same QSAR study? An example with HIV-integrase inhibitors. In: 6th International Congress of Pharmaceutical Sciences - CIFARP 2007, 2007, Ribeirão Preto. 6th International Congress of Pharmaceutical Sciences - program and abstracts, 2007. v. Único.

Takahashi, H.T.; Bunhak, E.J.; Reis, R.R.; Melo, E.B. Quantitative paracetamol assays in fifteen brands of oral solution available in the Brazilian market. In: 6th International Congress of Pharmaceutical Sciences - CIFARP 2007, 2007, Ribeirão Preto. 6th International Congress of Pharmaceutical Sciences - program and abstracts, 2007. v. único.

Bruni, A.T.; Melo, E.B.; Friozi, M.C. Estrutura-atividade por PCA de dibenzoilmetanos quanto à inibição da atividade mutagênica induzida em *Salmonella typhimurium*. In: 29ª Reunião Anual da Sociedade Brasileira de Química (RASBQ), 2006, Águas de Lindóia - SP. Programa da 29ª Reunião Anual da Sociedade Brasileira de Química, 2006. v. Único.

Pereira, F.S. ; Martins, J.P.A. ; Melo, E.B.; Ferreira, M.M.C.. 2D-QSAR analysis of aziridinyl-1,4-naphthoquinone antimalarials using Partial Least Square (PLS). In: The 3rd Brazilian Symposium on Medicinal Chemistry, 2006, São Pedro - SP. Abstract Book of The 3rd Brazilian Symposium on Medicinal Chemistry, 2006. v. único.

Melo, E.B.; Olian, C.A. ; Fortuna, J.A. ; Tolentino, M.L. ; Godoi, M.F.S.; dal Piva, R.; Sbrussi, V. Avaliação in vitro da ação de medicamentos antiácidos. In: 1 Congresso de Ciências Farmacêuticas de Cascavel - PR, 2004, Cascavel - PR. Anais do 1 Congresso de Ciências Farmacêuticas de Cascavel - PR e do 1 Simpósio em Ciência e Tecnologia de Alimentos do Mercosul, 2004. v. Único.

Melo, E.B.; Carvalho, I. Síntese de (+)-(2R,3S,4R)-2,3,4-tribenziloxi-cicloexanol via redução direta de grupamento carbonila para grupamento metileno. In: 1 Congresso de Ciências Farmacêuticas de Cascavel - PR, 2004, Cascavel - PR. Anais do 1 Congresso de Ciências Farmacêuticas de Cascavel - PR e do 1 Simpósio em Ciência e Tecnologia de Alimentos do Mercosul, 2004. v. Único.

Melo, E.B.; Lima Neto, Q.A.; Nogueira, M.A.; Reis, R.R. Dibenzoilmetanos: modelagem molecular e avaliação inicial da correlação propriedades x atividade anti-câncer de mama. In: XIII EAIC - Encontro Anual de Iniciação Científica, 2004, Londrina. Anais do XIII EAIC, 2004. v. CD-ROM.

Melo, E.B.; Lima Neto, Q.A. Nogueira, M.A.; Reis, R.R. Systematic Conformational Analysis of Dibenzoilmethane and Two Classics Derivatives, Parsol and Eusolex. In: The 2nd Brazilian Symposium on Medicinal Chemistry, 2004, Rio de Janeiro. Abstract Book, 2004. v. Único. p. 32.

Melo, E.B.; Ferreira, M.M.C.; Bruni, A.T. Systematic Conformational Analysis of 5CITEP, a Lead Compound for the Development of HIV Integrase Inhibitors. In: The 2nd Brazilian Symposium on Medicinal Chemistry, 2004, Rio de Janeiro. Abstract Book, 2004. v. Único. p. 40.

Melo, E.B.; Lima Neto, Q.A.; Nogueira, M.A.; Reis, R.R. Molecular Modeling and Basic QSAR of 2-Allyl-1,3-Diphenyl-1,3-Propanodione (DMA2) and Derivatives. In: The 2nd Brazilian Symposium on Medicinal Chemistry, 2004, Rio de Janeiro. Abstract Book, 2004. v. Único. p. 40.

Melo, E.B.; Carvalho, I. Facile Functionalization of O-Benzyl Cyclitol Derivative Using p-Toluenesulphonylhydrazide Protective Group. In: 12th European Carbohydrate Symposium, 2003, Grenoble, França. Anais of 12th European Carbohydrate Symposium, 2003.

Melo, E.B.; Caldeira, L.F.; Giordani, F.; Guerreiro, P.; Jorge, A.S.; Lenartovicz, V.; Miyata, D.F. O papel do profissional farmacêutico: uma visão da população. In: Congresso Brasileiro de Farmácia, 2003, São Paulo. Anais do Congresso Brasileiro de Farmácia, 2003.

Melo, E.B.; Carvalho, I. Conversão de O-benzil e O-benzoil glicopiranosídeos em derivados carbocíclicos, mediada por diferentes catalisadores de mercúrio. In: 25a Reunião Anual da Sociedade Brasileira de Química, 2002, Poços de Caldas. Anais da 25a Reunião Anual da Sociedade Brasileira de Química, 2002.

Melo, E.B.; Carvalho, I. Síntese de Pseudodissacarídeo, via condensação de (-)-ceto-conduritól e 3-amino-3-desoxi-alfa-D-glicopiranosídeo de metila. In: 24a Reunião Anual da Sociedade Brasileira de Química, 2001, Poços de Caldas. Anais da 24a Reunião Anual da Sociedade Brasileira de Química, 2001.

Melo, E.B.; Carvalho, I. Syntheses of (-)-keto-conduritól, via Ferrier Rearrangement. In: 20th International Carbohydrate Symposium, 2000, Hamburgo. Anals of 20th International Carbohydrate Symposium, 2000.

Melo, E.B.; Villafanha, J. Interações medicamentosas: análise de prescrições médicas no Hospital de Base de São José do Rio Preto. In: 45a Jornada Farmacêutica UNESP Araraquara, 1998, Araraquara-SP. Anais da 45a Jornada Farmacêutica UNESP Araraquara, 1998.

Melo, E.B.; Bogdan, G.B.; Ribeiro, R.A.; D'angelo, E.O.; Guerreschi, F.; Ribeiro, A.; Buzatto, É.; Santos, F.A.; Brizoti, P.H.; Alves, S.; Chiaravallotti Neto, F.; Carloni Filho, J. Determinação de propriedades físico-químicas e bacteriológicas da água em localidades rurais de São José do Rio Preto. In: 45a Jornada Farmacêutica da UNESP de Araraquara, 1998, Araraquara-SP. Anais da 45a Jornada Farmacêutica UNESP Araraquara, 1998.

4.5 Resumos simples de congressos publicados em revistas

Melo, E.B.; Bruni, A.T.; Ferreira, M.M.C. Validation of semi-empirical methodology for geometry optimization of HIV-IN inhibitors. Revista Brasileira de Ciências Farmacêuticas, São Paulo, v. 41, n. Sup 1, p. 265, 2005.

Friozi, M.C. ; Lima Neto, Q.A.; Reis, R.R.; Nogueira, M.A.; Melo, E.B. A HCA study for antimutagenic activity of dibenzoylmethanes in *Salmonella typhimurium*. *Revista Brasileira de Ciências Farmacêuticas*, São Paulo, v. 41, n. Sup 1, p. 268, 2005.

Melo, E.B.; Lorenzoni, R. Azasugars: a correlation study between alfa-glucosidase inhibition vs electronic descriptors. *Revista Brasileira de Ciências Farmacêuticas*, São Paulo, v. 41, n. Sup 1, p. 268, 2005.

Lima Neto, Q.A.; Reis, R.R.; Nogueira, M.A.; Melo, E.B. Dibenzoilmetanos: Hansch analyses for the inhibition of MCF7 cell lines. *Revista Brasileira de Ciências Farmacêuticas*, São Paulo, v. 41, n. Sup 1, p. 267, 2005.

Mânica, G.C.; Vieira, J.J.T; Melo, E.B. Commerced medicines in non-pharmaceutical establishments located in the district of Jd Universitário and Nova Cidade (Cascavel, PR).. *Revista Brasileira de Ciências Farmacêuticas*, São Paulo, v. 41, n. Sup 1, p. 403, 2005.

Stoef, P.R. ; Mânica, G.C.; Rosa, J.T.; Melo, E.B. Analysis of captopril and ASA samples distributed by Farmácia Básica de Cascavel, PR. *Revista Brasileira de Ciências Farmacêuticas*, São Paulo, v. 41, n. Sup 1, p. 410, 2005.

Reis, R.R.; Nogueira, M.A.; Melo, E.B. Basic Computational Analysis of Dibenzoylmethanes. *Revista Brasileira de Ciências Farmacêuticas*, São Paulo, v. 39, n. supl 2, p. 186, 2003.

Melo, E.B.; Borges, Á.L.; Carvalho I. HPLC Analyses of Conduritol Derivatives as a Potential D-Glycosidase and HIV Inhibitor. *Revista Brasileira de Ciências Farmacêuticas*, São Paulo, v. 39, n. supl 2, p. 267, 2003.

4.5 Resumos expandidos de congresso publicados em anais

Martins, J.P.A.; Melo, E.B.; Ferreira, M.M.C. 2D-QSAR study of antimutagenic flavonoids using Ordered Predictors Selection (OPS). In: 4th Brazilian Symposium on Medicinal Chemistry, 2008, Porto de Galinhas. 4th Brazilian Symposium on Medicinal Chemistry, 2008.

Bunhak, E.J.; Stoef, P.R.; Melo, E.B. Ensaios da qualidade físico-química de comprimidos de AAS distribuídos pela farmácia básica de Cascavel, PR, Brasil. In: 3º Congresso de Ciências Farmacêuticas e 3º Simpósio em Ciência e Tecnologia de Alimentos do Mercosul, 2008, Cascavel, PR. Anais do 3º Cosimp, 2008. v. único.

Martins, J.P.A. ; Melo, E.B.; Pereira, F.S.; Nogueira, M.A.; Ferreira, M.M.C. QSAR multivariado de dibenzoilmetanos (DBMs) alfa substituídos com atividade anti câncer de mama. In: 47º Congresso Brasileiro de Química, 2007, Natal. Anais do 47º Congresso Brasileiro de Química, 2007. v. único.

Melo, E.B.; Ferreira, M.M.C. QSAR multivariado de inibidores de HIV-integrase: carboxamidas. In: XIV Simpósio Brasileiro de Química Teórica, 2007, Poços de Caldas. Livro de Resumos, 2007.

Pannizon, G.; Lorscheider, J.A.; Oliveira, S.M.M.; Melo, E.B. Análise da qualidade físico-química de matérias-primas farmacêuticas de farmácias de manipulação de Cascavel PR. In: I Congresso Paranaense de Pesquisa em Saúde, 2007, Cascavel. Anais do evento, 2007.

Stoef, P.R.; Melo, E.B.; Reis, R.R. Controle de qualidade físico-químico de medicamentos: acompanhamento da qualidade de comprimidos de captopril distribuídos pela Farmácia Básica de Cascavel - PR.. In: XV EAIC - Encontro Anual de Iniciação Científica, 2006, Ponta Grossa. Anais do XV EAIC, 2006.

Melo, E.B.; Friozi, M.C. Quimioprevenção do câncer: estudo qualitativo das relações estrutura atividade antimutagênica de dibenzoilmetanos e flavonóides em *Salmonella typhimurium*.. In: XV EAIC - Encontro Anual de Iniciação Científica, 2006, Ponta Grossa. Anais do XV EAIC, 2006.

Melo, E.B.; Friozi, M.C. Estudo das relações estrutura-atividade antimutagênica (SAR) de dibenzoilmetanos e flavonóides em *Salmonella typhimurium*. In: 14º SBQ Sul - Encontro da Sociedade Brasileira de Química - Seccional Sul, 2006, Erechim - RS. Livro de resumos, 2006. v. único.

Engles, R.J.; Reis, R.R.; Melo, E.B.; Nogueira, M.A. Síntese de beta-fosfato de cálcio para biocerâmicas por precipitação e estado sólido. In: XI Encontro de Química da Região Sul, 2003, Pelotas. Anais do XI Encontro de Química da Região Sul, 2003.

Shimabuku Jr, R.S.; Reis, R.R.; Melo, E.B.; Nogueira, M.A. Modelagem Molecular do 1,3-difenil-2-(2-propinil)-1,3-propanodiona (DMP2) e derivados. In: XI Encontro de Química da Região Sul, 2003, Pelotas. Anais do XI Encontro de Química da Região Sul, 2003.

Melo, E.B.; Cristani, C.; Reis, R.R.; Nogueira, M.A. Modelagem Molecular do 2-benzil-1,3-difenil-1,3-propanodiona (DMB2) e derivados. In: XI Encontro de Química da Região Sul, 2003, Pelotas. Anais do XI Encontro de Química da Região Sul, 2003.

Melo, E.B.; Ogassawara, G.S.; Carvalho, I; Nogueira, M.A.; Reis, R.R. Modelagem Molecular de Potenciais Inibidores de beta-glicosidase. In: XI Encontro de Química da Região Sul, 2003, Pelotas. Anais do XI Encontro de Química da Região Sul, 2003.

4.6 Orientações

Trabalho de conclusão de curso de especialização. André Leandro Knappmann. Avaliação das embalagens, bulas e da qualidade físico-química de amostras de medicamentos contendo dipirona em solução oral comercializadas em uma drogaria na cidade de Cascavel - PR. 2006. Especialização Lato Sensu em Controle de Qualidade, Unioeste.

Trabalho de conclusão de curso de especialização. Ana Maria Pugens. Equivalência farmacêutica em três apresentações de captopril como controle de qualidade total do medicamento. 2006. Especialização Lato Sensu em Controle de Qualidade, Unioeste.

Trabalho de conclusão de curso de especialização. Cátia Panizzon Dal Curtivo. Análise do estudo de bioequivalência de comprimidos de norfloxacino 400 mg perante a política de medicamentos genéricos no Brasil. 2006. Especialização Especialização Lato Sensu em Farmacologia, Unioeste.

Trabalho de conclusão de curso de graduação. Graciele Mânica. Levantamento e análise dos medicamentos vendidos em estabelecimentos não-farmacêuticos dos bairros Faculdade e Nova Cidade de Cascavel, PR. 2005.

Trabalho de conclusão de curso de graduação. Qurino Alves da Lima Neto. Dibenzoilmetanos: modelagem molecular e avaliação inicial da correlação propriedades x atividade anti-câncer de mama. 2005.

Trabalho de conclusão de curso de graduação. Clair Cristani. Estudo químico-computacional de 2-benzil-1,3-difenil-1,3-propanodiona (DMB2). 2003.

Trabalho de conclusão de curso de graduação. Gilmara Sayuri Ogassawara. Estudo químico-computacional básico de derivados de conduritóis potencialmente ativos. 2003.

Orientação Pibic/Unioeste. Marcelo C Friozi. Estudo da relação estrutura-atividade teórica de potenciais candidatos a fármacos antitumorais e antivirais. 2006.

Co-orientação Pibic/Unioeste. Paulo Stoef. Controle de qualidade de medicamentos - análise de captopril utilizado pela Farmácia Básica do município de Cascavel, PR. 2006.

5. Prêmios recebidos

2008 - **Most Cited Paper 2005 - 2008 Tetrahedron Award**, pela editora Elsevier Ltd, pelo artigo " α and β -Glucosidase inhibitors: structure and biological activity".

2006 - **Menção Honrosa** - revista Tetrahedron e Editora Elsevier, pelo artigo " α and β -Glucosidase inhibitors: structure and biological activity".

1995 - **Mérito Acadêmico** - Faculdade de Farmácia e Nutrição de São José do Rio Preto (atual Unip São José do Rio Preto).

“Muitas lições puderam ser apreendidas neste intenso período inicial, e a maioria sugere que a ciência requer muita modéstia”.

Robert C. Gallo e Luc Montagne, sobre os estudos que levaram a descoberta do HIV.

“As pessoas pensam que um pesquisador é alguém que pratica um ofício como qualquer outro, seja um bombeiro, um carpinteiro, um industrial, um engenheiro ou um médico. A diferença entre estes ofícios e o de pesquisador é que nenhum dos profissionais mencionados publica sobre sua maneira de trabalhar e que, talvez, a maioria deles esquece o ofício, quando se retira de seu lugar habitual de trabalho. O pesquisador (...) 'não separa seu trabalho de sua vida privada'. Se ele pode levar a sério ambas as coisas e não dissociá-las, poderá enriquecer sua imaginação para acrescentar todos os dias idéias úteis a sua pesquisa. Escolher o caminho da pesquisa é escolher um estilo de vida”.

Frederico Tobar.
Margot Romano Yalour.

“Nós quisemos permanecer juntos depois que nossos pais e avós morreram do AIDS. Eu queria voltar para a escola, mas não tenho dinheiro... eu preciso trabalhar duro para ter uma boa vida e olhar para trás para não pegar à doença meus mãe e pai tiveram”.

Félix, 15 anos, sul-africano, único trabalhador em uma casa que inclui seus cinco irmãos mais novos e um tio de 80 anos.

“O farmacêutico é, antes e acima de tudo, o profissional do medicamento”.

Andrejus Korokolvas.

RESUMO

Apesar da implantação da HAART, existe uma necessidade contínua de novos agentes anti-HIV. Os inibidores da enzima HIV-1 integrase (HIV-IN) constituem um dos mais recentes avanços na luta contra a AIDS. A principal abordagem utilizada nessas pesquisas são os métodos relacionados ao planejamento de fármacos auxiliados por computador (CADD). Neste trabalho, foram realizados três estudos QSAR (2D, 4D e híbrido) utilizando um conjunto de treinamento formado por 85 compostos descritos como inibidores da reação de transferência de fita catalisada pela HIV-IN. No estudo QSAR-2D, foram utilizados 1291 descritores físico-químicos obtidos por diversos programas. Para o estudo QSAR-4D, perfis de amostragem conformacionais (PACs) foram obtidos com o programa de dinâmica molecular Gromacs 4, e 65.856 descritores de campo (Coulomb e Lennard-Jones) foram obtidos a partir do programa LQTA-QSAR. As seleções de variáveis foram realizadas pela metodologia *Ordered Predictors Selection* (OPS), e os modelos foram construídos utilizando regressão por quadrados mínimos parciais (PLS). Na etapa de QSAR-2D, foi realizado um estudo preliminar com 33 compostos com baixa variabilidade estrutural e 167 descritores de mais simples interpretação. O modelo obtido foi formado por duas variáveis latentes e quatro descritores. Esse modelo apresentou uma relação direta com o mecanismo de inibição mais aceito. Já para o modelo com o conjunto completo, foram selecionados quatro descritores, porém de difícil interpretação, provavelmente devido à grande variabilidade estrutural do conjunto de treinamento. Já para o modelo QSAR-4D, uma relação direta com o mecanismo de inibição, com descritores correspondentes à interação com os co-fatores metálicos e com a alça hidrofóbica do sítio de ligação da HIV-IN, também pôde ser traçada. Todos os modelos apresentaram qualidade estatística aceitável, com boas capacidades de predição interna e robustez, além de não apresentarem correlação ao acaso. Já o modelo híbrido, construído com alguns dos descritores selecionados nos estudos anteriores, possui alta qualidade estatística, mas é inferior ao modelo QSAR-4D. Logo, ao serem considerados os resultados obtidos, conclui-se que os objetivos da tese foram alcançados e que os modelos obtidos apresentaram grande potencial para proposição de novos inibidores da HIV-IN.

Palavras-chaves: QSAR; Modelagem Molecular; HIV-integrase; LQTA-QSAR; AIDS; Planejamento de Fármacos Auxiliado por Computador; Química Medicinal.

ABSTRACT

Despite the HAART implantation, there is a continuous need to search for new anti-HIV agents. The HIV-1 integrase (HIV-IN) inhibitors are one of the most recent breakthrough in AIDS research. So, the computer aided-drug design (CADD) related methods have been the main approach used in the research of such class of drugs. In this work, three QSAR studies (2D, 4D and hybrid), with a training set, consisted of 85 inhibitors of strand transfer (ST) reaction catalyzed by HIV-IN. In the 2D-QSAR study, 1,291 physicochemical descriptors were obtained by several programs. For the 4D-QSAR study, the conformational ensembles profiles (CEPs) were obtained by the molecular dynamic program Gromacs 4. With the LQTA-QSAR program, 65,856 descriptors (Coulomb and Lennard-Jones) were obtained. In both the studies, the variable selections were carried out according to the *Ordered Predictors Selection* (OPS) method while the models were composed with Partial Least Squares (PLS) regression. In the 2D-QSAR step, a preliminary study with 33 compounds with low structural variability and 167 descriptors of more simple interpretation was developed. The obtained model was based on two latent variables and four descriptors. But, for the model with a complete set, there were four selected descriptors, although the difficult interpretation, probably due to the great structural variability of the training set. On the other hand, a direct relation with the inhibition mechanism could be traced for the 4D-QSAR model, including descriptors related with the interaction with the metallic co-factors and with the hydrophobic loop, placed in the binding site of HIV-IN. All the models showed an acceptable statistic quality, with good capacity of internal prediction and robustness. Moreover, the models did not present any randomized correlation. But, the hybrid model, built with some of descriptors selected in both studies, although it also has high statistic quality, is inferior to the 4D-QSAR model. Hence, considering the good obtained results, it can be concluded that the purposes of this thesis were achieved and that the models present a great potential to propose new HIV-IN inhibitors.

Keywords: QSAR; Molecular Modeling; HIV-integrase; LQTA-QSAR; AIDS; Computer Aided Drug-Design; Medicinal Chemistry.

SUMÁRIO

Lista de abreviaturas	xviii
Lista de tabelas	xx
Lista de figuras	xxii
1 INTRODUÇÃO	1
1.1 <i>AIDS/HIV: SURGIMENTO, DESCOBERTA E DOENÇA</i>	1
1.2 <i>MORFOLOGIA E CICLO REPLICATIVO DO HIV</i>	4
1.3 <i>EPIDEMIOLOGIA E IMPACTO SOCIO-ECONÔMICO DA AIDS – BRASIL E MUNDO</i>	7
1.4 <i>FÁRMACOS ANTI-HIV E A NECESSIDADE DE NOVOS AGENTES</i>	11
1.5 <i>A HIV-INTEGRASE COMO ALVO PARA FÁRMACOS ANTI-HIV</i>	17
1.6 <i>COMPOSTOS PROTÓTIPOS E PLANEJAMENTO DE FÁRMACOS AUXILIADO POR COMPUTADOR</i>	24
1.7 <i>ESTUDOS DE RELAÇÃO ESTRUTURA-ATIVIDADE QUANTITATIVA E QUALITATIVA</i>	30
1.7.1 <i>Regressão linear múltipla (MLR)</i>	32
1.7.2 <i>Métodos de projeção: PCR e PLS</i>	33
1.7.3 <i>Métodos não-lineares: ANN</i>	34
1.7.4 <i>Outras metodologias em QSAR</i>	36
1.7.5 <i>Validação de modelos QSAR</i>	39
1.8 <i>MODELAGEM MOLECULAR</i>	41
1.8.1 <i>Métodos clássicos</i>	44
1.8.2 <i>Métodos quânticos</i>	46
1.8.2.1 <i>Métodos ab initio</i>	48
1.8.2.2 <i>Métodos semiempíricos</i>	50
1.8.2.3 <i>Métodos DFT</i>	52
1.8.2.4 <i>Conjuntos (ou funções) de base</i>	54
1.10 <i>INIBIDORES DE HIV-INTEGRASE</i>	57
2 OBJETIVOS	69
2.1 <i>OBJETIVO GERAL</i>	69
2.2 <i>OBJETIVOS ESPECÍFICOS</i>	69
3 METODOLOGIA	71
3.1 <i>CONJUNTO DE TREINAMENTO</i>	71
3.2 <i>MODELAGEM MOLECULAR</i>	71
3.3 <i>DESCRITORES FÍSICO-QUÍMICOS</i>	81
3.4 <i>QSAR</i>	84
3.5 <i>QUALIDADE DOS MODELOS</i>	90
4 RESULTADOS E DISCUSSÃO	98
4.1 <i>QSAR-2D</i>	98

4.1.1 <i>Estudo QSAR-2D preliminar</i>	98
4.1.1.1 E_{HOMO}	102
4.1.1.2 α_{yy}	104
4.1.1.3 Energia total, E_{T}	105
4.1.1.4 SeaC2C2aa	106
4.1.2 <i>Estudo QSAR-2D completo</i>	107
4.1.2.1 Descritor BCUT: BELe8.....	124
4.1.2.2 Descritores GETAWAY: RTe e R1u+.....	127
4.1.2.3 Descritor WHIM: G2e.....	130
4.1.2.4 <i>Outliers</i>	133
4.2 <i>QSAR-4D</i>	139
4.3 <i>QSAR-2D/4D</i>	162
5 CONCLUSÕES	172
6 REFERÊNCIAS BIBLIOGRÁFICAS	176
ANEXO 1. ESTRUTURAS E VALORES DE pIC50 DOS COMPOSTOS DO CONJUNTO DE TREINAMENTO.	197
ANEXO 2. ALGORITMOS DESNVOLVIDOS POR MEMBROS DO GRUPO LQTA E UTILIZADOS NESSE TRABALHO.	201
ANEXO 3. TABELAS COM OS VALORES DE DESCRITORES OBTIDOS PARA OS MODELOS APRESENTADOS NESTE TRABALHO.	207

LISTA DE ABREVIATURAS

- ♦ 3'P: *3'-processing* – reação de processamento 3';
- ♦ 5CITEP: 1-(5-cloroindol-3-il)-3-hidroxi-3-(2H-tetrazol-5-il)-propenona;
- ♦ AIDS: *acquired immunodeficiency syndrome*; síndrome da imunodeficiência adquirida;
- ♦ AM1: Austin Model 1;
- ♦ ANN: *artificial neural networks* – redes neurais artificiais;
- ♦ ARC: *Aids related complex* – complexo relacionado a Aids
- ♦ ARE_{pred} : *average relative error of prediction* - erro relativo médio da validação externa
- ♦ ATM: sistema ataxia-telangiectasia-mutante;
- ♦ B3LYP: Becke, three-parameter, Lee-Yang-Parr;
- ♦ BCUT: *Burden Chemical Abstract Service of University of Texas*;
- ♦ BMS: Bristol-Myers Squibb;
- ♦ CADD: *computer-aided drug design* – planejamento de fármacos auxiliado por computador;
- ♦ CASSCF: *complete active space SCF*;
- ♦ CC: *coupled cluster*
- ♦ CDC: *Centers of Disease Control and Prevention* – Centros de Controle e Prevenção de Doenças;
- ♦ cDNA: ácido desoxirribonucleico complementar;
- ♦ ChelpG: *CHarges from ELectrostatic Potentials using a Grid based method*;
- ♦ CNDO: *complete neglect of differential overlap*;
- ♦ CoMFA: *Comparative Molecular Field Analysis*;
- ♦ CoMSIA: *comparative molecular similarity indices analysis*;
- ♦ CPI: complexo de pré-integração;
- ♦ CSD: Cambridge Structural Database;
- ♦ CTD: ligação C-terminal;
- ♦ DCA: β -dicetoácido;
- ♦ DFT: *density functional theory* – teoria do funcional de densidade;
- ♦ DM: dinâmica molecular;
- ♦ E_{HOMO} : energia do orbital molecular ocupado mais alto;
- ♦ ESP: *electrostatic potential* – potencial eletrostático;
- ♦ E_T : energia total;
- ♦ FDA: *food and drugs administration*;
- ♦ GETAWAY: *Geometry, Topology And Atom-Weights Assembly*;
- ♦ GGA: *Generalized Gradient Approximation*;
- ♦ GSK: GlaxoSmithKline;
- ♦ GTO: *Gaussian type orbitals*;
- ♦ HAART: *highly active anti-retroviral therapy* – terapia antiretroviral de alta eficácia;
- ♦ HCV: *hepatite C virus* – vírus da hepatite C;
- ♦ HDAC: histona deacetilase-1;
- ♦ HF: Hartree-Fock;
- ♦ HIV: *human immunodeficiency virus* – vírus da imunodeficiência humana;
- ♦ HIV-IN: HIV-integrase;
- ♦ HOMO: *highest occupied molecular orbital* - orbital molecular ocupado mais alto;
- ♦ HTLV-1: *human T lymphotropic virus type 1* – vírus humano linfotrópico T humano;
- ♦ IA: inteligência artificial;
- ♦ IL-2: interleucina-2;
- ♦ INSTI: *integrase strand transfer inhibitors* – inibidores da reação de transferência de fita da integrase;
- ♦ IP: inibidores de protease;
- ♦ ITRN: inibidores de transcriptase reversa nucleosídeo-nucleotídeo;
- ♦ ITRNN: inibidores da transcriptase reversa não-nucleosídeo;
- ♦ LCAO: *Linear Combination of Atomic Orbitals* combinações lineares de OA;
- ♦ LDA: *Local Density Approximation*;
- ♦ LJ: Lennard-Jones;
- ♦ LNO: validação cruzada leave-N-out;
- ♦ LQTA: laboratório de quimiometria teórica e aplicada;
- ♦ LV: *latent variables* – variáveis latentes;

-
- ♦ MIC: *minimal inhibitory concentration* - concentração inibitória mínima;
 - ♦ MIM: *molecular influence matrix* - matriz de influência molecular;
 - ♦ MLR: *multiple linear regression* - regressão linear múltipla;
 - ♦ MNDO: *modified neglect of diatomic overlap*;
 - ♦ MPn: Møller-Plesset;
 - ♦ mRNA; RNA mensageiro;
 - ♦ NDDO: *neglect of diatomic differential overlap*;
 - ♦ NTD: domínio N-terminal;
 - ♦ OECD: *Organization for Economic Co-Operation and Development* - Organização para a Cooperação Econômica e Desenvolvimento;
 - ♦ OPS: *ordered predictors selection* - seleção por preditores ordenados;
 - ♦ PCA: *principal component analysis* - análise de componentes principais;
 - ♦ PCR: *principal component regression* - regressão por componentes principais;
 - ♦ PDB: *Protein Data Bank*;
 - ♦ PEOE: *partial equalization of orbital electronegativity*;
 - ♦ PLS: *Partial Least Squares* - quadrados mínimos parciais;
 - ♦ PM3: *Parameter Model 3*;
 - ♦ $PRESS_{cal}$: *predictive residual sum of squares of calibration* - soma dos resíduos de predição ao quadrado da calibração;
 - ♦ $PRESS_{val}$: *predictive residual sum of squares of calibration of validation* - soma dos resíduos de predição ao quadrado da validação cruzada LOO;
 - ♦ Q^2_{LNO} : *correlation coefficient of leave-N-out cross-validation* - coeficiente de correlação da validação cruzada leave-N-out;
 - ♦ Q^2_{LOO} : *correlation coefficient of leave-one-out crossvalidation* - coeficiente de correlação da validação cruzada leave-one-out;
 - ♦ QSAR: *quantitative structure-activity relationship* - relações estrutura-atividade quantitativa;
 - ♦ QSPR: *quantitative structure-property relationship* - relações estrutura-propriedade quantitativa;
 - ♦ R^2 : *coefficient multiple determination* - coeficiente de determinação múltipla;
 - ♦ R^2_{pred} : *correlation coefficient of prediction* - coeficiente de correlação da validação externa;
 - ♦ RDF: *radial distributiton function*;
 - ♦ RMN: ressonância magnética nuclear;
 - ♦ RNA: ácido ribonucléico;
 - ♦ SEC: *standard error of calibration* - rro padrão de calibração;
 - ♦ SEP: *standard error of prediction* - rro padrão de predição externa;
 - ♦ SEV: *standard error of cross-validation* - erro padrão da validação cruzada leave-one-out;
 - ♦ SIV: *simian imunodeficiency vírus* - vírus da imunodeficiência humana;
 - ♦ SNC: sistema nervoso central;
 - ♦ SSy: *sum of squares of the response values* - soma dos quadrados dos valores de resposta;
 - ♦ ST: strand transfer - transferência de fita;
 - ♦ STO: *Slater Type Orbitals*;
 - ♦ TR: transcriptase reversa;
 - ♦ UNAIDS: *Joint United Nations Programme on HIV/Aids* - Programa Internacional de Combate à Aids;
 - ♦ VDW: van der Waals;
 - ♦ WHIM: *Weighted Holistic Invariant Molecular index*.

LISTA DE TABELAS

Tabela 1. As Químicas Medicinais de “antes” e “agora”. Tabela baseada em Lombardino e Lowe III (2004).....	26
Tabela 2. Classificação dos descritores físico-químicos que podem ser utilizados em QSAR, e em alguns exemplos. Baseado na tabela apresentada por Ferreira e Kiralj (2008), segundo a classificação de Dunn III (1989).....	31
Tabela 3. Funcionais de densidade. Extraído de Yong (2001).....	53
Tabela 4. Informações sobre os programas e descritores utilizados.....	82
Tabela 5. Parâmetros estatísticos utilizados para avaliar a qualidade dos modelos e suas respectivas equações.....	91
Tabela 6. Parâmetros estatísticos utilizados para avaliar a qualidade externa dos modelos e suas respectivas equações.....	92
Tabela 7. Descritores utilizados para formulação do Modelo I, atividades observadas e valores previstos pela validação LOO (<i>outliers</i> já excluídos).....	100
Tabela 8. Resultados obtidos para a validação externa do Modelo I.....	102
Tabela 9. Contribuição dos descritores do Modelo I em cada LV, informação de cada LV, coeficientes de correlação de Pearson e coeficientes autoescalados do modelo.....	102
Tabela 10. Coeficientes de correlação de Pearson e coeficientes autoescalados dos descritores que não apresentaram auto consistência nos Modelos III, V e VI.....	116
Tabela 11. Contribuição dos descritores do Modelo IV em cada LV, informação de cada LV, coeficientes de correlação de Pearson e coeficientes autoescalados do modelo.....	116
Tabela 12. Resultados da etapa de validação externa realizada com o Modelo Real VIII a partir do conjunto teste selecionado do Modelo Auxiliar IV.....	121
Tabela 13. Valores não escalados da eletronegatividade de Sanderson dos átomos que formam o conjunto de treinamento (SANDERSON, 1983).....	126
Tabela 14. <i>Outliers</i> retirados do Modelo IV.....	136
Tabela 15. Contribuição dos descritores do Modelo XI em cada LV, informação de cada LV e coeficientes autoescalados do modelo.....	142
Tabela 16. <i>Outliers</i> retirados do modelo XI.....	143
Tabela 17. Resultados da etapa de validação externa realizada com o modelo real XII a partir do conjunto teste selecionado do modelo auxiliar XI.....	147

Tabela 18. Relação dos artigos de QSAR-3D com inibidores de HIV-IN publicados entre 1995 e 2009.....	159
Tabela 19. <i>Outliers</i> retirados do modelo XIII.....	164
Tabela 20. Resultados da etapa de validação externa realizada com o Modelo Real XIV a partir do conjunto teste selecionado do Modelo Auxiliar XIII.....	167
Tabela 21. Contribuição dos descritores do Modelo XIII em cada LV, informação de cada LV e coeficientes autoescalados do modelo.....	170

LISTA DE FIGURAS

Figura 1. Morfologia do HIV. Adaptado de Avert (2009) e Ramalho (2006). O nucleocapsídeo não é apresentado.....	6
Figura 2. Ciclo replicativo do HIV. Adaptado de Souza e Almeida (2003) e LaBonte, Lebbos e Kirkpatrick (2003), já publicado pelos autores em Melo, Bruni e Ferreira (2006): 1: adsorção/fusão; 2: desencapamento; 3: transcrição reversa; 4: formação do complexo DNA viral/integrase; 5: integração; 6: síntese do RNA viral genômico; 7: síntese do RNAm viral; 8: tradução das proteínas virais; 9: montagem; 10: brotamento; 11: maturação.....	6
Figura 3. Inibidores da transcriptase reversa nucleosídeo-nucleotídeo (ITRN).....	13
Figura 4. Inibidores da transcriptase reversa não-nucleosídeo (ITRNN).....	13
Figura 5. Inibidores peptídicos da protease (IP).	14
Figura 6. Estruturas do tipranavir, maraviroc, T-20 e etavirine.....	15
Figura 7. Resumo gráfico do processo de integração do DNA proviral ao DNA hospedeiro. Parcialmente adaptada e reproduzida de Kawasuji e colaboradores (2006) com permissão da Elsevier® licença nº 2275380640114 obtida <i>on-line</i> via http://www.copyright.com	18
Figura 8. Estrutura do KU-55933.....	20
Figura 9. Representação das estruturas secundárias das cadeias A dos complexos cristalográficos PDB 1QS4, 1HYV, 1HYZ e do modelo teórico PDB 1WKN. Proteínas coloridas de acordo com suas cadeias secundárias. Os ligantes são apresentados como “bolas-e-bastões”.....	22
Figura 10. Exemplos de compostos orgânicos isolados de fontes naturais que apresentaram a capacidade de inibir a HIV-IN.....	23
Figura 11. Esquemas ilustrativos dos modelos farmacofóricos tridimensionais, construídos com base nos trabalhos de Carlson e colaboradores (2000), Dayam e colaboradores (2005) e Barreca e colaboradores (2005). DLH: doador de ligação hidrogênio; ALH: acceptor de ligação hidrogênio; ALHFBAR: acceptor de ligação hidrofóbica aromática; ALHFB: acceptor de ligação hidrofóbica.....	27
Figura 12. Esquema básico clássico da organização de um algoritmo ANN, baseado em Livingstone (2003).....	35
Figura 13. Representação de um perfil de amostragem conformacional, formado dentro de uma grade virtual. Nos processos de QSAR-3D, é utilizada uma grade semelhante. Figura cedida por Euzébio Guimarães Barbosa e João Paulo Athaide Martins.....	38

Figura 14. Representação da SEP de uma molécula com apenas dois ângulos torsionais livres. Exceto pelas indicações para mínimos global e local, a figura foi mantida como na fonte. Extraída de http://chem.syr.edu/faculty/kortner.html . Acesso: 06 de abril de 2009.....	43
Figura 15. Uma representação comum de um <i>split valence basis set double-zeta</i> . Se há mais algum número após “c”, estas funções são denominadas <i>triple-zeta</i> , <i>quadruple-zeta</i> , etc.....	56
Figura 16. Outra representação da estrutura do complexo cristalográfico 5CITEP e HIV-1 IN. Este complexo apresenta três subunidades idênticas, A, B e C. A região expandida apresenta o sítio onde o inibidor complexou-se (face da subunidade A). Figura já publicada pelo autor em Melo, Bruni e Ferreira (2006).....	57
Figura 17. Estrutura dos quatro dicetoácidos da Merck descritos por Hazuda e colaboradores (2000).....	59
Figura 18. Estrutura do ácido benzopirúvico e representação do equilíbrio ceto-enólico deslocado a favor da forma enólica por meio da formação de ligação hidrogênio intramolecular, e diânion que é formado quando na presença de íons metálicos.....	60
Figura 19. Representação do farmacóforo dos DCAs (Ar= cadeias laterais aromáticas diversas) e de alguns dos resultados de alguns dos diferentes estudos realizados para avaliar os possíveis modos de ligação dos inibidores de HIV-IN com o domínio principal. Baseado nas informações de Goldgur e colaboradores (1999) Sotriffer, Ni e McCammon (2000), Ni, Sotriffer e McCammon (2001), Keserü e Kolossvary (2001), Chen e colaboradores (2008), Healy e colaboradores (2009).....	61
Figura 20. Estrutura do S-1360, primeiro inibidor da HIV-integrase a entrar em testes clínicos.....	62
Figura 21. Estrutura do GSK-364735.....	63
Figura 22. Estruturas das naftiridinas L-870,810, L-870,812 e L-900,564.....	64
Figura 23. Estruturas do ácido nalidixico, do ciprofloxacino, do GS-9137, e do protótipo GS-9160.....	65
Figura 24. Estrutura do raltegravir e estruturas básicas das 4,5-diidroxipirimidinas carboxamidas e das N-Me pirimidonas.....	66
Figura 25. Estrutura do BMS-707035.....	67
Figura 26. Estrutura do MK-2048.....	68

Figura 27. Compostos descritos por Petrocchi e colaboradores (2007). Pertencem a classe das 4,5-diidroxipirimidinas carboxamidas.....	73
Figura 28. Compostos descritos por Summa e colaboradores (2006). Pertencem a classe das 4,5-diidroxipirimidinas carboxamidas.....	74
Figura 29. Compostos descritos por Guare e colaboradores (2006). Pertencem a classe das 8-hidroxi-1,6-naftiridinas.....	74
Figura 30. Compostos descritos por Egbertson e colaboradores (2007). Pertencem a classe das 8-hidroxi-1,6-naftiridinas.....	75
Figura 31. Compostos descritos por Hazuda e colaboradores (2000). Pertencem a classe dos β -dicetoácidos.....	75
Figura 32. Compostos descritos por Zhuang e colaboradores (2003). O conjunto é formado por um dicetoácido, duas 8-hidroxi-1,6-naftiridinas e quatro derivados relacionados.....	76
Figura 33. Compostos descritos por Wai e colaboradores (2000). Pertencem a classe dos β -dicetoácidos.....	76
Figura 34. Estruturas cristalográficas selecionadas do banco de dados CSD que foram utilizadas para como base para a construção do conjunto de treinamento. Os nomes de cada estrutura correspondem ao seu código específico no CSD.....	77
Figura 35. Esquema apresentando as duas formas de se obter as microespécies predominantes de uma molécula em um determinado pH utilizando o programa Marvin 4.1.8.....	78
Figura 36. Representação da caixa periódica com solvente explícito utilizada na etapa de dinâmica molecular.....	81
Figura 37. Interface gráfica do programa LQTA-QSAR.....	83
Figura 38. Exemplo de perfil de amostragem conformacional (composto E21b).....	83
Figura 39. Representação esquemática do funcionamento do algoritmo <i>in-house</i> planejado para selecionar descritores com tendência linear e distribuição normal em relação a atividade biológica. Algoritmo desenvolvido por Euzébio Guimarães Barbosa (grupo LQTA). Figuras cedidas pelo mesmo e adaptadas pelo autor.....	87
Figura 40. Gráficos e tabelas do Pirouete 4 utilizadas para auxiliar na redução de variáveis e eliminação de outliers.....	89

Figura 41. Representação gráfica da análise dos resultados de um teste de randomização de y segundo os critérios de Eriksson e colaboradores (2003). Os pontos distantes dos demais (valores = 1) correspondem aos valores obtidos para o modelo. O intercepto também pode ser observado na equação da reta. Teste feito com valores hipotéticos e dez randomizações, apenas como exemplo.....	96
Figura 42. Gráficos dos resultados obtidos no teste de randomização de y e na validação LNO para o modelo I.....	101
Figura 43. Orbitais moleculares ocupados mais altos (HOMOs) dos compostos P19 , P28 e P38	103
Figura 44. Representação dos eixos cartesianos para os compostos P19 , P28 e P38	105
Figura 45. Aspecto do <i>QSARmodelling</i> , programa escrito em JAVA para seleção de variáveis e construção de modelos utilizando o método PLS. O programa também é capaz de realizar o teste de randomização de y e a validação LNO.....	108
Figura 46. Resultados do teste de randomização de y e da validação LNO para o modelo III.....	113
Figura 47. Resultados do teste de randomização de y e da validação LNO para o modelo IV.....	113
Figura 48. Resultados do teste de randomização de y e da validação LNO para o modelo V.....	114
Figura 49. Resultados do teste de randomização de y e da validação LNO para o modelo VI.....	114
Figura 50. Gráfico dos pIC_{50} 's observados (eixo x) pelos previstos (eixo y), obtido no processo de validação cruzada Q^2_{LOO} do modelo VI (sem outliers e antes da retirada do conjunto teste).....	117
Figura 51. Dendograma (dados autoescalados) do conjunto de treinamento completo para o modelo VI. O conjunto teste encontra-se destacado (pontos pretos).....	120
Figura 52. Exemplo dos eixos internos de alguns compostos do conjunto de treinamento (E5b , $pIC_{50} = 8,15$; P18 , $pIC_{50} = 4,300$; W2 , $pIC_{50} = 5,25$). O terceiro eixo encontra-se paralelo ao plano. Figura interpretada no programa HyperChem 7 (HYPERCUBE, 2002).....	132
Figura 53. Representação do farmacóforo DCA, apresentando os átomos que foram identificados como 1 a 8 para a etapa de alinhamento dos conformeros obtidos por DM.....	140
Figura 54. Resultados do teste de randomização de y e da validação LNO para o modelo XI.....	145

Figura 55. Dendograma (dados autoescalados) do conjunto de treinamento completo para o Modelo XI. O conjunto teste encontra-se destacado (pontos pretos).....	146
Figura 56. Descritores relativos ao modelo QSAR-4D (Modelo XI) plotados no espaço tridimensional. Esquerda: apresentados como grade. Direita: forma pontual e suas respectivas indicações no formato x.y.z.sonda.potencial.(sinal do coeficiente). Azul: descritor Zn ²⁺ .LJ positivo; verde: descritor Zn ²⁺ .C negativo; cinza: descritor Ar(C-H).LJ positivo; rosa: descritor Ar(C-H) negativo. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).....	148
Figura 57. Plotagem dos descritores 20.17.19.Zn ²⁺ .C (verde) e 14.15.21.Zn ²⁺ .LJ (azul) no espaço tridimensional em torno do compostos E5b . Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).....	150
Figura 58. Plotagem do descritor 20.17.19.Zn ²⁺ .C no espaço tridimensional em torno do perfil de amostragem conformacional (formato arame) do inibidor L731942 . Devido a grande flexibilidade conformacional do inibidor, uma conformação isolada (formato bastões) é apresentada para facilitar a visualização. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).....	150
Figura 59. Plotagem do modelo completo no espaço tridimensional dos perfis de amostragem conformacional (A, em formato arame) e de uma conformação isolada (B, em formato bastões) do inibidor E5b . Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).....	151
Figura 60. Plotagem do modelo completo no espaço tridimensional dos perfis de amostragem conformacional (A, em formato arame) e de conformações isoladas (uma em B e duas em C, em formato bastões) do inibidor L731942 . Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).....	152
Figura 61. Interações ligante-receptor entre uma naftilpiridina e o sítio de ligação da HIV-IN propostas por Healy e colaboradores (2009). Região destacada da Figura 20.....	153
Figura 62. Plotagem dos descritores Ar(C-H).LJ que localizam-se no espaço tridimensional a cadeia lateral aromática (perfil de amostragem conformacional do inibidor E5b). Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).....	154
Figura 63. Plotagem dos descritores 21.15.16.Ar(C-H).LJ e 15.15.18.Ar(C-H).LJ no espaço tridimensional em torno do inibidor E5b , e da distância entre ambos. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).....	156
Figura 64. Plotagem do descritores sobre o sítio de ligação da HIV-IN (PDB 1QS4), sem 5CITEP. A alça hidrofóbica é representada apenas como uma região (direita), pois, nesta estrutura cristalográfica, ela não foi resolvida. A sobreposição foi construída utilizando o programa HyperChem 7.1 (HYPERCUBE, 2002) e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).....	157

Figura 65. Estrutura básica das estirilquinolinas.....	157
Figura 66. Mapas de contorno do modelo CoMFA relativos à inibição da reação ST, propostos por Raghavan e colaboradores (1995). A imagem corresponde à Figura 7 do artigo original. Reproduzida com permissão da American Chemical Society® licença nº 2224490959169 obtida <i>on-line</i> via http://www.copyright.com).....	160
Figura 67. Mapas de contorno (para duas conformações, A e B) dos modelos CoMFA relativos à inibição da reação 3’P (porém similares aos mapas do modelo ST) propostos por Kuo e colaboradores (2004). A imagem corresponde a uma parte da Figura 4 do artigo original. Parcialmente adaptada e reproduzida com permissão da American Chemical Society® licença nº 2224491199544 obtida <i>on-line</i> via http://www.copyright.com).....	160
Figura 68. Mapas de contorno do modelos CoMFA relativos à inibição da reação ST propostos por Makhija e Kulkarni (2002b). A imagem corresponde a uma parte da Figura 11 do artigo original. Parcialmente adaptada e reproduzida com permissão da SpringerLink® (licença nº 2233250553421 obtida <i>on-line</i> via http://www.copyright.com).....	161
Figura 69. Plotagem dos descritores 26.18.14.Ar(C-H).LJ, 26.20.19.Ar(C-H).LJ e 24.13.12 Ar(C-H).LJ no espaço tridimensional em torno do compostos E5b . Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYIS, 2001).....	162
Figura 70. Resultados do teste de randomização de <i>y</i> e da validação LNO para o modelo XIII.....	166
Figura 71. Dendograma (dados autoescalados) do conjunto de treinamento completo para o Modelo XIV. O conjunto teste encontra-se destacado (pontos pretos).....	168

1. INTRODUÇÃO

1.1 AIDS/HIV: SURGIMENTO, DESCOBERTA E DOENÇA.

Apesar das evidências de mortes causadas pela síndrome da imunodeficiência adquirida (*Aids: acquired immunodeficiency syndrome*) há mais de 50 anos (AVERT, 2009a), os primeiros relatos oficiais de pacientes com quadros patológicos compatíveis datam de junho de 1981 em Los Angeles e Nova Iorque (EUA) (LEWIS *et al.*, 1997).

Essa doença caracteriza-se por ser de longa duração, com um grande espaço de tempo entre a exposição ao agente etiológico e um profundo estado de imunossupressão devido a uma drástica redução no número de linfócitos T (células CD4⁺). Tais células são responsáveis pela ativação de outros linfócitos que realizam a defesa do organismo. A diminuição da razão entre células “T helper” e células T supressoras pode facilitar a ocorrência de infecções oportunistas que causam doenças como tuberculose, pneumonia (sendo esses dois quadros patológicos os principais causadores de óbitos em pacientes imunodeprimidos), histoplasmose, toxoplasmose, pneumocistose e outras. Também pode ocorrer o surgimento de câncers, em especial carcinoma anal, linfoma não-Hodgkin e o Sarcoma de Kaposi (GALLO e MONTAGNIER, 2002; PRUSIER, 2002; SOUZA *et al.*, 2008; NADAL *et al.*, 2007). Outros quadros patológicos associados podem ser encontrados na literatura: astenia, perda de peso acentuada, infarto cerebral, doenças oculares e complicações neurofisiológicas (YAKASAI, 2004, ROCHA *et al.*, 2003; MATOS, SANTOS e MUCCIOLLI, 1999).

O curso da doença pode ser relatado pela concentração de células T no sangue, cuja contagem normal varia de 800 a 1200 células/mm³. Os indivíduos que apresentam contagem média de 500 células/mm³ são definidos como infectados pelo agente etiológico

(soropositivos) ou portadores[#]. Apesar de aparentemente saudáveis, podem transmitir a infecção. Quando a contagem fica entre 200 e 500 células/mm³, o quadro é dado como complexo relacionado à Aids (ARC: *Aids related complex*). Finalmente, é considerado doente de Aids[#] aquele indivíduo cuja contagem está abaixo de 200 células/mm³. Com prosseguimento da redução, o agente etiológico passa a infectar outros tipos de células, como musculares ou do sistema nervoso central (SNC) (GREGORY e TOLMAN, 2004).

Apesar desta doença estar relacionada principalmente à infecção e depleção dos linfócitos T, outras células que expressam o receptor CD4 também podem ser infectadas, como monócitos e macrófagos. Os monócitos são importantes para a patogenia da Aids, pois não são rapidamente destruídos pelo agente etiológico, ao contrário do que ocorre com os linfócitos. Desta forma, eles disseminam o agente para o cérebro, medula óssea, e outros órgãos. Esta pode ser a principal causa dos quadros patológicos relacionados à Aids que não são decorrentes de infecções oportunistas. Por exemplo, a demência relacionada à Aids provavelmente é causada ou promovida por mais citocinas liberadas por macrófagos e monócitos (GREGORY e TOLMAN, 2004).

A descrição da doença em populações muito diferentes, como homens jovens e homossexuais sexualmente ativos, usuários de drogas injetáveis, hemofílicos, haitianos, recém-nascidos e receptores de transfusões sanguíneas, levou à conclusão da existência de algum agente infeccioso ou contaminante, como fungos e produtos químicos, ou até mesmo autoimunidade contra linfócitos (PRUSINER, 2002). A maioria das possíveis causas foram descartadas quando os centros de prevenção e controle de doenças (CDC: *Centers of Disease Control and Prevention*) descreveram casos de Aids em pacientes hemofílicos que receberam

[#] Segundo o Ministério da Saúde e outros órgãos, desde 2005, a utilização do termo *aidético* para definir os portadores de HIV ou doentes de Aids é incorreta, inadequada e preconceituosa. Todos os termos corretos podem ser encontrados em http://www.roche.com.br/TherapeuticAreas/doencas_infecciosas/aids/hiv_aids/curiosidades.

somente fatores sanguíneos filtrados, sem a possibilidade de presença de agentes maiores que um vírus (GALLO e MONTAGNIER, 2002).

Na década de 1960, determinou-se que alguns retrovírus animais (como o vírus de sarcoma em aves) possuíam a capacidade de causar cânceres. Isto levou, na década de 1970, à formação de grupos de pesquisa com o objetivo de identificar vírus e retrovírus oncogênicos para os seres humanos. Porém, apesar dos esforços, até meados da década nenhum havia sido descrito, o que causou o encerramento das atividades de muitos desses grupos por uma conclusão errônea de que não existiam retrovírus humanos* (GALLO, 2002; MONTAGNIER, 2002).

Porém, os grupos de pesquisa de Robert C. Gallo, nos EUA, e Luc Montagnier, na França, continuaram os estudos. Utilizando dados obtidos durante 15 anos de pesquisa básica de retrovírus leucomogênicos de animais, foi possível o desenvolvimento de ensaios bioquímicos baseados na enzima transcriptase reversa (TR), já conhecida e essencial para os retrovírus. O grupo de Gallo também descobriu o fator de crescimento das células T, a interleucina-2 (IL-2). Assim, em 1979, este grupo conseguiu isolar o primeiro retrovírus humano, o vírus de linfomas de células T (HTLV-1: *human T lymphotropic virus type I*), que podia ser transmitido por meio de transfusões sanguíneas, contato sexual e de mãe para filho. Em 1982, o mesmo grupo isolou o HTLV-2 (GALLO, 2002; MONTAGNIER, 2002; GALLO e MONTAGNIER, 2002).

A possibilidade de que o agente etiológico da Aids fosse um vírus, levou tais pesquisadores à hipótese de que o HTLV poderia ser o agente, já que as vias de transmissão sugeridas eram as mesmas. Além disto, o mesmo marcador biológico, uma diminuição nos

*Robert C. Gallo comenta um fato interessante. Além do fracasso inicial na identificação de retrovírus humanos, a afirmação de que doenças epidêmicas sérias eram limitadas ao 'terceiro mundo' levou ao fechamento de um número indeterminado de departamentos de microbiologia de escolas de medicina norteamericanas e a uma perturbadora falta de suporte para os DCDs.

níveis dos linfócitos T CD4⁺, estava presente (GALLO e MONTAGNIER, 2002). Porém, no início de 1983, o grupo de Gallo encontrou sequências de DNA relacionadas ao HTLV em somente dois pacientes de um grupo de 33 pessoas com Aids. Já o grupo de Luc Montagnier encontrou traços de TR no sobrenadante da cultura de linfócitos de pacientes, mas não observou a precipitação imune dos anticorpos para HTLV (GALLO, 2002; MONTAGNIER, 2002). Essas e outras ocorrências levaram os dois grupos à conclusão de que o agente etiológico era um vírus um pouco diferente do HTLV. Gallo determinou que 5 a 10% dos pacientes com Aids apresentavam infecção dupla pelo HTLV e pelo novo vírus, ocorrência que dificultou seus estudos (GALLO, 2002; MONTAGNIER, 2002).

Novas evidências de que esse retrovírus era o agente etiológico da Aids foram descobertas, e, em meados de 1984, o grupo de Gallo conseguiu fazer crescer vírus isolados de pacientes em culturas de células T CD4⁺. Os dados confirmaram que um novo retrovírus era o agente etiológico da Aids (GALLO, 2002; MONTAGNIER, 2002). O novo vírus pertencia à subfamília dos *lentiretrovirus*, foi então denominado vírus da imunodeficiência humana (HIV: *human immunodeficiency viruses*) (PRUSIER, 2002; GALLO e MONTAGNIER, 2002).

1.2 MORFOLOGIA E CICLO REPLICATIVO DO HIV.

A origem do HIV é ainda desconhecida. Uma das hipóteses mais aceitas é a de que tenha surgido na África Central, como resultado de uma mutação, descendendo por via indireta de outro vírus, o vírus da imunodeficiência símia (SIV: *simian immunodeficiency virus*), identificado em algumas espécies de símios, como o *Pan troglodytes troglodytes* (AVERT, 2009a).

Os retrovírus possuem RNA (ácido ribonucléico) como material genético. No citoplasma, o RNA é convertido em cDNA (ácido desoxirribonucléico complementar) pela enzima transcriptase reversa (TR). Em seguida, o cDNA é integrado ao genoma do hospedeiro pela enzima integrase (IN), formando um pró-vírus maduro capaz de sintetizar novas fitas de RNA genômico viral e de mRNA (RNA mensageiro) viral (VOISSET *et al*, 2000; SETHI, 2002). Como o processo de integração causa danos ao material genético do hospedeiro, os retrovírus podem causar tumores em animais e humanos (SETHI, 2002).

A estrutura morfológica dos HIV, tipos 1 e 2 (o último descrito em 1986), inclui proteínas estruturais e funcionais e um genoma de RNA protegidos pelo envelope viral. O envelope viral é constituído por uma bicamada lipídica e contém uma proteína complexa, conhecida como *env*. Tal proteína é constituída pelas glicoproteínas gp41, transmembrana, e gp120, exposta à camada externa do envelope (CHALLAND e YOUNG, 1997).

Na face interior, o HIV possui a proteína viral denominada p17 (matriz), que envolve um capsídeo composto pela proteína p24. Na parte mais interna, encontram-se os elementos mais importantes: dois filamentos simples de RNA protegidos por um nucleocapsídeo (p7) e três enzimas essenciais, a HIV-transcriptase reversa (HIV-TR), HIV-protease (HIV-PR) e HIV-integrase (HIV-IN) (PEÇANHA, ANTUNES e TANURI, 2002; AVERT, 2009b) (Figura 1).

O ciclo replicativo do HIV, resumido na Figura 2, ocorre no meio intracelular e é constituído de várias etapas, e cada uma tem um alvo em potencial para o desenvolvimento de estratégias terapêuticas contra a Aids (PEÇANHA, ANTUNES e TANURI, 2002; MITSUYA, SHIRASAKA e BRODER, 1990).

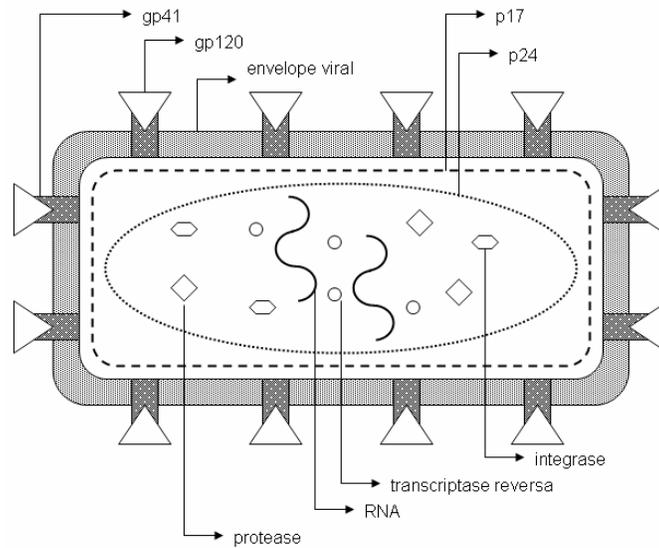


Figura 1. Morfologia do HIV. Adaptado de Avert (2009) e Ramalho (2006). O nucleocapsídeo não é apresentado.

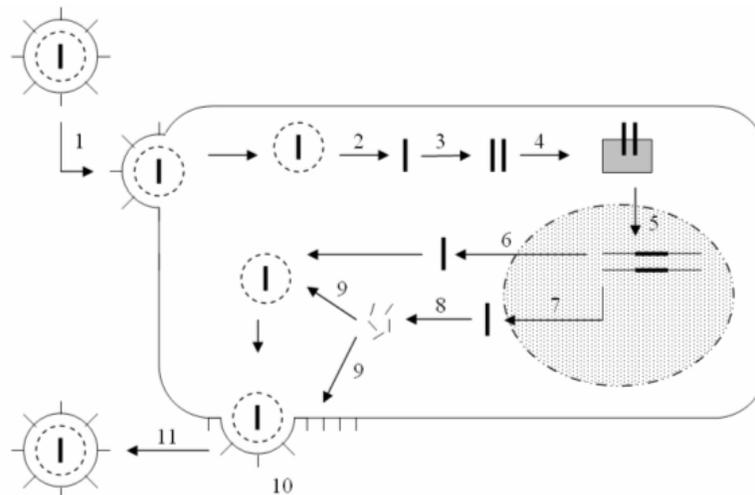


Figura 2. Ciclo replicativo do HIV. Adaptado de Souza e Almeida (2003) e LaBonte, Lebbos e Kirkpatrick (2003), já publicado pelos autores em Melo, Bruni e Ferreira (2006): 1: adsorção/fusão; 2: desencapamento; 3: transcrição reversa; 4: formação do complexo DNA viral/integrase; 5: integração; 6: síntese do RNA viral genômico; 7: síntese do RNAm viral; 8: tradução das proteínas virais; 9: montagem; 10: brotamento; 11: maturação.

1.3 EPIDEMIOLOGIA E IMPACTO SOCIOECONÔMICO DA AIDS – BRASIL E O MUNDO

A Aids é considerada uma doença pandêmica (PEÇANHA, ANTUNES e TANURI, 2002). Dados da Un aids (Programa Internacional de Combate à Aids das Nações Unidas) estimam em 33 milhões o número de pessoas contaminadas pelo HIV (UNAIDS, 2008). Até 2005, 3,1 milhões de mortes já haviam sido registrados, sendo 570 mil crianças (AIDS, 2005). A situação é mais grave na África Subsaariana, onde a Aids é a principal causa de mortes. Na virada do século, a expectativa de vida naqueles países havia caído de 62 para 47 anos (MINISTÉRIO, 2002a; UNAIDS, 2002). No Brasil, de 1980 até julho de 2008, foram registrados 506.499 casos da doença, com 205.409 óbitos registrados (MINISTÉRIO, 2008; BRASIL, 2008). Segundo estimativas da Un aids (2008), o Brasil possui 730.000 de um total de 1.700.000 soropositivos da América Latina, com tendência de estabilização da epidemia entre os homens, redução entre os consumidores de drogas injetáveis e aumento entre as mulheres e populações menos favorecidas.

A epidemia mundial de Aids evoluiu em três etapas. A fase inicial caracterizou-se pela infecção entre homens que fazem sexo com outros homens, com alto nível de escolaridade. Na fase secundária, ocorreu o aumento do número de casos devido ao uso de drogas injetáveis, a diminuição da faixa etária dos infectados, e o aumento da disseminação entre indivíduos heterossexuais. A terceira e atual fase é marcada pela acentuação da tendência de disseminação entre os heterossexuais, principalmente as mulheres. Esta última observação merece destaque, pois vários países vêm observando uma ‘feminização’ da epidemia: enquanto na década de 1980, a média foi de 6,5 homens para uma mulher infectada, atualmente a razão é de 1,7 homens para uma mulher (TEIXEIRA, 2003).

A epidemia global de Aids apresenta um comportamento complexo, dinâmico e instável, pois ocorre de formas diferentes de acordo com a região e com o comportamento humano individual e coletivo, configurando-se como um mosaico de subepidemias regionais. Por exemplo, a doença apresenta um baixo nível de incidência em países como a Argélia, Chile, Costa Rica, Cuba, Suriname, Rússia e Polônia, enquanto é generalizada na África subsaariana e Haiti (BRITO, CASTILHO e SZWARCOWALD, 2000). Na maior parte da África subsaariana, a disseminação do HIV ocorre principalmente devido a transmissão pelas profissionais do sexo para clientes do sexo masculino, os quais transmitem a doença para suas parceiras ou esposas, sendo que elas também podem ter mais de um parceiro sexual. Já na Ásia Central, outra região de alta incidência, a transmissão é virtualmente linear e está mais associada aos usuários de drogas injetáveis, e a profissionais do sexo que contaminam clientes do sexo masculino e esses o repassam às suas parceiras e esposas, as quais são normalmente monogâmicas (COOVADIA e HADINGHAM, 2005).

Uma queda da letalidade passou a ser observada em meados da década de 1990, coincidindo com a garantia de acesso ao tratamento antiretroviral de alta eficácia (HAART: *highly active antiretroviral therapy*) (AIDS, 2003b). A HAART é uma combinação de fármacos capazes de inibir a replicação viral e age em mais de um ponto do sítio replicativo; é capaz de diminuir em até 100 vezes o ritmo de produção do vírus quando comparada à monoterapia (SOUZA e ALMEIDA, 2003; La BONTE, LEBLOS e KIRKPATRICK, 2003). No Brasil, a Lei 9113/96 garantiu o acesso livre ao tratamento (NATIONAL, 2001).

Porém, além de ser um grande problema de saúde pública e um desafio à ciência, a Aids apresenta graves consequências sociais, bem como um grande impacto sobre a economia de diversos países. A epidemia brasileira, por exemplo, caracterizou-se na última década pelo processo de *pauperização*, ou seja, cresceu entre as populações socioeconomicamente mais

vulneráveis. Em 1985, 76,0% dos casos diagnosticados apresentavam nível superior ou médio. Em 1999/2000, ocorreu uma inversão, com 74,0% de analfabetos ou que tinham completado apenas o nível fundamental. Isto provavelmente deve-se ao fato de que a baixa escolaridade está diretamente relacionada a um maior comportamento sexual de risco. A frequência do uso de preservativo também aumenta de acordo com o grau de escolaridade, enquanto o uso de drogas diminui da mesma forma. Assim, jovens e adolescentes (especialmente mulheres) com baixos níveis socioeconômicos e de instrução são, atualmente, mais susceptíveis às infecções sexualmente transmissíveis. Em 1996, a doença foi a quarta causa de mortes entre os adultos brasileiros de ambos os sexos na faixa produtiva de 20 a 49 anos (BRITO, CASTILHO e SZWARCOWALD, 2000). O fenômeno mais recente é o aumento observado na incidência de Aids entre indivíduos acima dos 60 anos de idade, tendo como causas principais à existência de tabu sobre sexualidade na terceira idade e facilidade de, entre aqueles idosos com maiores recursos pessoais, terem uma vida sexual mais ativa (ARAÚJO *et al.*, 2006).

Globalmente, a epidemia diminui o crescimento econômico devido ao seu impacto no trabalho e na produção: indivíduos doentes perdem seus empregos, passando a consumir e investir menos. Nos países mais atingidos, ocorre uma redução da força de trabalho, e setores como fazendas, fábricas, prestação de serviços e órgãos governamentais tornam-se menos produtivos. Isto também leva à queda da qualidade do trabalho, pois trabalhadores infectados ou que vêm a óbito são substituídos por jovens com menos experiência, problema que se torna ainda mais crítico devido à perda de professores e treinadores. A Organização Internacional do Trabalho estima uma queda de 5 a 35% na força de trabalho em 38 países até 2020 por causa da Aids (UNAIDS, 2005).

A mortalidade dentro da faixa economicamente ativa também causa um fenômeno que se agrava desde o início da epidemia: o aumento do número de órfãos (BRITO, CASTILHO e

SZWARCWALD, 2000). Na África, o número de órfãos decorrentes da Aids é estimado em 12 milhões, o que demonstra a proporção da tragédia humana que a doença significa. Mesmo em países onde a prevalência do HIV estabilizou ou diminuiu, o número de órfãos continua a aumentar devido ao espaço de tempo entre quando os pais tornam-se infectados e quando eles morrem.

Na África do Sul, espera-se que em 2010 haja 3,1 milhões de órfãos decorrentes da AIDS, cerca de 18% de todas as crianças. Além do abandono, a situação desses indivíduos acaba se agravando pelo desenvolvimento de problemas psicológicos, como a síndrome de estresse pós-traumático, abuso de álcool e drogas, agressividade e suicídio. Muitos sofrem com o preconceito por serem ‘órfãos da AIDS’, um estigma em alguns países e deixam as escolas para trabalharem e auxiliarem suas famílias. Também ficam mais sujeitos à: exploração do trabalho, exploração sexual, violência, vida nas ruas e, em casos extremos, ao sequestro e uso como “crianças-soldados” em campos de batalha da África (UNAIDS, 2005).

No Brasil, a epidemia pode ser considerada como concentrada¹. Haja vista as profundas desigualdades sociais, a propagação da infecção pelo HIV no País sofre transformações significativas ao longo do tempo. Inicialmente restrita às metrópoles nacionais, onde homens com comportamento homossexual e indivíduos hemofílicos eram os principais alvos, a epidemia passou por processos de heterossexualização, feminização, interiorização e pauperização. O número de órfãos decorrentes da morte de pais aidéticos era estimado, ainda em 1999, em cerca de 29.928 pessoas (BRITO, CASTILHO e SZWARCOWALD, 2000).

Os avanços tecnológicos e o melhor conhecimento da etiopatologia da AIDS permitiram o surgimento de novas propostas de intervenções diagnósticas, profiláticas e terapêuticas, as quais podem ser atribuídas ao expressivo aumento da sobrevivência dos doentes

¹Epidemia concentrada: prevalência da infecção mostra-se superior a 5% em uma ou mais subpopulações com comportamento de auto risco, mas a prevalência entre gestantes atendidas em clínicas de pré-natal revela-se menor do que 5%.

(BRITO, CASTILHO e SZWARCOWALD, 2000). A introdução da HAART ajudou muito no aumento da sobrevivência, além de promover uma redução significativa do risco de progressão da AIDS (HAGMAM, 2003). Nos últimos anos da década de 1990, a introdução de novos fármacos anti-retrovirais trouxe alguma perspectiva de melhoria no que se refere à mortalidade por AIDS. Um aumento na sobrevivência dos pacientes foi descrito em Nova Iorque, E.U.A., o qual, em 1996, coincidiu com a disponibilidade ampla de tais medicamentos, fato observado também na Itália, Inglaterra, Espanha e Brasil (BRASIL, 2005; MENESIA *et al.*, 2001).

O Brasil ganhou destaque internacional na luta contra a Aids e foi o primeiro país em desenvolvimento a garantir acesso livre e universal ao tratamento antiretroviral para todos os pacientes (CHEQUER *et al.*, 2005). Atualmente, a doença assumiu um caráter crônico com evolução prolongada, onde longos períodos assintomáticos são interrompidos pelo surgimento de infecções oportunistas (BRITO, CASTILHO e SZWARCOWALD, 2000).

1.4 FÁRMACOS ANTI-HIV E A NECESSIDADE DE NOVOS AGENTES.

Apesar do enorme progresso científico obtido em mais de duas décadas de pesquisa, ainda não se dispõe de uma vacina profilática capaz de eliminar, ou mesmo controlar a epidemia de AIDS, o que permanece um grande desafio para a ciência. Várias formulações antigênicas já foram testadas, entretanto, até dezembro de 2007, as formulações testadas em humanos não se mostraram eficazes. O desenho de uma formulação única de uma vacina profilática anti-HIV efetiva, para uso em todo o mundo, também parece improvável, já que o vírus está em constante modificação e adaptação para se evadir à resposta imunológica (BRASIL, 2008b).

Assim, enquanto as abordagens visando à prevenção (uso de preservativos, educação sexual, controle de sangue e derivados sanguíneos, etc) e à redução de danos (distribuição de seringas descartáveis) continuam sendo as melhores abordagens para evitar a contaminação pelo HIV, enquanto a terapia antiretroviral continua sendo a melhor abordagem para o tratamento dos pacientes.

Dependendo do estado e da idade do paciente (MINISTÉRIO, 2002a; MINISTÉRIO, 2002b), a HAART pode ser composta por dois ou mais inibidores de transcriptase reversa nucleosídeo-nucleotídeo (ITRN) (Figura 3), inibidores da transcriptase reversa não-nucleosídeo (ITRNN) (Figura 4), ou inibidores de protease (IP) (Figura 5).

Nos últimos anos, novos fármacos das classes já existentes e de novas classes foram lançados (Figura 6): o T-20 (Roche), um polipeptídeo (36 aminoácidos) que interage com a glicoproteína de superfície viral gp41 e age impedindo a fusão do vírus com a membrana da célula hospedeira (SOUZA e ALMEIDA, 2003; LA BONTE, LEBLOS e KIRKPATRICK, 2003; CUNICO, GOMES e VELLASCO JR, 2008); o tipranavir (Boehinger), um inibidor não-peptídico da protease (INPP) (FLEXNER, BATE e KIRKPATRICK, 2005), e o maraviroc (Pfizer), um bloqueador da fusão que age como antagonista do co-receptor de membrana CCR5. Também foram lançados um novo IP, o darunavir, e um novo ITRNN, o etavirine (CUNICO, GOMES e VELLASCO JR, 2008), ambos da Tibotec Inc.

Mesmo com o surgimento da HAART, que é capaz de promover a diminuição do índice de partículas virais a níveis indetectáveis, o vírus normalmente persiste como próvirus integrados ao material genético de linfócitos T adormecidos, que provêm uma fonte para replicação viral quando da interrupção da terapia (LEHRMAN *et al.*, 2005). Essas células podem ser ativadas pela administração de interleucina-2 (IL-2) em pacientes que recebem a

HAART. Mas, a ativação pela IL-2 não é seletiva, e os reservatórios são refeitos (BARTON, 2005; LEHRMAN et al., 2005).

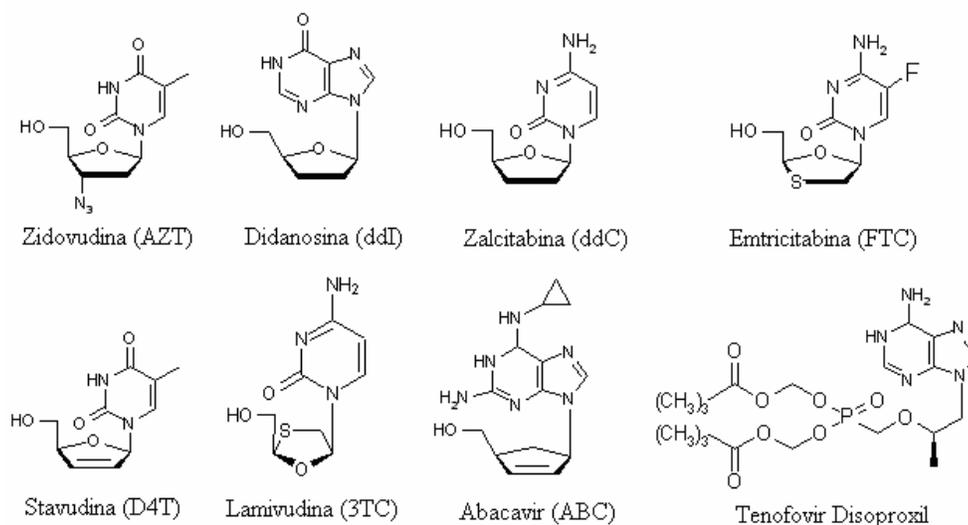


Figura 3. Inibidores da transcriptase reversa nucleosídeo-nucleotídeo (ITRN).

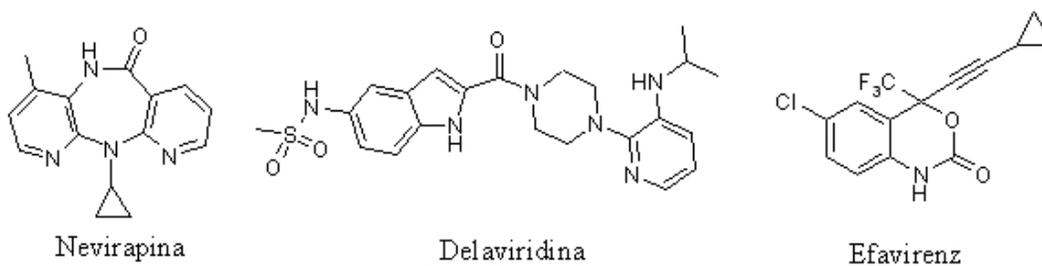


Figura 4. Inibidores da transcriptase reversa não-nucleosídeo (ITRNN).

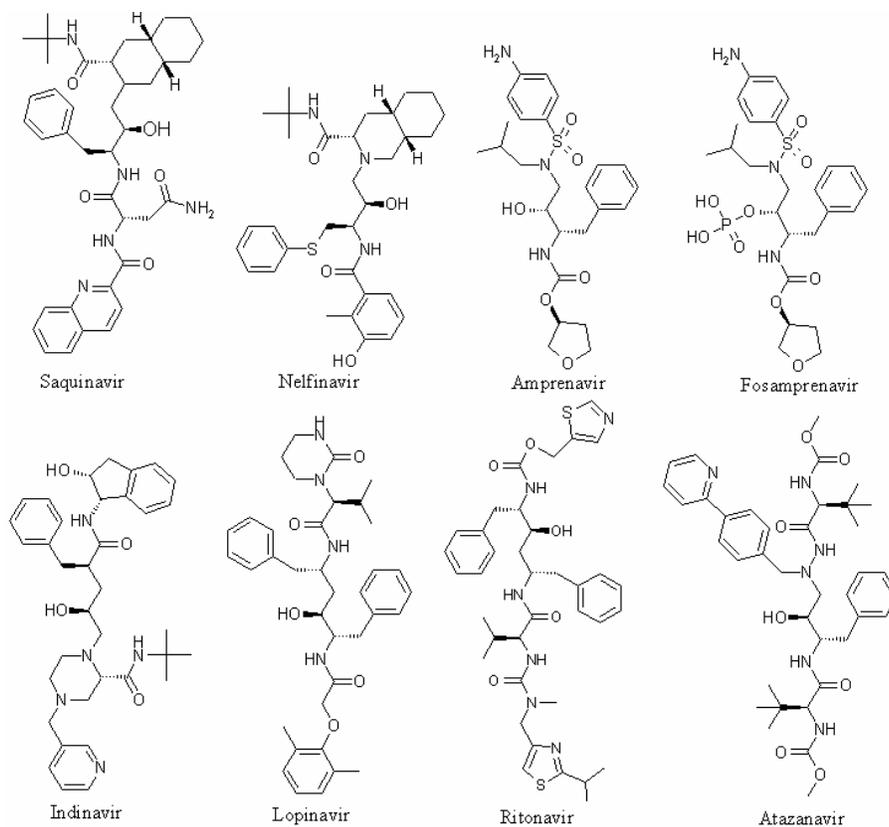


Figura 5. Inibidores peptídicos da protease (IP).

Estudos têm sido realizados para que se encontrem agentes que induzam uma forma seletiva de expressão dos genes próvirais latentes nessas células. O ácido valpróico, um conhecido anti-convulsivante, apresenta esta capacidade, pois é capaz de inibir a histona deacetilase-1 (HDAC1), uma enzima envolvida na remodelagem de cromatina e importante para a manutenção da latência do HIV-1 integrado (BARTON, 2005; LEHRMAN *et al.*, 2005).

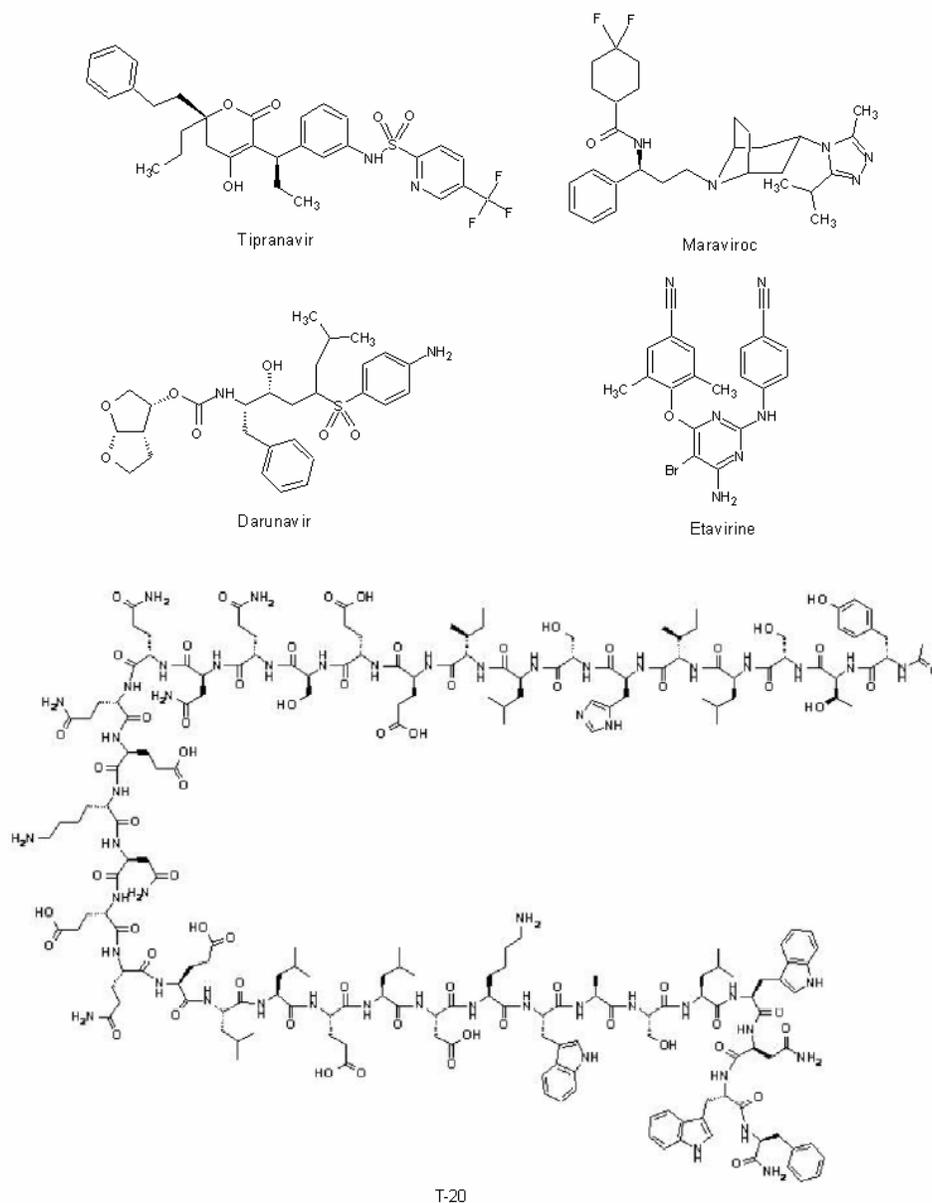


Figura 6. Estruturas do tipranavir, maraviroc, T-20 e etavirine.

O surgimento e a velocidade de aparecimento de cepas resistentes a diferentes combinações de fármacos disponíveis no mercado, inclusive pela resistência cruzada entre fármacos de uma mesma classe, é outro fator limitante. O surgimento de cepas multi-fármaco resistentes, inclusive altamente virulentas e de progressão acelerada, ocorrem principalmente por falhas na farmacoterapia e, mais raramente, de forma natural (BERKHOUT, de RONDE e

van der HOEK, 2005; PEÇANHA, ANTUNES e TANURI, 2002; SOUZA e ALMEIDA, 2003; La BONTE, LEBLOS e KIRKPATRICK, 2003; GOLDGUR *et al.*, 1999). Uma ocorrência recente foi o denominado *vírus de Nova Iorque*. Essa variante, altamente virulenta (progrediu em cerca de apenas cinco meses), mostrou-se resistente a todos os medicamentos disponíveis na época. Porém, apesar desse caso ter ganhado destaque, não é um caso isolado (SMITH, 2005; BERKHOUT, de RONDE e van der HOEK, 2005).

A alta incidência de efeitos colaterais causados pelos fármacos anti-HIV disponíveis é outro fator importante. Um estudo realizado por Lignani Jr, Greco e Carneiro (2001) concluiu que ocorrências adversas devido ao uso de medicamentos antiretrovirais (vômitos, diarreias, náuseas, dores abdominais, dores de cabeça e alterações da coloração da pele) são as principais causas de falhas na aderência dos pacientes ao tratamento (20,5% dos casos), o que aumenta a probabilidade de surgimento de cepas resistentes. Waltz (2006) mostrou que os indivíduos que fazem uso constante desses medicamentos possuem uma chance 50% maior de risco de morte por doenças cardiovasculares, hepáticas e renais. Já Farhi, Lima e Cunha (2008) mostraram que o tempo de uso de antiretrovirais relaciona-se diretamente ao aumento do risco de dislipidemia (aumentos das concentrações de triglicérides e colesterol no sangue) entre os indivíduos em tratamento. Além disso, observou-se nos últimos anos o surgimento da *síndrome lipodistrófica do HIV*, uma distribuição irregular de gordura no corpo decorrente do uso prolongado dos medicamentos antiretrovirais (ALENCAR, NEMES e VELLOSO, 2008).

Os fatores apresentados fomentam a contínua necessidade de descoberta e/ou desenvolvimento de novos e eficientes fármacos que atuem sobre a HIV-TR, HIV-PR e em outros pontos do ciclo de replicação viral (PEÇANHA, ANTUNES e TANURI, 2002; SOUZA e ALMEIDA, 2003; La BONTE, LEBLOS e KIRKPATRICK, 2003). Outra possibilidade é o desenvolvimento de fármacos que atuem sobre proteínas e enzimas celulares

utilizadas na replicação viral, como os co-receptores de superfície CCR5 e CXCR4, que o vírus utiliza para auxiliar na penetração da célula hospedeira (local de ação do maraviroc) (KURITZKES e KIRKPATRICK, 2008) ou as α -glucosidases I e II, enzimas celulares utilizadas no ciclo de replicação viral para a síntese da glicoproteína viral de superfície gp120 (MELO e CARVALHO, 2006). O grande desafio são as descobertas de possíveis alvos que efetivamente interrompam o ciclo do vírus, causando o menor dano possível aos tecidos.

Um dos possíveis novos alvos terapêuticos que têm sido estudados é a enzima HIV-integrase (HIV-IN), uma das três enzimas virais essenciais para a replicação do HIV (etapa 4, Figura 2) (POMMIER, MARCHAND e NEAMATI, 2000; YOUNG, 2001; NEAMATI, SUNDER e POMMIER, 1997). Até o começo da década de 2000, essa enzima havia recebido pouca atenção em comparação com a HIV-TR e HIV-PR (TAKTAKISHVILLI *et al.*, 2000). Esta situação mudou com a descoberta dos primeiros protótipos seletivos e potentes para a mesma (GOLDGUR *et al.*, 1999; HAZUDA *et al.*, 2000).

1.5 A HIV-INTEGRASE COMO ALVO PARA FÁRMACOS ANTI-HIV

A enzima HIV-IN (que faz parte da família das nucleotidiltransferases, E.C. 2.7.7.49, e possui massa de 32 kDa^{*}) é codificada pela terminação 3' do gene viral *pol*, assim como a HIV-TR e a HIV-PR, e é a chave para um processo existente apenas nos vírus de RNA da família *Lentivirae*: a inserção do cDNA, produzido pela TR a partir do RNA viral, no genoma do hospedeiro (CRAIGIE, 2001). O processo, conhecido como integração, é dividido em duas fases: a reação de processamento 3' (3'P) e a de transferência de fita (ST: *strand transfer*). Esta é a única enzima viral responsável por esse processo, que pode ocorrer tanto *in vitro*

* A unidade de massa atômica unificada (u), ou dalton (Da) é uma unidade de massa utilizada para expressar massas atômicas e moleculares. É a massa aproximada de um átomo de hidrogênio, um próton ou um nêutron.

quando *in vivo* (PLUYMERS *et al.*, 2002; GUPTA e NAGAPPA, 2003). O processo de integração encontra-se resumido na Figura 7.

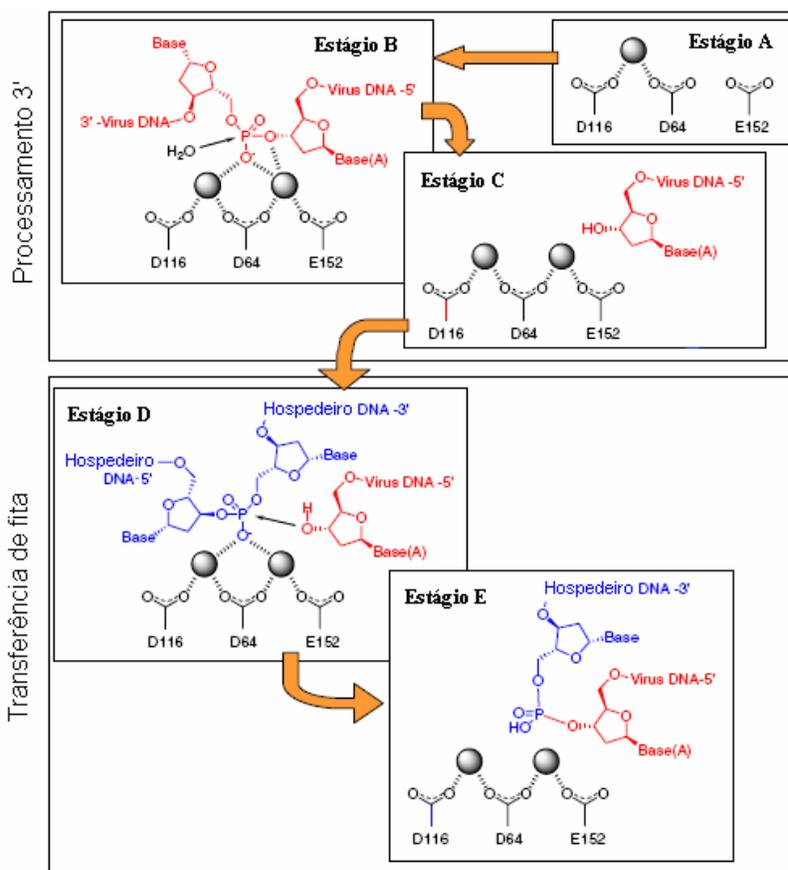


Figura 7. Resumo gráfico do processo de integração do DNA proviral ao DNA hospedeiro. Parcialmente adaptada e reproduzida de Kawasuji e colaboradores (2006) com permissão da Elsevier® licença nº 2275380640114 obtida *on-line* via <http://www.copyright.com>.

A HIV-IN se dobra em três domínios funcionais (ADESOKAN *et al.*, 2004). O domínio principal, formado pelos resíduos 50 - 212, contém a tríade catalítica responsável pelas reações 3'P e ST. Esta tríade é composta por dois aspartatos, D64 e D116, e um glutamato, E152, sendo então chamada "DDE" (POMMIER, MARCHAND e NEAMATI, 2000), e são os responsáveis pela ligação de dois átomos de íons metálicos (provavelmente

Mg⁺²), essenciais para a sua função biológica (YOUNG, 2001; THOMAS e BRADY, 1997). Este domínio é o local de ação dos inibidores da HIV-IN (GOLDGUR *et al.*, 1999).

Os resíduos 213 a 288 formam o domínio de ligação C-terminal (CTD) da HIV-IN, que aparenta ter a função de ligar e orientar o DNA proviral durante o processo de integração. Já os resíduos 1 a 55 formam o domínio N-terminal (NTD). Essa porção contém um par de resíduos de histidina (Hys) e cisteína (Cys) que formam o domínio de ligação HHCC modificado, responsável por coordenar átomos de zinco que seriam importantes, porém não essenciais, para a atividade biológica. Apesar de sua função biológica ainda não ter sido determinada com exatidão, o mais aceito é que esse domínio interaja com o principal, formando um multímero essencial para as reações 3'P e ST. É também sabido que a ligação com o zinco metálico auxilia a manter essa estrutura. Estudos *in vitro* mostraram que, apesar da atividade catalítica ocorrer devido a ação do domínio principal, tanto o CTD quanto o NTD são necessários para ambas as reações (ZHENG, JENKINS e CRAIGIE, 1997; YOUNG, 2001; ZEINALIPOUR-LOIZIDOU *et al.*, 2007).

Durante sua ação biológica, a HIV-IN forma um complexo estável com regiões de repetição terminal longa (RTL) de sequências específicas do cDNA. Em seguida, ocorre o processamento endonucleolítico da terminação 3' (a reação 3'P) de ambas fitas. Neste passo, a sequência dinucleotídeo terminal GT na sequência 5'-ACTG---CAGT-3' é clivada e forma duas terminações ---CA-3'-OH. Origina-se então um DNA dupla-fita onde ambas as terminações ---CA-5' encontram-se desemparelhadas (YOUNG, 2001; TAKTAKISHVILLI *et al.*, 2000; GUPTA e NAGAPPA, 2003).

O complexo de pré-integração (CPI), formado pelo DNA proviral 3' processado e pela HIV-IN, é transportado através da membrana nuclear. No interior do núcleo ocorre a reação ST, e a HIV-IN exerce sua função catalítica final, cortando o DNA hospedeiro e ligando

covalentemente as terminações ---CA-3'-OH do DNA proviral a uma terminação fosfato-5' formada, dando origem a um produto de reação aberto. Em seguida, ocorre uma reação de ligação da terminação ---CA-5', com a remoção de dois nucleotídeos desemparelhados da terminação 5' do DNA proviral e a reparação dos espaços entre as sequências virais e do hospedeiro. Após a reparação, inicia-se a produção de novas partículas virais (YOUNG, 2001; TAKTAKISHVILLI *et al.*, 2000; GUPTA e NAGAPPA, 2003).

Acredita-se que a integrase dissocia-se por si mesma do DNA e que o reparo da falha entre a terminação 5' do DNA proviral e a terminação 3' do DNA hospedeiro ocorre pela ação de nucleotídeo polimerases endógenas do hospedeiro, de forma semelhante aos reparos realizados no DNA quando este é danificado e que podem levar a morte celular caso não ocorram. O mecanismo biológico pelo qual se processa este reparo ainda não foi determinado com exatidão. Uma possibilidade é a ATM quinase (ou sistema ataxia-telangiectasia-mutante dependente de dano ao DNA). Lau e colaboradores (2005) demonstraram que o inibidor KU-55933 (Figura 8), um potente inibidor seletivo da ATM quinase, foi capaz de impedir a replicação do HIV-1 em linfócitos T humanos. Isto prova que, de alguma forma, essa enzima celular é importante para o ciclo de replicação viral (YOUNG, 2001; THOMAS e BRADY, 1997).

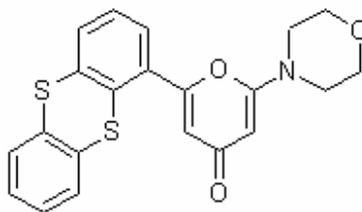


Figura 8. Estrutura do KU-55933.

Uma das vantagens de se utilizar a HIV-IN como alvo de novos agentes antiretrovirais é que esta não apresenta qualquer homólogo celular. Isto pode, em teoria, favorecer o desenvolvimento de fármacos com menor índice de efeitos colaterais (DAYAM e NEAMATI, 2003). Outra possível vantagem é que a HIV-IN apresenta um menor potencial para o desenvolvimento de resistência, já que essas enzimas são altamente conservadas nos retrovírus e utilizam um único sítio de ligação para acomodar duas diferentes configurações dos substratos de DNA (DNA hospedeiro e cDNA) (POMMIER, MARCHAND e NEAMATI, 2000). Mas, a falta de um equivalente celular também pode ser considerada uma desvantagem: sem um mediador endógeno, não existe estrutura química que possa ser utilizada como base para o desenvolvimento de um inibidor. As poucas pesquisas realizadas com retrovírus em anos anteriores à descoberta do HIV também contribuíram para o atraso nas pesquisas de inibidores da HIV-IN (POMMIER, MARCHAND e NEAMATI, 2000).

Atualmente, estão disponíveis* no *Protein Data Bank* (PDB; BERMAN *et al.*, 2000; <http://www.pdb.org>) 65 arquivos de estruturas cristalográficas de integrases, sendo 25 de HIV-1 IN e duas de HIV-2 IN. Porém, há apenas três complexos ligante/HIV-IN: 1QS4 (complexo 5CITEP/HIV-1 IN) (GOLDGUR *et al.*, 1999), e os complexos 1HYV e 1HYZ (ambos tetrafenil-arsênio, porém o segundo encontra-se incompleto) (MOLTENI *et al.*, 2001).

Apenas o complexo 1QS4 tem sido realmente utilizado em estudos, já que é o único que apresenta um ligante co-cristalizado na região da tríade catalítica. Além disso, não estão disponíveis complexos experimentais HIV-IN/substratos de DNA. Modelos teóricos, como o PDB 1WKN (De LUCA *et al.*, 2005), procuram fornecer informações sobre como ocorre a ligação do material genético com a HIV-IN (Figura 9) e que possam ser utilizadas no desenvolvimento de novos protótipos ou fármacos anti HIV-IN.

* Pesquisa realizada em 3 de abril de 2009 no site www.pdb.org utilizando a opção *advanced search* → *molecule* → palavras-chave (integrase).

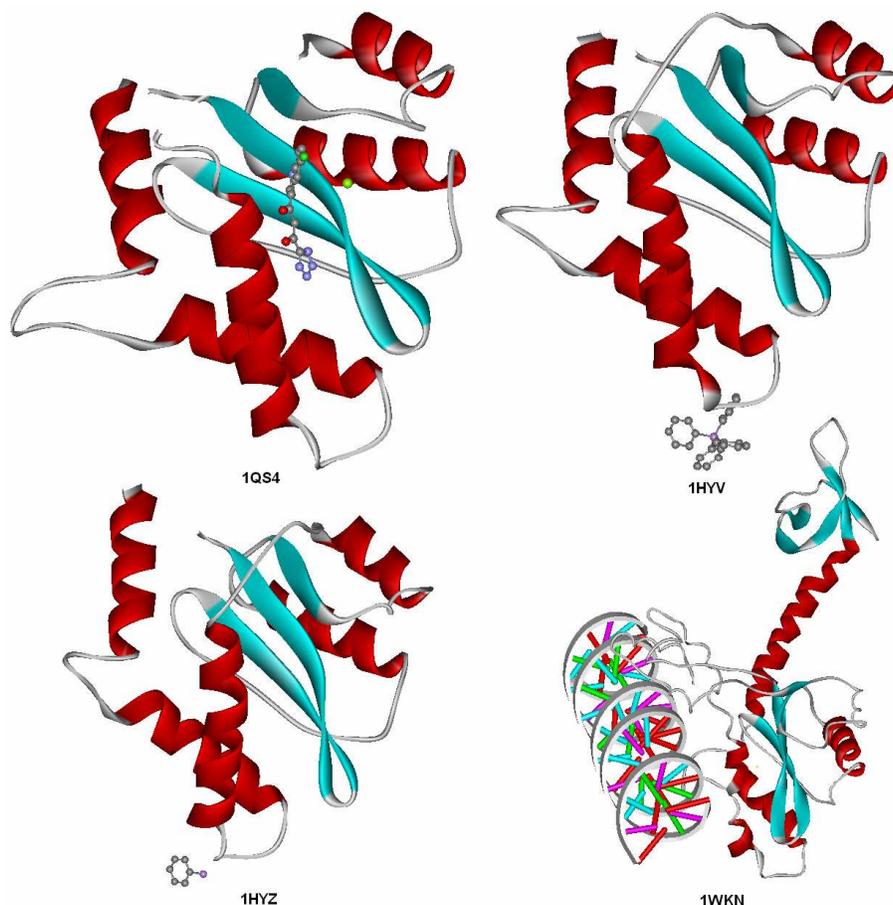


Figura 9. Representação das estruturas secundárias das cadeias A dos complexos cristalográficos PDB 1QS4, 1HYV, 1HYZ e do modelo teórico PDB 1WKN. Proteínas coloridas de acordo com suas cadeias secundárias. Os ligantes são apresentados como “bolas-e-bastões”.

Os esforços para a obtenção de inibidores da HIV-IN são focados em dois pontos: busca, isolamento e modificação de compostos naturais bioativos (Figura 10) (EICH *et al.*, 1996; SINGH, FELOCK e HAZUDA, 2000; SINGH *et al.*, 1999, 2001, 2002, 2005; JAYASURIYA *et al.*, 2003; MITCHELL, 2000; KIM *et al.*, 1998; HERALTH *et al.*, 2004; SHIOMI *et al.*, 2005; ONDEYKA, 2003) e em estratégias de planejamento de fármacos, auxiliadas por computador.

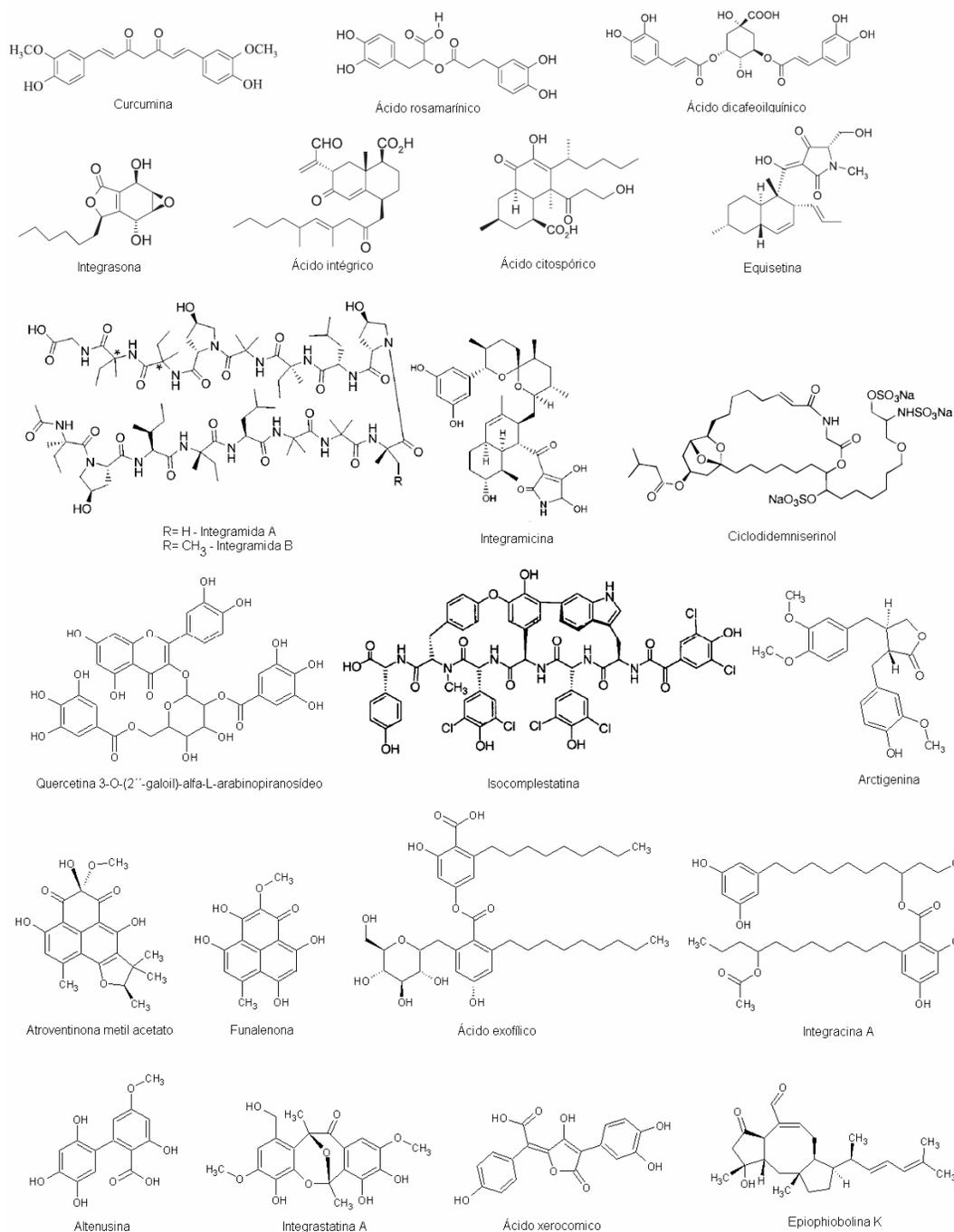


Figura 10. Exemplos de compostos orgânicos isolados de fontes naturais que apresentaram a capacidade de inibir a HIV-IN.

1.6 COMPOSTOS PROTÓTIPOS E PLANEJAMENTO DE FÁRMACOS AUXILIADOS POR COMPUTADOR

O planejamento racional de um novo fármaco depende da disponibilidade de informações sobre o mecanismo pelo qual se processa o efeito biológico que sofrerá intervenção. Portanto é necessário que os *alvos terapêuticos*, as biomacromoléculas envolvidas na função biológica, sejam conhecidos. Estas informações, cruciais para o processo de planejamento racional, normalmente são obtidas por meio de estudos básicos (BARREIRO, 2009).

Quando a estrutura do alvo terapêutico não é conhecida, o planejamento de novos fármacos pode ser conduzido por outros meios, como a partir da estrutura química do substrato natural envolvido na ação biológica, ou por meio de estudos como relação estrutura-atividade de séries congêneres que apresentam a atividade biológica/terapêutica em estudo (BARREIRO, 2009).

Diversas estratégias podem ser utilizadas para o planejamento molecular de um novo candidato a fármaco. De modo geral, este processo inicia-se com o isolamento ou síntese de um *composto protótipo*, que pode ser definido como aquele composto que originalmente apresenta uma determinada atividade farmacológica de interesse (BARREIRO e FRAGA, 2001). Em seguida, utilizando técnicas experimentais de modificações moleculares (bioisosterismo, simplificação, hibridação molecular, etc), este composto pode ser utilizado como base para a síntese de derivados otimizados (p.ex., mais potentes, mais seletivos, menos tóxicos, ou com comportamento farmacocinético aperfeiçoado). Assim, um novo fármaco pode ser obtido (PATANI e LAVOIE, 1996; BARREIRO *et al.*, 2001; BARREIRO, 2002; CHUNG *et al.*, 2005).

A obtenção de um composto protótipo e o desenvolvimento de novos fármacos envolvem estudos altamente complexos e interdisciplinares. Os químicos medicinais por muito tempo trabalharam apenas na realização de modificações sistemáticas de um protótipo, utilizando, principalmente, a facilidade sintética, a experiência, a intuição e o acaso. Isto mudou de forma considerável nos últimos 40 anos. Embora as técnicas clássicas de síntese e ensaio biológico ainda sejam muito importantes, grandes contribuições vieram de pesquisas bioestruturais, incluindo cristalografia de proteínas, bio-RMN multidimensional e modelagem molecular. As Químicas Mediciniais de “antes” e a “agora” estão resumidas na Tabela 1.

O planejamento de fármacos auxiliado por computador (CADD: *computer-aided drug design*) envolve todas as técnicas computacionais utilizadas para descobrir, planejar e aperfeiçoar compostos biologicamente ativos com a possibilidade de serem utilizados como fármacos (SANT’ANNA, 2002). Algumas técnicas também podem auxiliar na elucidação de informações estruturais sobre o sítio ativo (WATERBEEND, 1994), como a modelagem de proteínas por homologia. Essa técnica utiliza uma proteína homóloga com estrutura tridimensional conhecida como molde na construção de modelos de proteínas-alvo que têm apenas a estrutura primária (sequencia de aminoácidos) conhecida (SANTOS FILHO e ALENCASTRO, 2002; SILVA e SILVA, 2007).

Outra metodologia que pode ser utilizada com este objetivo, mas de forma indireta, são os estudos de *abordagem farmacofórica*. O termo grupo farmacofórico refere-se ao arranjo tridimensional de átomos ou grupos funcionais que é necessário para que um composto (ou conjunto de compostos) ligue-se a uma enzima ou a um receptor específico e promova determinado efeito biológico. Os modelos criados por esta abordagem podem ser utilizados para pesquisa em quimiotecas de compostos ou para o direcionamento de estudos de síntese orgânica (HONG *et al.*, 1997; MILNE *at al.*, 1994; WATERBEEND, 1994; BARRECA *at al.*,

Tabela 1. As Químicas Mediciniais de “antes” e “agora”. Tabela baseada em Lombardino e Lowe III (2004).

Antes (1950-1980)	Agora (1980-atualmente)
Técnicas limitadas para a síntese e isolamento de novos compostos	Desenvolvimento da Química Combinatória, permitindo a obtenção de muitos protótipos em pouco tempo; avanço da tecnologia (RMN, I.V., CG, espectrometria de massa, HPLC, cristalografia de raios-X, etc)
Impossibilidade de testes biológicos <i>in vitro</i> ; estes realizados principalmente <i>in vivo</i>	Possibilidade de avaliação <i>in vitro</i> – identificação, clonagem e produção de alvos moleculares (receptores) em grande quantidade; desenvolvimento do HTS (<i>High Throughput Screening</i>), permitindo o ensaio biológico <i>in vitro</i> de grande quantidade de substância em pouco tempo
Necessidade de síntese ou isolamento de grande quantidade de substâncias-teste	Necessidade de pequenas quantidades da substância em estudo para teste de atividade biológica (microescala)
Informática pouco desenvolvida – armazenamento e consulta manual de dados	Avanço da informática como um todo
Tentativa e erro – intuição química; estudos relação estrutura-atividade experimentais	Avanço da química computacional; surgimento e aperfeiçoamento do CADD; desenvolvimento de fármacos baseado na estrutura do receptor (<i>structure-based drug design</i>)
Exigências regulatórias brandas; <i>tragédia da talidomida</i> .	Aumento das exigências regulatórias e do tempo de lançamento de um novo fármaco

2006). Finalmente, diversas abordagens baseadas apenas na estrutura de ligantes podem ser utilizadas quando não se conhece o alvo terapêutico, tendo como base apenas as variações estruturais apresentadas pelo conjunto de treinamento em estudo e as atividades/potências biológicas determinadas experimentalmente (WATERBEEND, 1994).

Abordagens de CADD vêm sendo utilizadas há anos para pesquisar os vários aspectos necessários para o planejamento racional de inibidores da HIV-IN. Por exemplo, estão disponíveis na literatura muitos estudos sobre abordagem farmacofórica de inibidores de HIV-

IN. Podem ser apresentados como exemplos o modelo dinâmico de Carlson e colaboradores (2000), os estudos baseados no S-1360 (primeiro inibidor submetido a estudos clínicos) de Dayam e colaboradores (2005), e o estudo de Barreca e colaboradores (2005), utilizado para síntese (Figura 11).

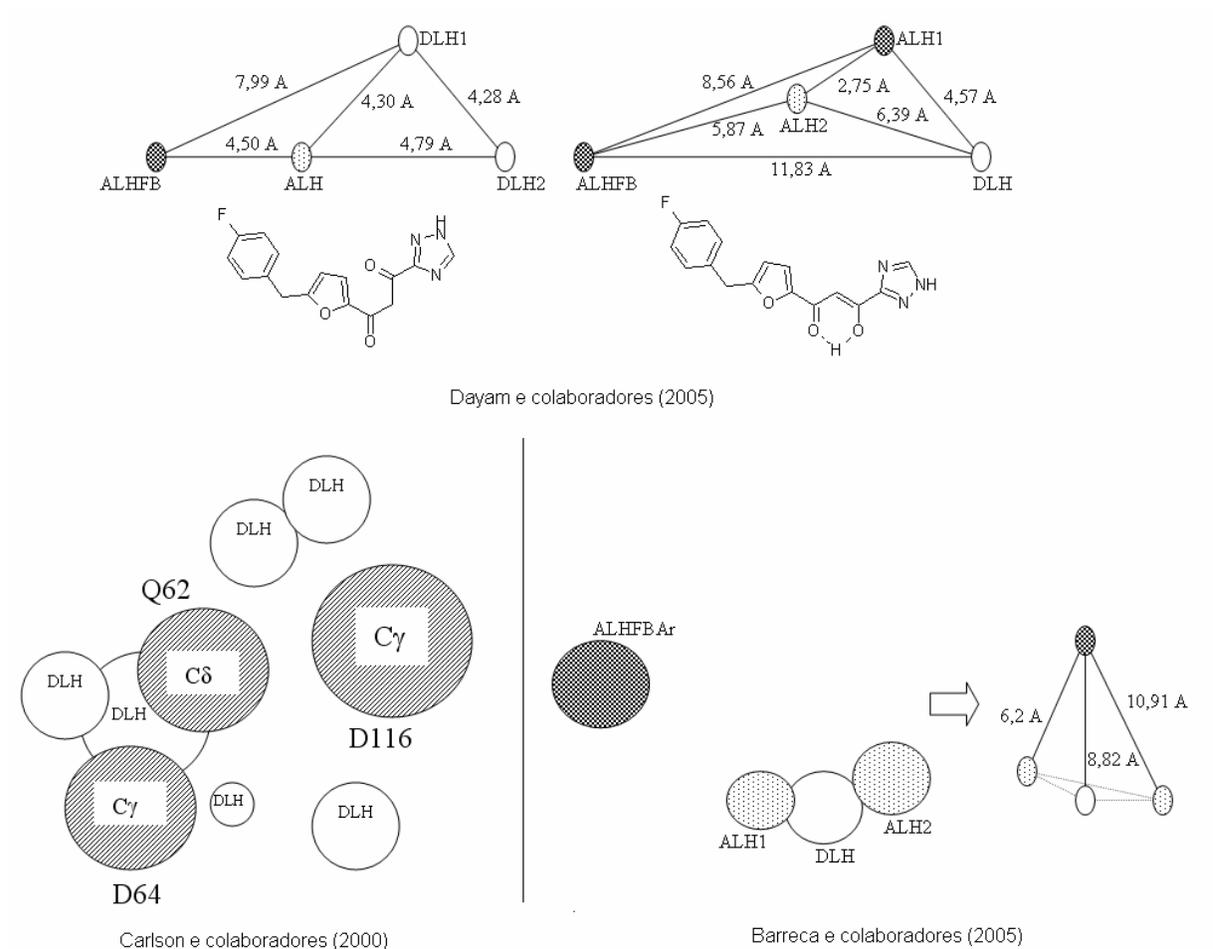


Figura 11. Esquemas ilustrativos dos modelos farmacofóricos tridimensionais, construídos com base nos trabalhos de Carlson e colaboradores (2000), Dayam e colaboradores (2005) e Barreca e colaboradores (2005). DLH: doador de ligação hidrogênio; ALH: acceptor de ligação hidrogênio; ALHFBAR: acceptor de ligação hidrofóbica aromática; ALHF: acceptor de ligação hidrofóbica.

As pesquisas de CADD sobre inibidores de HIV-IN evoluíram após a cristalização e depósito no PDB do complexo 1QS4 (Figura 9) (GOLDGUR *et al.*, 1997). Além de ter sido muito útil para confirmar a região onde se localiza a tríade catalítica, este complexo ajudou a

umentar a compreensão sobre a forma como esta enzima age e sobre o possível mecanismo de ação.

Utilizando o 1QS4, Sotriffeer, Ni e McCammon (2000) puderam levantar a hipótese de que o sítio ativo formado pela tríade catalítica não era o único ponto essencial para a atividade biológica da HIV-IN. Já Dayam e Neamati (2003) utilizaram o 1QS4 em técnicas de *docking* para identificar a conformação ativa e possíveis orientações de compostos protótipos no sítio de ligação, melhorando a compreensão das interações moleculares entre a HIV-IN e seus inibidores. Assim, Long e colaboradores (2004) puderam identificar uma “trincheira” localizada no laço desordenado formado pela sequência de aminoácidos Ile141 – Asn144, como um sítio de ligação.

No intuito de determinar quais descritores químicos são importantes para o planejamento de novos protótipos baseado apenas na estrutura de ligantes e aumentar a compreensão sobre o mecanismo de inibição, alguns estudos QSAR foram realizados com diversas classes de compostos (YUAN e PARRIL, 2000, 2002, 2005; RAGHAVAN *et al.*, 1995; MAKHIJA e KULKARNI, 2001; MAKHIJA e KULKARNI, 2002a; BUOLAMWINI e ASSEFA, 2002; DAYAM e NEAMATI, 2003; DAEYAERT *et al.*, 2004; SAÍZ-URRA *et al.*, 2007).

Por exemplo, Raghavan e colaboradores (1995) realizaram um estudo QSAR-3D com um grupo de 15 flavonas que inibem as reações 3’P e ST *in vitro* e demonstraram que a capacidade desses compostos inibirem a HIV-IN é 79% relacionada a propriedades eletrônicas e 21% a propriedades estéricas. Yuan e Parril (2000; 2002; 2005) realizaram uma série de estudos de SAR/QSAR com metodologias e classes estruturalmente diversas de inibidores na tentativa de encontrarem características estruturais importantes para os inibidores da HIV-IN,

relacionando as informações encontradas com o(s) possível(is) sítio(s) de ligação de cada classe.

Os estudos acima citados propuseram a existência de, pelo menos, dois sítios de interação para os inibidores da HIV-IN. Daeyaert e colaboradores (2004) realizaram um estudo QSAR-3D com base em 201 ésteres de ketoamidas com a metodologia CoMSIA (*comparative molecular similarity indices analysis*), e observaram que o melhor modelo apresentava contribuições relacionadas principalmente à formação de interações ligante-receptor dependentes da polarizabilidade e das cargas atômicas. Já Saíz-Urra e colaboradores (2007), em um estudo QSAR-2D apenas com descritores GETAWAY (CONSONNI, TODESCHINI E PAVAN, 2002), obtiveram modelos pesados por parâmetros relacionados à distribuição eletrônica (polarizabilidade e eletronegatividade), e os relacionaram com interações cátion- π ou carga-carga com os íons metálicos localizados no sítio ativo da enzima.

Uma das principais desvantagens dos métodos de CADD reside no fato da química ser uma ciência essencialmente experimental. Assim, os métodos *in silico* somente logram sucesso se forem baseados em dados experimentais, previamente obtidos, (como a atividade biológica de um conjunto de compostos previamente sintetizados ou a identificação e determinação da estrutura tridimensional de um alvo terapêutico) e sobre certas condições (como uma atividade biológica relacionada com propriedades químicas, a qual seja medida por um protocolo em comum e a utilização de conjuntos de treinamento com um grau mínimo de homogeneidade estrutural) (WATERBEEND, 1994).

1.7 ESTUDOS DAS RELAÇÕES ESTRUTURA-ATIVIDADE QUANTITATIVAS E QUALITATIVAS

Um estudo de relação estrutura-atividade quantitativa (QSAR: *quantitative structure-activity relationship*), uma das abordagens mais utilizadas em CADD, baseia-se na hipótese de que o comportamento de uma classe de compostos análogos em um sistema biológico *in vitro* ou *in vivo* pode ser quantitativamente descrito por modelos matemáticos multiparamétricos (TAVARES, 2004; DAYAM e NEAMATI, 2003; FERREIRA, MONTANARI E GAUDIO, 2002). Estes modelos devem ser capazes de explicar as complexas relações entre as variáveis independentes e dependentes em estudo, no caso descritores que codificam aspectos estruturais e propriedades físico-químicas de interesse, e uma ou mais atividades biológicas (ABs) de uma classe de compostos, respectivamente. Assim, as equações obtidas podem ser utilizadas para realizar a predição das ABs de outros compostos, assim como uma atividade toxicológica ou uma propriedade de interesse (JURS, DIXON e EGOLF, 1994; FERREIRA, MONTANARI e GAUDIO, 2002; ERIKSSON *et al.*, 2003; CARBÓ-DORCA e GIRONÉS, 2004). Esta idéia pode ser resumida pela Equação 1.

$$AB = f(\text{estrutura molecular}) = f(\text{descriptor(es)}) \quad (1)$$

Onde: AB = IC₅₀, pIC₅₀, DL₅₀, pDL₅₀

Os descritores físico-químicos são as variáveis independentes de um modelo QSAR. Estes podem ser obtidos utilizando diversas abordagens teóricas, como a teoria de grafos, modelagem molecular, por meio de metodologias desenvolvidas com base em dados experimentais, ou mesmo diretamente, via estudos experimentais. Aqueles capazes de codificar propriedades diretamente relacionadas às forças intermoleculares envolvidas em processos de interações fármaco-receptor, no transporte e na distribuição de fármacos, apresentam importância em especial (KUBINYI, 1993). Na literatura, os descritores são

classificados das mais diversas formas. Para este trabalho, foi utilizada a classificação geral apresentada por Dunn III (1989) (Tabela 2).

Tabela 2. Classificação dos descritores físico-químicos que podem ser utilizados em QSAR, e em alguns exemplos.

Classe	Exemplos
Descritores físico-químicos	Constante eletrônica de Hammet (σ) Constante substituinte de campo de Swain Lupton (\mathcal{S}) Constante substituinte de ressonância Swain Lupton (\mathcal{R}) Constante hidrofóbica de Hansch (π) Constante estérica de Taft (E_s) Refratividade molar (RM)* Logarítimo do coeficiente de partição octanol/água (Log P) Ponto de fusão
Descritores eletrônicos (mecânico-quânticos)	Momento dipolar (D) Cargas atômicas parciais (q) Polarizabilidade molecular (α) Refratividade molar (RM)* Energia do orbital molecular ocupado mais alto (HOMO) Energia do orbital molecular desocupado mais baixo (LUMO) Cargas atômicas parciais Parâmetros termodinâmicos (calor de formação e outros)**
Descritores geométricos	Volume molecular (VM)*** Área molecular (AM) Comprimentos e ângulos de ligação Ovalidade (Ov)
Descritores topológicos	Índice de Wiener (W) Índice de Randic (R) Índice de Balaban (B) Índices de conectividade (χ^y) Índices de forma Kappa (κ)
Descritores constitucionais	Massa molar (MM)*** Número de elétrons de valência Número de átomos (nAT) Número de ligações (nBT)
Combinados	Índices de estado eletrotológico (<i>E-state indices</i>) Densidade de carga superficial

*a refratividade molar é, em muitos trabalhos, definida como um descritor misto estérico/eletrônico; ** são considerados eletrônicos, pois são obtidos por meio de cálculos de orbital molecular, utilizando programas de modelagem molecular; ***descritores como área, volume e relacionados são comumente chamados de *estéricos*, pois, relacionam-se com o grau de importância do tamanho de uma molécula com a atividade biológica. Alguns trabalhos também fazem essa relação com a massa molar.

Já as variáveis dependentes que podem ser modeladas em QSAR são as atividades biológicas obtidas por meio de estudos farmacológicos experimentais. As mais comumente utilizadas IC_{50} (dose necessária para inibir ou estimular um determinado processo biológico ou atividade enzimática em 50%), DL_{50} (dose letal necessária para matar 50% das cobaias, células, etc), K_i (coeficiente de inibição enzimática), e MIC (concentração inibitória mínima).

Geralmente, dados de atividades biológicas não possuem uma distribuição normal, podendo apresentar faixas de variação numéricas largas (p.ex., entre 0,1 e 100.000) e concentradas em determinados intervalos. Isto é normalmente resolvido transformando cada valor em seu logaritmo inverso ($-\text{Log } IC_{50}$ ou pIC_{50}) (KUBINYI, 1993; van de WATERBEEMD, 1996; NANTASENAMAT *et al.*, 2009).

Dentre as diversas abordagens matemáticas que podem ser utilizadas em estudos QSAR, as mais utilizadas são a regressão linear múltipla, os métodos de projeção, e métodos não-lineares baseados em redes neurais artificiais.

1.7.1 Regressão linear múltipla (MLR)

Uma das grandes contribuições nesta área foi feita por Corvin Hansch e seus colaboradores, os quais reconheceram que a explicação da potência biológica pode requerer o uso de mais de uma propriedade química. A partir disso originou-se a denominada ‘equação’ ou ‘abordagem de Hansch’ (Equação 2) (LIVINGSTONE, 2003).

$$\text{Log } (1/C) = a\pi + b\pi^2 + c\sigma + dEs + \text{constante} \quad (2)$$

Onde: $\text{Log } (1/C) = -\text{Log } C$; $C = IC_{50}$

Os termos π , σ e Es (Tabela 2) correspondem a valores tabelados para substituintes específicos de uma estrutura básica. Em termos matemáticos, tal abordagem utiliza nada

menos que a regressão linear múltipla (MLR: *multiple linear regression*) (LIVINGSTONE, 2003). Nessa abordagem, as variáveis preditoras devem ser matematicamente independentes (ou ortogonais), pois uma das limitações da MLR é sua sensibilidade a descritores altamente correlacionados. Com a solução desse problema, essa metodologia fornece resultados satisfatórios em QSAR (ERIKSSON, 2003).

De uma maneira geral, a MLR é geralmente utilizada para ajustar um modelo de regressão linear onde a variável de resposta representada por y (AB) é uma combinação linear de m variáveis (descritores), com coeficientes b . O desvio entre y e o modelo é denominado resíduo (e) (ERIKSSON, 2003). Assim, uma equação genérica de MLR pode ser representada como a equação 3:

$$y = x_1b_1 + x_2b_2 + \dots + x_mb_m + e \quad (3)$$

1.7.2 Métodos de projeção: PCR e PLS

Atualmente são muito utilizados em QSAR os denominados métodos de projeção multivariada. Esses métodos de regressão linear são baseados nos princípios aplicados na chamada análise de componentes principais (PCA: *principal component analysis*), que é uma transformação matemática de um conjunto de dados em um novo conjunto de dados z dimensional com um número de variáveis independentes muito menor que o original ($\ll z$). Cada nova variável é obtida por combinações lineares das variáveis originais (LIVINGSTONE, 2003). Uma das grandes vantagens é que as novas variáveis são ortogonais entre si, o que é uma solução para o problema da colinearidade em MLR (LIVINGSTONE, 2003).

Dois métodos de regressão utilizam esse princípio: regressão por componentes principais (PCR: *principal components regression*) e por quadrados mínimos parciais (PLS: *partial least squares*). No primeiro, as novas variáveis são chamadas componentes principais (PC: *principal components*), e são formadas por combinações lineares das variáveis originais. No segundo, são denominadas variáveis latentes (LV: *latent variables*). Enquanto no método PCR as PCs são construídas primeiro, e apenas depois é realizada a regressão perante a variável dependente y , no método PLS as LVs são obtidas utilizando-se y no processo de decomposição. Assim, espera-se que a covariância entre as coordenadas das amostras no novo sistema de eixos x e y seja maximizada, com a obtenção de resultados superiores àqueles que seriam obtidos com PCR (FERREIRA e KIRALJ, 2008; ROY e ROY, 2008).

1.7.3 Métodos não lineares: ANN.

Enquanto MLR, PCR e PLS são indicados para a construção de modelo QSAR lineares, outros métodos podem ser utilizados quando se desejam estudar relações quantitativas não-lineares. Um método muito popular, nestes casos, são as redes neurais artificiais (ANN: *artificial neural networks*). Esta metodologia utiliza pesquisas de inteligência artificial (IA) que buscam replicar a estrutura do cérebro, emulando suas funções. Assim, uma ANN consiste em camadas de neurônios artificiais conectados um aos outros (LIVINGSTONE, 2003).

As camadas consistem em uma camada de entrada, uma de saída e algumas intermediárias, denominadas camadas intermediárias (*hidden layers*) (Figura 12). Cada unidade na rede é influenciada por aquelas unidades as quais estão conectadas, logo, o grau de influência é determinado pelos valores das ligações ou conexões. O comportamento geral do sistema pode ser modificado pelo ajuste de valores das conexões ou pesos, através de uma

aplicação repetida de um algoritmo de aprendizagem (como o *back-propagation*) (YU, YI e WANG, 2008).

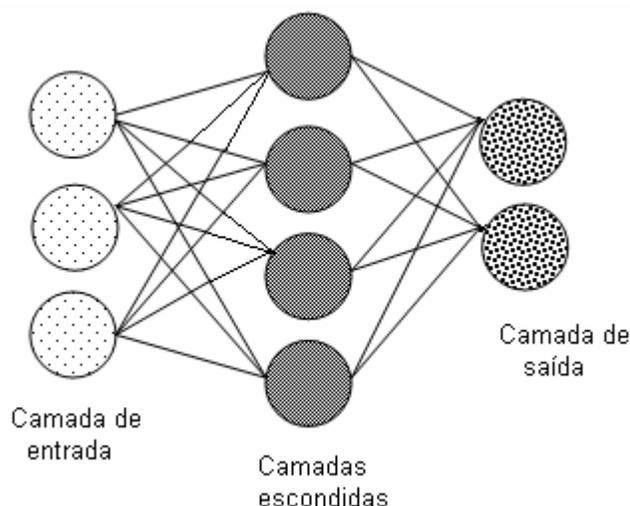


Figura 12. Esquema básico clássico da organização de um algoritmo ANN, baseado em Livingstone (2003).

Porém, ao contrário dos métodos lineares tradicionais utilizados em estatística, em ANN não existe um método para determinar automaticamente a rede neural ótima para o ajuste de um conjunto de dados. Assim, normalmente os algoritmos são utilizados durante certo número de vezes até a seleção da melhor (ou melhores) rede (AGATONOVIC-KUSTRIN, TURNER e GLASSA, 2008).

1.7.4 Outras metodologias em QSAR

Estudos QSAR não se limitam apenas à utilização de metodologias de regressão para a construção de modelos com variáveis independentes que codificam propriedades físico-químicas. Enquanto estas metodologias podem ser denominadas de QSAR-2D, há outras que podem ser classificadas como QSAR-3D a 6D. Todas podem utilizar os mesmos métodos de

regressão anteriormante citados para a obtenção dos modelos de predição. A diferença encontra-se nos tipos de descritores e nas ‘filosofias’ utilizadas.

A principal abordagem de QSAR-3D é conhecida como CoMFA (*Comparative Molecular Field Analysis*) (CRAMER, 1988). Nesta metodologia, assume-se que a atividade biológica é decorrente de interações ligante-receptor, relacionadas com forças estéricas e eletrostáticas. A estrutura tridimensional de uma molécula é colocada dentro de uma caixa virtual que representa um receptor hipotético rígido, onde uma ‘sonda’ (*probe atom*) simula determinados tipos de átomos, avaliando as propriedades eletrostáticas e estéricas do composto, e traduzindo-as em energias de interações em cada ponto da grade que forma a caixa virtual. Para medir as energias relacionadas a possíveis interações de van der Waals, é utilizada a função de Lennard-Jones (LJ, Equação 5), e para as interações entre grupos carregados e polarizados, é utilizada a função de Coulomb (C, Equação 6). Em seguida, utilizando o método de regressão PLS (devido à quantidade de descritores que são obtidos – na casa dos milhares), o modelo de predição é construído.

Os métodos QSAR-3D baseiam-se na hipótese de que os descritores importantes para a atividade biológica podem ser obtidos de uma única conformação. Porém, caso a conformação utilizada não seja aquela que representa a molécula em seu estado ligado ao receptor (ou *conformação bioativa*), o estudo pode ser prejudicado (MASUNARI e TAVARES, 2007). Visando solucionar esta limitação, Hopfinger e colaboradores (1997) propuseram a abordagem QSAR-4D, que insere uma ‘quarta dimensão’ ao estudo: liberdade conformacional. Isto é obtido através do perfil de amostragem conformacional (PAC), obtido através de dinâmica molecular. Assim, é possível avaliar o quanto a flexibilidade do conjunto de treinamento pode ser importante para a atividade biológica (Figura 12) (HOPFINGER *et al.*, 1997; VEDANI e DOBLER, 2002).

$$E_{vdW} = \sum_{i=1}^{n_c} \sum_{j=1}^{n_A} \left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \right) \longrightarrow C_{ij}^{(12)} = \left(\frac{1}{n_c} - C_{ii}^{(12)} \times C_{jj}^{(12)} \right)^{\frac{1}{2}} \quad (5)$$

$$C_{ij}^{(6)} = \left(\frac{1}{n_c} - C_{ii}^{(6)} \times C_{jj}^{(6)} \right)^{\frac{1}{2}}$$

$$E_c = \sum_{i=1}^{n_c} \sum_{j=1}^{n_A} \frac{q_i q_j}{4\pi\epsilon_0 n_c r_{ij}} \quad (6)$$

Onde: r_{ij} = distância entre a sonda i e o átomo j ;
 ϵ_0 = permissividade no vácuo;
 n_A = número de átomos em cada conformação do PAC;
 n_c = número de conformações de cada PAC;
 q_i = carga da i -ésima sonda;
 q_j = carga do j -ésimo átomo;
 C_{ii} = parâmetros de Lennard-Jones para a sonda i ;
 C_{jj} = parâmetros de Lennard-Jones para o átomo j ;
 $C_{ij}^{(12)} / r_{ij}^{12}$ = termo que descreve a repulsão de van der Waals entre a sonda i e o átomo j ;
 $C_{ij}^{(6)} / r_{ij}^6$ = termo que descreve a atração de van der Waals entre a sonda i e o átomo j .

As abordagens QSAR-5D e QSAR-6D foram propostas por Vedani e colaboradores (2005a; 2005b; 2007). A primeira introduz como quinta dimensão o encaixe induzido, considerando as diferentes adaptações que um sítio de ligação pode sofrer devido à topologia de uma pequena molécula que nele se encaixa (VEDANI *et al.*, 2005a). Já o QSAR-6D é uma complementação do 5D, onde o encaixe induzido pode ser realizado em diferentes tipos de meios solventes (VEDANI, DOBLER e LILL, 2005b; VEDANI *et al.*, 2007).

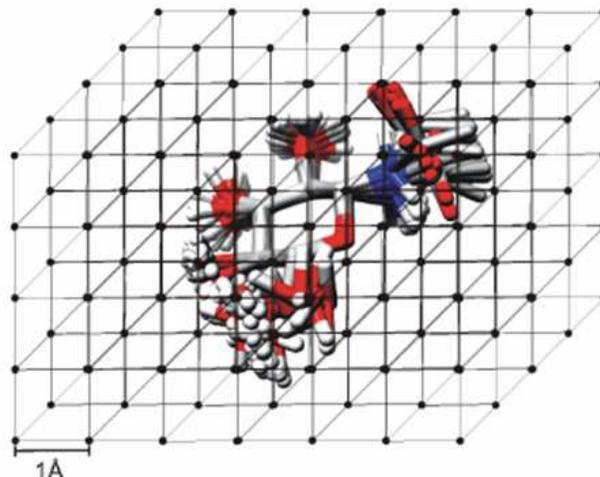


Figura 13. Representação de um perfil de amostragem conformacional, formado dentro de uma grade virtual. Nos processos de QSAR-3D, é utilizada uma grade semelhante. Figura cedida por Euzébio Guimarães Barbosa e João Paulo Athaide Martins.

Em estudos QSAR, a atividade biológica de um inibidor é decorrente de interações com sítios específicos de estruturas protéicas. Para que isto ocorra, um composto deve apresentar o farmacóforo necessário para a atividade. Apesar de ser possível a utilização de conjuntos heterogêneos em estudos QSAR e, especialmente, da relação estrutura-propriedade quantitativa (QSPR: *quantitative structure-property relationships*), descritores derivados de conjuntos de treinamento que apresentam grande variabilidade estrutural podem não explicar de modo adequado as relações entre a estrutura e a atividade em estudo (CONSONNI, TODESCHINI e PAVAN, 2002; van DRIE, 2007). Uma das vantagens dos métodos QSAR-3D a 6D é que um modelo pode ser traduzido em modelos farmacofóricos tridimensionais, mesmo que os compostos em estudo não possuam um mesmo esqueleto molecular. Isto permite o estudo de conjuntos de treinamento com razoável grau de heterogeneidade estrutural (MASUNARI e TAVARES, 2007; VEDANI e DOBLER, 2002; SANT'ANNA, 2002; RAGHAVAN, 1995; DEBNATH, 2005; KUBINYI, 1997).

1.7.5 Validação de modelos QSAR

Após a obtenção de um modelo QSAR, o mesmo deve ser validado antes de sua interpretação e utilização para que atividades biológicas ou outras características de interesse sejam previstas (WOLD e ERIKSSON, 1998). Segundo o guia da Organização para a Cooperação e Desenvolvimento Econômico (OECD: *Organisation for Economic Co-operation and Development*) para estudos QSAR, a validação pode ser definida como “o processo pelo qual a confiabilidade e a relevância de uma abordagem, método, processo ou avaliação específica, são estabelecidas com um objetivo definido” (OECD, 2007).

No processo de validação interna de um modelo avalia-se: (i) a capacidade do modelo ajustar os dados, ou a quantidade de variância explicada, utilizando para este fim a correlação linear univariada (correlação de Pierson, r) ou o coeficiente de determinação múltipla (R^2) com a atividade biológica em estudo, o desvio padrão do ajuste dos dados ou de calibração (SEC) e o grau de significância expresso pelo teste de Fisher (F) (KUBINYI, 1994); e (ii) a qualidade da predição interna, ou quantidade de variância predita, determinada pelo processo de validação cruzada (ou validação LOO, *leave-one-out*: ‘deixe-um-de-fora’). Neste teste, uma das amostras é retirada do modelo, o qual é reconstruído e utilizado para estimar o valor da amostra excluída. O resultado predito deve desviar o mínimo possível do valor experimental, e o resultado é avaliado através dos resultados obtidos para o coeficiente de correlação de validação cruzada LOO (Q^2_{LOO}) e para o erro padrão de validação (SEV : *standard error of validation*) (GÁUDIO e ZANDONADE, 2001; OECD, 2007).

Um processo que passou a ser considerado essencial para o desenvolvimento de um modelo QSAR é a *validação externa*, a utilização do modelo para prever a atividade de um conjunto de compostos conhecidos e não utilizados para a construção do modelo, cuja atividade de interesse foi avaliada pelo mesmo protocolo de ensaio farmacológico. Além do

menor desvio possível entre os valores observados e preditos, alguns critérios estatísticos podem ser utilizados para avaliar a predição externa, em especial coeficiente de correlação de predição (R^2_{pred} ou Q^2_{pred}) e o erro padrão de predição (*SEP: standard error of prediction*). Estes parâmetros possuem funções similares a R^2/Q^2_{LOO} e SEC/SEV , porém agora aplicados ao conjunto de validação externa (WOLD e ERIKSSON, 1998; GOLBRAIKH e TROPSHA, 2002; APTULA *et al.*, 2005). Gobraikh e Tropha (2002) e Gobraikh e colaboradores (2003) também sugerem a realização de dois testes para confirmar a qualidade preditiva externa: a avaliação das inclinações das retas, obtidas em uma regressão feita entre os valores observados e preditos (k) e entre os preditos e observados (k'), e a avaliação do valor absoluto da diferença entre os coeficientes de determinação centrados na origem das duas mesmas regressões (R^2_0 e R'^2_0). Os coeficientes de determinação podem ser obtidos realizando uma regressão simples entre os valores experimentais e preditos na validação externa, mas com a reta centrada arbitrariamente na origem dos dados.

Outra importante característica de um modelo é sua robustez, a capacidade de um método ou modelo não ser afetado por uma pequena e deliberada modificação em seus parâmetros (JORNADA e PIZZOLATO, 2007). Em QSAR, o método mais utilizado para a avaliação desse critério é a validação cruzada na qual mais de uma amostra é retirada (conhecida como *leave-N-out* - LNO, *leave-many-out* - LMO ou *leave-some-out* - LSO). O processo é semelhante ao da validação LOO: retira-se um número de compostos do modelo (o “ N ”), o qual é reconstruído sem os mesmos, e o novo modelo é utilizado para prever a atividade das amostras (FERREIRA e KIRALJ, 2009).

Outro teste importante para um modelo QSAR é a análise da possibilidade de que as variabilidades explicadas e preditas pelo modelo devem-se a ocorrências ao acaso (OECD, 2007). Na estatística, são descritos alguns casos evidentes deste fenômeno. Por exemplo, a alta

correlação entre o número de execuções de condenados a morte nos EUA e o crescimento populacional naquele país, e entre o número de cegonhas e o número de nascimentos na Inglaterra (ambos com $R^2 = 0,99$) (DOWEYKO, 2008). Porém, quando se tratam de correlações espúrias entre uma atividade biológica e descritores físico-químicos, esta possibilidade pode não ser tão aparente. Assim, é necessário analisar essa possibilidade por meio do teste de randomização de y (*y-randomization* ou *y-scrambling*) onde, enquanto a parte da matriz que contém os valores dos descritores físico-químicos permanece inalterada, a coluna y , que contém os valores das atividades, tem seus valores ‘misturados’. Após esse processo, espera-se que os modelos construídos nessas condições sejam de péssima qualidade, uma forte indicação de que os descritores que formam o modelo estão realmente relacionados com o mecanismo de ação (FERREIRA e KIRALJ, 2009; ZHANG *et al.*, 2007; RÜCKER, RÜCKER e MERINGER, 2007).

1.8 MODELAGEM MOLECULAR.

A elucidação dos mecanismos de ação dos fármacos pode ser compreendida pela integração de conhecimentos fundamentais de Química Orgânica, Bioquímica, Biologia Molecular e Farmacologia (CARVALHO *et al.*, 2003). A modelagem molecular compreende um grande conjunto de métodos computacionais que auxiliam nesta compreensão, tornando mais eficiente o processo de planejamento racional de compostos bioativos (SANT’ANNA, 2009).

Segundo a Iupac (União Internacional de Química Pura e Aplicada), Modelagem Molecular pode ser definida como a investigação de estruturas e propriedades moleculares pelo uso de Química Computacional e técnicas de visualização gráfica, visando fornecer uma representação tridimensional, sob um dado conjunto de circunstâncias (SANT’ANNA, 2002).

Utilizando esta abordagem, podem ser obtidas propriedades moleculares específicas ou padrões de similaridade estrutural entre diferentes moléculas que, por sua vez, podem apresentar relação com uma determinada atividade biológica. Outra utilidade da modelagem molecular, como apresentado em sua própria definição, é a visualização tridimensional de moléculas, isoladas, na forma de complexos ligante-receptor, etc., permitindo a obtenção de informações sobre requisitos estruturais necessários para a interação de uma molécula bioativa com seu bioreceptor (RODRIGUES, 2001).

Através da modelagem molecular, é possível desenhar a estrutura tridimensional de uma molécula e realizar cálculos para a obtenção de geometrias de mínimas energias (RODRIGUES, 2001). Porém, a geração das estruturas é um dos problemas fundamentais na Química Computacional. Quando se trabalha com planejamento de fármacos, geralmente é recomendado utilizar a conformação bioativa, já que muitas propriedades de interesse podem variar em função da conformação adotada (SADOWSKI, SCHWAB e GASTEIGER, 2004).

Porém, normalmente são investigadas grandes quantidades de moléculas para as quais não estão disponíveis informações experimentais sobre sua conformação bioativa. Atualmente, o PDB é a principal fonte de estruturas tridimensionais de ligantes em conformações bioativas complexadas às suas macromoléculas alvo. Mesmo assim, a quantidade é muito pequena, e muitos desses arquivos podem apresentar erros, baixas resoluções, e outros problemas (SADOWSKI, SCHWAB e GASTEIGER, 2004)

Assim, é possível derivar estruturas tridimensionais a partir de bidimensionais, utilizando programas como o CORINA (disponível *on-line* em <http://www.molecular-networks.com>), ou desenhar diretamente a estrutura tridimensional, como pode ser feito no programa HyperChem (HYPERCUBE, 2002; <http://www.hyper.com>). Outro processo é o desenho a partir de estruturas de raio-X similares a estrutura desejada (SADOWSKI,

SCHWAB e GASTEIGER, 2004), que podem ser obtidas no PDB, ou em bancos específicos para moléculas pequenas, como o Cambridge Structural Database (CSD) (CAMBRIDGE, 2009; <http://www.ccdc.cam.ac.uk>).

Após a obtenção da estrutura tridimensional de uma molécula, é necessário obter a conformação bioativa. Como não há como saber qual é esta conformação sem estudos experimentais, geralmente é utilizada uma conformação de baixa energia (SADOWSKI, SCHWAB e GASTEIGER, 2004). Em geral utiliza-se a conformação que corresponde ao menor valor, ou *mínimo global*, de sua superfície multidimensional de energia molecular (ou *superfície de energia potencial*, SEP). Teoricamente, esta será a conformação em maior número durante o processo de interação com o receptor. Porém, quanto maior o número de ângulos diedros livres, mais difícil será de se obter o mínimo global. Outras regiões da SEP que apresentam baixos valores são os *mínimos locais* (Figura 14) (SADOWSKI, SCHWAB e GASTEIGER, 2004; RODRIGUES, 2001).

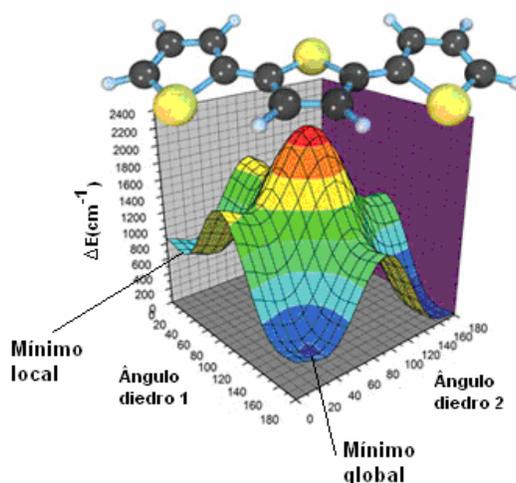


Figura 14. Representação da SEP de uma molécula com apenas dois ângulos torsionais livres. Exceto pelas indicações para mínimos global e local, a figura foi mantida como na fonte. Extraída e adaptada de <http://chem.syr.edu/faculty/kortner.html>. Acesso: 06 de abril de 2009.

As duas aproximações disponíveis para estudos de modelagem molecular são os métodos clássicos, que englobam a mecânica molecular e a dinâmica molecular, e os métodos quânticos, que englobam as teorias *ab initio*, semi-empíricas e a teoria do funcional de densidade (DFT: *density functional theory*). A escolha da abordagem depende das propriedades que se deseja estudar, da precisão desejada e da capacidade computacional disponível para a realização dos cálculos (SANT'ANA, 2009).

1.8.1 Métodos clássicos

A mecânica molecular é um método que calcula a estrutura e a energia das moléculas com base no movimento dos núcleos. Assim, as moléculas são descritas como um conjunto de “átomos conectados”, ao invés de núcleos e elétrons. Os núcleos são representados por “pontos”, e as ligações químicas por “molas” com constantes de força características. (SANT'ANNA, 2009; RODRIGUES, 2001; BARREIRO *et al.*, 1997). Assim, a energia é calculada apenas pela comparação entre ângulos e distâncias de ligação, ângulos diedros, distâncias de VDW, ligações hidrogênio, eletrostáticas, etc (CARVALHO *et al.*, 2003; BARREIRO *et al.*, 1997). Os parâmetros considerados dependerão do *campo de força* utilizado. Atualmente, muitos campos de força estão disponíveis, cada qual tendo diferentes objetivos.

Segundo Silva (2003), a *Equação de Westheimer* é a forma mais simples de representar um campo de força. Esta equação ($E = E_s + E_b + E_w + E_{nb}$) possui termos relacionados às energias envolvidas no estiramento ou compressão de uma ligação (E_s), deformação angular (E_b), torção de ligações (E_w) e de interações não-ligantes (E_{nb}). Cada um dos termos representa a diferença de energia entre uma molécula real e uma hipotética, sendo que todos os parâmetros estruturais apresentam valores de referência.

Porém, diferente do que ocorre nos estudos de otimização utilizando um campo de força, em sistemas reais a energia cinética dos átomos permite que mudanças conformacionais aconteçam a todo o momento (SANT'ANNA, 2009). Para estudos nessas condições surgiram os métodos de mecânica estatística. Esses métodos são capazes, em teoria, de determinar propriedades físicas associadas não a uma geometria específica, mas sim a uma amostragem das propriedades decorrentes de muitas conformações que podem ser adotadas (YOUNG, 2001).

Um dos métodos mais conhecidos e utilizados é a *dinâmica molecular* (DM), uma simulação tempo-dependente do comportamento de um sistema molecular. Para isto é necessária a utilização de uma forma de computar a energia do sistema, e isto é mais comumente realizado utilizando campos de força clássicos (YOUNG, 2001). Porém, uma simulação DM é um processo computacionalmente caro. Assim, o tempo considerado é em geral curto, da ordem de picossegundos ($1 \text{ ps} = 10^{-12} \text{ s}$) ou fentossegundos ($1 \text{ fs} = 10^{-15} \text{ s}$) (SANT'ANNA, 2009). Durante o processo, é determinado um intervalo de tempo em que ocorre transferência de calor para a molécula e outro intervalo em que a molécula equilibra a sua energia. Este processo é repetido um número de vezes especificado pelo usuário (SILVA, 2003).

A resolução das equações de movimento de DM representa a evolução temporal dos movimentos moleculares, ou trajetória, que pode ser usada para estudos que dependem do tempo. Assim, uma molécula com certa energia cinética pode superar barreiras de energia potencial, o que permite a exploração da SEP na busca de outras conformações (SANT'ANNA, 2009). As simulações de DM podem ser realizadas em diversos meios (fase gasosa, solução, o receptor biológico, etc), considerando as influências de forças estéricas e eletrostáticas nas conformações. (SADOWSKI, SCHWAB e GASTEIGER, 2004).

1.8.2 Métodos quânticos

No começo do século XX, os pesquisadores observaram que a mecânica clássica não era capaz de descrever corretamente o comportamento de partículas muito pequenas (SILVA, 2003). Assim, foi necessária a evolução dos conceitos para que toda uma nova teoria fosse desenvolvida com o propósito de se conseguir explicar o comportamento das partículas subatômicas (ATKINS e PAULA, 2004a).

Inicialmente, Niels Bohr observou que certos aspectos dos espectros eletrônicos dos átomos de hidrogênio e hidrogenóides (como Li^{++} e He^+) podiam ser explicados se fosse admitido que a energia de átomos fosse restrita a valores discretos, sendo emitida ou absorvida em determinadas quantidades. Ou seja, a energia era *quantizada* (do latim *quantum*, quantidade), (ATKINS e PAULA, 2004a; SOUZA, 2000). Esta observação fortaleceu uma antiga hipótese de que a luz se manifesta como se fosse constituída de partículas (ATKINS e PAULA, 2004a).

A *mecânica quântica* parte da admissão que, em lugar de se deslocar ao longo de uma trajetória perfeitamente definida, uma partícula se distribui através do espaço como uma onda. A representação matemática dessa distribuição foi proposta por Erwin Schrödinger e é denominada *função de onda* (ψ). A interpretação deste parâmetro baseia-se em uma sugestão de Max Born, pela qual o quadrado da função de onda (ou o quadrado do módulo, $|\psi|^2 = \psi^*\psi$, se ψ for uma função complexa) pode ser interpretado como uma medida de probabilidade de se encontrar uma partícula num certo ponto x . Assim, a probabilidade de se encontrar a partícula entre x e $x + dx$ é proporcional a $|\psi|^2 dx$. Assim, $|\psi|^2$ é denominado *densidade de probabilidade*. Esta função contém todas as informações necessárias sobre um sistema em estudo (ATKINS e PAULA, 2004a).

Assim, o cálculo de propriedades eletrônicas tornou-se possível, superando uma das principais limitações da mecânica molecular. A função de onda é um dos termos da *Equação de Schrödinger* (Equação 12), onde são aplicados operadores matemáticos de energia cinética e potencial na mesma, obtendo como resposta a energia total do sistema (LEVINE, 1992; BARREIRO *et al.*, 1997; SILVA, 2003; SOUZA, 2000; ATKINS e PAULA, 2004a; SANT'ANNA, 2009).

$$H\psi(R, r) = E\psi(R, r) \quad (12)$$

$$H = T + V$$

$$T = -\sum_{i=1}^n \frac{\hbar^2}{2m_i} \nabla_i^2$$

$$V = \frac{1}{4\pi\epsilon_0} \left(-\sum_i^{\text{elétrons}} \sum_I^{\text{núcleos}} \left(\frac{Z_I e^2}{\Delta r_{iI}} \right) + \sum_i^{\text{elétrons}} \sum_{j>i}^{\text{elétrons}} \left(\frac{e^2}{\Delta r_{ij}} \right) + \sum_I^{\text{núcleos}} \sum_{J>I} \left(\frac{Z_I Z_J e^2}{\Delta R_{IJ}} \right) \right)$$

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$$

Onde: H = operador Hamiltoniano

E = energia

V = energia potencial da partícula

∇^2 = operador Laplaciano

T = energia cinética da partícula

R_{IJ} = posição de um núcleo em relação a outro

m = massa da partícula

r_{ij} = posição de um elétron em relação a outro

Z = número atômico do núcleo

r_{iI} = posição de um elétron em relação a um núcleo

$$h = \frac{\hbar}{2\pi} = 1,05457 \times 10^{-34} \text{ J s}$$

$h = 6,62608 \times 10^{-34} \text{ J s}$ = constante de Planck

A utilização de métodos mecânico quânticos permite maior precisão nos resultados. Além disso, muitas propriedades de interesse em QSAR (como energias de orbitais moleculares, algumas propriedades termodinâmicas e momento dipolar) só podem ser obtidas com a utilização destes métodos (BILLINGS, 2002). Porém, a aplicação desta abordagem implica em um custo computacional (tempo de computação e capacidade de memória

necessária) muito maior que a mecânica molecular (BARREIRO *et al.*, 1997). Quanto maior a molécula ou mais complexo o sistema em estudo e/ou mais complexa é a abordagem utilizada, maior este custo. Apesar dos avanços na informática e no desenvolvimento de novas metodologias, a complexidade de determinados cálculos ainda constituem limitações.

Assim, torna-se necessário a utilização de métodos que, de alguma forma, forneçam resultados aproximados de dados que poderiam ser obtidos experimentalmente (CASTRO e CANUTO, 2007; LEVINE, 1991). Uma das principais aproximações na mecânica quântica é a *aproximação de Born-Oppenheimer*, muito útil para o auxílio na solução deste problema. Por esta abordagem, admite-se que os núcleos, muito mais pesados que os elétrons, têm movimentos relativamente lentos e podem ser tratados como estacionários, enquanto os elétrons se movem uns em relação aos outros. Assim, a equação de Schrödinger pode ser resolvida apenas para os elétrons (ATKINS e PAULA, 2004b).

Os métodos mecânico quânticos podem ser agrupados em três grandes classes: *ab initio*, semi-empíricos e teoria do funcional de densidade (DFT: *density functional theory*).

1.8.2.1 Métodos *ab initio*

Os métodos *ab initio* são aqueles em que se resolve a equação de Schrödinger com maior exatidão (SILVA, 2003). A origem do termo é o latim, e significa “do início” ou “dos primeiros princípios”, já que os cálculos realizados por estes métodos não utilizam dados experimentais para auxiliar na resolução da equação (YOUNG, 2001).

O procedimento computacional *ab initio* mais utilizado é a solução de campo autoconsistente ou solução SCF (*self-consistent field*), utilizada para resolver a equação de Hartree-Fock (HF). Nesta abordagem, a partir de geometrias específicas utilizadas como ponto de partida, resolve-se a forma matricial da equação HF (Equação 13, também conhecida como

equação secular), utilizando aproximações sucessivas. O método é dito autoconsistente porque, no final de um processo*, as soluções da equação devem ser os mesmos a partir dos quais se obteve **F**.

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad (13)$$

Além de ser capaz de fornecer uma boa solução aproximada para um problema de muitos elétrons, as geometrias obtidas podem ser utilizadas como um ponto de partida para outros métodos. Uma das vantagens dos métodos HF é que eles dividem uma equação de muitos elétrons em muitas equações de um elétron. Cada equação de um elétron é resolvida para fornecer uma função de onda individual (o orbital) e sua energia (YOUNG, 2001). A Equação 12 e os termos apresentados correspondem à mais simples representação da equação de Schrödinger utilizada pelos métodos HF (BILLINGS, 2002; SILVA, 2003; CASTRO e CANUTO, 2007; YOUNG, 2001).

Porém, o método HF possui algumas limitações. Para que uma solução exata da equação HF seja obtida, seria necessário utilizar um número infinito de funções matemáticas denominadas *funções* ou *conjuntos de base* (*basis-set*). Porém, na prática, isto é impossível, o que leva uma imprecisão associada a este método. Assim, quanto maior o número de funções utilizadas, mais próximo da energia extrapolada para um número infinito de base (ou *limite HF*) estará o resultado (CASTRO E CANUTO, 2007).

Outra limitação importante do método HF é partir da suposição que o movimento de um elétron não é influenciado pelos seus pares. Na verdade, cada elétron está sujeito a um potencial efetivo que considera suas interações com os outros elétrons através de uma média, mas os detalhes das interações particulares entre cada par de elétrons ficam perdidos. A

* Processo que se repete diversas vezes para se chegar a um resultado e a cada vez gera um resultado parcial que será usado na vez seguinte.

diferença entre a energia exata não relativística e a energia de HF (Equação 14) é chamada *energia de correlação eletrônica* (E_{corr}) (CASTRO e CANUTO, 2007).

$$E_{\text{corr}} = E_{\text{exata}} - E_{\text{HF}} \quad (14)$$

Embora E_{corr} seja frequentemente pequena em comparação com E_{exata} , ela é essencial para a descrição exata dos processos eletrônicos. Assim, foram propostos métodos “pós-HF” que consideram o efeito de correlação eletrônica nos resultados (CASTRO e CANUTO, 2007; BILLINGS, 2002). Os mais conhecidos entre eles são os *métodos perturbativos*, como a teoria de Møller-Plesset (MPn, n = 2, 3, 4) e o método de “*coupled cluster*” (CC), e os que utilizam a *superposição de configurações*, como interação de configurações e os métodos de campo autoconsistente multireferencial (como o CASSCF: *complete active space SCF*) (CASTRO e CANUTO, 2007).

1.8.2.2 Métodos semiempíricos.

O custo computacional foi outro fator que por muito tempo limitou a utilização de cálculos HF para moléculas com cerca de uma dezena de átomos. Quanto mais complexo o sistema em estudo, maior a quantidade de cálculos necessários, e assim maior será o custo computacional (basicamente, tempo de computação, necessidade de memória computacional, e velocidade de processamento matemático). Ao mesmo tempo, a necessidade da realização de estudos com conjuntos formados por várias dezenas ou centenas de moléculas ou sistemas moleculares complexos estimulou o desenvolvimento de metodologias que fornecessem resultados confiáveis a um custo computacional menor. Desta procura surgiram os denominados *métodos semiempíricos* (BREDOW, 2004; SIMAS e ROCHA, 2007).

Esses métodos possuem a mesma estrutura geral dos cálculos HF. Porém, certas informações podem ser omitidas total ou parcialmente. Por exemplo, os elétrons das camadas mais internas (conhecidas como *caroço*) e integrais de dois elétrons (utilizadas no método HF) podem ser omitidos. Além disso, parâmetros experimentais (o motivo da denominação recebida) ou calculados por *ab initio* podem ser utilizados para ajustar os resultados. Porém, caso uma molécula em estudo apresente diferenças significativas em relação àquelas originalmente utilizadas para parametrizar o método, resultados muito ruins poderão ser obtidos. Mas o inverso também é verdade (YOUNG, 2001).

A primeira metodologia proposta foi o *método de Hückel*, útil somente para os estudos de moléculas que apresentam elétrons de valência π em hidrocarbonetos conjugados cíclicos, mas que serviu de base para o *método estendido de Hückel*, parametrizado para considerar os elétrons de valência dos átomos de H, C, N, O e F de sistemas cíclicos (BREDOW, 2004).

Procurando expandir a utilização dos métodos semi-empíricos, John Pople e seus colaboradores propuseram os métodos CNDO (*complete neglect of differential overlap*) e NDDO (*neglect of diatomic differential overlap*). Ambos os métodos são aplicáveis a qualquer sistema, porém o primeiro desconsidera todas as integrais de repulsão eletrônica que possuam produtos envolvendo orbitais distintos, enquanto o segundo calcula uma série de integrais, podendo ser considerada superior. Já Michael Dewar lançou, em 1969, o MNDO (*modified neglect of diatomic overlap*), um método capaz de calcular propriedades de interesse químico (especialmente o calor de formação) com boa exatidão. Este método é baseado na metodologia NDDO, porém utiliza 22 integrais diatômicas de repulsão eletrônica para cada par de átomos pesados (ou seja, desconsidera o hidrogênio) (SIMAS e ROCHA, 2007; HYPERCUBE, 2002).

Um dos métodos mais utilizados atualmente foi desenvolvido a partir do MNDO: o AM1 (*Austin Model 1*) (SIMAS e ROCHA, 2007). Este método supera algumas limitações que o anterior apresentava em relação à repulsão eletrônica, e é muito utilizado para cálculos de estrutura eletrônica de compostos orgânicos, incluindo macromoléculas. Suas funções Gaussianas esféricas, incluídas nas integrais de repulsão núcleo-núcleo, diminuem a tendência de superestimar a repulsão entre átomos e permitem descrever ligações hidrogênio de forma mais correta (HYPERCUBE, 2002; BREDOW, 2004; SIMAS e ROCHA, 2007). Já o método PM3 (*Parameter Model 3*) é uma reparametrização do AM1. Ambos são praticamente idênticos, porém os parâmetros do PM3 são ajustados para reproduzir dados experimentais. Isto leva a obtenção de ótimos resultados para calores de formação, embora apresente erros quanto as cargas parciais de átomos de nitrogênio (YOUNG, 2001; SIMAS e ROCHA, 2007).

Os métodos semiempíricos AM1 e PM3 estão entre os mais utilizados. Apesar de suas deficiências, eles são capazes de gerar bons resultados para uma série de propriedades moleculares que podem ser utilizadas como descritores. Este fato, aliado ao menor custo computacional, são os motivos dessas duas metodologias serem amplamente utilizadas em QSAR. Além disso, novas metodologias baseadas em AM1 e PM3 continuam sendo desenvolvidas (p. ex., sparkle/AM1, PM5, PM6 e o RM1 – *Recife Model 1*) com o intuito de corrigir falhas e acrescentar utilizações (NETZEVA *et al.*, 2005; PUZYN *et al.*, 2008).

1.8.2.3 Métodos DFT

Considerando as desvantagens apresentadas pelos métodos *ab initio* e semiempíricos, uma terceira abordagem surgiu como uma alternativa a ambos os métodos: a teoria do funcional de densidade (DFT) (MORGON e CUSTÓDIO, 1995), atualmente um método muito importante para o estudo de propriedades eletrônicas (DUARTE e ROCHA, 2007).

Em relação aos métodos *ab initio* padrões, a abordagem DFT apresenta considerável ganho computacional e de memória. Quanto aos métodos semiempíricos, o hamiltoniano DFT é bem definido e nenhum parâmetro experimental precisa, em princípio, ser utilizado para ajustar um “funcional”, o que torna esta abordagem basicamente um tipo de teoria *ab initio* (MORGON e CUSTÓDIO, 1995). Porém, existem funcionais desenvolvidos com a utilização de parâmetros que auxiliam na reprodução de determinados resultados experimentais. Assim, também pode ser dito que determinados funcionais são semi-empíricos (SANT’ANNA, 2009). Os principais funcionais DFT são apresentados na Tabela 3.

Tabela 3. Funcionais de densidade. Extraído de Yong (2001).

Sigla	Nome	Tipo
X α	X-alfa	Somente troca
HFS	Hartree-Fock Slater	HF com troca LDA
VWN	Vosko, Wilks e Nusair	LDA
BLYP	Funcional de correlação de Becke com troca de Lee, Yang e Parr	Correção de gradiente
PW91	Perdue e Wang 1991	Correção de gradiente
G96	Gill 1996	Troca
P86	Perdew 1986	Correção de gradiente
B96	Becke 1996	Correção de gradiente
B3P86	Troca de Becke, correlação de Perdew	Híbrido
B3PW91	Troca de Becke, correlação de Perdew e Wang	Híbrido

A *densidade eletrônica*, $\rho(r)$, é a entidade básica em DFT, assim como a função de onda nos métodos *ab initio* (MORGON e CUSTÓDIO, 1995). Nesta abordagem, considera-se que a energia de um conjunto de elétrons sob influência de um campo externo é um funcional único da densidade eletrônica (SANT’ANNA, 2009). Assim, independente da forma como o método é apresentado, todas as propriedades são funcionais de $\rho(r)$, e a energia do estado fundamental de um sistema multieletrônico sob um dado potencial externo, $V(r)$, pode ser descrito como apresentado na Equação 14 (MORGON e CUSTÓDIO, 1995):

$$E_V[\rho(r)] = \int v(r)\rho(r) dr + F[\rho] \quad (14)$$

Onde F= funcional universal de ρ

Outra grande vantagem dos métodos DFT é a incorporação de efeitos de correlação eletrônica no cálculo da energia. Os diferentes funcionais disponíveis usam diferentes formas de calcular a correlação. No esquema de Kohn-Sham (KS), por exemplo, é definido um potencial de troca e correlação (Equação 15), que é a derivada funcional da energia total de troca-correlação (MORGON e CUSTÓDIO, 1995; DUARTE e ROCHA, 2007). Os melhores funcionais DFT podem fornecer resultados de qualidade similar aos métodos *ab initio* que consideram a correlação eletrônica, porém a um custo computacional muito menor (YOUNG, 2001).

$$V_{XC}(r) = \frac{\partial E_{XC}[\rho(r)]}{\partial \rho(r)} \quad (15)$$

1.8.2.4 Conjuntos (ou funções) de base

Um conjunto de base é um conjunto de funções matemáticas utilizado para descrever a forma dos orbitais atômicos (OA) onde podem localizar-se os elétrons de um átomo. Os elétrons de uma molécula, por sua vez, poderão localizar-se nos orbitais moleculares (OM), uma região finita do espaço. Os OM e a função de onda correspondente são criados utilizando combinações lineares das funções de base ou, como é mais conhecido, combinações lineares de OA (LCAO: *Linear Combination of Atomic Orbitals*) (YOUNG, 2001; SOUZA, 2000).

Uma decisão inicial que deve ser tomada em estudos que utilizam métodos DFT ou *ab initio* é a escolha do conjunto de base a ser utilizada. Em geral, há duas categorias: conjuntos de bases mínimas, que descrevem somente os aspectos básicos dos orbitais, e os estendidos,

mais detalhados. Já os métodos semi-empíricos utilizam bases pré-definidas (SHODOR, 2000; YOUNG, 2001; SIMAS e ROCHA, 2007).

Uma base mínima apresenta um conjunto de base individual para cada OA dos átomos em uma molécula. As bases mínimas mais comuns são os orbitais STO (*Slater Type Orbitals*) dos tipos STO- n G ($n = 3, 4, 6$), sendo n integrais cujo valor determina o número de funções tipo Gaussianas (GTO: *Gaussian type orbitals*) primitivas que formam a função de base individual (DAVIDSON e FELLER, 1986).

As funções STO (Equação 16) têm este nome porque foram desenvolvidas por J.C. Slater. Como o cálculo destas funções apresenta um alto custo computacional, Frank Boys desenvolveu como alternativa as funções GTO (Equação 17), comparativamente mais simples. Assim, as funções de base GTO são as mais importantes na química computacional (DAVIDSON e FELLER, 1986; SHODOR, 2000). A expressão geral de uma função de base é dada pela equação 18.

$$\text{STO} = \frac{\zeta}{\pi} \cdot e^{(-\zeta \cdot r)} \quad (16)$$

$$\text{GTO} = \frac{2\chi}{\pi} \cdot e^{(-\zeta \cdot r^2)} \quad (17)$$

$$\text{Funções de base} = N \cdot e^{(-\alpha \cdot r)} \quad (18)$$

Onde: N = constante de normalização

α = expoente do orbital

r = raio em angstroms

As bases extendidas, por sua vez, são aquelas que consideram os orbitais mais altos da molécula e o tamanho e forma de distribuição de carga. As mais conhecidas são as *split valence basis sets*, que representa os elétrons do caroço, como orbitais STO simples. Estes

conjuntos sempre são representados pela sequência “número – número número G”. Por exemplo, 3-21G e 6-31G. O significado desta sequência é apresentado na Figura 15. (SHODOR, 2000).

$$6^a - 3^b 1^c G$$

(a) o número de funções Gaussianas que descrevem os orbitais internos.
 (b) o número de funções Gaussianas primitivas que são combinadas para formar a primeira função de base que descreve os orbitais de valência.
 (c) segunda função Gaussianas primitiva também utilizada para formar a segunda função de base que descreve os orbitais de valência.

Figura 15. Uma representação comum de um *split valence basis set double-zeta*. Se há mais algum número após “c”, estas funções são denominadas *triple-zeta*, *quadruple-zeta*, etc.

Outra característica comum na seleção de um conjunto de base é a adição de *funções de polarização* (geralmente representadas por asteriscos ou pelas letras “d” e “p” após o “G”) e *funções difusas* (representadas pelo símbolo matemático “+” entre o último número e o “G”). As funções de polarização têm por objetivo adicionar certa “flexibilidade” ao conjunto de base utilizado, permitindo que os orbitais moleculares que envolvem os átomos sejam mais assimétricos, permitindo uma representação mais correta dos orbitais. Um único asterisco significa que uma função do tipo *d* é adicionada aos átomos pesados (aqueles com número atômico maior que 2), enquanto dois asteriscos correspondem ao mesmo processo acrescido da adição de funções tipo *p* aos átomos leves (H and He). Já as funções difusas são versões “maiores” das funções do tipo *s* e *p*, e sua utilização permite que os orbitais ocupem uma região maior do espaço. A utilização de um “+” significa que está sendo levado em consideração o orbital *p*, e dois que ambos os orbitais *p* e *s* são considerados (FOESMAN e FRISCH, 1996; SHODOR, 2000; HYPERCUBE, 2002).

1.9 INIBIDORES DE HIV-INTEGRASE

Até o ano 2000, muitos compostos com capacidade de inibir a HIV-IN *in vitro* haviam sido descritos. Porém, estes resultados não se traduziam em atividade antiviral em culturas celulares. Neste período, dois estudos foram considerados fundamentais para o desenvolvimento dos inibidores da HIV-IN. Em 1999 foi publicado o trabalho de Goldgur e colaboradores (1999) que descreveu o primeiro complexo cristalográfico de um inibidor da HIV-IN com o sítio-ativo. Este complexo (PDB 1QS4) (Figura 16) foi considerado uma plataforma importante para o planejamento reacional desta classe de fármacos, pois foi um dos primeiros a introduzir os β -dicetoácidos (DCAs) como protótipo, pois apresentou o inibidor 5CITEP [1-(5-cloroindol-3-il)-3-hidroxi-3-(2H-tetrazol-5-il)-propenona], da Shionogi e Co. Este inibidor foi co-cristalizado exatamente entre a tríade catalítica descrita como importante para a atividade, em acordo com resultados de estudos experimentais realizados antes da descrição do 1QS4; e (iii) ocorre a presença de magnésio (Mg^{+2}) como íon divalente, provavelmente o co-fator metálico importante para a atividade catalítica *in vivo*. O 5CITEP é, na verdade, um dicetoderivado, pois possui, no lugar da função ácido carboxílico um bioisómero, o grupamento cíclico 1H-tetrazol ou ácido tratrazóico, que são também apresentam estrutura planar e ionizam-se em pH 7,4, mas são 10 vezes mais lipofílicos que a função ácido carboxílico e são mais resistentes a metabolização (HERR, 2002).

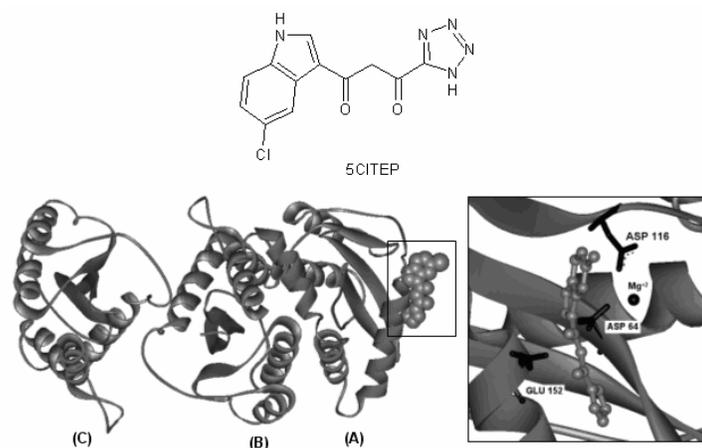


Figura 16. Outra representação da estrutura do complexo cristalográfico 5CITEP e HIV-1 IN. Este complexo apresenta três subunidades idênticas, A, B e C. A região expandida apresenta o sítio onde o inibidor complexou-se (face da subunidade A). Figura já publicada pelo autor em Melo, Bruni e Ferreira (2006).

Os cristais do domínio catalítico presentes no arquivo 1QS4 contêm três moléculas do núcleo da integrase (unidade A, B e C) dentro de uma unidade cristalográfica assimétrica. O 5CITEP ligou-se na face do sítio A, entre a tríade catalítica. Ao mesmo tempo, vários outros resíduos conhecidos como importantes para a catálise ou ligação do substrato de DNA localizam-se próximo ao 5CITEP (GOLDGUR *et al.*, 1999).

Porém, vários autores, como Sotriffer, Ni e McCammon (2000), têm chamado a atenção para as deficiências deste complexo, em especial devido aos efeitos prejudiciais de empacotamento cristalino que podem ter prejudicado sua obtenção. Chem e colaboradores (2008) observaram que o 5CITEP apresenta contatos limitados com a integrase, mas interage de forma significativa com outra molécula de 5CITEP localizada na célula cristalina adjacente. Este seria um forte indício de que o empacotamento cristalino influenciou seu posicionamento, limitando seu potencial como guia em estudos de CADD.

No começo de 2000, Hazuda e colaboradores (2000) apresentaram o primeiro trabalho com DCAs propriamente ditos como inibidores da HIV-IN. Os autores realizaram um

processo de ensaios aleatórios de mais de 250.000 compostos, que faziam parte de uma quimioeteca da Merck Co., quanto a capacidade de inibir especificamente a reação ST. Desta forma, foi possível a identificação de quatro DCAs potentes, específicos e capazes de inibir a replicação viral em culturas celulares (Figura 17). Estudos bioquímicos mostraram a ocorrência de mutações localizadas em aminoácidos próximos ao sítio ativo da enzima, o que ajudou a estabelecer os DCAs da Merck como os primeiros inibidores da reação ST biologicamente validados (HAZUDA *et al.*, 2000).

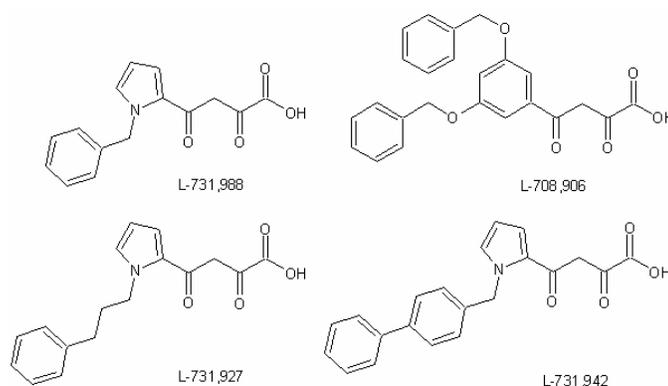


Figura 17. Estrutura dos quatro dicetoácidos da Merck descritos por Hazuda e colaboradores (2000).

Com as informações obtidas em mais de uma década sobre a HIV-IN e os diferentes tipos de classes de compostos capazes de inibir a função biológica da mesma, em 2007 cerca de seis compostos já haviam sido submetidos ou encontravam-se em ensaios clínicos.

De forma geral, os DCAs são β -dicetonas derivadas do ácido benzopirúvico. Normalmente, cetonas apresentam duas formas tautoméricas, a própria cetona e seu enol correspondente, sendo a primeira mais estável. Porém, em algumas situações o equilíbrio pode ser deslocado a favor da forma enólica (COSTA *et al.*, 2003). Um outro caso em que este deslocamento do equilíbrio ocorre é exatamente quando uma cadeia alifática não-substituída apresenta dois grupamentos carbonílicos em posição 1,3 um em relação ao outro. Quando esta

estrutura, denominada β -dicetona, está presente os compostos em solução tem o equilíbrio tautomérico deslocado a favor do enol (no caso, ceto-enol), pois ocorre a formação de uma ligação hidrogênio intramolecular com caracter de ligação covalente entre o oxigênio carbonílico e o hidrogênio enólico (RUSTICI, CARAMORI e GALEMBECK, 2006) (Figura 18). Este grupamento, quando na presença de íons metálicos (como o Mg^{+2}), pode ser facilmente desprotonado, gerando um diânion que pode interagir com dois íons divalentes (COTELLE, 2006).

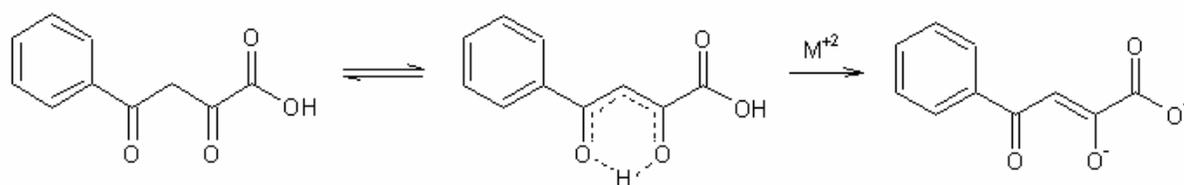


Figura 18. Estrutura do ácido benzopirúvico e representação do equilíbrio ceto-enólico deslocado a favor da forma enólica por meio da formação de ligação hidrogênio intramolecular, e diânion que é formado quando na presença de íons metálicos.

Outro atributo estrutural importante para a inibição da reação ST pelos DCAs é a presença de uma cadeia lateral aromática ligado ao grupamento DCA por uma cadeia flexível. A estrutura mais comum entre os inibidores mais potentes é um grupamento benzil ($-CH_2-Ph$) substituído por um ou mais átomos de halogênio, inclusive sendo encontrado na maioria dos compostos que estão ou já foram submetidos a ensaios clínicos (COTELLE, 2006). Segundo a literatura, esta importância reside no fato de que esta subestrutura provavelmente liga-se a HIV-1 IN interagindo com uma alça desordenada formada pelos resíduos identificados como 140 a 149* e localizada na parte superior do núcleo da enzima. Esta interação provavelmente ocorre por meio de uma interação hidrofóbica aromática ou do tipo π -stacking com o resíduo Tyr143 (BARRECA *et al.*, 2006; PAIS *et al.*, 2002; PARRIL, 2003).

* Gly140, Ile141, Pro142, Tyr143, Asn144, Pro145, Gln146, Ser147, Gln148 e Gly149.

Considerando estas informações, pode ser considerado como farmacóforo dos DCAs a subestrutura apresentada na Figura 19. Porém, é possível que outras interações também sejam necessárias para que ocorra a inibição da reação ST. Outras interações, especialmente ligações hidrogênio, puderam ser observadas em estudos de *docking*, dinâmica molecular e mesmo no complexo cristalográfico PDB 1QS4 (GOLDGUR *et al.*, 1999; SOTRIFFER, NI e MCCAMMON, 2000; NI, SOTRIFFER e MCCAMMON, 2001; KESERÛ e KOLOSSVARY, 2001; CHEN *et al.*, 2008; HEALY *et al.*, 2009).

Com base na estrutura do 5CITEP e em uma série de compostos relacionados (COTELLE, 2006), a Shionogi Co, em parceria com a GlaxoSmithKline, colocou em ensaios clínicos o fármaco S-1360 (BILLICH, 2003) (Figura 20). Porém, apesar da grande expectativa gerada em torno deste composto por volta de 2005-2006, este falhou nos ensaios clínicos de fase II, pois se mostrou pouco efetivo *in vivo* devido à conversão em um metabólito inativo seguido de uma rápida eliminação do organismo. Os estudos mostraram que enzimas da família das aldo-ceto redutases, presentes no fígado humano, eram as principais responsáveis pela sua rápida eliminação, pois o grupamento DCA alifático comporta-se como um substrato destas enzimas, sendo rapidamente reduzido para, em seguida, sofrer o processo de glucoronidação. Assim, apesar do grupamento DCA ser o principal responsável pela inibição da reação de transferência de fita, o mesmo é uma limitação quanto à eficácia terapêutica em humanos, o que levou a descontinuidade dos estudos com o S-1360 e compostos estruturalmente semelhantes (DAYAM, AL-MAWSAWI e NEAMATI, 2007).

Outra tentativa da Shionogi Co. e da GlaxoSmithKline em obter um inibidor da HIV-IN, e que atualmente se encontra em fase II de ensaios clínicos, é a naftiridona GSK-364735 (Figura 21), desenvolvida com base no modo de ligação dos inibidores da reação de ST no receptor (GARVEY *et al.*, 2008). Este composto que apresenta um farmacóforo de três pontos

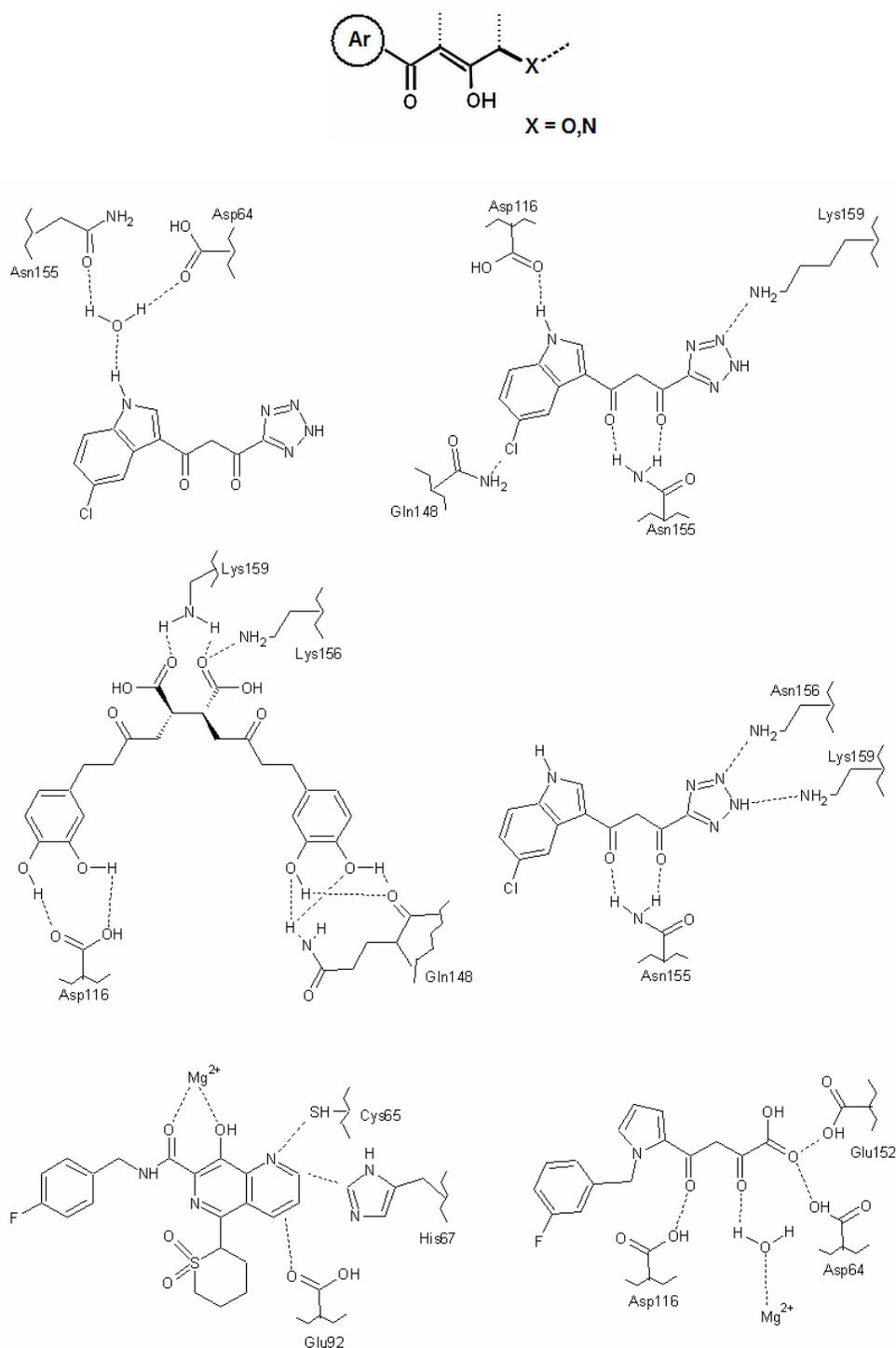


Figura 19. Representação do farmacóforo dos DCAs (Ar: cadeias laterais aromáticas diversas) e de alguns dos resultados de alguns dos diferentes estudos realizados para avaliar os possíveis modos de ligação dos inibidores de HIV-IN com o domínio principal. Baseado nas informações de Goldgur e colaboradores (1999) Sotriffer, Ni e McCammon (2000), Ni, Sotriffer e McCammon (2001), Keserü e Kolossvary (2001), Chen e colaboradores (2008), Healy e colaboradores (2009).

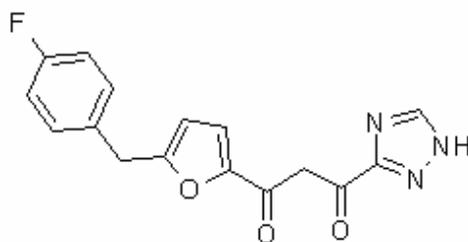


Figura 20. Estrutura do S-1360, primeiro inibidor da HIV-integrase a entrar em testes clínicos.

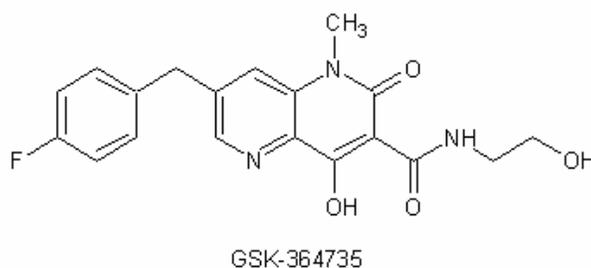


Figura 21. Estrutura do GSK-364735.

similar as naftiridinas L-870,810 e L-870,812 (Figura 22), porém “invertida” em relação a cadeia lateral aromática. Estudos voltados para identificar as mutações que ocorrem na HIV-IN com a utilização deste e de outros inibidores que estão ou passaram por ensaios clínicos indicaram algumas diferenças, o que foi considerado como uma possibilidade de que o GSK-364735 se ajuste no sítio de ligação de uma forma um pouco diferente do que os demais inibidores conhecidos (KOBAYASHI *et al.*, 2008).

Aprimoramentos dos compostos protótipos da Merck levaram as 8-hidroxi-1,6-naftiridinas (Figura 22), com estruturas metabolicamente mais estáveis (ZEINALIPOUR-LOIZIDOU *et al.*, 2007). O estudo desta classe levou a validação *in vivo* dos DCAs, já que o composto L-870,812 foi capaz de suprimir a viremia em Macacos Rhesus (HAZUDA *et al.*, 2004). Junto com este composto, também foi colocado em ensaios clínicos o L-870,810. Estudos indicaram que o mecanismo de ação destes compostos era o mesmo dos DCAs alifá-

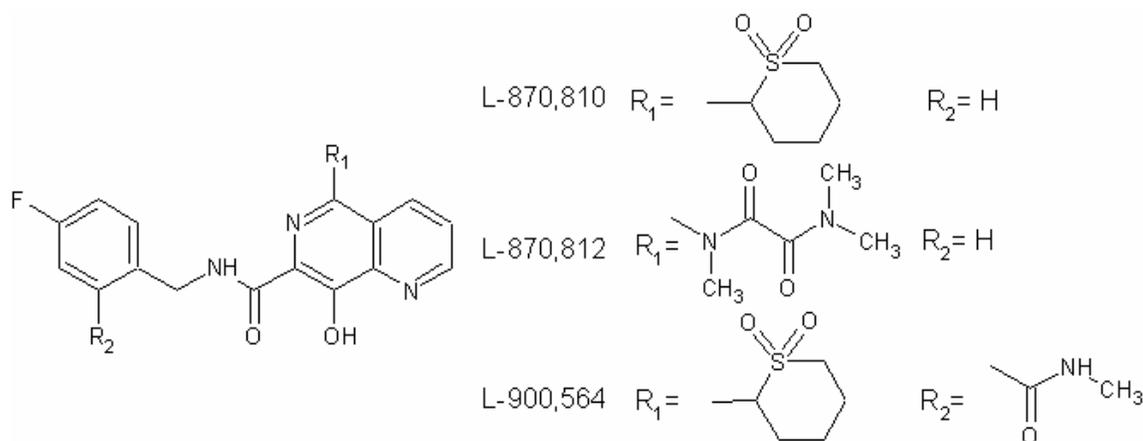


Figura 22. Estruturas das naftiridinas L-870,810, L-870,812 e L-900,564.

ticos. Apesar da potência apresentada tanto *in vitro* quanto *in vivo*, o desenvolvimento do L-870,812 foi suspenso devido a toxicidade renal e hepática em cães após administração por longos períodos e a alta afinidade a proteínas do soro. Com a intenção de corrigir estes problemas, a Merck obteve a naftiridina L-900,564. Este derivado apresentou excelente potência antiviral em cepas não-resistentes e resistentes e um ótimo perfil farmacocinético, sendo então selecionado para avaliação pré-clínica (DAYAM, AL-MAWSAWI e NEAMATI, 2007; EGBERTSON, 2007). Já a situação atual dos ensaios clínicos do L-870,810 é desconhecida, porém, devido a alta similaridade estrutural com a L-870,812, este composto provavelmente também deve apresentar alta afinidade por proteínas do soro (DAYAM, AL-MAWSAWI e NEAMATI, 2007).

Outra classe relacionada com os DCAs é a dos ácidos 4-quinolona-3-carboxílicos. Estes inibidores foram obtidos por meio de modificações de antibióticos da classe das quinolonas, onde se enquadram o ácido nalidixico e o ciprofloxacino (Figura 23) (MITSCHER, 2002). A partir de modificações nestas estruturas foi possível obter uma classe de compostos com capacidade de inibir a reação ST *in vitro* e *in vivo*, além de apresentar considerável atividade antiviral. Assim, foi possível chegar ao GS-9137. Este composto, cujos

direitos foram adquiridos pela Gilead Sciences da Japan Tobacco (o motivo de sua antiga denominação, JTK-303), entrou em ensaios clínicos no Japão. Além de potente, este composto mostrou ser capaz de inibir a replicação de cepas de HIV-1 resistentes a todos os antiretrovirais atualmente disponíveis, apresentou bons resultados em tratamentos combinados com outros antiretrovirais, mostrou boa bioavaliabilidade oral e bom perfil farmacocinético, além de um baixo índice de efeitos adversos. Atualmente, este fármaco encontra-se na fase III de estudos clínicos (DAYAM, AL-MAWSAWI e NEAMATI, 2007).

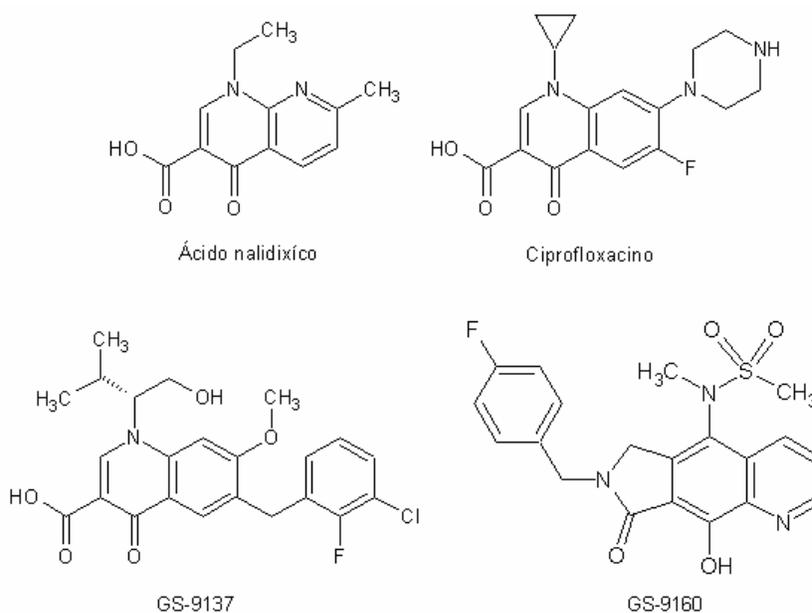


Figura 23. Estruturas do ácido nalidixíco, do ciprofloxacino, do GS-9137, e do protótipo GS-9160.

A Gilead Sciences os estudos relacionados ao GS-9137 como base para mais dois inibidores, o GS-9160 (um derivado tricíclico de nafilpiridinas, também apresentado na Figura 23) e o GS-9224 (cuja estrutura não foi divulgado até o momento). Porém, enquanto os estudos com o primeiro foram cancelados devido ao baixo grau de biodisponibilidade oral, o estágio de desenvolvimento do segundo é atualmente desconhecido (DAYAM, AL-MAWSAWI e NEAMATI, 2007). Outro fator negativo, segundo Cotelle (2006), é que, por

não possuir o farmacóforo três pontos organizado como nas demais classes de DCAs, estes compostos podem apresentar baixa seletividade pela reação de transferência de fita.

O primeiro sucesso da busca de inibidores de HIV-IN como fármacos anti-HIV veio apenas em 2007, com a aprovação pelo FDA e lançamento do raltegravir (Isentress®) (Figura 24) pela Merck (OPAR, 2007). Este composto, uma N-Me pirimidona, é um derivado da classe das 4,5-diidroxipirimidinas carboxamidas, e foi obtido por meio do aprimoramento estrutural de protótipos que foram inicialmente descritos como inibidores da HCV NS5B polimerase (HCV: *hepatite C virus*) (DAYAM *et al.*, 2008). Nos estudos clínicos, este fármaco apresentou excelente eficácia terapêutica in vivo, tanto para cepas não-resistentes quanto para aquelas multi-resistentes. Estudos mostraram que o raltegravir inibe seletivamente a reação de transferência de fita e apresentou sinergismo com outros fármacos atualmente em uso clínico. Nenhum efeito adverso grave foi observado (apenas dores de cabeça, tonturas e outros efeitos leves) e o perfil farmacocinético foi considerado adequado para administração oral (DAYAM, AL-MAWSAWI e NEAMATI, 2007).

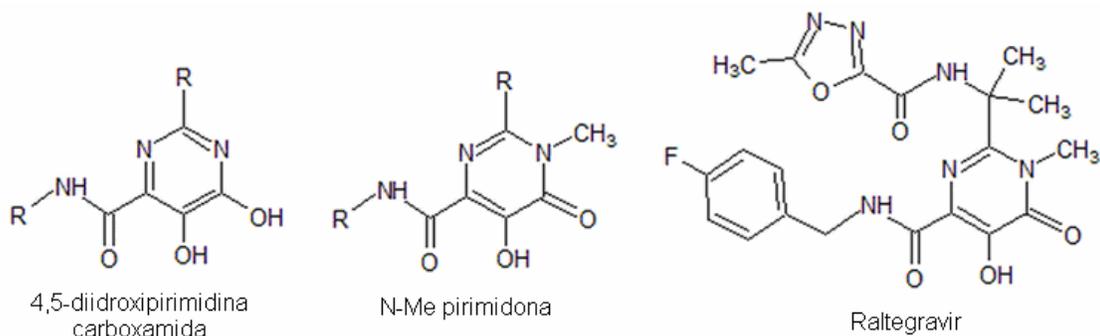


Figura 24. Estrutura do raltegravir e estruturas básicas das 4,5-diidroxipirimidinas carboxamidas e das N-Me pirimidonas.

Após o sucesso da pesquisa com o raltegravir, uma tentativa de desenvolvimento de um fármaco “eu-também” (*me-too*) foi feita pela Bristol-Myers Squibb. Este composto, o BMS-707035 (Figura 25), difere do raltegravir apenas pela substituição do grupo 1,3,4-

oxadiazol por uma sulfonamida cíclica, mesmo grupamento presente no L-870,810 e L-900,564. Como esta região da molécula está relacionada apenas a propriedades farmacocinéticas, a potência *in vitro* dos dois compostos é praticamente a mesma. Porém o BMS-707035 apresentou um índice de resistência viral muito maior que o raltegravir, provavelmente devido ao fato de causar mutações em diversos pontos da enzima. Provavelmente este foi o motivo de seu desenvolvimento ter sido interrompido durante a fase II de ensaios clínicos (DAYAM, AL-MAWSAWI e NEAMATI, 2007; SERRAO *et al.*, 2009).

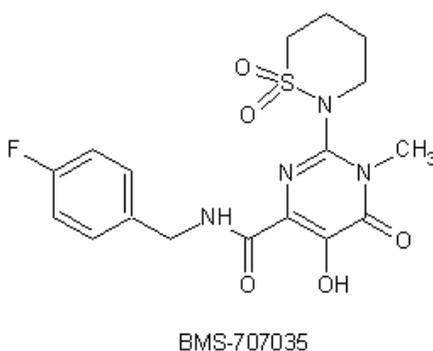


Figura 25. Estrutura do BMS-707035.

A própria Merck encontra-se desenvolvendo um inibidor de integrase cujo protótipo é o próprio raltegravir, o MK-2048 (Figura 26). Considerado um inibidor de integrase de segunda geração, este composto, um hidroxipirrol tricíclico, mimetiza o farmacóforo DCA e, além de potência adequada e bom perfil farmacocinético, espera-se que ele apresente um índice de desenvolvimento de resistência viral menor que o observado para o raltegravir e outros inibidores de integrase (SERRAO *et al.*, 2009; VACCA *et al.*, 2007).

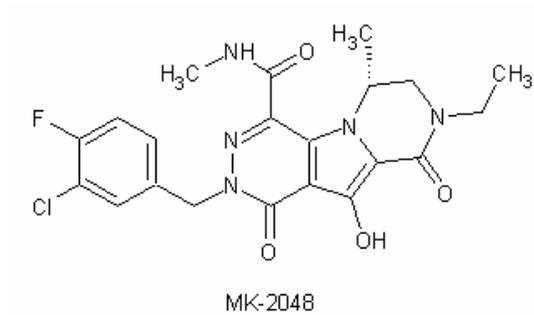


Figura 26. Estrutura do MK-2048.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Este trabalho teve por objetivo desenvolver novos modelos QSAR que possam vir a ser utilizados para o planejamento de novos inibidores da HIV-IN. Para isto, foi determinado que os modelos deveriam: (i) apresentar uma qualidade estatística mínima, segundo critérios e limites recomendados pela literatura; (ii) apresentar a robustez adequada; e (iii) ser capazes de prever a potência de compostos análogos que não foram utilizados na construção dos modelos.

Outro ponto importante que foi levado em consideração foi à existência de uma relação entre as informações codificadas pelos descritores selecionados e o mecanismo de ação. O cumprimento deste quesito aumentaria a credibilidade de um modelo, fortalecendo sua possível aplicação em pesquisas envolvendo o sistema biológico em estudo.

2.2 OBJETIVOS ESPECÍFICOS

Para que o objetivo geral do trabalho pudesse ser alcançado, os seguintes objetivos específicos precisaram ser cumpridos:

- 1) Seleção de um conjunto de treinamento que apresentasse o grupamento farmacofórico DCA, com uma boa quantidade de compostos, um grau de variabilidade estrutural razoável e, obrigatoriamente, a determinação da potência de cada composto contra a enzima HIV-IN por um mesmo protocolo de ensaio farmacológico;
- 2) Desenho da estrutura tridimensional de cada amostra do conjunto de treinamento e aplicação de métodos de modelagem molecular para a obtenção das geometrias mais estáveis,

que foram consideradas como conformações bioativas. Além de serem utilizadas como fonte de parte dos descritores do estudo QSAR-2D, estas conformações podem ser utilizadas como pontos de partida para estudos de dinâmica molecular relativos ao estudo QSAR-4D;

3) Obtenção de descritores moleculares eletrônicos e geométricos a partir das geometrias mais estáveis obtidas por modelagem molecular, e por outras formas de representação estrutural.

4) Realização de um estudo QSAR-2D utilizando regressão por quadrados mínimos parciais (PLS) a partir dos descritores moleculares obtidos na etapa 3;

5) Realização de um estudo QSAR-4D utilizando descritores de campo obtidos a partir de um perfil de amostragem conformacional obtido através da aplicação de um processo de dinâmica molecular em cada amostra;

6) Construção de um modelo híbrido QSAR-2D/4D a partir dos descritores selecionadas nas etapas 4 e 5;

7) Validação dos modelos obtidos nas etapas 4 a 6 quanto à sua qualidade estatística interna (ajuste e validação cruzada) e externa (validação ou capacidade de predição externa), assim como verificação das respectivas robustezes;

8) Interpretação dos melhores modelos perante os mecanismos de inibição da reação de transferência de fita catalizada pela HIV-integrase.

3. METODOLOGIA

3.1 CONJUNTO DE TREINAMENTO

Durante o processo de revisão bibliográfica, foi selecionado da literatura um conjunto de treinamento formado por 85 inibidores da HIV-IN do subtipo viral 1 (HIV-1 IN) (Figuras 27 a 33, e Anexo 1) (PETROCCHI *et al.*, 2007; SUMMA *et al.*, 2006; GUARE *et al.*, 2006; EGBERTSON *et al.*, 2007; HAZUDA *et al.*, 2000; ZHUANG *et al.*, 2003; WAI *et al.*, 2000). Todos os compostos apresentam em comum o farmacóforo de três pontos apresentado na Figura 19. Por meio de uma consulta cuidadosa em cada fonte, pode ser confirmado que a atividade biológica de cada substância quanto à capacidade de inibir a reação ST foi determinada por um mesmo protocolo farmacológico, com o objetivo de medir a capacidade de uma droga inibir a reação ST (HAZUDA *et al.*, 1997). Em cada referência, este valor foi apresentado como IC_{50} , a dose mínima necessária, em M^* , para inibir em 50% a reação ST. Os valores foram convertidos para o logaritmo negativo ($-\text{Log } IC_{50}$ ou pIC_{50}), resultando em um intervalo de variação de 4,30 a 8,15 unidades logarítmicas, que são apresentados em cada figura logo abaixo da identificação do respectivo composto.

3.2 MODELAGEM MOLECULAR

As estruturas tridimensionais do conjunto de treinamento foram construídas no programa HyperChem 7 com base em estruturas cristalográficas extraídas do banco de dados Cambridge Structural Database (Figura 34) (CAMBRIDGE, 2007). Geometrias de mínimos de energia das estruturas neutras foram calculadas na sequência MM+ (HyperChem) \rightarrow HF/6-31G(d) (Gaussian 03) \rightarrow B3LYP/6-31G(d,p) (Gaussian 03). A otimização inicial foi realizada em MM+ porque os campos de força da família MM (MM2, MM3, MM4, MM+,

* Concentração molar.

etc) são considerados os mais adequados para prever mínimos locais e globais em relação a valores experimentais, além de serem mais indicados para a previsão correta de conformações de molécula orgânicas pequenas (LILJEFORS *et al.*, 2004). Já o funcional B3LYP (*Becke's three-parameter hybrid exchange functional and the Lee-Yang-Parr correlation functional*, ou *Becke, three-parameter, Lee-Yang-Parr*), método englobado pela Teoria do Funcional de Densidade (DFT: *density functional theory*) (AYRES, 2004), foi escolhido como método principal. Este é, atualmente, o funcional de maior destaque, e é formado por um funcional de troca exato HF, por um funcional de correlação LDA (*Local Density Approximation*), pelos funcionais de correlação GGA (*Generalized Gradient Approximation*) de Lee, Yang e Parr (LYP) e funcional Becke 88, e por três parâmetros empíricos determinados pelo ajuste de valores preditos para um conjunto de energias de atomização, potenciais de ionização, e afinidades por prótons (DUARTE e ROCHA, 2007; BECKE, 1993; FORESMAN e FRISCH, 1996). Estudos comparativos realizados por diversos autores (WAN *et al.*, 2004; YAN *et al.*, 2006; BASAK, 2003; ZHANG e XIAO, 2002) indicaram que os descritores eletrônicos obtidos utilizando o funcional B3LYP permitiram a obtenção de melhores modelos do que aqueles obtidos por métodos *ab initio* e semiempíricos. Nesta etapa, as cargas parciais foram calculadas utilizando a análise populacional de Mulliken (OLIVEIRA e ARAÚJO, 2007).

A etapa de QSAR-4D também depende de simulações de DM. Os estudos de DM foram realizados em meio solvente utilizando o programa GROMACS (van der SPOEL *et al.*, 2005; <http://www.gromacs.org>). Este meio foi utilizado para que a influência do mesmo sobre as conformações obtidas, devido as forças estéricas e eletrostáticas decorrentes da presença explícita de moléculas de água, fosse considerada no estudo. Além disso, o próprio ensaio biológico foi originalmente realizado em meio aquoso (HAZUDA *et al.*, 1997).

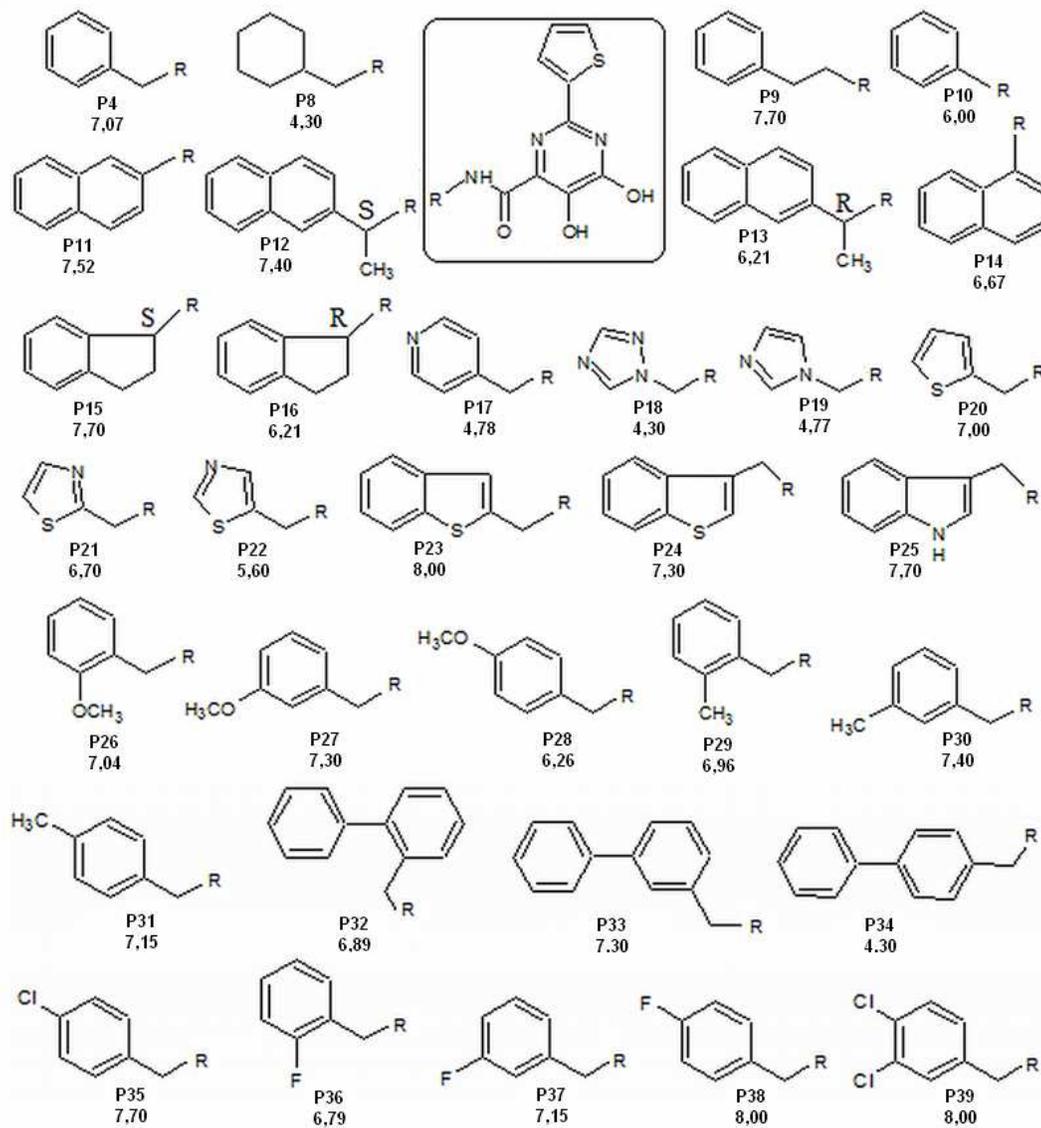


Figura 27. Compostos descritos por Petrocchi e colaboradores (2007). Pertencem a classe das 4,5-diidroxipirimidinas carboxamidas.

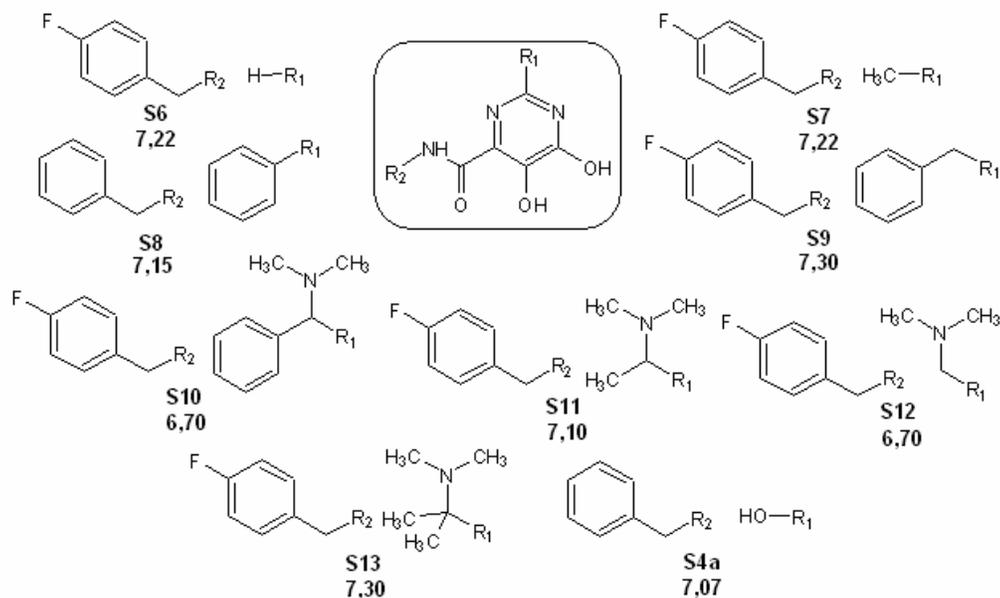


Figura 28. Compostos descritos por Summa e colaboradores (2006). Pertencem a classe das 4,5-dihidropirimidinas carboxamidas.

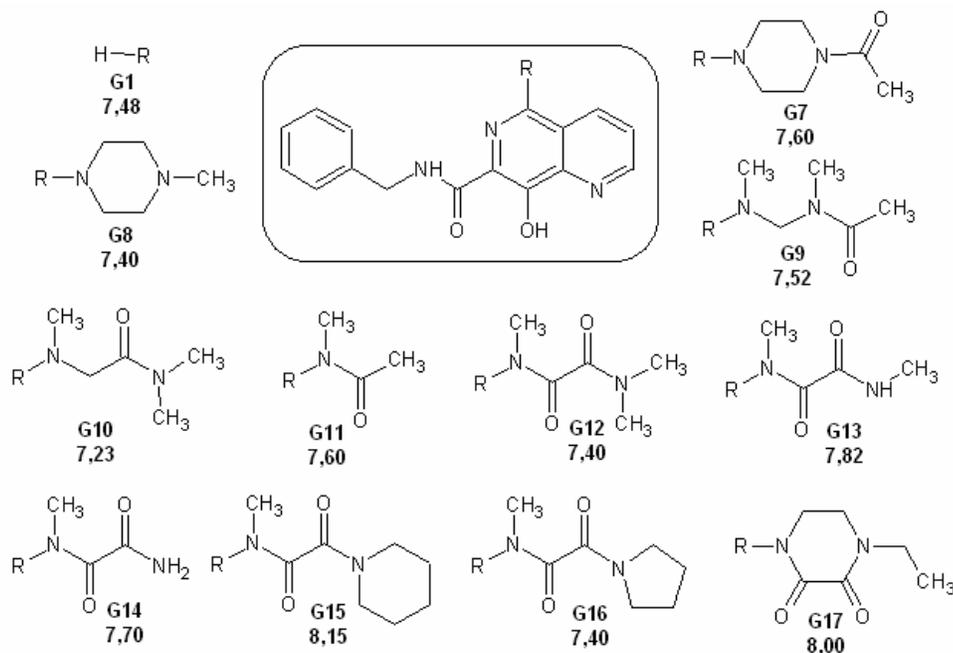


Figura 29. Compostos descritos por Guare e colaboradores (2006). Pertencem a classe das 8-hidroxí-1,6-naftiridinas.

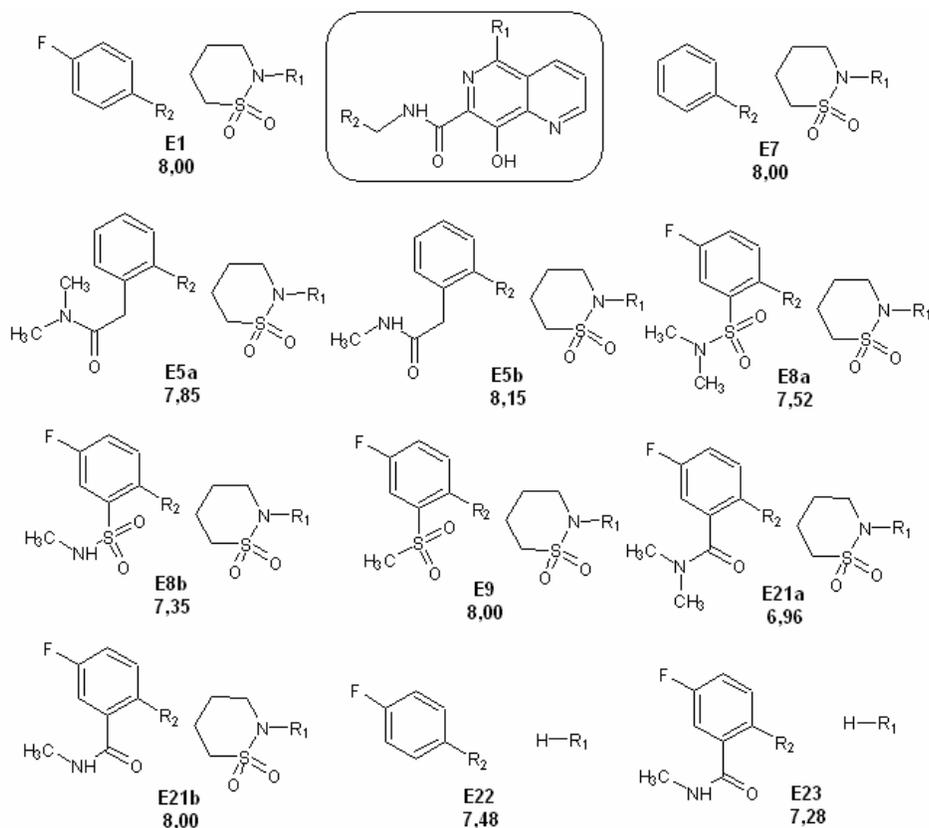


Figura 30. Compostos descritos por Egbertson e colaboradores (2007). Pertencem a classe das 8-hidroxi-1,6-naftiridinas.

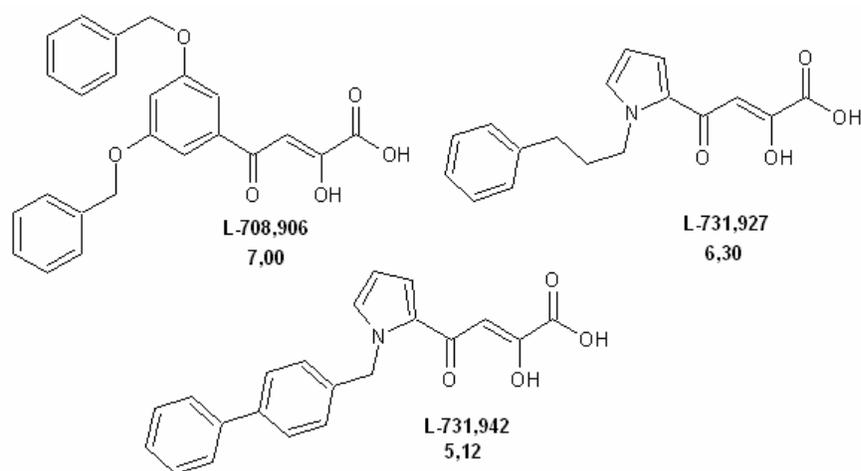


Figura 31. Compostos descritos por Hazuda e colaboradores (2000). Pertencem a classe dos β -dicetoácidos.

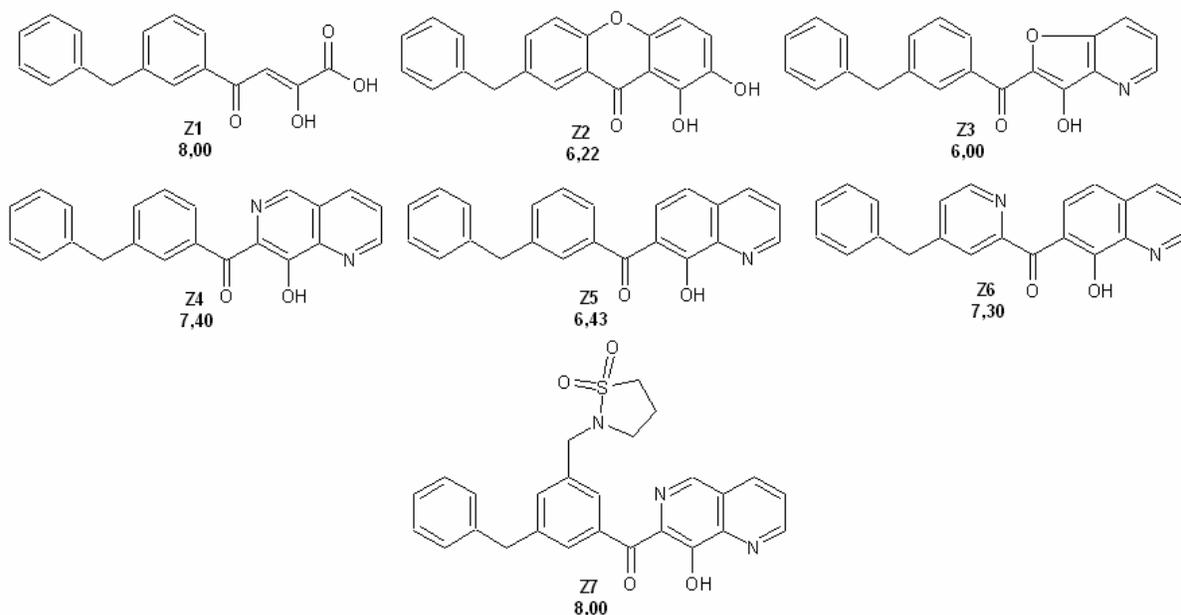


Figura 32. Compostos descritos por Zhuang e colaboradores (2003). O conjunto é formado por um dicetoácido, duas 8-hidroxi-1,6-naftiridinas e quatro derivados relacionados.

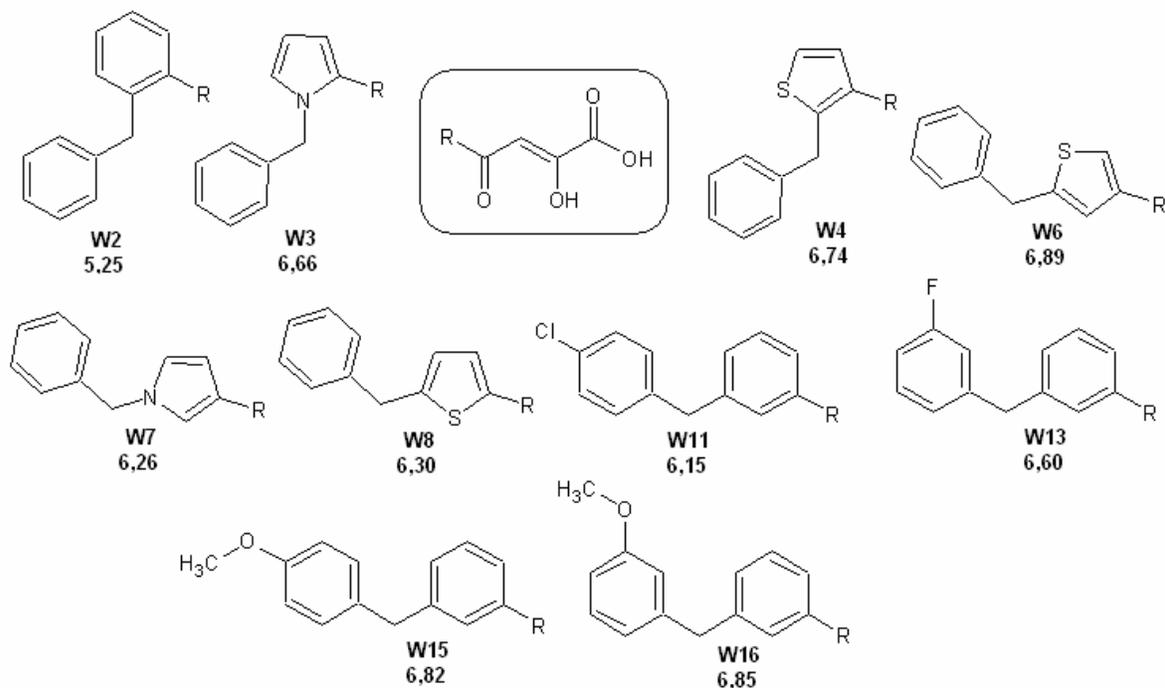


Figura 33. Compostos descritos por Wai e colaboradores (2000). Pertencem a classe dos β -dicetoácidos.

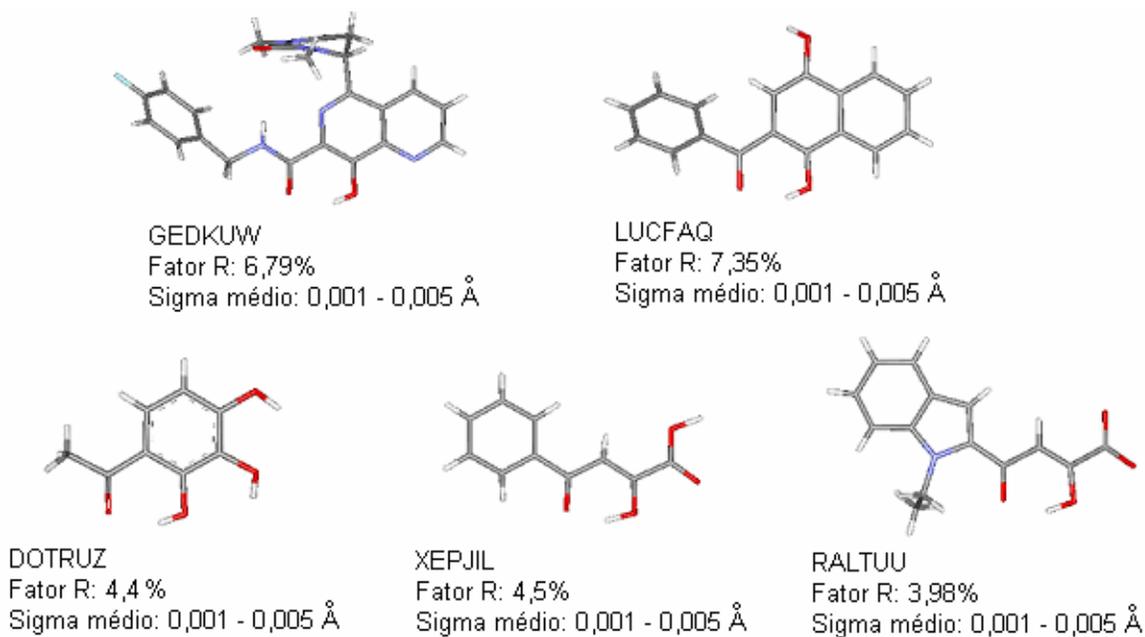


Figura 34. Estruturas cristalográficas (e respectivas resoluções) selecionadas do banco de dados CSD que foram utilizadas para como base para a construção do conjunto de treinamento. Os nomes de cada estrutura correspondem ao seu código específico no CSD.

Considerando que todas as moléculas do conjunto de treinamento apresentam grupamentos capazes de se ionizarem em meio aquoso, a forma correta de realizar a etapa DM é com a utilização das microespécies predominantes de cada molécula no pH utilizado para o ensaio. Para que cada microespécie fosse determinada, foi utilizado o programa Marvin 4.1.8 (ChemAxon Inc.; <http://www.chemaxon.com>). Este possui a ferramenta *pKa plugin* (Figura 35), que possui exatamente o objetivo de prever microespécies a partir da estrutura 2D de um composto. Fornecendo um valor específico para a ferramenta, esta pode devolver ou uma relação das microespécies que estarão presentes em uma determinada faixa de pH, ou apenas a microespécie predominante naquele pH, conforme desejo do usuário. Segundo Hazuda e colaboradores, o pH utilizado no ensaio foi de 7,5. Como resultado, 75 estruturas foram utilizadas na forma ionizada e 10 na forma neutra.

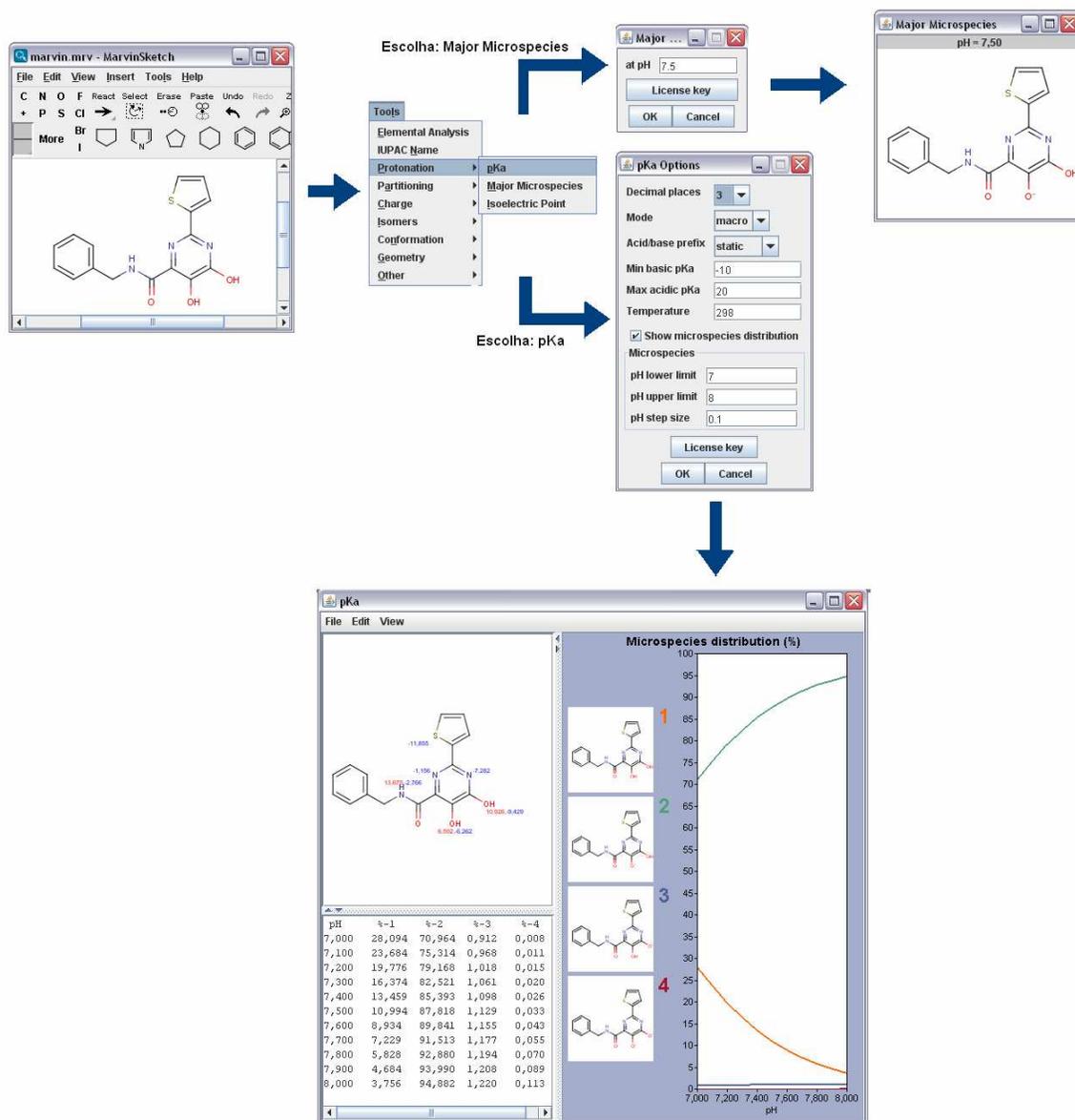


Figura 35. Esquema apresentando as duas formas de se obter as microespécies predominantes de uma molécula em um determinado pH utilizando o programa Marvin 4.1.8.

Cargas de potencial eletrostático (ESP: *electrostatic potential*) ChelpG (*CHarges from E*lectrostatic Potentials using a Grid based method) foram calculadas para cada microespécie. Cargas ESP podem ser consideradas mais adequadas para o estudo de QSAR-4D porque elas são capazes de reproduzir de forma mais correta o potencial eletrostático da molécula

(YOUNG, 2001), o que favorecerá a descrição das interações de Coulomb entre as sondas e as moléculas em estudo.

A partir dos arquivos de saída do programa Gaussian 03 (OUT) foram obtidos arquivos MOL2 utilizando o programa Open Babel 2.1.1 (MORLEY, 2006). As codificações em texto destes arquivos foram utilizadas no servidor PRODRG 2.5 (SCHUETTELKOPF e van AALTEN, 2004; <http://davapc1.bioch.dundee.ac.uk/prodrg>) para obter os arquivos de geometria (GRO) e de topologia (TOP) moleculares para serem utilizados como arquivos de entrada para o programa GROMACS.

Como o campo de força do GROMACS (ffG43a1) é parametrizado para cargas empíricas de Gasteiger e Marsili (1980), houve a necessidade de obtenção de novos valores de cargas parciais. Para isto foi utilizado o programa HyperChem 7 (opção *QSAR properties*).

A abordagem de Gasteiger e Marsili (1980) para o cálculo de cargas parciais é denominada *partial equalization of orbital electronegativity* (PEOE). Este método considera a identidade de cada átomo junto com suas conectividades para reproduzir a distribuição de carga nas moléculas, garantindo assim que átomos de um mesmo tipo, mas que estejam em vizinhanças diferentes (p.ex, oxigênio do $-OH$, $-OCH_3$ e $-O^-$), apresentem cargas diferentes. Como apenas a topologia da molécula determina a natureza da distribuição eletrônica, a conformação de uma molécula não é levada em consideração, o que garante uma forma rápida de obtenção de valores de cargas parciais. Da mesma forma, este é um método insensível a variações de cargas parciais induzidas por alterações conformacionais, e os valores para cada átomo serão sempre os mesmos, independente se for utilizada uma conformação de máxima ou mínima energia (GASTEIGER e MARSILI, 1980; HYPERCUBE, 2002).

O método PEOE é iterativo. O cálculo inicia-se de um conjunto inicial de cargas, até que as eletronegatividades de todos os átomos estejam balanceadas, resultando em um valor de

carga parcial. Assim, no lugar de se partir de cargas de valor zero para todos os átomos, as cargas ChelpG obtidas anteriormente puderem ser utilizadas como um ponto de partida quimicamente razoável. Finalmente, a implementação deste método no HyperChem apresenta parâmetros para os elementos C, H, N, O, S e halogênios (HYPERCUBE, 2002), exatamente os tipos de átomos que forma as moléculas do conjunto de treinamento.

Finalmente, as simulações DM de todos os compostos foram realizadas em uma caixa virtual cúbica com propriedades periódicas (Figura 36) específica para cada molécula, tendo como único parâmetro a distância mínima de 1 nm, ou 10 Å, entre a molécula e a borda da caixa. Água explícita foi utilizada como meio solvente. Para as moléculas ionizadas foram utilizados íons sódio (Na^+) ou cloro (Cl^-) como contra-íon para satisfazer a condição de neutralidade do sistema. Cada molécula foi otimizada utilizando os algoritmos de *máxima descida* (*steepest descent*) e gradiente conjugado, com convergência em 50 N de força máxima aplicada aos átomos. O algoritmo máxima descida promove uma convergência mais rápida do sistema e é útil para eliminar grandes interações não-ligadas que são normalmente encontradas em estruturas iniciais, mas apresenta menores chances de encontrar o mínimo de energia. Já o gradiente conjugado considera resultados encontrados em outros ciclos de otimização, e assim aquele ponto da superfície de energia potencial não será mais explorado. Isto aumenta as chances de sucesso em encontrar o mínimo de energia, mas aumenta o custo computacional (HYPERCUBE, 2002).

Após a etapa de minimização, o volume do sistema foi equilibrado utilizando um esquema de aquecimento dividido em etapas de 50 K, 100 K, 200 K e 350 K (20 ps cada). Em seguida, o sistema foi resfriado para 300 K e simulado por 500 ps. Um arquivo de trajetória foi salvo a cada 1000 passos da simulação. As conformações obtidas de cada ligante foram organizadas em um mesmo arquivo GRO. Para a construção *perfis de amostragem conforma-*

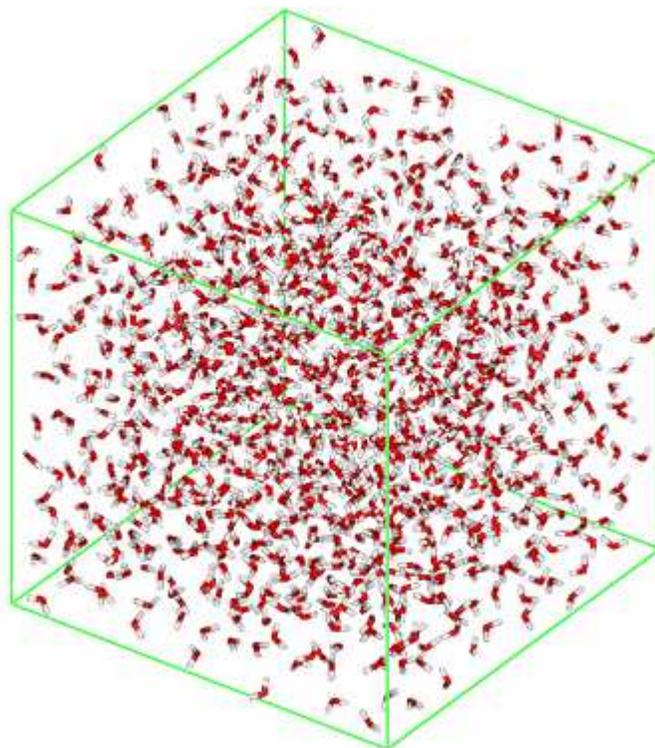


Figura 36. Representação da caixa periódica com solvente explícito utilizada na etapa de dinâmica molecular.

cional foram eliminados os movimentos rotacionais e translacionais que cada conformação apresentou durante a DM e todas as conformações obtidas foram alinhadas com base no farmacóforo, sem considerar as cadeias laterais aromáticas (Figura 19).

3.3 DESCRITORES FÍSICO-QUÍMICOS

A Tabela 4 apresenta um resumo dos programas, classes e quantidade de descritores de cada classe que foram calculados para utilização nos estudos de QSAR-2D. No total, foram utilizados 1291 descritores para esta etapa. Diversos tipos de descritores foram utilizados, com grande destaque para os topológicos, baseados na teoria de grafos, e os topológicos tridimensionais disponibilizados pelo programa Dragon Web Version 3.0 (TODESCHINI *et*

al., 2003). Devido a grande quantidade de descritores, será apresentada a teoria apenas daqueles que forem selecionados que formaram os melhores modelos.

Tabela 4. Informações sobre os programas e descritores utilizados.

Programa	Número de descritores	Classes	Representação estrutural
Marvin 4.1.8	40	Topológicos, constitucionais, solubilidade (Log $D_{7,5}$), estéricos e eletrônicos empíricos.	Estrutura 2D
ALOGPS 2.1	13	Solubilidade (Log P e Log S calculados por várias teorias)	SMILES**
E-Dragon 1.0*	430	Topológicos, constitucionais, estéricos, solubilidade.	SMILES
ETState*	12	Índices de estado eletrotológico	SMILES
Dragon Web Version 3.0	702	Topológicos tridimensionais	Estrutura 3D mais estável
ACD/ChemSketch 10.0	10	Estéricos, constitucionais, solubilidade (ACD/Log P, constante hidrofóbica de Hansch para cadeia lateral aromática)	Estrutura 2D
Gaussian 03/Gauss View***	84	Termodinâmicos, eletrônicos	Estrutura 3D mais estável

*via interface Parameter Client (<http://www.vcclab.org>); ***Simplified Molecular Input Line Entry System*: é uma representação tipográfica utilizada para representar moléculas e reações. Muitos programas utilizados para o cálculo de descritores em QSA(P)R utilizam esta abordagem; *** devido a razoável variação estrutural observada, foram utilizadas para o estudo 2D descritores de cargas de Mulliken, ChelpG e NBO (*Natural Bond Orders*) correspondentes apenas aos átomos que formam a subestrutura DCA do farmacóforo. Estes valores foram obtidos tanto para as estruturas neutras quanto microespécies ionizadas de cada molécula.

Já a base do estudo QSAR-4D é o programa LQTA-QSAR (Figura 37; MARTINS *et al.*, 2009), desenvolvido pelo Laboratório de Quimiometria Teórica e Aplicada da Unicamp (LQTA; <http://lqta.iqm.unicamp.br>). Este programa, escrito na linguagem de programação JAVA (<http://www.java.com>), utiliza o perfil de amostragem conformacional (Figura 38) obtido como resultado do estudo de DM, colocando-o no centro de uma grade virtual. Esta grade pode ter qualquer tamanho definida pelo usuário. As coordenadas da grade utilizada neste trabalho foram obtidas com o auxílio do programa AutoDock tools versão 1.5.2-2007 (SANNER *et al.*, 1998) utilizando a função *Grid* → opção *Grid box* para criar uma caixa

virtual para a molécula com maior volume molecular do conjunto de treinamento, a **P35**. Como resultado, foi construída uma caixa cúbica $28 \times 28 \times 28 \text{ \AA}^3$, com centro nas coordenadas cartesianas $x = 6 \text{ \AA}$, $y = 4 \text{ \AA}$ e $z = 3 \text{ \AA}$. A seguir, cada ponto desta grade é explorado por uma sonda determinada pelo usuário. As *energias de interações tridimensionais* entre estas sondas e o PAC em cada ponto da grade serão medidas, utilizando para isso os potenciais de Coulomb e Lennard-Jones (Equações 5 e 6). Estas energias são tratadas como descritores de campo, que serão utilizados para o estudo QSAR.

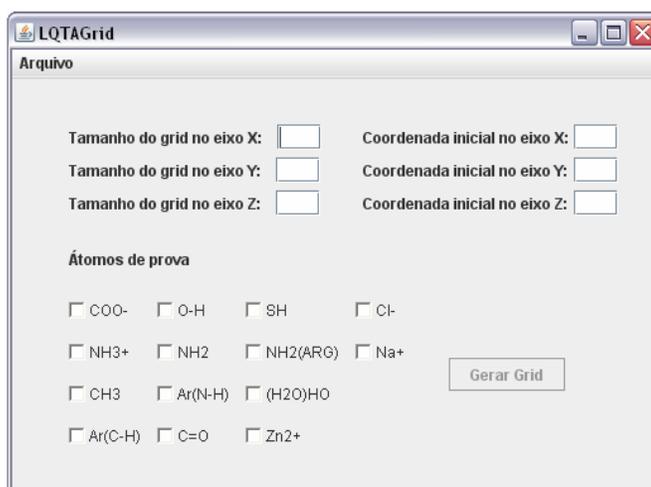


Figura 37. Interface gráfica do programa LQTA-QSAR.

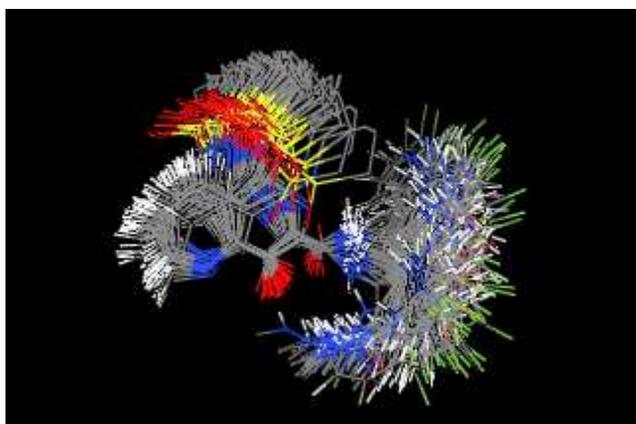


Figura 38. Exemplo de perfil de amostragem conformacional (composto **E21b**).

As coordenadas GRO do perfil de amostragem conformacional e o arquivo TOP contendo as cargas ChelpG foram utilizados como arquivos de entrada para o programa LQTA-QSAR. Considerando as informações disponíveis sobre o mecanismo de ação, foram utilizados descritores obtidos por duas sondas: uma sonda de carbono aromático Ar(C-H), devido à importância da cadeia lateral aromática e de sua interação com o sítio hidrofóbico da HIV-IN, e uma sonda de zinco metálico divalente (Zn^{+2}), que nesta parte do estudo representa o átomo de Mg^{+2} . Como a primeira sonda é neutra, apenas descritores de Lennard-Jones são obtidos. Para a sonda Zn^{+2} , o programa gera ambos os tipos de descritores. Considerando as dimensões da grade virtual utilizada, foram obtidos 21.952 descritores de Coulomb e 21.952 descritores de Lennard-Jones para a sonda Zn^{+2} , e 21.952 descritores de Lennard-Jones para a sonda Ar(C-H). Assim, o estudo QSAR-4D foi realizado com 65.856 descritores.

3.4 QSAR

Em geral, considera-se que quando a correlação linear de Pearson (r), ou coeficiente de correlação, entre um descritor e a atividade biológica em estudo é muito baixa, este não apresenta informação suficientemente relevante para ser utilizado em um estudo QS(P)R. Desta forma, um processo muito utilizado em estudos destes tipos é a eliminação daqueles descritores cujo valor absoluto de r esteja abaixo de um determinado valor, normalmente definido pelo pesquisador com base em critérios próprios. Isto auxilia na redução de tempo de trabalho, pois reduz o número total de variáveis de centenas ou milhares para dezenas ou centenas. Além disso, as variáveis restantes são aquelas com maior relevância estatística, apresentando maior chance de auxiliar a originar bons modelos. Assim, para o estudo QSAR-

2D foram excluídos os descritores que apresentaram $r < |0,3|$. Para o estudo QSAR-4D foi utilizado $r < |0,2|$, pois poucas variáveis apresentavam $r < |0,3|$.

Considerando as vantagens da metodologia PLS, assim como a quantidade de descritores tanto do estudo QSAR-2D quanto, e principalmente, do QSAR-4D, este método de regressão pode ser considerado o mais adequado para a construção dos modelos de predição deste trabalho. Porém, no estudo 2D os diferentes descritores apresentavam escalas numéricas muito diferentes, mas a importância relativa de cada um era desconhecida. Assim, foi realizado o pré-processamento dos dados denominado *autoescalamento*: cada valor de um descritor é subtraído pelo valor médio obtido e dividido pelo desvio padrão. O mesmo processo é realizado com a variável dependente. Este processo faz com que todas as variáveis passem a apresentar a mesma escala numérica, e assim a importância de cada descritor em uma eventual relação com a variável dependente não será superestimada ou subestimada (WOLD, SJÖSTRÖM e ERIKSSON, 2001; FERREIRA e KIRALJ, 2008).

Porém, dependendo dos tipos de variáveis presentes, autoescalar os dados pode mascarar a importância relativa de um ou mais descritores. Isso pode ocorrer quando os valores alguns valores apresentam ordens de magnitude muito maiores, e em especial quando são do mesmo tipo (ou seja, possuem a mesma unidade). Assim, o alto desvio padrão utilizado no processo de autoescalamento pode prejudicar os resultados obtidos. Para contornar este problema, pode ser utilizado o pré-processamento conhecido como *escalamento em blocos* (do inglês *block-scaling*). De modelo geral, nesse procedimento os dados são pré-processados separadamente (os “blocos”) de modo a permitir que sejam selecionadas aqueles descritores com variância maximizada em relação a atividade biológica, sem o risco dessa importância ser mascarada por outros descritores (ORTIZ *et al.*, 1994; KNEKTA *et al.*, 2004).

Nesse trabalho, o processo de escalamento em blocos foi aplicado aos descritores do estudo QSAR-4D, pois muitos dos descritores de Lennard-Jones obtidos (assim como valores individuais para alguns compostos) apresentavam ordens de magnitude muito superiores aos descritores de Coulomb. Assim, autoescalar os dados poderia dar a mesma importância para todos os descritores dentro da caixa virtual cúbica, tanto os distantes quanto os próximos dos inibidores, o que poderia prejudicar a interpretação do modelo em relação ao mecanismo de inibição da reação ST. Para isso, a matriz completa com os 65.856 descritores foi tratada no Matlab 7 (MATHWORKS, 2006), onde foi dividida em duas, uma com os descritores de Lennard-Jones e outra com os descritores de Coulomb. Em seguida, foi utilizado um algoritmo *in-house*, denominado *LQTAcutoff* (ver *script* no Anexo 1), que utilizou um critério de corte de 30,00 kcal/mol (125,52 KJ/mol) nos descritores de Lennard-Jones, adequando a ordem de magnitude, porém sem prejudicar a importância relativa entre um e outro descritor desta mesma classe. Após esta etapa, foram realizados os processos de seleção de variáveis (ver a seguir) com as duas matrizes em separado e com dados *centrados na média* (pré-processamento onde a média de cada variável x é subtraída de cada valor individuais da mesma). Após a seleção das variáveis mais relevantes, essas foram unidas em uma única matriz. A partir deste ponto, os dados foram autoescalados.

Para o QSAR-4D, foi utilizado outro algoritmo *in-house* (Anexo 2), também implementado no programa Matlab 7 (MATHWORKS, 2006), para excluir aqueles descritores com dispersão anormal em relação a faixa de variação da atividade biológica. Este processo foi necessário, pois muitos dos descritores de Coulomb e Lennard-Jones apresentavam ajuste de dados anormais (Figura 39), o que poderia levar a um modelo com capacidade de predição ruim. O algoritmo divide o intervalo de variação dos dados x (nesse caso, os descritores) em subintervalos menores, mantendo apenas aqueles que apresentavam as maiores dispersões. O

quando menor a f_{\max} (frequência máxima) calculada, mais disperso é o descritor em relação à faixa de variação da atividade biológica.

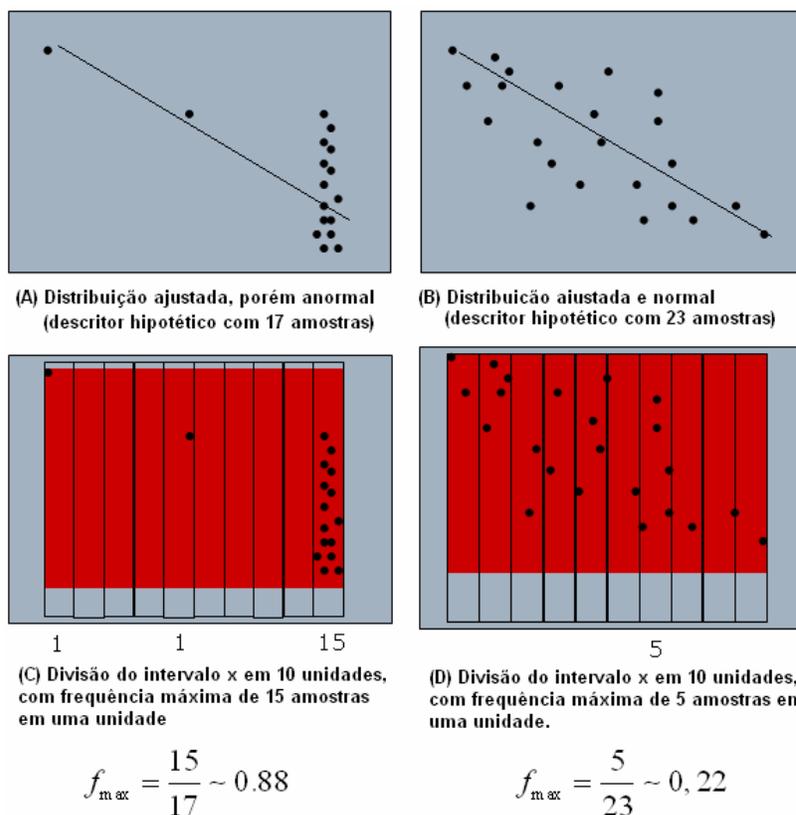


Figura 39. Representação esquemática do funcionamento do algoritmo *in-house* planejado para selecionar descritores com tendência linear e distribuição normal em relação a atividade biológica. Algoritmo desenvolvido por Euzébio Guimarães Barbosa (grupo LQTA). Figuras cedidas pelo mesmo e adaptadas pelo autor.

Após as etapas iniciais de redução de variáveis de cada estudo, as matrizes resultantes foram submetidas à seleção de variáveis utilizando o método *Ordered Predictors Selection – OPS* (TEÓFILO, MARTINS e FERREIRA, 2009). Esta metodologia, também desenvolvida no LQTA, opera em três etapas: (i) obtenção de um vetor informativo (o vetor de regressão, o vetor dos coeficientes de correlação entre os descritores e a atividade biológica, e o produto de ambos); (ii) ordenamento decrescente das variáveis de acordo com os valores dos elementos do vetor; e (iii) investigação das variáveis ordenadas por meio da construção de modelos

utilizando regressão por PLS. Os modelos obtidos devem ser ordenados segundo algum critério estatístico que indique a qualidade dos mesmos, como o SEV ou o Q^2_{LOO} (ver adiante). Para esta etapa, foram utilizadas versões do método OPS implementados em dois programas: no Matlab (MATHWORKS, 2006), e em um programa desenvolvido pelo grupo LQTA *QSARmodelling* (ainda não publicado).

Para facilitar a interpretação físico-química e melhorar a qualidade estatística, os modelos selecionados foram refinados no programa Pirouette 4 (INFOMETRIX, 2007). Aqueles descritores que pouco contribuíam para a qualidade estatística do modelo ou, com base nas informações sobre o mecanismo de ação, que não fizessem sentido físico, foram excluídos. Amostras *outliers* foram detectadas por consulta no gráfico de resíduos de *Student*, σ_i , pelos valores de *leverage* (calculados no Pirouette 4 pelas Equações 20 e 21, respectivamente), h_i , de cada amostra (Figura 40). Neste gráfico, os limites dos resíduos (linhas horizontais) correspondem a duas vezes o σ_i ($2\cdot\sigma_i$, correspondente a 95% de significância, $\alpha = 0,05$) enquanto o limite do leverage (linha vertical) é calculado pelo programa Pirouette 4 utilizando a relação $h_{limite} = 2k/n$, onde k é o número de variáveis latentes do modelo PLS e n é o número de compostos utilizado para construí-lo. Foram retirados especialmente aqueles compostos que se localizavam nas regiões superior ou inferior direita (ou seja, resíduos e leverage altos). Também foram retiradas amostras onde apenas os valores dos resíduos de Student fossem muito altos (muito acima do limite de $2\cdot\sigma_i$). Ainda houve cuidado para que a retirada de *outliers* fosse feita apenas nos casos necessários, para que a redução do tamanho do conjunto de treinamento fosse reduzido apenas o suficiente para a obtenção de cada modelo (Figura 40A).

$$\sigma_i = \frac{f_i}{\sqrt{\left(\frac{1}{n-p}\right) \sum f_i^2 \times \sqrt{(1-h_i)}}} \quad (20)$$

$$h_i = \frac{1}{n} + t_i^T \left(\mathbf{T}_p^T \mathbf{T}_p \right)^{-1} t_i \quad (21)$$

Onde: h_i = leverage da amostra i ;
 t_i = escore da amostra i ;
 \mathbf{T}_p = matriz de loadings;
 f_i = resíduo de predição da amostra i .
 p = número de LVs do modelo;
 n = número de amostras do modelo.

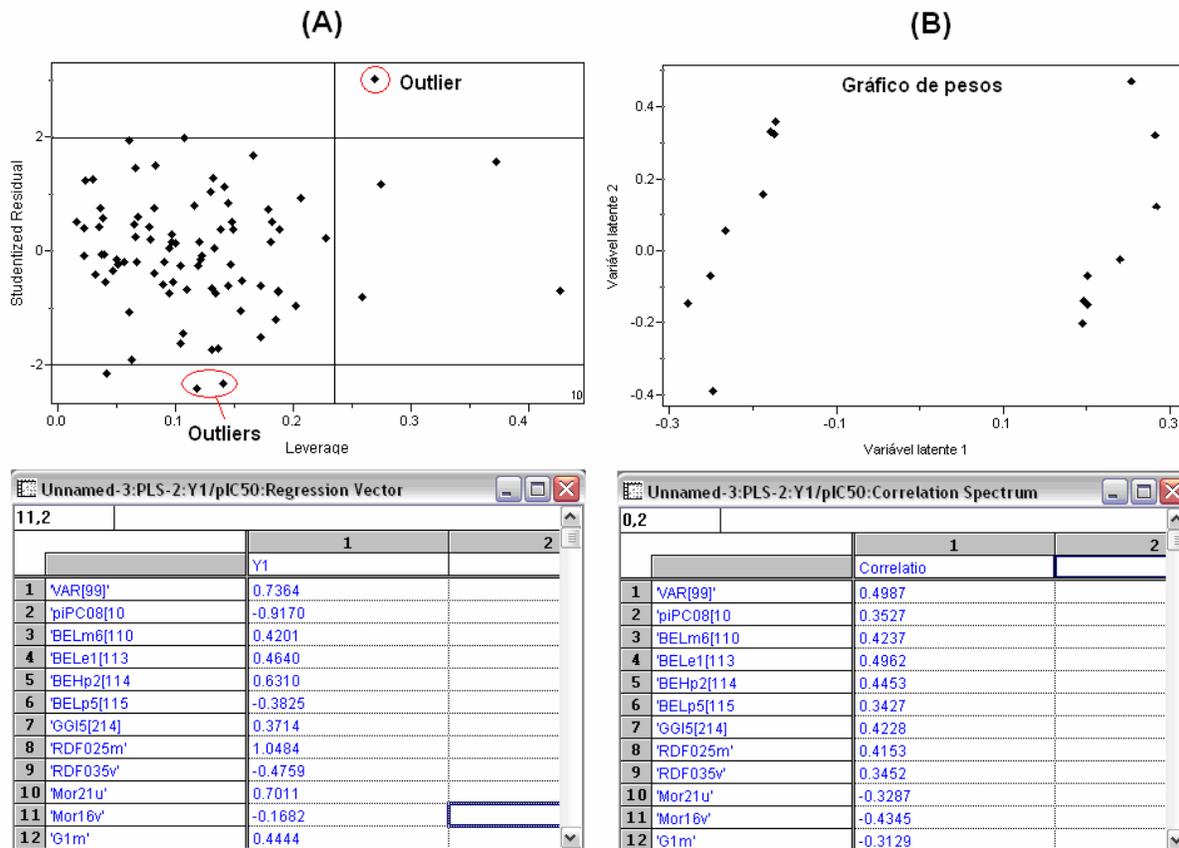


Figura 40. Gráficos e tabelas do Pirouete 4 utilizadas para eliminação de outliers (A) e redução de variáveis (B).

A redução do número de descritores foi realizada utilizando como guia principal seus coeficientes de regressão autoescalados. Segundo Wold (1998) e Roy e Roy (2008), a

eliminação daqueles descritores que apresentam os menores coeficientes auxilia no descarte daqueles que contribuam de forma insignificante no modelo, podendo até mesmo maximizar os resultados de R^2 e Q^2_{LOO} e minimizar o SEV . Mas uma variável importante para a modelagem da variável dependente também pode ser identificada pelos resultados de seus pesos das LVs do modelo PLS (WOLD, 1998), e estes resultados também foram consultados (Figura 41B). Em caso de dúvidas, os valores dos coeficientes de correlação entre cada descritor e a variável dependente foram consultados.

3.5 QUALIDADE DOS MODELOS

As Tabelas 5 e 6 apresentam os parâmetros utilizados para avaliar a qualidade estatística dos modelos. Para cada parâmetro foram utilizados os limites máximos e/ou mínimos recomendados pela literatura.

Segundo a literatura, o ajuste de um modelo ser aprovado, ele deve apresentar $R^2 > 0,6$ e o resultado de seu teste- F maior que seu valor de referência (ou F -crítico) encontrado em tabelas de estatística (sendo este determinado por $F_{p,n-p-1}$, onde p é o número de LVs do modelo e n é o número de compostos). Isto significa que um modelo deve ser capaz de explicar pelo menos 60% ($R^2 \times 100$) da variabilidade da atividade biológica (ou seja, deve ajustar 60% da informação disponível no modelo). Já o teste F pode ser explicado com uma razão entre a variabilidade explicada pelo modelo e a variabilidade que permanece sem explicação. Quanto mais acima de seu valor crítico for o resultado deste teste, mais significativo será o modelo, sendo mais comumente utilizado o nível de 95% ($\alpha = 0,05$). Quanto ao SEC , quanto mais próximo de zero, mais variabilidade ou informação pode ser explicada pelo modelo (WOLD e ERIKSSON, 1998; GÁUDIO e ZANDONADE, 2001; FERREIRA, MONTANARI e GAUDIO, 2002; ERIKSSON *et al.*, 2003; GOLBRAIKH *et al.*, 2003; ZHANG *et al.*, 2007; FER-

Tabela 5. Parâmetros estatísticos utilizados para avaliar a qualidade interna dos modelos e suas respectivas equações.

Parâmetro	Significado	Equação
R^2	Coefficiente de determinação múltipla ^a	$1 - \frac{\sum_i (y_{obsi} - y_{ci})^2}{\sum_i (y_{obsi} - \bar{y}_{ci})^2}$
SEC	Erro padrão de calibração ^a	$\sqrt{\frac{\sum_i (y_{obsi} - y_{ci})^2}{n - p - 1}}$
$PRESS_{cal}$	Soma dos resíduos de predição ao quadrado da calibração ^a	$\sum_i (y_{obsi} - y_{ci})^2$
F	Teste F (com 95% de confiança, $\alpha = 0,05$) ^a	$\frac{\frac{\sum_i (y_{obsi} - y_{ci})^2}{k}}{\frac{\sum_i (y_{obsi} - \bar{y}_{obs})^2}{n - p - 1}}$
Q^2_{LNO}	Coefficiente de correlação da validação cruzada (“leave- N -out”, LNO) ^{b,c}	$1 - \frac{\sum_i (y_{obsi} - y_{vi})^2}{\sum_i (y_{obsi} - \bar{y}_{obs})^2}$
SEV	Erro padrão da validação cruzada “leave-one-out” ^c	$\sqrt{\frac{\sum_i (y_{obsi} - y_{vi})^2}{n}}$
$PRESS_{val}$	Soma dos resíduos de predição ao quadrado da validação cruzada LOO^c	$\sum_i (y_{obsi} - y_{vi})^2$

^aqualidade do ajuste; ^bvalidação cruzada, se $N = 1$, e teste de robustez *leave- N -out*, se $N > 1$; ^cqualidade da predição interna; ^dvalidação externa; y : atividade biológica; \bar{y} : atividade biológica média; obs : valores experimentais; c : atividade estimada no modelo de regressão; v : valor obtido na validação cruzada; n : número de compostos do conjunto de treinamento; p : número de variáveis latentes; \bar{y}_{obs} : atividade média observada para o conjunto de treinamento completo.

REIRA e KIRALJ, 2009).

Quanto aos resultados da validação cruzada LOO , o critério para o SEV é o mesmo do que SEC : quanto menor, melhor. Quanto à quantidade de informação ou variabilidade predita, é de consenso geral que o mínimo aceitável é $Q^2_{LOO} > 0,5$ (capacidade de prever pelo menos

Tabela 6. Parâmetros estatísticos utilizados para avaliar a qualidade externa dos modelos e suas respectivas equações.

R^2_{pred}	Coefficiente de correlação de predição ^d	$1 - \frac{\sum_i (y_{obsi} - y_{evi})^2}{\sum_i (y_{obsi} - \bar{y}_{obs})^2}$
ARE_{pred}	Erro relativo médio da validação externa ^d	$\frac{\sum_i (y_{obsi} - y_{evi})^2}{y_{obsi}} * 100$ n
SEP	Erro padrão de predição externa ^d	$\sqrt{\frac{\sum_i (y_{obsi} - y_{evi})^2}{n}}$
k e k'	Inclinações das linhas das regressões lineares de predição ^d	$\frac{\sum_i (y_{obsi} - y_{evi})}{\sum_i y_{evi}} \text{ e } \frac{\sum_i (y_{obsi} - y_{evi})}{\sum_i y_{obsi}}$

^aev: valor obtido na validação externa; n: número de compostos do conjunto de treinamento; p: número de variáveis latentes; \bar{y}_{obsi} : atividade média observada para o conjunto de treinamento completo; test: conjunto-teste (validação externa). Demais legendas podem ser encontradas na Tabela 5.

50% da variabilidade) (FERREIRA, MONTANARI e GAUDIO, 2002; GOLBRAIKH e TROPSHA, 2002; ERIKSSON *et al.*, 2003; GOLBRAIKH *et al.*, 2003; ZHANG *et al.*, 2007). Também deve ser observada a diferença entre os valores de R^2 e Q^2_{LOO} . Diferenças maiores que 0,2 – 0,3 unidades podem constituir um forte indício de sobreajuste dos dados (*overfitting*) (FERREIRA e KIRALJ, 2009).

Seguindo uma sugestão de Wold e Eriksson (1998), um terceiro parâmetro foi utilizado para auxiliar a avaliar a capacidade de predição interna do modelo e assim aumentar o rigor sobre esta etapa do estudo: a soma dos quadrados dos resíduos de predição obtidos na validação cruzada ($PRESS_{val}$: *predictive residual sum of squares of validation*). Este parâmetro pode conter contribuições dos resultados obtidos para cada composto no processo de validação cruzada LOO. Observando a equação Q^2_{LOO} (Tabela 5), pode-se observar que este termo corresponde ao numerador da equação. Assim, quando ele é comparado com o

denominador (a soma dos quadrados dos valores de resposta, SS_y), fica claro que, quanto menor o $PRESS_{val}$, maior o valor de Q^2_{LOO} . Assim, quanto menor $PRESS_{val}$ em relação a seu SS_y , o nível de significância do modelo será maior e a probabilidade de estar ocorrendo correlação por chance será menor (WOLD e ERIKSSON, 1998). Por analogia, o mesmo pode ser dito para a soma dos quadrados dos resíduos de predição na calibração ($PRESS_{cal}$: *predictive residual sum of squares of calibration*): quanto menor este valor, melhor o ajuste do modelo.

A divisão do conjunto de dados em um conjunto de treinamento e um conjunto teste foi realizada após a construção de um modelo com todas as amostras, denominado *modelo auxiliar*. O conjunto de compostos utilizados na validação externa é denominado conjunto de teste (*test set*). A subdivisão de um conjunto de compostos em um conjunto de treinamento e seu respectivo conjunto teste pode ser feita de várias formas (GOBRAIKH *et al.*, 2003), mas alguns critérios básicos devem ser seguidos: o conjunto de teste deve ser formado por pelo menos cinco compostos, os quais devem estar distribuídos da forma mais uniforme possível pelo intervalo de atividade biológica do conjunto de treinamento. Também devem ser estruturalmente similares e representativos do mesmo (WOLD e ERIKSSON, 1998; GOBRAIKH e TROPSHA, 2002). Após a divisão, é obtido o *modelo real*, aquele que efetivamente será utilizado para a validação externa (FERREIRA e KIRALJ, 2009). Este procedimento foi realizado para que um modelo mais representativo fosse construído com maior quantidade possível de amostras. Por regra, o modelo real deve apresentar diferença estatística mínima em relação a seu modelo auxiliar. Assim, para auxiliar a dividir as amostras também foi utilizado o gráfico dos resíduos de Student pelos valores de Leverage fornecido pelo programa Pirouette 4 (Figura 40) (INFOMETRIX, 2007). Assim, foi possível determinar para cada modelo quais amostras apresentavam os menores leverages, pois era esperado que a

retirada destes compostos não levasse a alterações drásticas da qualidade estatística do modelo. Mas também foram avaliadas quais amostras eram as mais representativas de todo o intervalo de atividade biológica e da variabilidade estrutural do conjunto de treinamento. Desta forma, caso um composto apresentasse um leverage relativamente alto, mas se enquadrasse melhor nestas duas últimas condições, este era o escolhido (desde que sua retirada não alterasse consideravelmente o modelo).

Os limites mínimos para o R^2_{pred} de um modelo não são bem definidos na literatura, apesar destes parâmetros serem amplamente utilizados. Em alguns casos, é possível até mesmo encontrar um quadrado do coeficiente de correlação de predição, R^2_{pred} ou Q^2_{pred} , superior ao Q^2_{LOO} , situação que não ocorre quando comparamos R^2 e Q^2_{LOO} . Resultados como estes, apesar de aparentemente contraditórios, podem ocorrer porque os resultados obtidos na validação externa não estão relacionados em nada aos da validação cruzada (ROY e ROY, 2008). Este fenômeno é conhecido como *Paradoxo de Kubinyi* (DOWEYKO, 2008). As melhores sugestões vêm dos estudos de Gobraikh e Tropsha (2002) (R^2_{pred} ao menos entre 0,5 – 0,6), Melagraki e colaboradores (2007) ($R^2_{pred} > 0,6$) e Roy e Roy (2008a, 2008b) ($R^2_{pred} > 0,5$), sendo que este último será o adotado neste trabalho. Porém, segundo os autores, utilizar apenas o R^2_{pred} para avaliar a qualidade da predição externa não é suficiente porque o resultado desejado é um ajuste exato dos dados, e não apenas uma correlação linear. Assim, caso os dois testes permaneçam dentro de determinados limites, o ajuste pode ser considerado aceitável e o modelo apresenta boa capacidade de predição. Por outro lado, os limites para os testes com as inclinações k/k' e com os coeficientes de determinação são muito bem definidos (GOBRAIKH e TROPSHA, 2003): o resultado de pelo menos uma das inclinações deve estar dentro do intervalo 0,85 – 1,15, e o valor absoluto da diferença entre R^2_0 e R'^2_0 deve ser menor que 0,3 ($|R^2_0 - R'^2_0| < 0,3$). Considerando a importância da etapa de validação externa, a definição

inambígua destes critérios constitui um auxílio importante para estudos QSAR. Finalizando, um quarto parâmetro foi introduzido no estudo por idéia dos autores deste trabalho: a avaliação do erro relativo médio da validação externa (ARE_{pred} : *average relative error of prediction*). Como este parâmetro é baseado nos erros entre os valores preditos e observados, seu resultado fornece, em teoria, a possibilidade de erro (em %) que se pode esperar do modelo quando se sua efetiva utilização para a predição de novos compostos.

A robustez dos modelos foi avaliada pela validação LNO. Este teste não possui limites mínimos. Porém, o Q^2_{LNO} médio deve ser alto, preferencialmente igual ou próximo ao Q^2_{LOO} , com mínimas oscilações. Caso o modelo atenda a este critério, ele pode ser considerado como sendo robusto (MELAGRAKI *et al*, 2007). Para que o teste seja realmente efetivo, N deve representar uma fração significativa das amostras (20 a 30%, no mínimo) (FERREIRA e KIRALJ, 2009). Assim, conforme o tamanho do conjunto de treinamento utilizado, a validação LNO foi realizada 6 ou 17 vezes (ou seja, $N = 1$ a 6 ou 1 a 17). Cada “ N ” foi realizado em triplicata, onde em cada replicata as linhas da matriz de dados (X e Y) foram aleatorizadas. Este procedimento é desejável, já que este teste é sensível à retirada de amostras em determinadas combinações (FERREIRA e KIRALJ, 2009).

Já quanto ao teste de randomização de y , não existe um consenso na literatura quanto aos resultados que devem ser buscados, nem quanto aos valores limites. Wold e Eriksson (1998) sugerem que os modelos randomizados devem apresentar resultados máximos de 0,4 para o R^2 e 0,3 para o Q^2_{LOO} . Hashemianzadeh e colaboradores (2008) não estabeleceram limites, mas consideraram aceitáveis os resultados máximos de $R^2 = 0,68$ e $Q^2 = 0,383$. Freitas (2008) procedeu de forma parecida, porém considerou aceitável o resultado de $R^2 = 0,7$, sendo que não foi realizado a avaliação dos valores de Q^2_{LOO} . A proposta mais clara vem de Eriksson e colaboradores (2003): segundo os autores, um bom procedimento é construir um gráfico

utilizando os valores absolutos das correlações lineares de Pearson, $|r|$, entre a coluna y original e cada uma daquelas obtidas após o processo de randomização no eixo das abscissas, e os valores de R^2 ou Q^2_{LOO} de cada um dos modelos (original e dos modelos obtidos após a randomização) no eixo das ordenadas, e realizar uma nova regressão linear, obtendo assim dois gráficos de reta. Caso os interceptos de ambas as regressões fiquem abaixo de 0,3 para o gráfico R^2 e 0,05 para o gráfico Q^2_{LOO} , (exemplo na Figura 41), o modelo pode ser caracterizado como livre de correlação ao acaso (FERREIRA e KIRALJ, 2009).

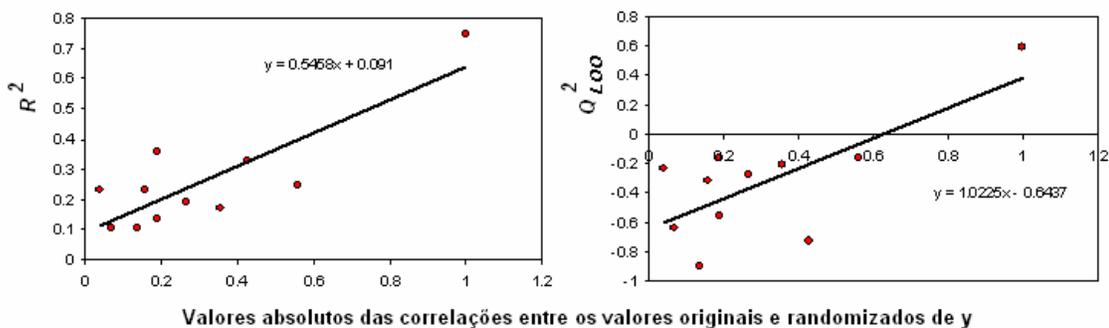


Figura 41. Representação gráfica da análise dos resultados de um de randomização de y segundo os critérios de Eriksson e colaboradores (2003). Os pontos distantes dos demais (valores = 1) correspondem aos valores obtidos para o modelo. O intercepto também pode ser observado na equação da reta. Teste feito com valores hipotéticos e dez randomizações, apenas como exemplo.

Também há diferenças no consenso quando a quantidade de randomizações que podem ou devem ser realizadas em um estudo. Em alguns trabalhos podem ser encontradas randomizações na casa das centenas, ou randomizações baseadas no tamanho do conjunto de treinamento. Apesar disto, Ferreira e Kiraj (2009) mostraram recentemente que os resultados de processos com dez ou com mil randomizações são praticamente idênticos quando é utilizado o método proposto por Eriksson e colaboradores (2003). Assim, baseando-se no tamanho do conjunto de treinamento, foram realizadas 50 randomizações.

Segundo a OECD (2004), tanto a validação LNO quanto o teste de randomização de y correspondem a medidas da qualidade interna de um modelo. Assim, estes testes foram realizados no modelo auxiliar antes da divisão do conjunto completo de amostras em conjunto de treinamento e conjunto teste. Para estas etapas também foi utilizada uma rotina do Matlab (MATHWORKS, 2006), desenvolvido *in-house* pelo grupo LQTA (Anexo 2).

4. RESULTADOS E DISCUSSÃO

4.1 QSAR-2D

4.1.1 Estudo QSAR-2D preliminar

Devido a uma preocupação quanto a grande variabilidade estrutural do conjunto de treinamento em estudo, e também para a preparação de um protocolo de estudo adequado, foi realizado um estudo preliminar com os 33 compostos descritos por Petrocchi e colaboradores (2007) (Figura 27 e Anexo 1). As 4,5-diidroxipirimidinas carboxamidas constituem o maior subconjunto entre os 85 compostos, apresentam pouca variabilidade estrutural (apenas a cadeia lateral aromática) e sua atividade biológica varia de 4,30 a 8,00 unidades logarítmicas. Considerando os motivos que levaram à realização desta etapa, foram utilizados apenas 162 descritores, todos considerados de interpretação mais simples. A maioria dos descritores topológicos baseados na teoria de grafos e todos os tridimensionais calculados pelo programa Dragon 3.0 não foram utilizados nesta etapa. A redução de descritores utilizando o critério de eliminação $r < |0,3|$ forneceu uma matriz 85x63, tratada com uma versão do método OPS para utilização no Matlab 2006 (MATHWORKS, 2006). Para a classificação dos modelos, foi utilizado o parâmetro Q^2_{LOO} .

O modelo selecionado pelo método OPS foi refinado no Pirouette 4. Com retirada de dois descritores e três compostos *outliers* (8, 21 e 34), detectados a partir do gráfico de Leverages pelos resíduos de *Studentized*, foi obtido o Modelo I. Esse foi formado por duas variáveis latentes com 59,54% da informação original. Os descritores selecionados foram a

energia do orbital molecular ocupado de mais alta energia (E_{HOMO}), a componente vetorial da polarizabilidade total no plano y (α_{yy}), a energia total (E_{T}), e a soma dos valores eletrotológicos de ligações aromáticas carbono-carbono em que os carbonos não estão substituídos (ou *bond-type E-state index* SeaC2C2aa) (Tabela 7). Os referidos descritores apresentam coeficientes de correlação de Pearson (r) com as atividades biológicas de 0,508, 0,571, -0,399 e 0,542, respectivamente. Ao formarem um modelo multivariado, tais propriedades foram capazes de explicar 68,10% ($R^2 \times 100$, $R^2 = 0,681$) e prever 57,67% ($Q^2_{\text{LOO}} \times 100$, $Q^2_{\text{LOO}} = 0,577$) da variância total. O resultado para o teste- F ficou muito acima do valor crítico tabelado para um intervalo de confiança de 95,0% ($\alpha = 0,05$). Finalmente, o resultado de $PRESS_{\text{val}}$ ficou bem abaixo que seu respectivo valor de referência SS_y .

$$\text{pIC}_{50} = +127,168(E_{\text{HOMO}}) + 0,062(\alpha_{yy}) - 0,001(E_{\text{T}}) + 0,100 (\text{SeaC2C2aa}) \quad (\text{I}) \\ + 28,665$$

$n = 30$; outliers: 3; LVs = 2; informação acumulada: 59,539%; $R^2 = 0,681$; $SEC = 0,568$; $PRESS_{\text{cal}} = 8,716$; $F_{(2,27)} = 28,967$ ($cF = 3,350$); $Q^2_{\text{LOO}} = 0,576$; $SEV = 0,622$; $PRESS_{\text{val}} = 11,623$ ($SS_y = 27,463$).

Os resultados da validação LNO e da randomização de y são apresentados na Figura 42. O modelo apresenta um Q^2_{LNO} médio estável em torno do Q^2_{LOO} , com pequenas flutuações dos desvios padrões, calculados para três repetições, para cada N . O maior valor encontrado foi referente a L7O (0,585), enquanto o menor refere-se a L2O (0,560). As randomizações ficaram dentro dos limites (interceptos das regressões correspondentes a R^2 e Q^2_{LOO} menores que 0,3 e 0,05, respectivamente), o que indica que a variância explicada pelos modelos não é devido à correlação ao acaso. Assim, o modelo pode ser considerado robusto.

Tabela 7. Descritores utilizados para formulação do Modelo I, atividades observadas e valores previstos pela validação LOO (*outliers* já excluídos).

	$E_{\text{HOMO}}^{(a)}$	$\alpha_{yy}^{(b)}$	$E_T^{(c)}$	SeaC2C2aa ^(d)	pIC ₅₀ obs	pIC ₅₀ pred	Resíduos
P4	-0,214	39,000	-1405,705	11,938	7,070	6,761	0,309
P9	-0,215	47,319	-1445,023	12,137	7,700	7,179	0,521
P10	-0,217	37,740	-1366,392	11,619	6,000	6,280	-0,280
P11	-0,213	50,826	-1520,039	14,159	7,520	8,125	-0,605
P12	-0,214	42,003	-1598,670	12,465	7,400	7,253	0,147
P13	-0,214	44,465	-1598,666	12,465	6,210	7,485	-1,275
P14	-0,211	30,463	-1559,327	14,306	6,670	7,364	-0,694
P15	-0,214	39,024	-1483,133	10,451	7,700	6,681	1,019
P16	-0,214	39,031	-1483,133	10,451	6,210	6,761	-0,551
P17	-0,218	38,296	-1421,739	7,604	4,780	5,936	-1,156
P18	-0,222	33,891	-1415,701	3,940	4,300	4,831	-0,531
P19	-0,222	34,629	-1399,664	5,626	4,770	4,894	-0,124
P20	-0,215	36,725	-1726,455	8,111	7,000	6,462	0,538
P22	-0,217	38,019	-1742,498	3,972	5,600	6,044	-0,444
P23	-0,215	49,411	-1880,110	10,300	8,000	7,659	0,341
P24	-0,217	51,029	-1880,110	10,301	7,300	7,625	-0,325
P25	-0,206	45,101	-1537,278	10,149	7,700	8,389	-0,689
P26	-0,212	44,565	-1520,232	9,817	7,040	7,333	-0,293
P27	-0,213	44,845	-1520,230	7,868	7,300	6,979	0,321
P28	-0,213	40,498	-1520,230	7,866	6,260	6,772	-0,512
P29	-0,214	42,056	-1445,023	10,123	6,960	6,834	0,126
P30	-0,214	42,880	-1445,026	8,095	7,400	6,636	0,764
P31	-0,214	41,426	-1445,026	8,143	7,150	6,568	0,582
P32	-0,214	47,645	-1636,765	7,896	6,890	7,239	-0,349
P33	-0,214	43,005	-1636,768	9,366	7,300	7,047	0,253
P35	-0,217	41,587	-1865,299	7,441	7,700	6,539	1,161
P36	-0,212	37,662	-1504,938	5,911	6,790	6,416	0,374
P37	-0,216	39,381	-1504,936	18,471	7,150	7,411	-0,261
P38	-0,216	38,968	-1504,936	16,427	8,000	6,889	1,111
P39	-0,219	44,985	-2324,888	16,505	8,000	8,402	-0,402

^(a) em e.V; obtido no programa Gaussian 03;

^(b) calculado com a opção Charge Plugin do programa Marvin 4.1.8;

^(c) em Hartress; obtido no programa Gaussian 03;

^(d) calculado com o programa *ESTate* implementado na interface PClient (<http://www.vcclab.org>).

Utilizando o gráfico de Leverage pelos resíduos de Student como guia, foram selecionados cinco compostos (**P10**, **P12**, **P22**, **P26** e **P36**) como conjunto teste. Todos apresentavam baixos *leverages* e valores de pIC₅₀ distribuídos de modo a representar toda a variação de potência. O modelo sem tais compostos (Modelo II) apresentou parâmetros estatísticos semelhantes aos do Modelo I, e assim ambas as equações podem ser considera-

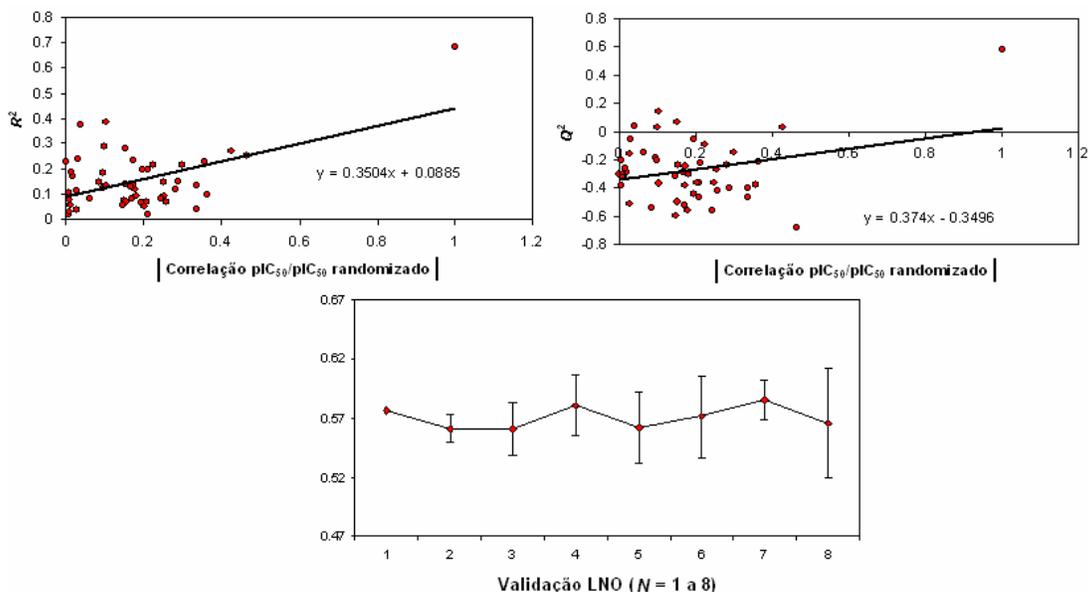


Figura 42. Gráficos dos resultados obtidos no teste de randomização de y e na validação LNO para o modelo I.

das equivalentes. Os resultados da validação externa (Tabela 8) mostram que o modelo apresenta alta capacidade de predição externa (GOLBRAIKH e TROPSHA, 2002). Os valores de SEP e ARE_{pred} também podem ser considerados baixos, o que pode indicar baixos erros de predição (ou seja, os valores previstos apresentam baixos desvios em relação ao valor real) para um derivado sintetizado com base neste modelo.

$$pIC_{50} = +106,517(E_{HOMO}) + 0,063(\alpha_{yy}) - 0,001(E_T) + 0,101(SeaC2C2aa) + 24,574 \quad (II)$$

$LVs = 2$; Informação acumulada = 60,49%; $R^2 = 0,667$; $SEC = 0,574$; $PRESS_{cal} = 7,253$; $Q^2_{LOO} = 0,523$; $SEV = 0,645$; $PRESS_{val} = 10,406$.

Considerando os coeficientes autoescalados do Modelo I (Tabela 9), todos os descritores podem ser considerados significativos. Mas, apesar do modelo apresentar boa qualidade estatística, boa capacidade de predição e ser robusto, é desejável que um paralelo entre as propriedades físico-químicas, representado pelos descritores moleculares selecio-

Tabela 8. Resultados obtidos para a validação externa do Modelo I.

Composto	pIC ₅₀ obs	pIC ₅₀ pred	Resíduos
P10	6,000	6,264	-0,264
P12	7,400	7,267	0,133
P22	5,600	6,007	-0,407
P26	7,040	7,308	-0,268
P36	6,790	6,480	0,310
R^2_{pred}			0,874
SEP			0,290
ARE _{pred}			4,366%
k			0,986
k'			1,012
$ R_0^2 - R'^2_0 $			0,175

Tabela 9. Contribuição dos descritores do Modelo I em cada LV, informação de cada LV, coeficientes de correlação de Pearson e coeficientes autoescalados do modelo.

Descritor	LV1	LV2	r	Coefficientes autoescalados
E _{HOMO}	0,499	0,746	0,508	0,405
α _{yy}	0,561	-0,649	0,571	0,314
E _T	-0,392	0,128	-0,399	-0,250
SeaC2C2aa	0,532	0,079	0,542	0,364
Quantidade de informação	38,407%	21,133%		

dados, e o mecanismo de ação em estudo (quando conhecido), possa ser traçado (OECD, 2007).

4.1.1.1 E_{HOMO}

O Modelo I apresenta grande influência de propriedades relacionadas à distribuição eletrônica. O descritor E_{HOMO} é normalmente relacionado à reatividade molecular, ao potencial de ionização, à capacidade de uma molécula realizar ataques nucleofílicos, e à estabilização de complexos de transferência de carga (KARELSON, LOBANOV e KATRITZKY, 1996).

A Figura 44 mostra que os HOMOs das moléculas concentram-se na estrutura

básica do conjunto de treinamento, incluindo o farmacóforo DCA. Ao se considerar o mecanismo de ação proposto (COTELLE, 2006), a presença do descritor E_{HOMO} no modelo era esperada, já que os inibidores da reação ST (*strand transfer*) provavelmente agem como bases de Lewis. Assim, os inibidores doariam elétrons localizados no HOMO para a formação das ligações com os íons metálicos. Moléculas com altas energias de E_{HOMO} são mais capazes de doar seus elétrons para o LUMO de uma outra molécula, sendo mais reativos quando comparados com moléculas com valores baixos para este parâmetro. O quão rápido este processo ocorre é refletido pelo valor de E_{HOMO} . Considerando que o descritor E_{HOMO} está relacionado como a nucleofilicidade de uma molécula, aquelas com elevados valores apresentam uma tendência maior em doar seus elétrons para o LUMO de outra molécula (SILAKARI, 2008), ou, no caso em estudo, de realizar ataques nucleofílicos em cátions.

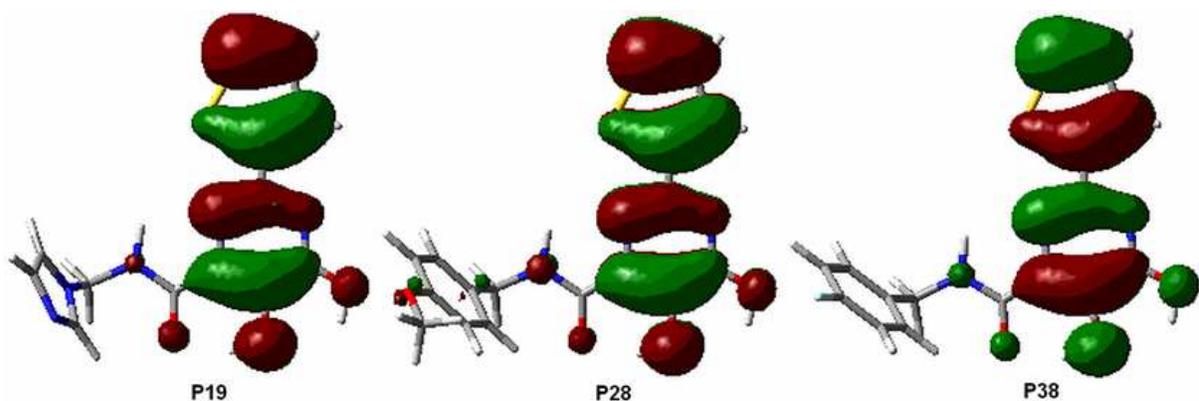


Figura 43. Orbitais moleculares ocupados mais alta energia (HOMOs) dos compostos **P19**, **P28** e **P38**.

Nesse caso, os compostos menos ativos, **P19** e **P20**, apresentam os menores valores, enquanto **P25**, um dos mais potentes, apresenta um alto valor. O coeficiente positivo do descritor dá suporte a tal relação: já que a afinidade eletrônica dos íons metálicos é constante, somente os valores de E_{HOMO} devem ser considerados como fatores

determinantes na reatividade entre o conjunto de treinamento e os íons. Outros estudos QSAR, em que E_{HOMO} ou E_{LUMO} (a energia do orbital molecular desocupado mais baixo) foram selecionados, sustentam tal hipótese (MORLEY e MATTHEWS, 2006; NIKOLIC, 2007; SILAKARI *et al.*, 2008).

O intervalo de valores desse descritor é muito pequeno (-0,210 a -0,220 eV), mas foi suficiente para descrever a importância desta propriedade para o conjunto de treinamento estudado. A tendência pode ser observada quando é considerado o coeficiente de correlação desse descritor com pIC_{50} sem os *outliers* ($r = 0,508$). O fato dos dados terem sido autoescalados provavelmente também colaborou para que este resultado fosse obtido.

4.1.1.1.2 α_{yy}

A polarizabilidade molecular é descrita como a facilidade de uma molécula ter sua nuvem eletrônica distorcida por um campo elétrico externo em uma determinada direção (GONÇALVES *et al.*, 2006). O descritor α_{yy} (direção y) foi calculado por um método baseado em um modelo empírico de Miller e Savihik (1979), disponível no programa Marvin 4.1.8 (CHEMAXON, 2008). Apesar de existir certo deslocamento no centro de massa, a direção y do conjunto de treinamento passa através do farmacóforo DCA, o local onde provavelmente ocorre a ligação com os íons (Figura 44). Desta forma, tal descritor pode estar relacionado com um possível ataque nucleofílico aos íons metálicos e à consequente mudança na distribuição de carga. O sinal positivo do coeficiente indica que o aumento da polarização neste plano é favorável à atividade. De fato, a tendência dos compostos localizados no intervalo pIC_{50} 4,30-7,07 (metade do conjunto de treinamento sem *outliers*) é apresentar baixos valores para α_{yy} .

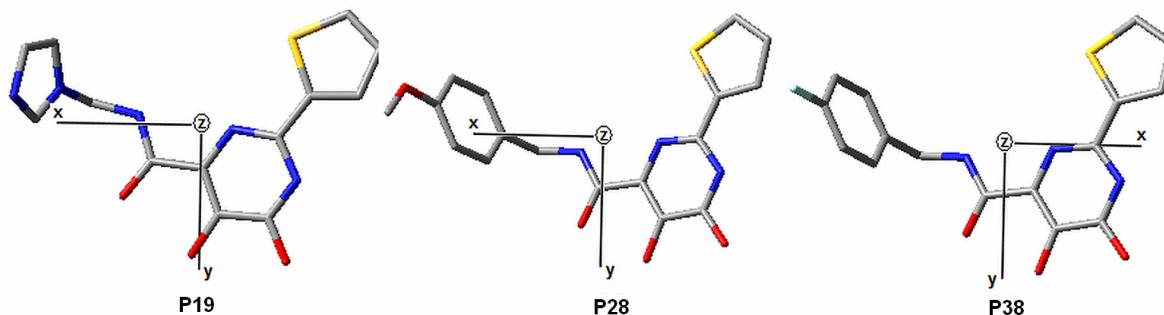


Figura 44. Representação dos eixos cartesianos para os compostos **P19**, **P28** e **P38**.

4.1.1.1.3 Energia total, E_T

O descritor energia total, E_T , corresponde a energia molecular de uma geometria otimizada, e pode ser calculado por meio da distribuição eletrônica utilizando modelagem molecular (BELLO-RAMÍREZ, CARREÓN-GARABITO e NAVA-OCAMPO, 2002; PAIS *et al.*, 2002). Estudos QSAR disponíveis na literatura sugerem que E_T (e outros descritores relacionados, como calor de formação, energia livre de Gibbs, etc) esteja relacionado à estabilidade molecular (TOIT *et al.*, 2005; LOHRAY *et al.*, 2006; PHILIPS, UDO E SAMUEL, 2008). Nas estruturas moleculares do conjunto de treinamento, a região que apresenta maior liberdade conformacional corresponde à cadeia lateral aromática. Considerando a proposta de que essa subestrutura estabelece uma interação do tipo π -stacking com a alça flexível da HIV-IN (PAIS *et al.*, 2002; BARRECA *et al.*, 2006; PARRIL, 2003), o sinal negativo do coeficiente pode significar que a ligação dos inibidores menos estáveis (menores valores absolutos de E_T) não é favorecida. Segundo Foresman e Frisch (1996), o valor obtido corresponde a um ponto estacionário na superfície de energia potencial da molécula. Assim, o aumento dos valores deste parâmetro significa que a molécula é mais estável, o que pode favorecer uma ligação com o sítio ativo e promover o aumento da potência.

4.1.1.1.4 SeaC2C2aa

O formalismo *E-state* considera que cada átomo ou ligação possui um estado intrínseco, o qual é perturbado por cada outro átomo ou ligação da molécula. Esse estado codifica informações referentes à distribuição eletrônica (na forma de uma variação causada por todos os demais átomos) e aspectos topológicos (maior/menor acessibilidade de átomos e ligações ao ambiente externo), e como tais informações podem influenciar as interações intermoleculares (KIER e HALL, 1999). Provavelmente, o descritor SeaC2C2aa também está relacionado à ligação π -stacking entre a cadeia lateral aromática e a alça flexível, já que ele varia apenas em função dessa subestrutura.

O coeficiente positivo indica que o aumento do referido parâmetro é favorável à atividade. Alguns dos compostos menos ativos - **P8** (*outlier*), **P18**, **P19** e **P22** – possuem os menores valores e alguns dos mais ativos – **P11**, **P37**, **P38** e **P39** – possuem os maiores valores. Esse descritor considera o ambiente de cada ligação, mas a eletronegatividade dos substituintes também é importante para o resultado. Por exemplo, os valores encontrados para os compostos **P4** e **P37** foram 11,94 e 18,46, respectivamente. Porém, a única diferença entre os dois compostos é um átomo de flúor, localizado no substituinte fenila. Desta forma, os índices *E-state* do tipo ligação codificam a acessibilidade eletrônica de um tipo específico de ligação (VOTANO *et al.*, 2004). Esta informação pode indicar uma relação entre o descritor SeaC2C2aa e a importância da interação π -stacking com o sítio ativo da HIV-IN (PAIS *et al.*, 2002; BARRECA *et al.*, 2006; PARRIL, 2003).

Os valores do descritor SeaC2C2aa calculados para os **P8**, **P34** e **P21** também podem ser os motivos deles terem sido classificados como *outliers*. O primeiro (SeaC2C2aa = 4,036) não possui ligações aromáticas na cadeia lateral, enquanto o segundo

tem valor (SeaC2C2aa = 7,532) próximo aos dos compostos **P32** e **P33** (SeaC2C2aa = 7,896 e 9,366), seus isômeros, mas é muito menos ativo. Já o composto **P21** (SeaC2C2aa = 5,781) pode ser um *outlier* porque o valor desse descritor é próximo ao do seu isômero **P22** (SeaC2C2aa = 3,972), mas é razoavelmente mais ativo.

4.1.2. Estudo QSAR-2D completo

Os resultados obtidos com o estudo preliminar mostraram que é possível obter um modelo QSAR-2D em que uma relação direta com o mecanismo de inibição da reação ST possa ser traçada. Assim, seguiu-se o estudo com o conjunto de treinamento em que foram utilizados os 85 inibidores selecionados, incluindo os utilizados no estudo preliminar. Os descritores utilizados nesta etapa foram calculados através de diferentes programas ou metodologias (Tabela 4).

Ao ser realizada a redução inicial da quantidade de variáveis, foram mantidos apenas os descritores cujos valores absolutos de r foram maiores que 0,3. Assim, de um total de 1291 descritores, restaram 500, os quais foram submetidos à seleção de variáveis pelo o método OPS. Nessa etapa, o *SEV* foi utilizado como critério de classificação. Após a obtenção de modelos contendo em torno de 100 variáveis, foi utilizado o Q^2_{LOO} . Assim, buscou-se a obtenção de modelos com erro minimizado e máxima capacidade de predição. Nessa etapa do estudo, foi utilizada uma versão do OPS disponível no programa *QSARmodelling* (MARTINS e FERREIRA, 2009) (Figura 45).

	SaaCH[873...	SaasC[873...	SdO[873.0...	SdssC[873...	SsOH[873...	SaasC.arf1...	y
Amostra1	14,757	0,302	12,244	-0,582	19,592	-0,391	8
Amostra2	5,474	-0,38	12,313	-0,483	19,664	-1,104	6,96
Amostra3	15,139	0,525	12,261	-0,557	19,629	-0,173	8
Amostra4	14,15	-0,125	12,24	-0,626	19,554	-0,809	7,48
Amostra5	18,548	-0,121	12,684	-0,63	19,939	-0,801	7,28
Amostra6	19,346	0,276	12,718	-0,589	20,004	-0,411	8
Amostra7	19,346	0,276	12,718	-0,589	20,004	-0,411	7,85
Amostra8	18,766	0,048	13,052	-0,542	20,165	-0,639	8,15
Amostra9	13,397	1,77	12,636	-0,533	19,897	1,066	7,52
Amostra10	13,397	1,77	12,636	-0,533	19,897	1,066	7,35
Amostra11	12,134	0,183	12,228	-0,602	19,574	-0,501	8
Amostra12	8,099	-0,798	12,104	-0,663	19,457	-1,465	7,48
Amostra13	10,162	-0,706	12,132	-0,623	19,485	-1,385	7,6
Amostra14	11,089	0,398	12,17	-0,568	19,524	-1,275	7,4
Amostra15	8,813	0,056	12,142	-0,608	19,496	-1,355	7,52
							7,23

Figura 45. Aspecto do *QSARmodelling*, programa escrito em JAVA para seleção de variáveis e construção de modelos utilizando o método PLS, além de também realizar o teste de randomização de y e a validação LNO.

Vários ciclos de utilização do método OPS foram realizados utilizando esses parâmetros, sempre a partir da matriz 85x500. Assim, foram selecionadas várias combinações de descritores, as quais, nem sempre correspondiam ao melhor modelo segundo o programa, pois todos os critérios fornecidos pelo OPS (SEV , Q^2_{LOO} , número de descritores e número de LVs) foram considerados no processo de escolha. Assim, procurou-se explorar a possibilidade de que modelos obtidos em ciclos intermediários de uso do método OPS, e não classificados como melhores, pudessem fornecer modelos finais de qualidade superior, já considerando o refinamento realizado no programa Pirouette 4 (INFORMETRIX, 2007).

A partir desse procedimento, quatro modelos (III a VI) apresentaram qualidade estatística satisfatória quanto à variância explicada e predita. Ao compararem-se os quatro modelos, é possível observar que a quantidade de LVs, descritores e *outliers*, varia consideravelmente: cinco LVs, sete descritores e seis *outliers* para o modelo III, dois LVs, quatro descritores e oito *outliers* para o modelo IV, cinco LVs, oito descritores e três *outliers* para o modelo V, e sete LVs, dez descritores e sete *outliers* para o modelo VI.

$$pIC_{50} = -0,0106(\text{VAR}) + 6,586(\text{BEHp2}) + 0,0309(\text{RDF025m}) - 1,454(\text{Mor16v}) - 59,052(\text{G2v}) + 0,550(\text{HGM}) - 10,129(\text{HATS1e}) - 7,739(\text{R1e+}) - 28,821 \quad (\text{III})$$

$n = 79$; *outliers*: 6; LVs = 5; Informação acumulada: 92,920%; $R^2 = 0,632$; $SEC = 0,545$; $PRESS_{cal} = 21,689$; $F_{(5,73)} = 25,129$ ($cF = 2,340$); $Q^2_{LOO} = 0,524$; $SEV = 0,596$; $PRESS_{val} = 28,082$ ($SS_y = 59,019$).

$$pIC_{50} = +4,757(\text{BELe8}) - 3,021(\text{RTe}) - 25,984(\text{G2e}) - 7,593(\text{R1u+}) - 5,501 \quad (\text{IV})$$

$n = 76$; *outliers*: 9; LVs = 2; Informação acumulada: 84,723%; $R^2 = 0,619$; $SEC = 0,520$; $PRESS_{cal} = 19,773$; $F_{(2,73)} = 59,239$ ($cF = 3,122$); $Q^2_{LOO} = 0,578$; $SEV = 0,536$; $PRESS_{val} = 21,866$ ($SS_y = 51,863$).

$$p\acute{I}C_{50} = -2,596(\text{BELm7}) + 6,844(\text{BEHv2}) + 2,172(\text{BEHe5}) + 0,0780(\text{RDF060m}) - 0,0630(\text{RDF030e}) - 1,308(\text{Mor16p}) - 34,741(\text{G2e}) - 1,073(\text{H4v}) - 17,018 \quad (\text{V})$$

$n = 82$; *outliers*: 3; LVs = 5; Informação acumulada = 91,510%; $R^2 = 0,614$; $SEC = 0,553$; $PRESS_{cal} = 23,234$; $F_{(5,76)} = 24,239$ ($cF = 2,335$); $Q^2_{LOO} = 0,501$; $SEV = 0,606$; $PRESS_{val} = 30,070$ ($SS_y = 60,286$).

$$pIC_{50} = -0,484(\text{RDF065u}) + 0,620(\text{VAR}) + 0,594(\text{HGM}) + 0,107(\text{BEHe5}) - 0,431(\text{BELm5}) - 0,787(\text{G1e}) - 0,300(\text{Mor16v}) - 0,314(\text{R1e+}) + 0,492(\text{BEHv2}) - 0,372(\text{qXMull_I}) - 10,025 \quad (\text{VI})$$

$n = 78$; *outliers*: 7; LVs = 7; Informação acumulada: 95,290%; $R^2 = 0,683$; $SEC = 0,504$; $PRESS_{cal} = 17,780$; $F_{(7,70)} = 21,559$ ($cF = 2,143$); $Q^2_{LOO} = 0,571$; $SEV = 0,555$; $PRESS_{val} = 24,044$ ($SS_y = 56,112$).

Apesar da diferença na quantidade de descritores e LVs, os modelos III, V e VI apresentaram praticamente a mesma quantidade de informação acumulada, com uma média de 92,670%. Já o modelo III apresentou menor quantidade de informação acumulada (84,723%, 7,947% a menos), mas estatisticamente é muito semelhante aos demais. Mesmo assim, todos apresentaram mais informação do que aquela acumulada pelo Modelo I. Foram selecionados apenas descritores topológicos baseados na teoria de grafos ou geométricos, com exceção para qXMull_I. Porém, considerando a heterogeneidade do

conjunto de treinamento, o resultado não pode ser considerado inesperado.

Com relação à quantidade de informação acumulada, o resultado indica que o método OPS auxiliou na seleção dos descritores que codificavam a maior quantidade de informação relevante para a construção dos modelos. Essa possibilidade é reforçada pela presença de alguns descritores em mais de um modelo. Além disso, quando os descritores são comparados, considerando a classe às quais pertencem, é possível observar que alguns deles (Mor16v e Mor16p; RDF060m e RDF065u; G1e e G2e; R1e+ e R1u+) codificam informações muito semelhantes. As Tabelas 3.1 a 3.4 (Anexo 3)* apresentam os valores dos descritores que compõem cada modelo, as atividades experimentais, seus coeficientes de correlação de Pearson individuais, as atividades preditas obtidas no processo de validação cruzada e os respectivos resíduos de predição.

Os quatro modelos apresentaram resultados estatísticos dentro dos limites esperados. Quanto ao ajuste dos dados, todos os resultados para o parâmetro R^2 ficaram acima do mínimo sugerido pela literatura (0,60). O modelo VI foi capaz de explicar a maior quantidade de variância (68,31%, $R^2 = 0,683$), enquanto o modelo V explica a menor quantidade (61,40%, $R^2 = 0,614$). O resultado do teste- F , com 95% de confiança ($\alpha = 0,05$), mostra que todos os modelos apresentaram valores bem superiores aos valores críticos tabelados. Porém, tal resultado mostrou que o ajuste de dados do modelo VI apresenta menor significância estatística. Neste caso, o melhor resultado foi obtido para o modelo IV (59,239). Tal resultado pode indicar que, apesar de apresentar um melhor ajuste, o modelo VI não será aquele que, necessariamente, poderá ser classificado como o melhor, fortalecendo a necessidade da realização de outros testes estatísticos.

* Algumas tabelas foram disponibilizadas no Anexo 3 apenas devido a tamanho das mesmas.

O modelo que apresenta a maior quantidade de variância predita também é o IV (57,80%, $Q^2_{LOO} = 0,578$). Assim, o valor foi praticamente o mesmo do Modelo VI (57,10%, $Q^2_{LOO} = 0,571$), uma indicação de que ambos os modelos apresentam a mesma capacidade de fazer previsões da atividade biológica de novos compostos. Porém, o *SEV* de VI é maior que o do Modelo IV (0,555 e 0,536, respectivamente). Isso indica que o erro associado à capacidade de predição interna do Modelo VI é maior. Se ainda for considerado que o Modelo IV apresenta menor quantidade de LVs, acumulando apenas 9,927% a menos de informação, menor quantidade de descritores, e um melhor resultado para o teste-*F*, esse modelo pode ser considerado superior ao Modelo VI. Além disso, ele também pode ser considerado superior aos modelos III e V pelos mesmos motivos.

Quanto aos resultados dos $PRESS_{val}$, todos ficaram bem abaixo de seus respectivos SS_y , uma primeira indicação de que a variância predita por cada modelo é real e não devido à correlação ao acaso. Finalmente, quando são comparados os valores de R^2 e Q^2_{LOO} de todos os modelos (ROY, LEONARD e ROY, 2008; FERREIRA e KIRALJ, 2009), a maior diferença encontrada foi 0,113 (Modelo V), diminuindo a possibilidade de que ocorra sobreajuste dos dados em algum dos modelos. O Modelo IV apresentou a menor diferença (0,041).

As Figuras 46 a 49 apresentam os resultados do teste de randomização de *y* e da validação LNO para os modelos III a VI. É possível observar que nenhum dos modelos apresenta correlação ao acaso. Nos quatro casos, os interceptos da regressão entre as correlações das atividades originais/randomizadas e dos valores de R^2 dos modelos randomizados e dos valores de Q^2_{LOO} ficaram abaixo do limite máximo recomendado por Eriksson e colaboradores (2003). Esse resultado mostra que os modelos randomizados não explicam ou predizem variância considerável. Assim, os modelos III a VI não apresentam

correlação ao acaso. Este é um resultado importante quando se considera que os resultados estatísticos da qualidade interna dos modelos foram apenas razoáveis.

Por outro lado, os resultados da validação LNO mostram um comportamento menos uniforme. Todos os modelos apresentaram pouca variação em torno de seus Q^2_{LNO} médios e pouca diferença em relação a seus respectivos Q^2_{LOO} . Porém, os valores médios específicos de cada N oscilaram consideravelmente de modelo para modelo, como pode ser observado pelos desvios padrões, expressos pelas barras de erros. No modelo III, os resultados oscilaram pouco entre LOO e L90, mas a situação alterou-se em L120 e L130, e entre L150 e L170. Contudo, a alteração foi pequena em relação a seu Q^2_{LNO} médio e Q^2_{LOO} . Para o modelo V, pôde-se observar uma alta oscilação já em L70, enquanto, a partir de L120, os desvios foram maiores em relação ao Q^2_{LNO} médio e a maioria dos valores médios N ficou abaixo do Q^2_{LOO} . Já para o modelo VI, os N oscilaram razoavelmente em relação ao seu Q^2_{LNO} médio desde L20, embora os desvios padrões até L100 tenham sido pequenos. O modelo IV apresentou maior estabilidade, sendo que os N médios apresentaram pequenos desvios em relação a Q^2_{LOO} e Q^2_{LNO} médio, e desvios padrões menores, com uma oscilação média um pouco maior entre L130 e L170. A partir desses resultados, pôde-se considerar o modelo V como inadequado em relação à validação LNO e o modelo IV como o mais robusto, enquanto os modelos III e VI podem ser considerados ‘aprovados com ressalvas’.

Considerando a análise da qualidade interna dos Modelos III a VI, é possível observar que todos explicam e predizem a quantidade de informação suficiente para serem aprovados, sem que as mesmas ocorram devido à correlação ao acaso. Além disso, todos são estatisticamente significativos. Porém, as diferenças mais acentuadas foram observadas para os resultados da validação LNO, por meio da qual o Modelo IV pode ser considerado

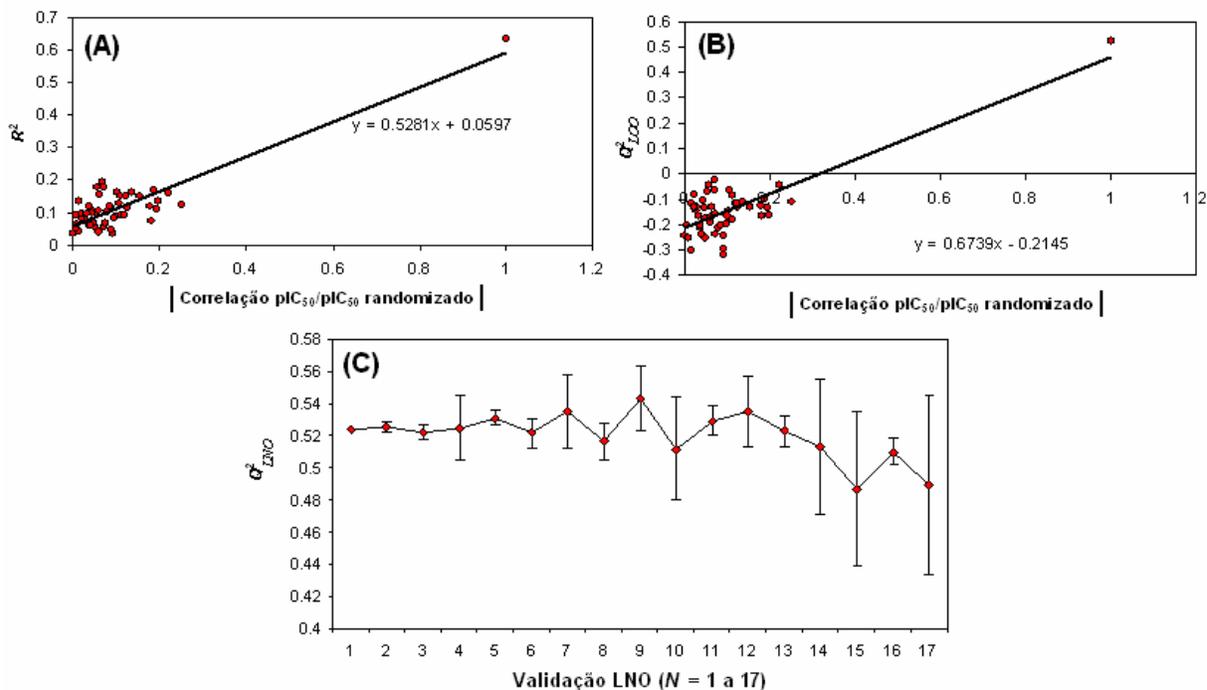


Figura 46. Resultados do teste de randomização de y (A e B) e da validação LNO (C) para o modelo III.

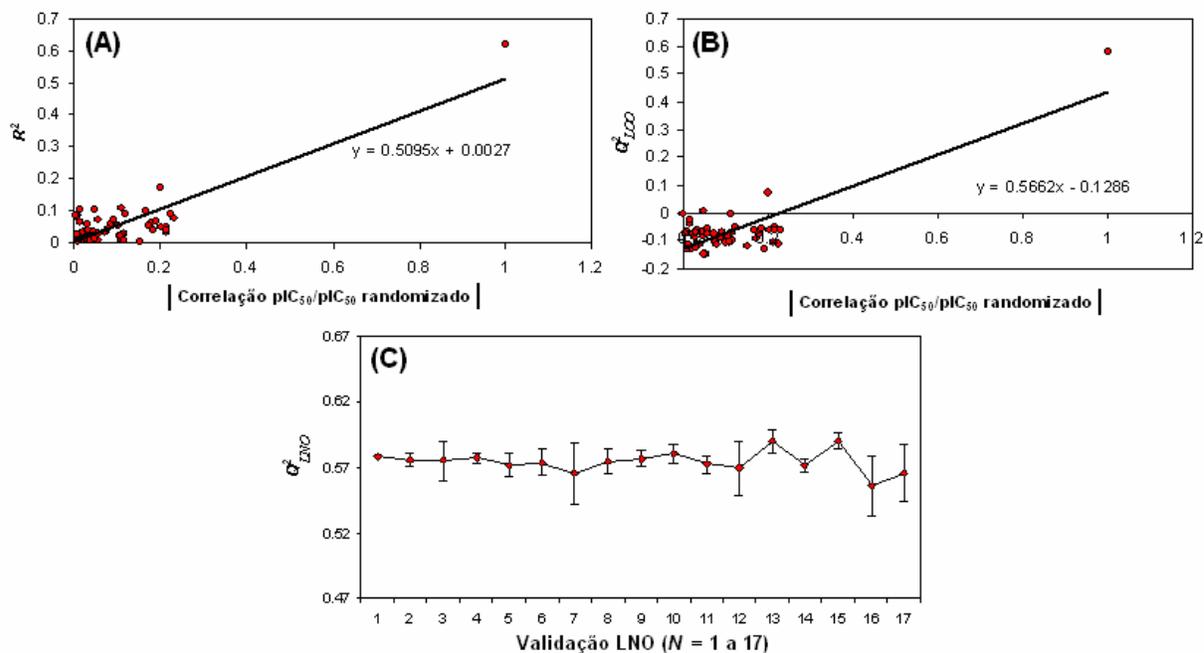


Figura 47. Resultados do teste de randomização de y (A e B) e da validação LNO (C) para o modelo IV.

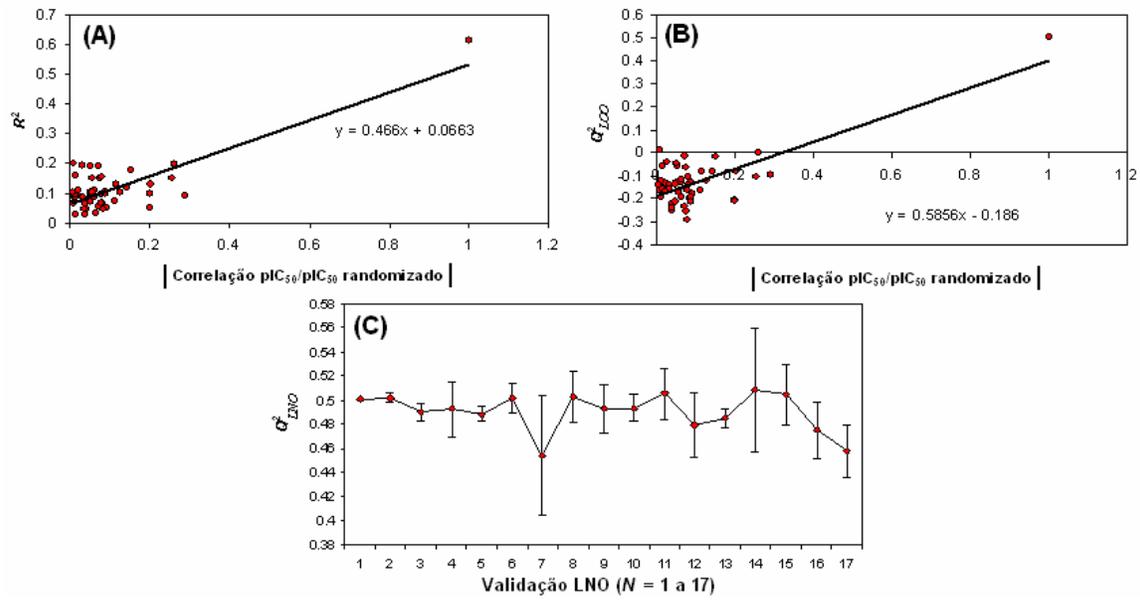


Figura 48. Resultados do teste de randomização de y (A e B) e da validação LNO (C) para o modelo V.

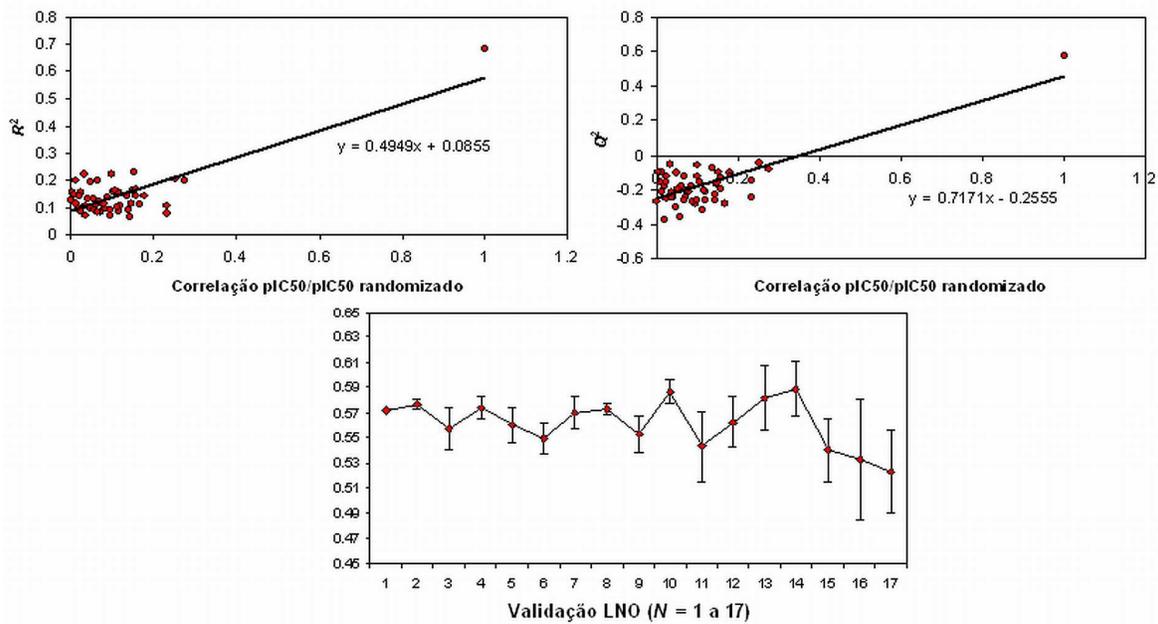


Figura 49. Resultados do teste de randomização de y (A e B) e da validação LNO (C) para o modelo VI.

o mais estável, enquanto os demais podem, por comparação, ser considerados menos robustos.

Outro fator que pode ser avaliado para se escolher o Modelo IV como o mais apropriado é a coincidência entre os sinais dos r de cada descritor com a variável dependente y e os sinais dos coeficientes b no modelo. Segundo Ferreira e Kiralj (2009), a não coincidência entre as contribuições desses dois fatores é uma indicação de falta de autoconsistência do modelo, fator importante tanto pelo ponto de vista estatístico quanto para a interpretação do modelo obtido perante o mecanismo de ação da atividade em estudo. Como pode ser visto na Tabela 10, todos os modelos apresentaram descritores onde os sinais de seus coeficientes não coincidiam com a informação fornecida pela correlação com a atividade biológica. O único modelo onde isso não ocorre é no Modelo IV (Tabela 11), e assim ele pode ser classificado como o melhor, pois, somada a estatística de maior qualidade e maior robustez indicada pela validação LNO, esse modelo também pode ser considerado o único onde os dados são autoconsistentes. É possível observar na Tabela 9 que o Modelo I também é autoconsistente.

Com a finalização da verificação da qualidade interna dos modelos, seguiu-se a etapa de validação externa. Apesar dos Modelos IV (em especial), III e VI terem sido aprovados em todos os testes, os resultados não são suficientes para garantir que os modelos obtidos apresentem a capacidade preditiva necessária. Como já comentado, apesar do fato de que baixos valores de Q^2_{LOO} possam ser utilizados como um indicador de uma baixa capacidade preditiva de um modelo, o oposto não é necessariamente verdade (GOLBRAIKH e TROPSHA, 2002; GRAMATICA, 2007; ROY e ROY, 2008). Assim, a validação externa dos modelos se faz necessária. Para isto, foram selecionados conjuntos-testes que representassem, da forma mais adequada possível, a faixa de atividade biológica

Tabela 10. Coeficientes de correlação de Pearson e coeficientes autoescalados dos descritores dos Modelos III, V e VI, com destaque para aqueles que não apresentaram autoconsistência.

Descritor	<i>r</i>	Coeficientes autoescalados
Modelo III		
VAR	0,487	0,327
BEHp2	0,454	0,519
RDF025m	0,464	0,178
Mor16v	-0,407	-0,387
G2v	-0,320	0,522
HGM	-0,305	0,647
HATS1e	-0,440	-0,470
R1e+	-0,486	-0,319
Modelo V		
BELm7	0,432	-0,488
BEHv2	0,562	0,473
BEHe5	0,571	0,454
RDF060m	0,489	0,493
RDF030e	0,357	-0,407
Mor16p	-0,399	-0,396
G2e	-0,543	-0,299
H4v	0,509	-0,229
Modelo VI		
RDF065u	0,336	-0,484
VAR	0,536	0,620
HGM	-0,401	0,594
BEHe5	0,549	0,107
BELm5	0,377	-0,431
G1e	-0,538	-0,784
Mor16v	-0,405	-0,300
R1e+	-0,548	-0,314
BEHv2	0,579	0,492
qXMull_I	0,389	-0,372

Tabela 11. Contribuição dos descritores do Modelo IV em cada LV, informação de cada LV, coeficientes de correlação de Pearson e coeficientes autoescalados do modelo.

Descritor	LV1	LV2	<i>r</i>	Coeficientes autoescalados
BELe8	0,433	0,808	0,529	0,384
RTe	-0,482	0,580	-0,589	-0,127
G2e	-0,511	0,095	-0,624	-0,239
R1u+	-0,564	0,039	-0,689	-0,278
Quantidade de informação	63,672%	21,051%		

e a variabilidade estrutural do conjunto de treinamento.

A seleção dos conjuntos-testes foi limitada por alguns fatores. Apesar da ampla faixa de atividade biológica (pIC_{50} 4,30 – 8,15), esta não é uniforme. É possível observar que a faixa de atividade é deslocada em direção aos mais potentes (situação ilustrada na Figura 50), com 33 compostos localizados na faixa de 4,30 a 6,96 (apenas 38% do conjunto de treinamento completo) e o restante está concentrado na faixa de 7,00 a 8,15. Isto levou à seleção de conjuntos testes concentrados em uma faixa de 1,15 unidades logarítmicas. Assim, todos os conjuntos-testes foram formados por 14 compostos. Isto correspondeu a 17,721%, 18,421%, 17,073% e 17,949% das amostras realmente utilizadas nos modelos III a VI, respectivamente.

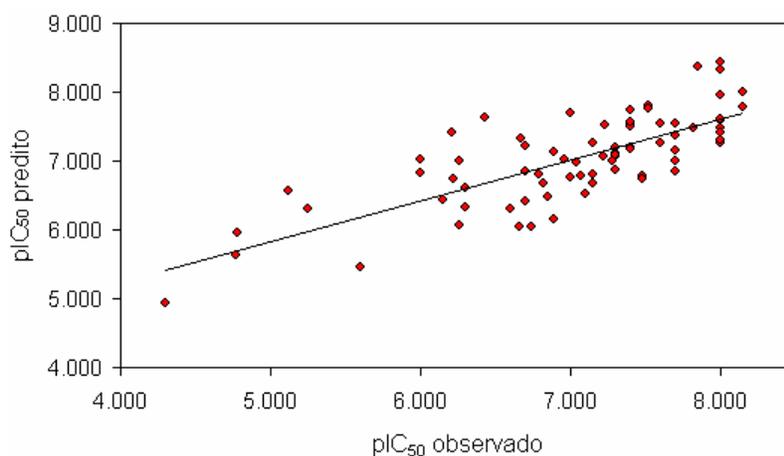


Figura 50. Gráfico dos pIC_{50} 's observados (eixo das ordenadas) pelos previstos (eixo das abcissas), obtido no processo de validação cruzada Q^2_{LOO} do modelo IV (sem *outliers* e antes da retirada do conjunto-teste).

Um último fator de peso para a seleção dos conjuntos testes foram as amostras selecionadas como *outliers* em cada modelo, assim como as quantidades dos mesmos. Isto impediu que fossem sempre utilizados conjuntos-testes formados pelos mesmos

compostos. Essas diferenças limitam a capacidade de comparação dos resultados obtidos na predição externa, especialmente para os parâmetros estatísticos cujos valores são muito próximos. Ao mesmo tempo, tal característica dos modelos pode ser considerada uma indicação de que, apesar de guardarem determinadas semelhanças quanto aos descritores selecionados e alguns parâmetros estatísticos, as informações codificadas por cada equação não são exatamente as mesmas. Caso contrário, os *outliers* seriam sempre os mesmos.

Com a retirada dos conjuntos-testes, específicos de cada modelo, quatro novos foram obtidos (VII a X). Porém, uma das exigências nesta etapa era de que a qualidade estatística de cada novo modelo, os *modelos reais*, fosse minimamente alterada em relação aos *Modelos Auxiliares* III a VI. Desta forma, é possível propor que ambos os modelos são equivalentes, e assim o modelo auxiliar, mais representativo da ‘população’ de inibidores da reação ST, pode ser utilizado para a predição da potência de novos inibidores no lugar do modelo real.

$$pIC_{50} = +0,00799(VAR) + 7,285(BEHp2) + 0,00947(RDF025m) - 1,639(Mor16v) \quad (VII) \\ + 54,736(G2v) + 0,445(HGM) - 9,780(HATS1e) - 7,584(R1e+) - 29,481$$

LVs = 5; Informação acumulada = 92,92%; $R^2 = 0,593$; $SEC = 0,537$; $PRESS_{cal} = 18,745$; $Q^2_{LOO} = 0,423$; $SEV = 0,639$; $PRESS_{val} = 26,588$.

Conjunto teste: **P10, P16, P19, P32, S9, S12, G10, G12, Z5, Z7, W2, W7, E1, L708906**

$$pIC_{50} = +4,952(BELe8) - 3,513(RTe) - 25,193(G2e) - 8,213(R1u+) - 6,132 \quad (VIII)$$

LVs = 2; Informação acumulada = 85,363%; $R^2 = 0,681$; $SEC = 0,485$; $PRESS_{cal} = 13,876$; $Q^2_{LOO} = 0,634$; $SEV = 0,561$; $PRESS_{val} = 15,912$.

Conjunto teste: **P10, P16, P21, P32, S9, S12, G10, G12, Z5, Z7, W2, W7, E1, L708906**

$$p\acute{I}C_{50} = -2.384(\text{BELm7}) + 6.902(\text{BEHv2}) + 2.108(\text{BEHe5}) + 0.078(\text{RDF060m}) - 0.062(\text{RDF030e}) - 1.234(\text{Mor16p}) - 40.937(\text{G2e}) - 1.631(\text{H4v}) - 15,986 \quad (\text{IX})$$

LVs = 5; Informação acumulada = 92,326%; $R^2 = 0,614$; $SEC = 0,541$; $PRESS_{cal} = 18,181$; $Q^2_{LOO} = 0,464$; $SEV = 0,609$; $PRESS_{val} = 25,210$.

Conjunto teste: **P10, P16, P19, P32, S9, S12, G10, G12, Z5, Z7, W2, W7, E1, L708906**

$$p\acute{I}C_{50} = -0,043(\text{RDF065u}) + 0,014(\text{VAR}) + 0,391(\text{HGM}) + 0,909(\text{BEHe5}) - 2,349(\text{BELm5}) - 88,070(\text{G1e}) - 0,860(\text{Mor16v}) - 7,734(\text{R1e+}) + 7,155(\text{BEHv2}) - 5,005(\text{qXMull_I}) - 10,225 \quad (\text{X})$$

LVs = 7; Informação acumulada = 95,382%; $R^2 = 0,666$; $SEC = 0,505$; $PRESS_{cal} = 14,313$; $Q^2_{LOO} = 0,498$; $SEV = 0,580$; $PRESS_{val} = 21,520$.

Conjunto-teste: **P13, P16, P19, P32, S9, S12, G10, G12, Z5, Z7, W2, W8, E1, L708906**

Como resultado, é possível observar que apenas o modelo VIII não sofreu alterações consideráveis em sua qualidade interna em relação ao seu predecessor IV, especialmente no que se refere aos Q^2_{LOO} : os novos valores desse parâmetro ficaram abaixo de 0,50 para os modelos VII, IX e X (0,423, 0,464 e 0,498, respectivamente). Por outro lado, a capacidade de predição interna do modelo VIII foi maximizada, com seu Q^2_{LOO} alterado para 0,632 (0,054 unidade). O resultado contrasta com os dados advindos da significância estatística, já que o modelo IV apresentou o melhor F entre todos, e com a validação LNO do mesmo, que apresentou a menor oscilação nos resultados. Por outro lado, o modelo VI, o qual apresentou os melhores R^2 e Q^2_{LOO} , variou 0,073 para o último parâmetro e ficou abaixo do limite mínimo de 0,50. Esta etapa também ajuda a confirmar a falta de robustez do modelo III, proposta com base nos resultados da validação LNO, pois foi o que sofreu maior alteração de Q^2_{LOO} (0,101 unidade).

Com esses resultados, a realização da etapa de validação externa é justificada apenas para o Modelo Real VIII. Na Figura 51 é apresentado um dendograma obtido

utilizando Análise Hierárquica de Agrupamentos (HCA: *Hierarchical Cluster Analysis*) centrado nas amostras do conjunto de treinamento completo. Este dendrograma mostra que o conjunto teste apresenta boa distribuição e é representativo do conjunto de treinamento completo.

Os resultados são apresentados na Tabela 12. É possível observar que o modelo possui capacidade preditiva dentro dos limites desejados. O resultado para R^2_{pred} ficou acima do limite mínimo de 0,50, sendo apenas 0,060 unidade menor que o Q^2_{LOO} . A diferença entre SEP e SEV pode ser um pouco superior (0,147 unidade), mas ainda pode ser considerada aceitável. Os resultados indicam que as capacidades de predição externa e interna são equivalentes. Já os resultados de k e k' localizaram-se dentro do intervalo proposto 0,85 – 1,15, e o parâmetro $|R^2_0 - R'^2_0|$ ficou abaixo de 0,3, como proposto por Roy e Roy (2008a, 2008b) e Gobraikh e Tropsha (2003), confirmando que a capacidade prediti-

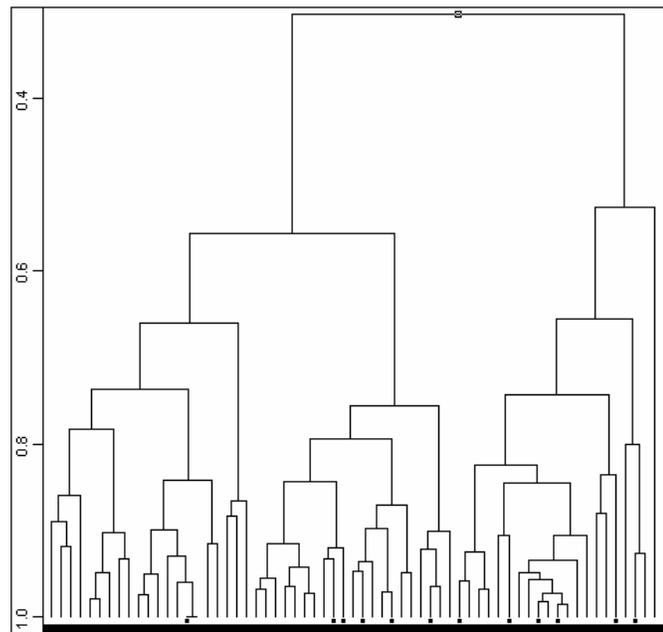


Figura 51. Dendrograma (dados autoescalados) do conjunto de treinamento completo para o modelo IV. O conjunto teste encontra-se destacado (pontos pretos).

Tabela 12. Resultados da etapa de validação externa realizada com o Modelo Real VIII a partir do conjunto teste selecionado do Modelo Auxiliar IV.

Composto	$pIC_{50\text{ obs}}$	$pIC_{50\text{ prev}}$	Resíduos
E1	6,000	7,061	-1,061
G10	6,210	7,360	-1,150
G12	4,770	5,733	-0,963
L708906	6,890	7,041	-0,151
P10	7,300	7,397	-0,097
P16	6,700	6,483	0,217
P21	7,230	7,542	-0,312
P32	7,400	7,667	-0,267
S12	6,430	7,751	-1,321
S9	8,000	8,334	-0,334
W2	5,250	6,015	-0,765
W7	6,260	6,269	-0,009
Z5	8,000	7,627	0,373
Z7	7,000	7,642	-0,642
R^2_{pred}			0,518
SEP			0,683
ARE_{pred}			8,622%
k			0,936
k'			1,062
$ R^2_0 - R'^2_0 $			0,251

va apresentada pelo modelo é real. Finalmente, o resultado do ARE_{pred} mostra que o potencial de erro médio, quando da efetiva utilização do modelo, deve ficar próximo a 10%, e pode ser considerado aceitável.

A partir dos resultados encontrados, apenas o modelo auxiliar IV e seu respectivo modelo real VIII podem ser considerados aprovados em todos os critérios estatísticos, pois apresentaram ajuste, significância estatística, capacidade preditiva interna, externa e robustez necessária para isto.

Porém, assim como foi feito no estudo preliminar, apenas a utilização de parâmetros estatísticos para classificar um modelo como adequado pode não ser suficiente (GAUDIO e ZANDONADE, 2001). Quando o mecanismo de ação de uma atividade

biológica em estudo é desconhecido e a obtenção de um modelo apenas com alta qualidade estatística é aceitável. Porém, quando esse mecanismo é conhecido, ou pelo menos existem informações sobre o mesmo, é desejável que os descritores que originaram um modelo descrevam-no da melhor forma possível (OECD, 2007). Isto pode ser observado claramente no estudo preliminar apresentado, em que pode ser traçada uma relação direta entre os descritores e tal mecanismo de ação seja o mais aceito. Por exemplo, em uma situação hipotética onde o descritor E_{LUMO} fosse selecionado, parâmetro relacionado com ataques eletrofílicos, o modelo poderia ser desqualificado.

Porém, o significado químico de alguns descritores utilizados em QSAR e QSPR não é tão claro como os que foram selecionados para o modelo preliminar. Dentre os de interpretação mais complexa, os descritores baseados em informações sobre a topologia e geometria molecular podem, de modo geral, ser considerados os mais complexos, mesmo quando o mecanismo de ação é conhecido. A literatura propõe que esses descritores sejam utilizados como uma medida quantitativa da falta de homogeneidade estrutural ou da diversidade de um grafo molecular, sendo relacionados à simetria associada às estruturas. Mas, os tipos de informações que podem estar contidos em um grafo não são únicos e dependem do quanto eles são equivalentes entre si. Normalmente, a literatura refere-se a este tipo de descritores como relacionados à importância da forma, tamanho, simetria e/ou ramificação de um conjunto de substâncias químicas para uma atividade biológica em estudo (KATRITZKY e GORDEEVA, 1993; NEVES, COSTA e NDUJAE, 1998; TODESCHINI e CONSONNI, 2000; PADRÓN, CARRASCO e PELLÓN, 2002; GARCIA *et al.*, 2005).

Apenas descritores topológicos e geométricos foram selecionados para o Modelo IV. Um deles, BELe8, é topológico baseado na teoria de grafos, e foi obtido pelo programa

E-Dragon 1.0, via interface *Parameter Client*, www.vcclab.org. Este descritor pertence à classe BCUT (*Burden Chemical Abstract Service of University of Texas* ou descritores de medida de *Burden*). Os demais são classificados como geométricos, pois dependem da geometria molecular, e por isto foram calculados no programa *Dragon 3.0 Web Version* a partir das geometrias de mínimas energias. Dois deles, RTe e R1u+, pertencem a classe GETAWAY (*Geometry, Topology And Atom-Weights Assembly*), enquanto G2e é classificado como WHIM (*Weighted Holistic Invariant Molecular index*).

Três dos descritores do Modelo IV (BELe8, RTe e G2e) são ponderados pela eletronegatividade atômica. A utilização de propriedades atômicas para ponderar determinadas classes de descritores objetiva verificar a influência de aspectos estruturais específicos sobre informações dependentes da geometria 3D (forma, simetria, etc) ou topologia (HEMMER, STEINHAEUER e GASTEIGER, 1999; CONSONNI, TODESCHINI e PAVAN, 2002). Como tais descritores são normalmente de difícil interpretação, muitos autores costumam relacionar um descritor deste tipo com a influência da propriedade atômica utilizada sobre a atividade biológica. Por exemplo, se é selecionado um descritor ponderado pela eletronegatividade, considera-se que a distribuição eletrônica influencia a atividade (SAÍZ-URRA *et al.*, 2007).

Segundo Consonni, Todeschini e Pavan (2002), tais descritores são mais adequados para estudos QSAR ou QSPR com conjuntos de treinamento estruturalmente heterogêneos. Essa situação é encontrada neste trabalho, o que pode explicar o motivo de nenhum descritor de mais simples interpretação (como E_{HOMO} , no Modelo I) ter sido selecionado no estudo com o conjunto completo.

Na Tabela 11 foram apresentadas as contribuições de cada descritor para cada LV,

a contribuição (em porcentagem) de cada LV no modelo e os coeficientes autoescalados de cada descritor. A partir dos valores absolutos desses coeficientes, é possível observar que a ordem de importância de cada descritor no modelo é $BELe8 > R1u+ > G2e > RTe$. O mais importante, $BELe8$, é ponderado pela eletronegatividade de Sanderson (SANDERSON, 1955). Isso pode ser uma indicação, junto com a presença de $G2e$ e RTe , de que a distribuição eletrônica é a característica mais importante para a atividade em estudo.

Devido à natureza complexa dos descritores selecionados, na próxima seção, serão apresentadas informações teóricas sobre cada descritor do Modelo IV e a melhor interpretação encontrada em relação ao mecanismo de inibição da reação ST.

4.1.2.1 Descritor BCUT: $BELe8$

Diversos descritores buscam capturar informações relacionadas à geometria molecular para utilização em estudos SAR/QSAR. Muitos procuram combinar duas ou mais propriedades derivadas de átomos em um único valor. Os descritores BCUT foram desenvolvidos com este objetivo e compõem uma classe de ‘descritores híbridos’. Seus valores são calculados a partir dos autovalores de uma matriz quadrática simétrica que representa um grafo molecular (TODESCHINI e CONSONNI, 2000; STANTON, 1999). Essa classe é uma extensão do método proposto por Burden (1989) para identificação de estruturas orgânicas utilizando matrizes em que os elementos b_{ii} podem ser utilizados para a codificação de informações sobre cargas atômicas, polarizabilidade, eletronegatividade, capacidade de formação de ligação hidrogênio, dentre outras características, enquanto b_{ij} podem incluir informações sobre a conectividade entre os átomos (por exemplo, distâncias interatômicas e ordens de ligação) (TODESCHINI e CONSONNI, 2000).

Segundo Todeschini e Consonni (2000), os descritores de Burden com os menores autovalores (*lowest eigenvalues*) codificam informações relacionadas principalmente à topologia molecular. Porém, alguns autores propõem os descritores BCUT são capazes de codificar outras informações, já que suas matrizes diagonais podem apresentar propriedades atômicas relacionadas a interações fármaco-receptoras (PIRARD e PICKETT, 2000; STANTON, 1999). Por exemplo, González e colaboradores (2005) relacionaram o descritor BEHe7 (*highest eigenvalue n° 7 of Burden Matrix weighted by atomic Sanderson electronegativities*) com a afinidade de um grupo de compostos agonistas pelo receptor A3, propondo que o aumento da eletronegatividade aumentaria essa afinidade. Por analogia, uma relação semelhante pode ser proposta para o descritor BELe8 (*lowest highest eigenvalue n° 8 of Burden Matrix weighted by atomic Sanderson electronegativities*).

A eletronegatividade utilizada para ponderar esse descritor é obtida a partir de um método proposto por Sanderson (1955; 1983), utilizando uma teoria denominada *equalização de eletronegatividade*: eletronegatividades atômicas são obtidas utilizando uma função inversa do volume de um átomo e, a partir desses valores, as eletronegatividades moleculares podem ser calculadas. Considerando que os átomos mais eletronegativos são aqueles que possuem maior quantidade de pares de elétrons livres, pode-se propor que uma maior eletronegatividade pode estar relacionada com uma mais fácil interação com os íons metálicos provavelmente localizados no sítio de ligação. Isso pode estar relacionado com o sinal positivo do coeficiente autoescalado desse descritor no modelo, considerando que a eletronegatividade de Sanderson também é um parâmetro positivo (Tabela 13). Os átomos mais importantes para essa interação são N e O, presentes no farmacóforo, dois dos mais eletronegativos. Os átomos de F e Cl (o primeiro mais

eletronegativo que O e N, e o segundo mais que N) também podem ser importantes nessa

Tabela 13. Valores não escalados da eletronegatividade de Sanderson dos átomos que formam o conjunto de treinamento (SANDERSON, 1983).

Átomo	Eletronegatividade de Sanderson
H	2,592
C	2,746
N	3,194
O	3,654
F	4,000
S	2,957
Cl	3,475

hipótese, já que os compostos de maior potencial de uso clínico (inclusive o próprio raltegravir) apresentam halogênios ligados à cadeia lateral aromática (DAYAM, AL-MAWSAWI e NEAMATI, 2007). Porém, no conjunto de treinamento em estudo, apenas o primeiro aparece com frequência, sendo 19 compostos com F e apenas dois com Cl.

Assim, é possível propor que o descritor BELe8 esteja relacionado a interações importantes para a ocorrência da inibição da reação ST, que são dependentes da distribuição eletrônica, ou seja, as interações com os co-fatores metálicos. Idéia semelhante foi apresentada por Saíz-Urra e colaboradores (2007) para um modelo baseado em outros inibidores da HIV-IN, construído com descritores GETAWAY pesados por eletronegatividade e polarizabilidade.

4.1.2.2 Descritores GETAWAY: RTe e R1u+.

Os descritores GETAWAY são baseados em uma matriz de *leverage* denominada ‘*molecular influence matrix*’ (MIM) **H**, proposta como uma representação molecular que pode ser facilmente calculada a partir das coordenadas espaciais dos átomos de uma

molécula em determinada conformação. A magnitude do *leverage* máximo depende do tamanho e da forma da molécula. Informações sobre relações entre dois átomos em uma mesma molécula também podem ser obtidas, visando à codificação de informações relacionadas à forma, e sua relativa dependência de mudanças conformacionais (CONSONNI, TODESCHINI e PAVAN, 2002).

Estes descritores são divididos em dois conjuntos: (i) H-GETAWAY, que é obtido utilizando-se informações originárias pela MIM; e (ii) R-GETAWAY, o qual combina essas informações com as distâncias geométricas interatômicas da molécula obtidas em uma matriz geométrica, denominada *matriz de distância G*. Tanto o descritor RTe (*R total index weighted by the Sanderson electronegativity*) quanto R1u+ (*R maximal autocorrelation of lag 1/unweighted*) se enquadram no segundo conjunto (CONSONNI, TODESCHINI e PAVAN, 2002; TODESCHINI *et al.*, 2003; GONZÁLEZ *et al.*, 2005).

O descritor RTe pode ser obtido por meio da Equação 22. O resultado do descritor é diretamente proporcional ao parâmetro w utilizado, a eletronegatividade de Sanderson. O coeficiente negativo pode indicar que o aumento desse parâmetro pela presença de maior quantidade de átomos eletronegativos nas moléculas (especialmente N e O) seja prejudicial à atividade. Isto vai contra a teoria sobre o mecanismo de ação e entra em conflito com a interpretação do descritor BELe8. Porém, RTe também sofre influência da distância r_{ij} .

$$RT(w) = 2 \cdot \sum_{i=1}^{A-1} \sum_{j>i} \frac{\sqrt{h_{ii} \cdot h_{jj}}}{r_{ij}} \cdot w_i \cdot w_j \quad (22)$$

r_{ij} = distância entre os átomos i e j

h_{ii} e h_{jj} = elementos da diagonal da MIM

w = propriedade atômica utilizada para pesar o descritor.

Quando as contribuições de RTe e BELe8 na primeira LV (Tabela 11), aquela com

maior quantidade de informação, são comparadas, é possível observar que elas são muito semelhantes em valores absolutos ($|0,482|$ e $|0,433|$, respectivamente). Porém, os tipos de informações são diferentes. Mais especificamente, enquanto a contribuição de BELe8 na primeira LV favorece a atividade em estudo, a contribuição de RTe desfavorece a mesma. Assim, é possível propor que RTe insere, no modelo, principalmente, informações relacionadas não à eletronegatividade, mas sim à importância da forma da molécula para a inibição da reação ST.

Se, pela Equação 22, a distância r_{ij} é inversamente proporcional ao valor do descritor, quanto menor for cada r_{ij} (ou seja, quanto mais compacta for a molécula), maior será o valor do descritor. Logo, como o sinal do coeficiente é negativo, isto significa que compostos com altos RTe tenderão a serem menos ativos. Ou seja, quanto maior r_{ij} , (ou menos compactas as moléculas), mais potentes serão os compostos. Já quanto à eletronegatividade, mesmo considerando que essa característica não é a principal informação inserida por RTe no modelo, caso ela esteja presente na segunda LV, esse pode ser o motivo da contribuição da segunda LV ser positiva.

O descritor R1u+ pode ser obtido por meio da Equação 23. É muito clara a semelhança com RTe, já que os descritores GETAWAY $R_k(w)_+$ são simplesmente o máximo valor (\max_{ij}) utilizado para o cálculo de RT(w). Assim como RTe, o resultado do descritor é diretamente proporcional ao parâmetro w utilizado e inversamente proporcional à distância r_{ij} . Na definição de R1u+, lag é a distância topológica ou todas as contribuições de cada caminho diferente no grafo molecular. Quanto maior o lag maior é a distância entre dois átomos. Assim, termos baixos, como R1 e R2, representam moléculas pequenas, onde é esperada uma baixa dependência de mudanças conformacionais nas informações codificadas, já que os pares de átomos estão muito próximos uns dos outros (CONSONNI,

TODESCHINI e PAVAN, 2002; GONZÁLEZ *et al.*, 2005). De forma geral, como R1u+ não é ponderado por alguma propriedade química, ele codifica apenas informações relacionadas à forma das moléculas, com baixa dependência de mudanças conformacionais.

$$R_k^{(w)+} = \max_{ij} \left(\frac{\sqrt{h_{ii} \cdot h_{jj}} \cdot w_i \cdot w_j \delta(k; d_{ij})}{r_{ij}} \right) \quad (23)$$

$$i \neq j \quad \text{e} \quad k = 1, 2, \dots, d$$

r_{ij} = distância entre os átomos i e j

d = diâmetro topológico

d_{ij} = distância topológica entre os átomos i e j

h_{ii} e h_{jj} = elementos da diagonal da MIM

w = propriedade atômica utilizada para pesar o descritor

Como o sinal do coeficiente de R1u+ é negativo, o significado desse descritor no Modelo IV pode ser considerado semelhante ao de RTe. Porém, a importância do primeiro é maior em 0,151 unidade. Isto que pode ser considerado uma grande diferença já que os coeficientes encontram-se autoescalados. Além disso, quase toda a informação fornecida por R1u+ ao modelo encontra-se na primeira LV, já que a segunda LV apresenta contribuição de apenas 0,039. O referido dado pode ser uma indicação de que realmente a primeira LV dos descritores GETAWAY codificam informações relacionadas à forma da molécula. E, da mesma forma que RTe, quanto maior a distância r_{ij} menor será o valor do descritor, o que tenderá a favorecer a atividade.

Ao observar a Tabela 3.2, Anexo 3, esta relação fica clara, já que os compostos menos potentes apresentam maiores valores para ambos os descritores. Por exemplo, considerando R1u+ para **P18** (pIC₅₀ = 4,300) é 0,205, e **P19** (pIC₅₀ = 5,620) é 0,200, e **P22**

($pIC_{50} = 5,600$) é 0,259, os maiores valores para esse descritor, enquanto para **G15** e **E5b** ($pIC_{50} = 8,15$ para ambos) os valores são muito menores, 0,089 e 0,079, respectivamente. Já para RTe, os valores são 0,194, 0,189, 0,245, 0,094 e 0,074, para os mesmos compostos e respectivamente.

Assim, pode ser levantada a hipótese de que ambos descritores GETAWAY selecionados para o modelo codificam principalmente informações quanto à forma da molécula. A partir dos sinais dos coeficientes, há uma indicação de que moléculas com certo grau de liberdade conformacional, que favorecem geometrias com maiores valores de r_{ij} , tenderão a ser mais ativas.

4.1.2.3 Descritor WHIM: G2e.

Os descritores 3D WHIM são calculados ao se realizar uma análise de componentes principais (PCA) com uma matriz de covariância obtida a partir das coordenadas moleculares ponderadas por propriedades atômicas, sendo essas as mesmas utilizadas por outras abordagens, como GETAWAY, mais os índices eletrotopológicos de Kier e Hall (TODESCHINI e CONSONNI, 2000; GONZÁLEZ *et al.*, 2005; GRAMATICA, 2006). Os elementos da matriz podem ser calculados a partir da Equação 24.

Os descritores WHIM objetivam obter informações relacionadas ao tamanho, à simetria, forma e distribuição dos átomos. Esse conjunto de descritores pode ser subdividido em dois grupos menores: direcionais e não-direcionais. O primeiro caracteriza uma propriedade em particular em relação a um eixo principal, enquanto o segundo fornece informações sobre a molécula como um todo (GRAMATICA, 2006).

$$s_{jk} = \frac{\sum_{i=1}^A w_i (q_{ij} - \bar{q}_j)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad (24)$$

s_{jk} = covariância ponderada entre as j -ésimas e k -ésimas coordenadas atômicas

A = número de átomos

w_i = fator de ponderação do i -ésimo átomo

q_{ij} e q_{ik} = j -ésima e k -ésima coordenada ($j, k = x, y, z$)

\bar{q} = valor médio das coordenadas

O descritor WHIM selecionado para o modelo foi o G2e (*2nd component symmetry directional weighted by atomic Sanderson electronegativities*). Como pode ser visto na Tabela 3.3, Anexo 3, os valores obtidos para esse descritor apresentaram pouca variação e diferem apenas a partir da segunda casa decimal. Isto ocorre porque os descritores WHIM podem variar de 0 a 1. Como o coeficiente autoescalado é negativo, esse descritor pode indicar que a simetria das moléculas em estudo, em função do segundo eixo interno, apresenta influência negativa na potência. O valor nulo corresponde a uma molécula assimétrica, enquanto a unidade corresponde a uma molécula simétrica (TODESCHINI e CONSONNI, 2000).

Na Figura 52, onde é apresentado um composto muito potente (**E5b**), um pouco potente (**P18**) e um intermediário (**W2**), é possível observar que a tendência de posicionamento do segundo eixo está relacionada com o deslocamento do primeiro que, por sua vez, varia em função do deslocamento do centro de massa das moléculas devido à conformação adotada pela cadeia lateral aromática. De modo geral, o segundo eixo encontra-se na região do grupamento ceto-enol do farmacóforo DCA. Porém, enquanto para os compostos mais e menos potentes ele passa exatamente através do mesmo, no intermediário ele encontra-se deslocado. Esta semelhança no posicionamento do segundo

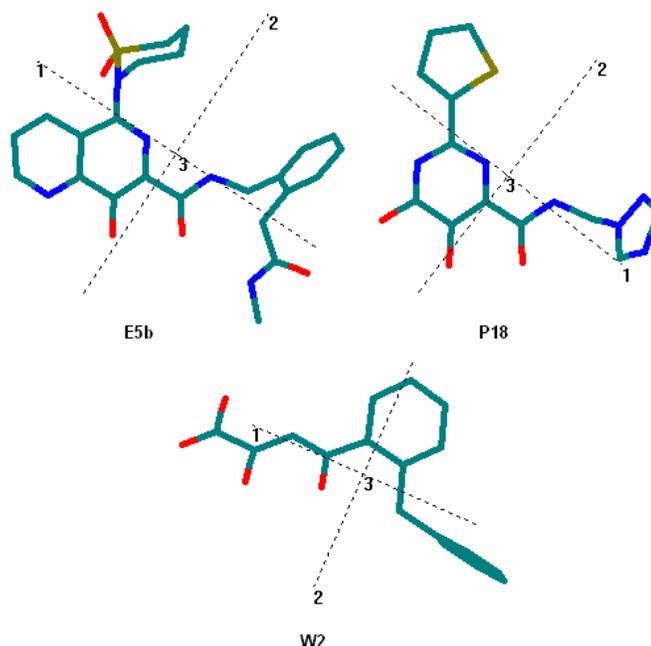


Figura 52. Exemplo dos eixos internos de alguns compostos do conjunto de treinamento (**E5b**, $pIC_{50} = 8,15$; **P18**, $pIC_{50} = 4,300$; **W2**, $pIC_{50} = 5,25$). O terceiro eixo encontra-se perpendicular ao plano da página.

eixo pode explicar o pequeno intervalo de variação dos valores de G_{2e} , apenas 0,04 unidade. Mesmo assim, há uma tendência de que os compostos com menores valores (ou seja, menor simetria em torno do eixo) sejam mais potentes. Por exemplo, o maior valor obtido foi de 0,184 para o composto menos potente, **P18** e o maior foi de 0,144 para **E5a**, um dos compostos mais potentes (Tabela 3.2, Anexo 3).

Quanto à eletronegatividade, considerando a Equação 24, a presença de átomos mais eletronegativos fará com que os termos da matriz sejam menores. A dependência de uma maior eletronegatividade para gerar termos menores pode significar que o parâmetro atômico realmente seja importante para a inibição da reação de transferência de fita. Esse fator também pode estar relacionado com a interação com os íons metálicos e/ou com a formação da interação tipo π -stacking com o sítio de ligação localizado na HIV-IN. Esta

tendência é expressa pelo próprio coeficiente autoescalado de G2e, assim como por seu r no modelo (Tabela 11). Caso esse fator de ponderação não fosse importante para a correlação entre esse descritor e pIC_{50} ($r = -0.468$, antes da retirada dos *outliers*), provavelmente, o descritor G2u (2^{nd} component symmetry directional unweighted) teria sido selecionado. Porém, com seu $r = -0,259$, fica clara a importância dessa propriedade.

4.1.2.4 *Outliers*.

Um problema típico no desenvolvimento de um modelo QSAR/QSPR é a presença de compostos atípicos que não se ajustam no modelo ou cuja previsão realizada é ruim por apresentarem alguma característica diferente em relação aos demais do conjunto de treinamento (os já citados *outliers*, ou, literalmente, ‘fora da linha’). Em geral, esses compostos apresentam tal comportamento devido: (i) à presença de algum parâmetro estrutural diverso; (ii) aos descritores selecionados não terem sido capazes de codificar algum aspecto estrutural relevante; (iii) aos dados experimentais incorretos; (iv) à interação com o mesmo receptor, mas de modo diferente; e/ou (v) a um mecanismo de ação diferente do restante do conjunto de treinamento.

O diagnóstico correto desses compostos é muito importante, pois ajuda a definir as limitações de um modelo (HELGUERRA *et al.*, 2008; FURUSJÖ *et al.*, 2006; VERMA e HANSCH, 2005; KUBINYI, 1993), bem como auxilia no desenvolvimento de modelos mais robustos e significantes quando realizados corretamente (CRONIN e SCHULTZ, 2003).

Apesar da importância, este é um ponto que, em geral, recebe menor atenção no desenvolvimento de um estudo QSAR/QSPR (KIM, 2007a; 2007b). A eliminação desses

compostos normalmente é feita pela avaliação da estatística referente ao seu valor predito, sendo aceito um erro máximo (CRONIN e SCHULTZ, 2003). Assim, um composto *outlier* geralmente é excluído sem uma justificativa específica do motivo desse comportamento.

Segundo Verma e Hansch (2005), modelos com grande e pequena quantidade de *outliers* podem ser classificados como ruins e bons, respectivamente e literalmente. O modelo obtido possui oito *outliers*, detectados ao se utilizar o gráfico de Leverages pelos resíduos de Studentized (GRAMATICA, 2007). Pode-se considerar essa quantidade de *outliers* aceitável (e assim classificá-lo como um bom modelo), pois sua proporção em relação ao conjunto de treinamento completo foi de apenas 10,6%, valor baixo ao se considerar a heterogeneidade dos compostos.

Outro fato que reforça tal conclusão é que o conjunto completo teve como fonte vários artigos da literatura. Segundo cada um deles, os ensaios de inibição da reação ST foram realizados sempre por um mesmo protocolo (HAZUDA *et al.*, 1997). Porém, esta é a única garantia de que esse processo foi realmente utilizado. Além disso, ainda devem ser considerados os potenciais erros que podem ter ocorrido quando da realização do ensaio pelos diferentes grupos ou em momentos diferentes.

Apesar da quantidade aceitável de *outliers*, procurou-se verificar se existiam justificativas estruturais para que cada composto apresentasse tal comportamento, assim como foi feito com o estudo preliminar. A melhor forma encontrada para realizar o processo foi comparar cada um dos *outliers* (Tabela 14) com os derivados estruturalmente mais relacionados, suas atividades, e os valores dos descritores calculados para cada um.

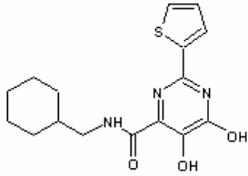
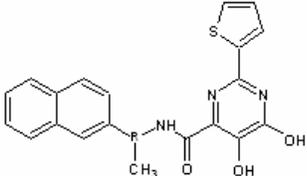
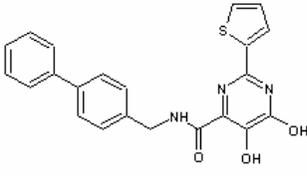
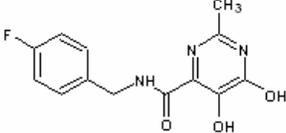
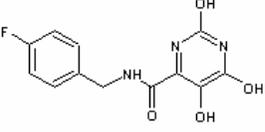
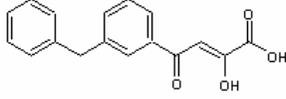
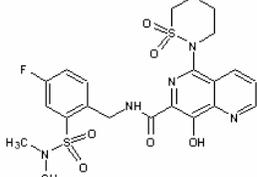
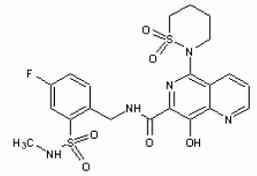
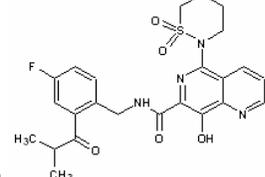
Porém, inicialmente, foi verificado o comportamento do conjunto de *outliers* como um todo. A primeira observação que pode ser feita em relação aos nove compostos é que

quatro deles apresentam pIC_{50} abaixo de 7,00, enquanto o restante está acima desse dado. Apesar da pequena diferença numérica, as quantidades equivalem a 14,288% da faixa até pIC_{50} 4,00 - 6,96 e 8,767% da faixa pIC_{50} 7,00 – 8,15. Como comentado anteriormente, apesar do intervalo razoável de potência apresentado pelo conjunto de treinamento (pIC_{50} 4,00 – 8,15), os compostos concentram-se na faixa acima de 7,00. Assim, essa maior representatividade provavelmente faz com que esse intervalo seja mais bem previsto no processo de validação cruzada, justificando a menor quantidade de *outliers*. Mesmo assim, quando os resultados da validação cruzada (Tabela 3.2, Anexo 3) e os da validação externa (Tabela 12) são consultados, esses apresentam erros bem distribuídos em relação ao intervalo de atividade, além de compostos pouco potentes com baixos erros e vice-versa.

Outro fato a ser considerado é que cinco *outliers* são 4,5-diidroxipirimidinas carboxamidas (**P8**, **P13**, **P34**, **S7** e **S4a**), três são 8-hidroxi-1,6-naftiridinas (**E8a**, **E8b** e **E21a**), e um é β -dicetoácido (**Z1**). Uma justificativa aceitável não pode ser encontrada, exceto o fato de que a maioria dos compostos (42) são 4,5-diidroxipirimidinas carboxamidas e as atividades desses compostos encontram-se bem distribuídas por todo o intervalo. Assim, por consequência da maior representatividade dentro do conjunto de treinamento completo e dos possíveis maiores erros de previsão da faixa de compostos menos potentes, a subestrutura representou a maioria dos *outliers*.

A primeira observação pode ser feita para o composto **P8**. Este composto também foi classificado como *outlier* no Modelo I, provavelmente por ser o único entre todos a não possuir uma cadeia lateral aromática. Apesar dessa justificativa também ser válida para o conjunto de treinamento completo, também é possível observar que ele não seguiu a tendência observada para o descritor G2e: sua atividade é uma das menores do conjunto de

Tabela 14. *Outliers* retirados do Modelo IV.

Outliers	BELe8	RTe	G2e	R1u+	pIC ₅₀
P8 	3,690	0,167	0,156	0,133	4,300
P13 	3,908	0,156	0,154	0,124	6,210
P34 	3,799	0,154	0,153	0,126	4,300
S7 	3,683	0,174	0,167	0,141	7,220
S4a 	3,659	0,275	0,169	0,219	7,070
Z1 	3,737	0,153	0,163	0,162	8,000
E8a 	3,985	0,096	0,145	0,102	7,520
E8b 	3,981	0,106	0,146	0,079	7,350
E21a 	3,845	0,097	0,147	0,080	6,960

treinamento (4,300), porém ele apresenta um valor baixo para esse descritor, situação contrária do esperado.

O composto **P34** também foi um *outlier* em ambos os estudos 2D. Esse composto apresenta potência muito menor que seus isômeros **P32** e **P33**. O fato de serem isômeros pode explicar o porquê dos descritores BELe8 e G2e apresentarem praticamente os mesmos valores nos três compostos. Porém, há uma considerável diferença nos valores dos descritores RTe e R1u+ entre **P34** e seus dois isômeros. O primeiro vale 0,138 para **P32** e 0,133 para **P33**, enquanto o segundo vale 0,146 para **P32** e 0,144 para **P33**. Como esses valores foram praticamente os mesmos entre os dois isômeros, esperava-se que o mesmo deveria ocorrer com **P34**. Porém, isso não foi observado (Tabela 13), o que pode explicar sua classificação como *outlier*.

O composto **P13** é um isômero ótico de **P12**. Os valores de seus descritores são praticamente idênticos, como a maior diferença observada entre seus respectivos RTe (0.016 unidades). Porém, suas potências diferem em 1.46 unidades logaritmicas. Assim, fica claro que os descritores selecionados não são capazes de detectar a influência dessa diferença na potência dos compostos.

O composto **S4a** possui um grupamento hidroxila no lugar de um grupamento aromático (fenila ou tiofenila), como ocorre com o composto mais estruturalmente semelhante a ele, **P4**. Ambos possuem a mesma potência ($pIC_{50} = 7,07$). Porém, novamente os descritores RTe e R1u+ apresentaram valores muito diferentes. Apesar de uma diferença de valores entre algum dos descritores ser algo esperado, considerando que nesse caso não se trata de isômeros, como as potências são idênticas, isso pode ter feito com que **S4a** não fosse bem previsto no processo de validação cruzada e pode ser o motivo

de ter sido classificado como um *outlier*.

A diferença de valores de descritores além do esperado também pode ser o motivo de **S7** ser um *outlier*. Esse composto apresenta a mesma potência de **S6** ($pIC_{50} = 7,22$), a amostra mais estruturalmente semelhante, diferindo por apenas um grupamento metila. Hipoteticamente, os valores de seus descritores deveriam ser quase os mesmos. Porém, novamente os valores dos descritores RTe e R1u+ foram muito diferentes entre ambos os compostos, sendo de 0,085 e 0,086 para **S6**, respectivamente, o que novamente pode ser a justificativa de um composto ser um *outlier*.

O contrário ocorreu com **Z1**. Esse composto também possui um isômero, **W2**, e todos os descritores possuem praticamente os mesmos valores, com uma pequena diferença de 0,025 unidade entre os valores de RTe. Porém, as potências são totalmente diferentes (8,00 e 5,25, respectivamente). Assim, em hipótese, as potências não deveriam diferir tanto e esse pode ser o motivo de **Z1** ser um *outlier*. A mesma situação semelhante pode ser atribuída ao *outlier* **E21a**, a 8-hidroxi-1,6-naftiridina menos potente utilizada nesse trabalho ($pIC_{50} = 6,96$). Esse composto é muito semelhante a **E21b**, diferindo apenas em uma metila, mas é muito mais potente do que o primeiro ($pIC_{50} = 8,00$). Porém, os valores de seus descritores também são equivalentes, sendo a maior diferença observada também entre os valores de RTe (0,021 unidade).

Já os compostos **E8a** e **E8b** não possuem um fator claro para terem sido classificados como *outliers*. Ambos são estruturalmente semelhantes ao composto **E9**, mantido no modelo. Porém, as diferenças de pIC_{50} não são grandes como entre **E21a** e **E21b** (0,48 unidade entre **E8a** e **E9**, e 0,65 **E8b** e **E9**). Os valores dos quatro descritores também são equivalentes, com diferença máxima de 0,02 unidade para os valores de R1u+

entre **E8b** e **E9**. Porém, a não exclusão desses dois compostos leva à uma variância explicada de 57,50% ($R^2 = 0,575$), abaixo do mínimo aceitável. A hipótese mais provável também deve estar relacionada a problemas nos valores das respectivas potências.

4.2 QSAR-4D

Ao se reputar a questão da heterogeneidade dos compostos, foi considerada a possibilidade de que a utilização de descritores que codificassem especificamente energias de interações ligante-receptor pudesse ser mais adequada para o estudo QSAR do conjunto de treinamento em questão. Esta situação é contemplada pelos métodos QSAR-3D a 6D. (MASUNARI e TAVARES, 2007; VEDANI e DOBLER, 2002; SANT'ANNA, 2002; RAGHAVAN, 1995; DEBNATH, 2005; HOPFINGER *et al.*, 1997; VEDANI, DOBLER e LILL, 2005; VEDANI *et al.*, 2007).

Como especificado no capítulo 3, nesta etapa foi utilizada a metodologia LQTA-QSAR (MARTINS *et al.*, 2009). As diferentes conformações obtidas para cada composto do conjunto de treinamento na etapa de DM foram alinhados com base na região mais comum entre todos os compostos: os oito átomos que formam a subestrutura DCA (Figura 53). Desta forma, procurou-se maximizar a semelhança entre os compostos para a etapa de

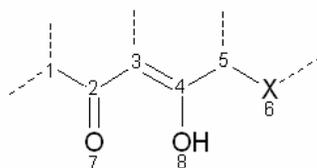


Figura 53. Representação do farmacóforo DCA, apresentando os átomos que foram identificados como 1 a 8 para a etapa de alinhamento dos conformeros obtidos por DM.

análise dos perfis de amostragem conformacionais pelas sondas Zn^{+2} e Ar(C-H) e diminuir uma possível influência de heterogeneidade nos resultados.

A geração dos descritores e a redução de variáveis foram realizadas separadamente para cada sonda e tipo de descritor gerado. Os 43904 descritores gerados pela sonda Zn^{+2} foram divididos em duas matrizes menores, uma 85×21952 com os descritores de Coulomb (C) e a 85×21952 com descritores de Lennard-Jones (LJ). Os descritores LJ gerados pela sonda Ar(C-H) originaram uma matriz 85×21952 . As três matrizes foram reduzidas, mantendo-se apenas aqueles onde $r > |0,2|$. Na matriz de descritores C da sonda Zn^{+2} restaram apenas vinte descritores.

As matrizes reduzidas dos descritores LJ das sondas Zn^{+2} (85×3174) e Ar(C-H) (85×3340) foram submetidas à redução de variáveis utilizando-se o algoritmo de tendências, gerando matrizes 85×1440 e 85×1616 , respectivamente. Essas matrizes foram então tratadas com o método OPS. Como os dados já haviam sido pré-processados com o esquema *blockscaling*, o qual registrou um peso semelhante para cada descritor, foi utilizada a opção *centrar na média* do programa *QSARmodeling*. De modo semelhante ao estudo completo QSAR-2D, O *SEV* foi utilizado para classificar os modelos. O melhor modelo correspondente a primeira matriz foi formado por 95 descritores, e o da segunda por 105 descritores.

As duas matrizes reduzidas (85×95 e 85×105) foram justapostas à matriz de 20 descritores C de Zn^{+2} , obtendo assim uma matriz 85×575 . Para a obtenção dos modelos finais, essa matriz foi submetida a uma nova etapa de seleção de variáveis com o método OPS. Porém, como agora os descritores de C e LJ estão em uma mesma matriz, foi necessário autoescalar os dados.

O melhor modelo obtido pelo método OPS foi formado por 34 descritores. Esse modelo foi transferido para o Pirouette 4 (INFOMETRIX, 2007) onde foi realizada a etapa de refinamento. A quantidade total de descritores foi reduzida para 14, sendo 12 correspondentes à sonda Ar(C-H) e dois à sonda Zn⁺². Tais descritores originaram oito LVs que acumularam 69,249% da informação (Tabela 15). Nove compostos (10,588% do conjunto de treinamento) foram identificados como *outliers* (Tabela 16), a mesma quantidade do Modelo IV. Um deles (**P34**) é um *outlier* comum ao estudo QSAR-2D completo, e a maioria são 4,5-diidroxipirimidinas carboxamidas descritas por Petrocchi e colaboradores (2007). Assim, foi obtido o Modelo XI.

Esse modelo é capaz de explicar 89,70% da informação acumulada, valor consideravelmente superior ao modelo QSAR-2D (Modelo IV). Seu *SEC* pode ser considerado muito baixo, especialmente quando comparado com o modelo 2D (0,270 contra 0,476, respectivamente). O resultado de seu teste *F* a 95% de confiança ($\alpha=0,05$) também ficou muito acima do valor crítico, e assim apresenta confiabilidade estatística superior aquela do modelo QSAR-2D. Assim, a utilização da metodologia LQTA-QSAR foi capaz de gerar descritores que originaram um modelo com capacidade de ajuste de dados muito superior ao melhor obtido utilizando-se QSAR-2D.

O Modelo XI é capaz de prever 84,20% da variância, quantidade bem superior ao do modelo 2D. Isso é uma indicação de que o Modelo XI é capaz de realizar melhores previsões da atividade de novos compostos. Consequentemente, o erro associado à sua capacidade de predição interna é muito menor, já que seu *SEV* também apresenta uma diferença de 0,209 em relação ao valor do parâmetro correspondente ao Modelo IV (0,318 contra 0,523, respectivamente). A diferença entre R^2 e Q^2_{LOO} foi de apenas 0,055 unidades,

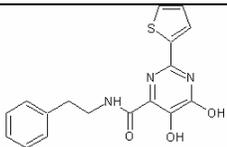
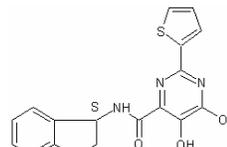
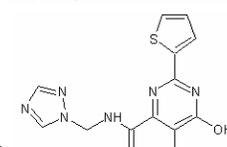
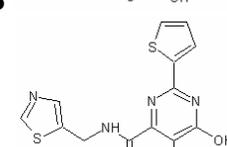
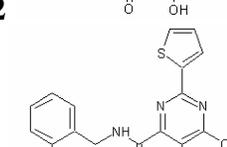
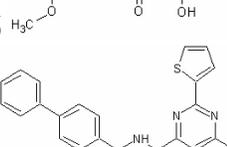
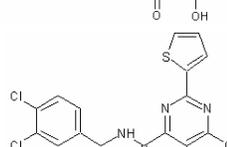
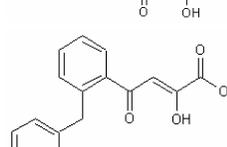
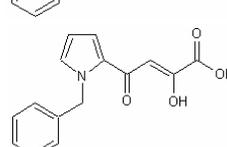
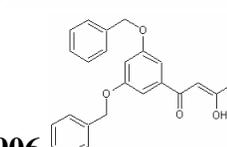
Tabela 15. Contribuição dos descritores do Modelo XI em cada LV, informação de cada LV e coeficientes autoescalados do modelo.

Descritor	LV1	LV2	LV3	LV4	LV5	LV6	LV7	LV8	<i>r</i>	Coeficientes autoescalados
24.13.12 Ar(C-H).LJ	0,196	0,226	0,103	0,298	0,118	0,186	-0,727	-0,349	0,236	0,251
18.18.19 Ar(C-H).LJ	0,227	-0,126	0,394	0,182	-0,457	0,204	0,046	0,304	0,272	0,188
21.15.16 Ar(C-H).LJ	-0,357	0,028	-0,263	-0,096	-0,645	-0,024	-0,299	-0,030	-0,429	-0,324
26.18.14 Ar(C-H).LJ	0,211	0,638	0,038	-0,054	-0,071	0,082	-0,150	0,411	0,254	0,349
19.21.24 Ar(C-H).LJ	0,211	0,005	-0,264	0,296	-0,106	0,318	0,347	-0,241	0,253	0,148
20.13.16 Ar(C-H).LJ	-0,258	0,258	-0,068	0,444	0,160	0,263	0,195	0,378	-0,310	-0,081
14.23.23 Ar(C-H).LJ	-0,249	0,154	0,541	-0,034	-0,292	0,108	0,188	-0,086	-0,299	-0,077
18.24.13 Ar(C-H).LJ	0,236	0,402	0,052	-0,498	-0,114	0,211	0,232	-0,404	0,283	0,266
20.17.19 Zn2+.C	-0,208	0,286	-0,500	-0,053	-0,007	0,199	0,129	-0,007	-0,250	-0,138
22.16.9 Ar(C-H).LJ	0,243	-0,304	-0,184	-0,436	-0,056	0,303	-0,137	0,360	0,291	0,037
20.18.17 Ar(C-H).LJ	-0,210	-0,190	-0,169	0,097	-0,238	0,326	-0,179	-0,074	-0,252	-0,233
15.15.18 Ar(C-H).LJ	-0,220	-0,156	0,206	-0,232	0,372	0,625	-0,111	0,047	-0,264	-0,179
26.20.19 Ar(C-H).LJ	0,416	0,001	-0,189	-0,003	-0,045	-0,101	-0,081	0,258	0,499	0,283
14.15.21 Zn2+.LJ	0,376	-0,193	-0,058	0,276	-0,144	0,227	0,155	-0,205	0,451	0,235
Quantidade de informação	12,404%	7,632%	10,658%	7,591%	6,726%	7,952%	10,441%	5,846%		

$$\begin{aligned}
 pIC_{50} = & 0,201 [24.13.12.Ar(C-H).LJ] + 0,147 [18.18.19.Ar(C-H).LJ] \\
 & - 0,260 [21.15.16.Ar(C-H).LJ] + 0,287 [26.18.14.Ar(C-H).LJ] \\
 & + 0,118 [19.21.24.Ar(C-H).LJ] - 0,064 [20.13.16.Ar(C-H).LJ] \\
 & - 0,062 [14.23.23.Ar(C-H).LJ] + 0,213 [18.24.13.Ar(C-H).LJ] \\
 & - 0,106 [20.17.19.Zn2+.C] + 0,030 [22.16.9.Ar(C-H).LJ] \\
 & - 0,211 [20.18.17.Ar(C-H).LJ] - 0,156 [15.15.18.Ar(C-H).LJ] \\
 & + 0,268 [26.20.19.Ar(C-H).LJ] + 0,185 [14.15.21.Zn2+.LJ] + 6,948
 \end{aligned}
 \tag{XI}$$

$n = 76$; $R^2 = 0,897$; $SEC = 0,270$; $PRESS_{cal} = 4,896$; $F = 72,827$ ($cF = 2,082$); $Q^2_{LOO} = 0,842$; $SEV = 0,314$; $PRESS_{val} = 7,484$ ($SS_y = 47,469$).

Tabela 16. *Outliers* retirados do modelo XI.

Outliers	pIC₅₀
P9 	7,700
P15 	7,7000
P18 	4,300
P22 	5,600
P26 	7,040
P34 	4,300
P39 	8,000
W2 	5,250
W3 	6,660
L708906 	7,000

contra 0,102 do Modelo IV. Isso mostra que a probabilidade de sobreajuste de dados é ainda menor (ROY, LEONARD e ROY, 2008; FERREIRA e KIRALJ, 2009). O resultado de $PRESS_{val}$ ficou ainda mais abaixo de seu valor SS_y de referência que o resultado deste parâmetro para o Modelo IV (diferença de 25,897 para o Modelo IV e 39,986 para o Modelo XI), uma primeira indicação de que a variância predita por esse modelo provavelmente é real, e não devido à correlação ao acaso. Assim, pode-se afirmar que o estudo QSAR-4D levou à obtenção de um modelo que também apresenta qualidade estatística satisfatória quanto a predita.

A Figura 54 apresenta os resultados do teste de randomização de y e da validação LNO para o Modelo XI. É possível observar que o modelo não apresenta correlação ao acaso (intercepto de 0,109 considerando os coeficientes dos R^2 e -0,398 considerando Q^2_{LOO}), confirmando o resultado do parâmetro $PRESS_{val}$. Assim, todos os modelos obtidos utilizando a variável dependente randomizada mostram total falta de ajuste e preditividade. Já o comportamento dos resultados da validação LNO foi altamente homogêneo para todos os N . A diferença entre o Q^2_{LOO} e o Q^2_{LLO} foi de apenas 0,007 unidade (0,838 e 0,831, respectivamente), e assim ambos podem ser considerados praticamente idênticos. A maior oscilação observada foi para L110, com um $Q^2_{L120} = 0,834 \pm 0,015$. Assim, o modelo XI também pode ser considerado robusto.

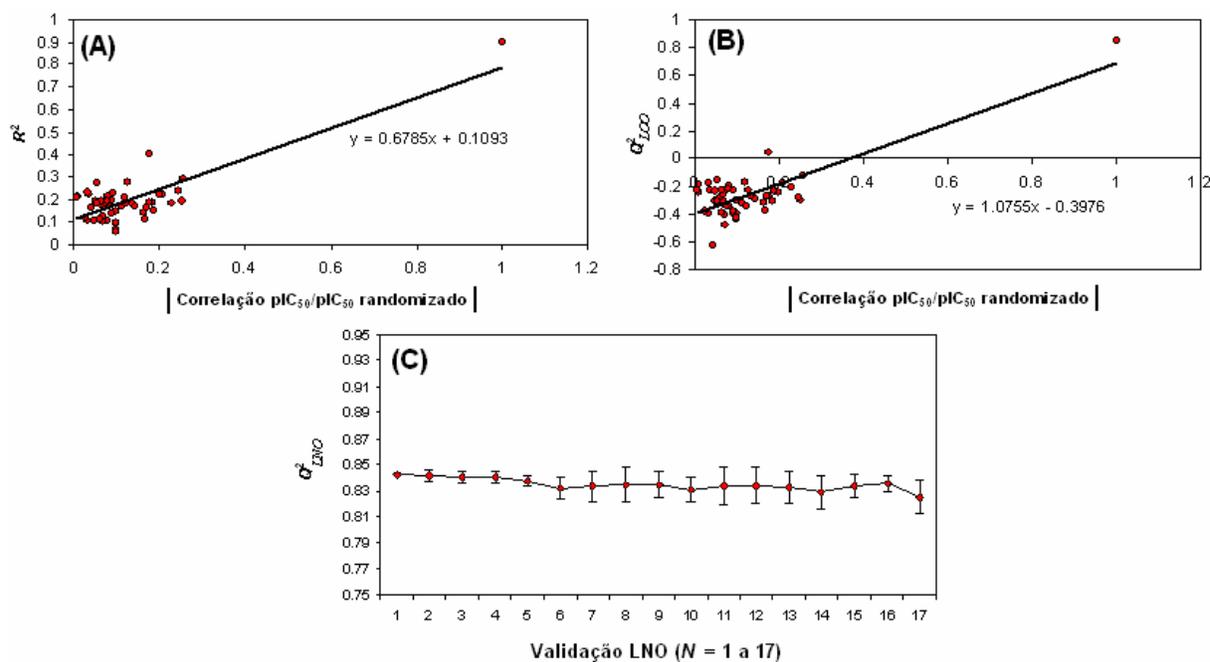


Figura 54. Resultados do teste de randomização de y (A e B) e da validação LNO (C) para o modelo XI.

O Modelo Real XII, obtido após a seleção do conjunto teste, pode ser considerado praticamente idêntico ao modelo XI. Na Figura 55 é apresentado o dendograma construído com o conjunto de treinamento completo, que mostra que o conjunto teste selecionado encontra-se ainda mais bem distribuído que o conjunto utilizado para o estudo QSAR-2D completo. Os resultados da validação externa são apresentados na Tabela 17. O resultado para R^2_{pred} ficou muito acima do limite mínimo de 0,5, sendo tão próximo de Q^2_{LOO} quanto o ocorrido com o Modelo IV. O valor do SEP é muito próximo ao valor do SEV , com diferença de apenas 0,07 unidades. Tais resultados indicam que as capacidades de predição externa e interna também são equivalentes. O resultado do ARE_{pred} mostra que o potencial de erro médio quando da efetiva utilização do modelo é razoavelmente menor que aquele referente ao Modelo IV (amplitude de 3,680%). Isso indica que os valores preditos de novos compostos tenderão a ser mais próximos da realidade do que aqueles obtidos com o

$$\begin{aligned}
 & pIC_{50} = 0,187 [24.13.12.Ar(C-H).LJ] + 0,167 [18.18.19.Ar(C-H).LJ] \\
 & -0,241 [21.15.16.Ar(C-H).LJ] + 0,320 [26.18.14.Ar(C-H).LJ] \\
 & +0,101 [19.21.24.Ar(C-H).LJ] - 0,029 [20.13.16.Ar(C-H).LJ] \\
 & -0,068 [14.23.23.Ar(C-H).LJ] + 0,166 [18.24.13.Ar(C-H).LJ] \\
 & -0,104 [20.17.19.Zn^{2+}.C] + 0,024 [22.16.9.Ar(C-H).LJ] \\
 & -0,252 [20.18.17.Ar(C-H).LJ] - 0,171 [15.15.18.Ar(C-H).LJ] \\
 & +0,265 [26.20.19.Ar(C-H).LJ] + 0,202 [14.15.21.Zn^{2+}.LJ] + 6,970
 \end{aligned}
 \tag{XII}$$

$n = 62$; informação acumulada: 67,995%; $R^2 = 0,906$; $SEC = 0,250$; $PRESS_{cal} = 3,314$; $F = 63,678$ ($cF = 2,094$); $Q^2_{LOO} = 0,835$; $SEV = 0,306$; $PRESS_{val} = 5,793$ ($SS_y = 35,169$).

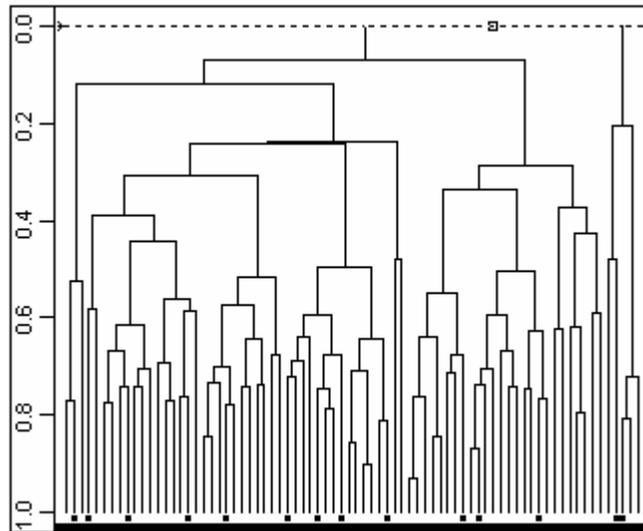


Figura 55. Dendrograma (dados autoescalados) do conjunto de treinamento completo para o Modelo XI. O conjunto teste encontra-se destacado (pontos pretos).

Modelo IV. Além dos valores desses parâmetros serem de qualidade muito superior àqueles do Modelo IV, os resultados de k e k' localizaram-se dentro do intervalo proposto e

o parâmetro $\left| R_0^2 - R'_0{}^2 \right|$ ficou abaixo de 0,3, como desejado.

Tabela 17. Resultados da etapa de validação externa realizada com o modelo real XII a partir do conjunto teste selecionado do modelo auxiliar XI.

Composto	pIC ₅₀ obs	pIC ₅₀ prev	Resíduos
E7	8,000	8,373	-0,373
E8B	7,350	7,626	-0,276
G12	7,400	6,930	0,470
S9	7,300	6,835	0,465
S12	6,700	7,059	-0,359
W7	6,260	6,054	0,206
W11	6,150	6,393	-0,243
Z5	6,430	7,344	-0,914
Z7	8,000	8,005	-0,005
P10	6,000	6,272	-0,272
P16	6,210	6,421	-0,211
P19	4,770	5,169	-0,399
P32	6,890	6,650	0,240
L731927	6,300	6,388	-0,088
			R^2_{pred} 0,839
			SEP 0,384
			ARE _{pred} 4,942%
			k 0,981
			k' 1,016
			$ R_0^2 - R'_0 ^2$ 0,0257

A partir dos resultados apresentados, o modelo QSAR-4D mostrou qualidade estatística muito superior ao QSAR-2D, mostrando ajuste, significância estatística, capacidade preditiva interna, externa e robustez muito superiores. O Modelo XII também mostrou ser autoconsistente, já que os coeficientes do modelo apresentaram o mesmo comportamento das correlações individuais entre cada descritor e a atividade biológica (FERREIRA e KIRALJ, 2009). Por último, ambos os modelos apresentaram a mesma quantidade de *outliers*, embora apenas um (P34) seja comum a ambos. Assim a utilização do primeiro é preferencial para a previsão da potência de novos derivados.

Como no caso dos modelos QSAR-2D, foi verificada a possibilidade de se traçar uma relação com o mecanismo de ação. Neste caso, o significado dos descritores de Coulomb e Lennard-Jones é bem mais claro do que aqueles selecionados para o Modelo

IV, pois ambos estão relacionados às energias de interações tridimensionais (MARTINS *et al.*, 2009).

A Figura 56 apresenta os descritores selecionados plotados no espaço tridimensional. A primeira observação que pode ser facilmente feita é quanto à quantidade de descritores relativos à sonda Ar(C-H) selecionada em relação à sonda Zn^{+2} . Os descritores da primeira sonda (12) encontram-se espalhados por todo o espaço, enquanto os da segunda (dois) concentram-se em uma pequena região.

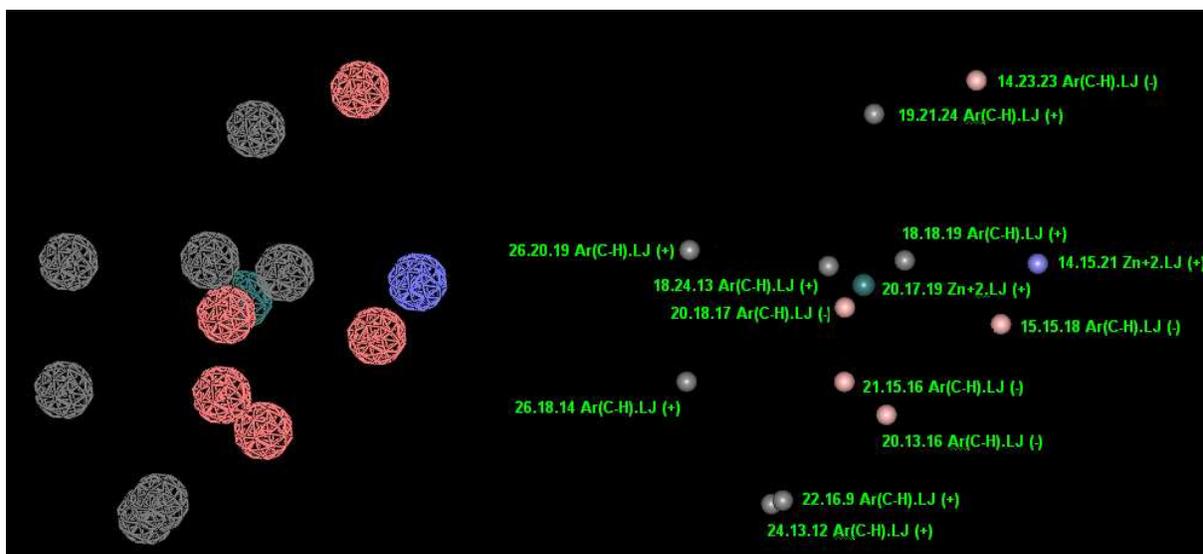


Figura 56. Descritores relativos ao modelo QSAR-4D (Modelo XI) plotados no espaço tridimensional. Esquerda: apresentados como grade. Direita: forma pontual e suas respectivas indicações no formato x.y.z.sonda.potencial.(sinal do coeficiente). Azul: descritor Zn^{2+} .LJ positivo; verde: descritor Zn^{2+} .C negativo; cinza: descritor Ar(C-H).LJ positivo; rosa: descritor Ar(C-H) negativo. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYS, 2001).

Considerando-se apenas os descritores da sonda Zn^{+2} , a qual representa o(s) co-fator(es) metálico(s) localizado(s) no sítio catalítico da HIV-IN, pode-se observar que os mesmos encontram-se razoavelmente próximos entre si (6,60 Å). Apesar dos descritores não serem representações exatas dos co-fatores metálicos, e sim apenas perturbações em pontos

específicos da grade tridimensional que se correlacionaram com a atividade biológica, essa distância é similar aquelas propostas por vários autores como distâncias inter-metálicas. Maurin e colaboradores (2004) utilizaram estudos de espectroscopia ESR (*electron spin resonance*) para a avaliação do comportamento do ácido benzopirúvico (Figura 18) na presença de íons Mg^{+2} em soluções com pH variando entre 7,50 e 10,00, e propuseram uma distância inter-metálica localizada na faixa de 4,00 – 6,00 Å. Todavia, Pais e colaboradores (2002) propuseram um modelo teórico HIV-IN/DNA em que os co-fatores apresentam uma distância de 7,00 Å entre si. O modelo teórico 1WKN (De LUCA *et al.*, 2005; Figura 9) apresenta distância inter-metálica de 3,86 Å. No complexo 1QS4 (GOLDGUR *et al.*, 1999), que não apresenta dois metais, a distância entre o átomo de Mg^{+2} e o resíduo Glu152 (um dos aminoácidos onde o segundo co-fator metálico liga-se) é um pouco superior (9,99 Å), mas ainda é aceitável para critério de comparação.

Tendo em vista a importância da interação inibidor-metal, faz-se necessária uma análise mais aprofundada desses dois descritores. Seus valores são apresentados na Tabela 3.6, Anexo 3, e na Figura 57 eles são apresentados no espaço localizado ao redor do inibidor **E5b**. Considerando o coeficiente do descritor 20.17.19.Zn2+.C, pode-se propor que o mesmo descreva o desfavorecimento de interações eletrostáticas entre os inibidores e o sítio de ligação. Isto fica confuso quando apenas a Figura 58 é analisada, pois esse descritor encontra-se mais próximo ao grupamento carbonila ceto enólica (1,00 Å). O mais esperado seria que esse descritor favorecesse a inibição da reação ST. Mas seu significado fica claro quando é utilizado o perfil de amostragem conformacional de um composto pouco potente (**L731942**, $pIC_{50} = 5,12$) (Figura 59). A cadeia lateral aromática desse composto é extremamente flexível, e devido a isso parte dela (mais especificamente, o anel pirrol) pode ocupar a coordenada $x = 16$, $y = 15$ e $z = 17$. Assim, é possível que o descritor

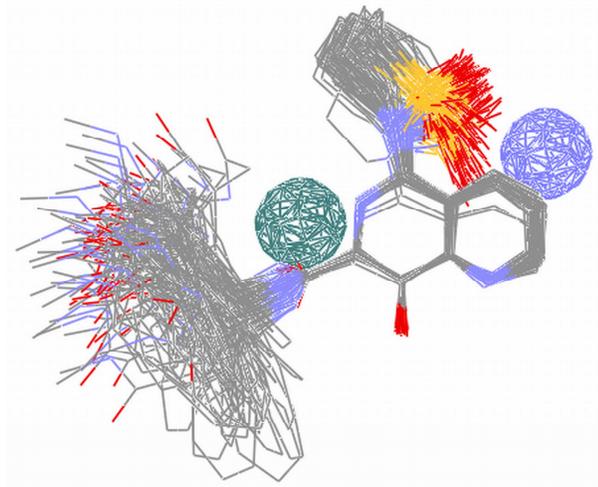


Figura 57. Descritores 20.17.19.Zn2+.C (verde) e 14.15.21.Zn2+.LJ (azul) no espaço tridimensional em torno do compostos **E5b**. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).

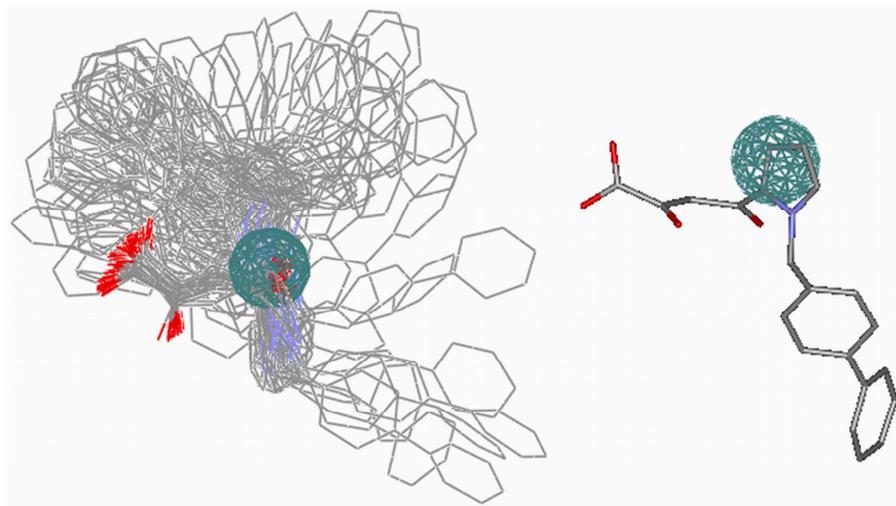


Figura 58. Descritores 20.17.19.Zn2+.C no espaço tridimensional em torno do perfil de amostragem conformacional (formato arame) do inibidor **L731942**. Devido a grande flexibilidade conformacional do inibidor, uma conformação isolada (formato bastões) é apresentada para facilitar a visualização. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).

esteja relacionado à perturbação de uma possível interação com um dos íons metálicos.

Já o descritor 14.15.21.Zn²⁺.LJ (azul), o mais importante entre ambos, localiza-se 3,62 Å a frente do perfil de amostragem conformacional de **E5b** (Figuras 57 e 59) e a 3,66 Å do perfil de **L731942** (Figura 60). Em ambos os perfis, esse descritor também se encontra nas vizinhanças de grupamentos hidrofóbicos. No primeiro, é o grupamento tetraidro-tiopirano de **E5b**. No segundo, a extremidade da cadeia lateral aromática de **L731942** pode ocupar essa região devido a sua grande flexibilidade, aproxima-se até 3,39 Å do descritor (Figura 60C). Como seu coeficiente autoescalado é positivo, pode-se propor que a ocupação dessa coordenada esteja relacionada com o aumento da potência.

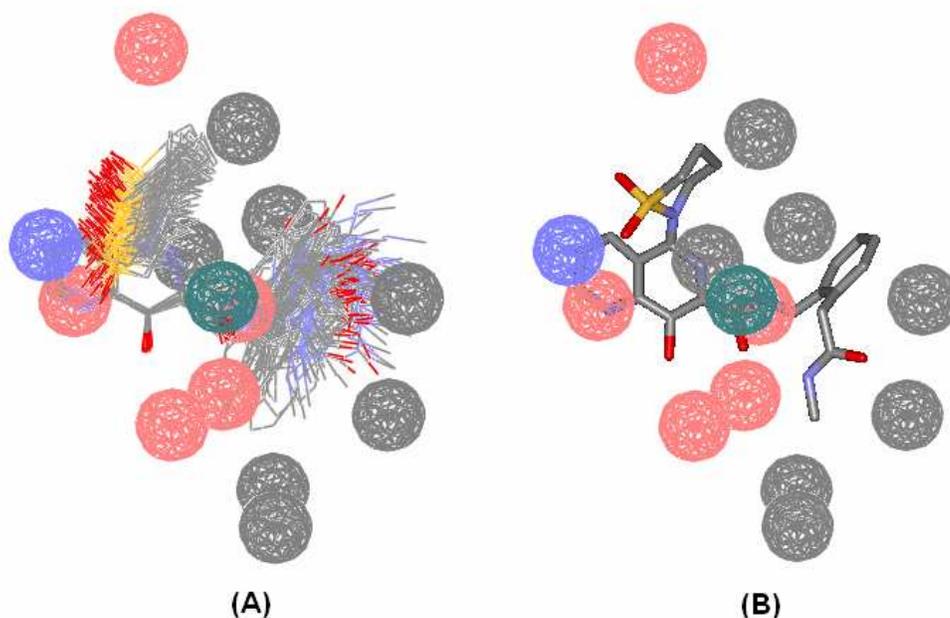


Figura 59. Modelo completo no espaço tridimensional em torno do PAC (A, formato arame) e de uma conformação isolada (B, formato bastões) do inibidor **E5b**. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYS, 2001).

As distâncias entre as duas estruturas e o descritor 14.15.21.Zn²⁺.LJ são razoavelmente pequenas (1,57 Å para **E5b** e 3,39 Å para **L731942**). Além disto, ao se considerar apenas o substituinte tetraidro-tiopirano de **E5b**, é fato aceito que esta posição

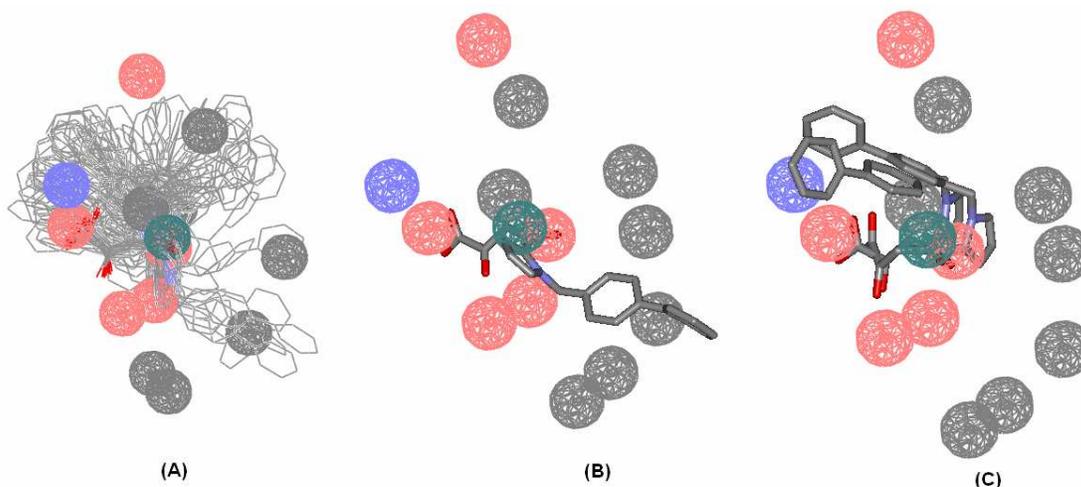


Figura 60. Modelo completo no espaço tridimensional em torno do PAC (A, formato arame) e de conformações isoladas (uma em B e duas em C, em formato bastões) do inibidor **L731942**. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRY, 2001).

tem a principal função de ajudar a adequar o comportamento farmacocinético dos inibidores (DAYAM, AL-MAWSAWI e NEAMATI, 2007), ou seja, não participa do mecanismo de ação. Outra possibilidade é a ocorrência de interações acessórias, ou seja, sem relação com as interações com os co-fatores metálicos ou com a alça flexível, hipótese levantada por vários autores e comentada no Capítulo 1. Algumas dessas ocorreriam exatamente nesta região ‘frontal’ dos inibidores. Por exemplo, as propostas para uma naftilpirina similar a **E5b** e os resíduos His67 e Glu92 (HEALY *et al.*, 2009; Figuras 19 e 61), são claramente do tipo van der Waals. Assim, também existe a possibilidade do descritor 14.15.21.Zn2+.LJ estar relacionado a algumas dessas interações.

As Figuras 60 e 61 também são úteis como comparações do grau de flexibilidade apresentado por um composto muito e outro pouco potente. É muito claro que o composto **L731942** apresentou uma variação conformacional nas condições utilizadas na etapa DM muito maior que **E5b**. Compostos conformacionalmente menos estáveis provavelmente são

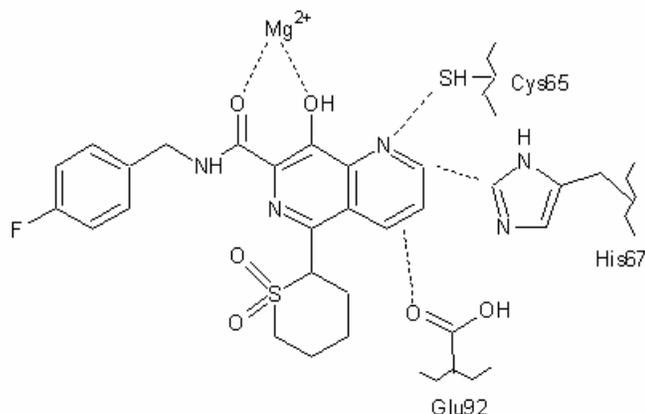


Figura 61. Interações ligante-receptor entre uma naftilpiridina e o sítio de ligação da HIV-IN propostas por Healy e colaboradores (2009).

menos propícios à formação de interações estáveis com o sítio de ligação da HIV-IN, o que os faria menos potentes. Informação semelhante foi atribuída ao descritor E_T (Energia Total), no Modelo I. A variação conformacional também pode ser considerada uma possível explicação para os descritores espalhados em todo o espaço localizado ao redor dos inibidores e não concentrados apenas em regiões próximas à cadeia lateral e à subestrutura DCA. Esta é a melhor explicação para a localização dos descritores 14.23.23.Ar(C-H).LJ e 19.21.24.Ar(C-H).LJ (Figura 57). Enquanto em **E5b** eles estão próximos ao substituinte tetraidro-tiopirano, que não possui função no mecanismo de inibição (DAYAM, AL-MAWSAWI e NEAMATI, 2007), em **L731942** eles encontram-se em regiões ocupadas pela cadeia lateral.

A interpretação dos descritores Ar(C-H) (Tabelas 3.7 e 3.8, Anexo 3) pode ser considerada um pouco mais complexa. A quantidade desses descritores selecionados para o modelo (12, apresentados nas Tabelas 3.8 e 3.9, Anexo 3) é provavelmente decorrente da grande flexibilidade da cadeia lateral. Mesmo nos compostos mais rígidos, como **E5b**, esta região das moléculas apresentou considerável variação conformacional durante a simulação DM. Apesar disto, é possível observar que oito descritores Ar(C-H)

concentraram-se na região em volta da cadeia lateral (Figura 62), sendo que quatro possuem coeficientes positivos e um negativo. Este fato pode ser relacionado com o posicionamento dos inibidores em relação à alça flexível. Entre os nove descritores, os três mais importantes (ver coeficientes autoescalados na Tabela 15) são 26.18.14.Ar(C-H).LJ, 21.15.16 Ar(C-H).LJ e 26.20.19.Ar(C-H).LJ, sendo o primeiro e o segundo os descritores mais importantes do modelo. Porém, apenas o primeiro e terceiro encontram-se posicionados em torno da cadeia lateral. Como seus coeficientes são positivos, é possível inferir que as interações codificadas por esses descritores favorecem o aumento da potência de inibidores da reação ST. Isto pode ser traduzido em um favorecimento da interação hidrofóbica aromática ou π -stacking com a alça flexível.

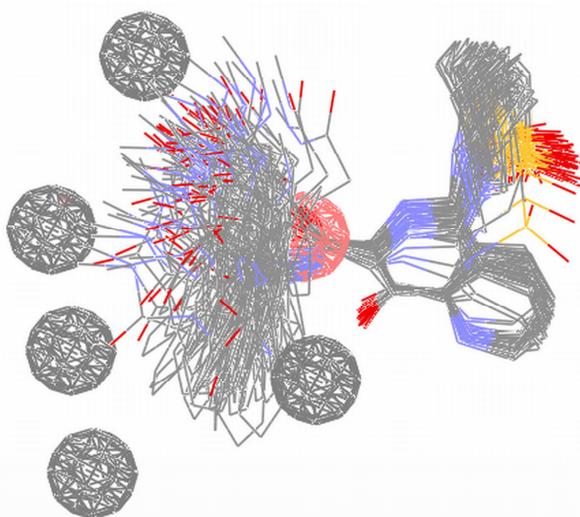


Figura 62. Descritores Ar(C-H).LJ localizados no espaço tridimensional a cadeia lateral aromática (perfil de amostragem conformacional do inibidor **E5b**). Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYS, 2001).

Porém, o posicionamento do descritor 21.15.16.Ar(C-H).LJ leva a uma interessante proposta. Este descritor apresenta um coeficiente negativo. Quando plotado no espaço tridimensional, esse descritor ocupa as coordenadas $x = 21$, $y = 15$ e $z = 16$, posicionando-

se muito próximo ao grupamento ceto enol (1,88 Å da carbonila e 3,61 Å da hidroxila). Já que este é um descritor de Lennard-Jones com coeficiente negativo, é possível propor que a potência dos inibidores é desfavorecida quando esta coordenada é ocupada. Além de sua importância no modelo (possui o segundo maior coeficiente autoescalado, em valores absolutos) e de sua proximidade ao grupamento ceto-enol, outro fator ajuda a dar suporte a esta hipótese: a distância de 6,32 Å com o descritor 15.15.18.Ar(C-H).LJ (Figura 63). Este descritor apresenta exatamente o mesmo comportamento de 21.15.16.Ar(C-H).LJ, possui razoável importância no modelo e, o que pode ser considerado mais importante, localiza-se ao lado do átomo identificado como X (X=N, O), que faz parte do farmacóforo DCA. Assim, a distância entre ambos encontra paralelo com aquelas propostas como as distâncias entre os co-fatores metálicos localizados no sítio de ligação da HIV-IN, em especial a distância de 4,00 – 6,00 Å porposta por Marin e colaboradores (2004). Assim, é bem possível que tal descritor esteja relacionado não a uma interação de van der Waals, mas à interação dos inibidores com um co-fator metálico.

Como o Modelo XI constitui um estudo QSAR-4D *independente do receptor* (QSAR-4D IR), a interpretação apresentada até aqui pode ser considerada suficiente e diretamente relacionada com o mecanismo de inibição da reação ST. Porém, como mostrado por Martins e colaboradores (2009), a construção de uma figura com os descritores localizados no sítio de ligação em estudo pode auxiliar na interpretação dos resultados obtidos. O resultado é apresentado na Figura 64, onde os descritores são superpostos à estrutura secundária da HIV-IN (direita) e apenas com os aminoácidos possivelmente mais importantes para a inibição (esquerda) (GOLDGUR *et al.*, 1999; SOTRIFFEER, NI e MCCAMMON, 2000; NI, SOTRIFFER e MCCAMMON, 2001; KESERÛ e KOLOSSVARY, 2001; CHEN *et al.*, 2008; HEALY *et al.*, 2009). A sobrepo-

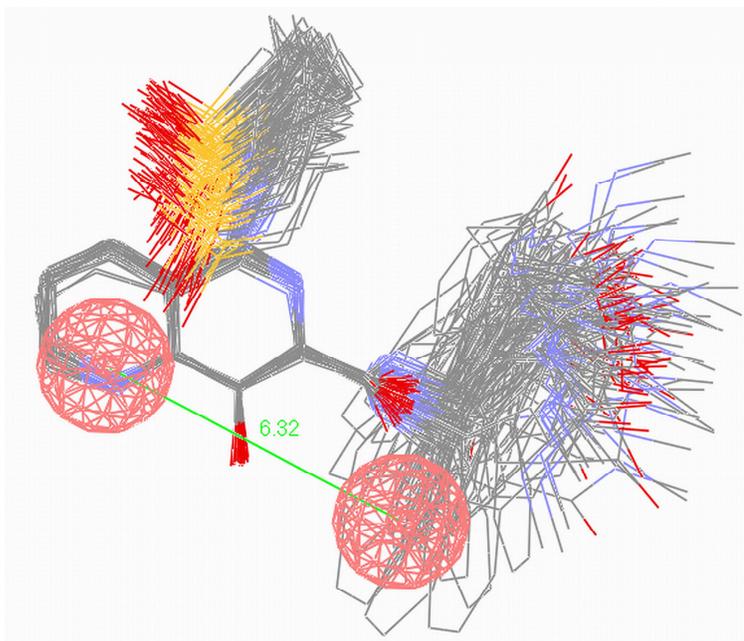


Figura 63. Descritores 21.15.16.Ar(C-H).LJ e 15.15.18.Ar(C-H).LJ no espaço tridimensional em torno do inibidor **E5b**, e da distância entre ambos. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRYS, 2001).

sição foi realizada com base na posição do átomo de Mg^{+2} . Assim, é possível observar que os descritores concentraram-se na região dos aminoácidos essenciais para a atividade catalítica (Asp64, Asp116 e Glu152) e no espaço onde, provavelmente, encontra-se a alça desordenada. Apesar desse resultado ter valor apenas qualitativo, ele fortalece os resultados do presente trabalho e corrobora com o possível mecanismo de inibição da reação ST.

Estes resultados não compõem o primeiro estudo QSAR-4D com inibidores de HIV-IN. Iyer e Hopfinger (2007) realizaram um estudo com 213 inibidores pertencentes a 12 compostos estruturalmente diversos, e, ao dividirem o conjunto em subconjuntos menores, avaliaram a possibilidade de descritores denominados impressões digitais 4D (4D

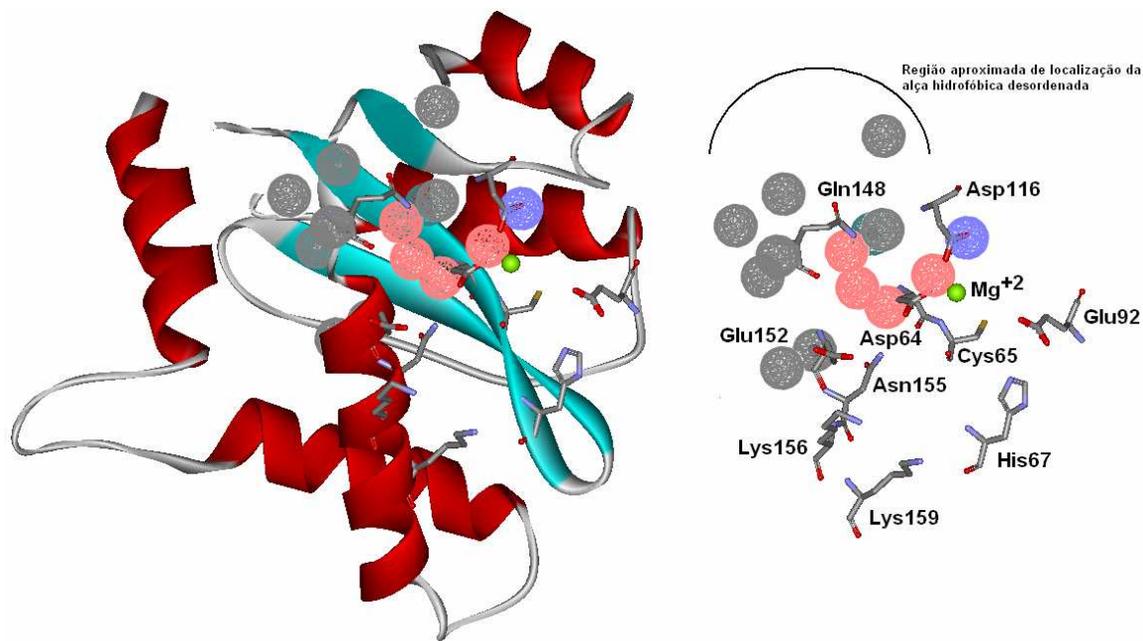


Figura 64. Plotagem dos descritores sobre o sítio de ligação da HIV-IN (PDB 1QS4), sem 5CITEP. A alça hidrofóbica é representada apenas como uma região (direita), pois, nesta estrutura cristalográfica, ela não foi resolvida. A sobreposição foi construída utilizando o programa HyperChem 7.1 (HYPERCUBE, 2002) e interpretada no programa ViewerLite 4.2 (ACCELRYYS, 2001).

fingerprints) gerarem bons modelos QSAR. Porém, dois motivos impossibilitam a utilização deste estudo para auxiliar na corroboração dos resultados do estudo aqui apresentados: (i) entre as 12 diversas classes, nenhum delas possui exatamente a subestrutura DCA. As que mais se aproximariam seriam as estirilquinolinas (Figura 65); e (ii) a atividade considerada foi a inibição da reação 3'P, a primeira etapa da ação biológica da HIV-IN.

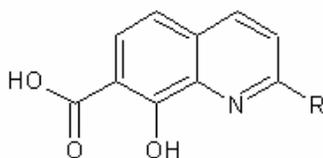


Figura 65. Estrutura básica das estirilquinolinas.

Considerando que a nova metodologia LQTA-QSAR utiliza os conceitos de QSAR-3D, mas especificamente do CoMFA (MARTINS *et al.*, 2009), é válido consultar a literatura em busca de dados de outros estudos que possam corroborar, de alguma forma, os resultados aqui apresentados. A partir de consultas às bases *Web of Science* (<http://www.isiknowledge.com>) e *Scopus* (<http://www.scopus.com>) e ao artigo de revisão publicado por Debnath (2005), foram encontrados 14 artigos de estudos QSAR-3D com inibidores de HIV-IN publicados entre 1995 e 2009 (Tabela 18). Desses, sete utilizaram CoMFA, mas apenas dois construíram modelos considerando a reação ST.

Raghavan e colaboradores (1995) construíram modelos CoMFA a partir de flavonóides, utilizando como variáveis dependentes ambas as reações catalisadas pela enzima. O modelo relativo à reação ST codificava 20,5% de informações relacionadas aos efeitos estéricos e 79,5% a efeitos eletrostáticos. Os descritores formaram campos que mostraram regiões relacionadas à importância de carga positiva (azul) localizada em região equivalente ao farmacóforo de três pontos (Figura 66).

Kuo e colaboradores (2004) também construíram modelos CoMFA para ambas as reações com um conjunto de inibidores de HIV-IN estruturalmente diversificado. Apesar do trabalho apresentar apenas os mapas relacionados ao modelo 3°P (Figura 67), os autores citam que esses são altamente similares ao modelo ST. A região em vermelho foi traduzida em uma região de alta densidade eletrônica próxima à região equivalente ao ceto enol do 5CITEP co-cristalizado no 1QS4, como determinado nos resultados de estudos de *docking* apresentados no mesmo trabalho. Esta parte do modelo favoreceria a potência dos inibidores. Além disso, grandes superfícies (amarelo) próximas ao anel aromático indicam que substituintes volumosos nesta região também favorecem a atividade. Os estudos de *docking* indicaram que a região encontra-se muito próxima à região do sítio hidrofóbico.

Tabela 18. Relação dos artigos de QSAR-3D com inibidores de HIV-IN publicados entre 1995 e 2009.

Autores	Metodologia utilizada	Referência	Atividade biológica
Raghavan, K. e colaboradores	CoMFA (<i>comparative molecular field analysis</i>)	<i>Journal of Medicinal Chemistry</i> 1995 , 38, 890-897.	Inibição das reações 3'P e ST
Makhija, M.T.; Kulkarni, V.M.	CoMSIA (<i>comparative molecular similarity indices analysis</i>)	<i>Journal Computer Aided Molecular Design</i> 2001 , 15, 961-978	Não especificado
Makhija, M.T.; Kulkarni, V.M.	CoMFA e CoMSIA	<i>Journal Computer Aided Molecular Design</i> 2002 , 16, 181-200	Inibição das reações 3'P e ST
Buolamwini, J.K.; Assefa, H.	CoMFA e CoMSIA	<i>Journal of Medicinal Chemistry</i> 2002 , 45, 841-852	Inibição da reação 3'P
Kuo e colaboradores	CoMFA e CoMSIA	<i>Journal of Medicinal Chemistry</i> 2004 , 47, 385-399	Inibição das reações 3'P e ST
Daeyaert, F.F.D. e colaboradores	CoMSIA	<i>Internet Electronic Journal of Molecular Design</i> 2004 , 3, 528-543	Não especificado
Mustata, G.I.; Brigo, A.; Briggs, J.M.	Catalyst/HypoGen	<i>Bioorganic and Medicinal Chemistry Letters</i> 2004 , 14, 1447-1454	Inibição da reação ST
Costi, R. e colaboradores	GRID/GOLPE	<i>Bioorganic and Medicinal Chemistry</i> 2004 , 12, 199-215.	Inibição da reação 3'P
Makhija, M.T.; Kasluwal, R.T.; Kulkarni, V.M.; Neamati, N. ^(a)	CoMFA	<i>Bioorganic and Medicinal Chemistry</i> 2004 , 12, 2317-2333	Não especificado
MA, X.; Zhang, X.; Tan, J.; Chen, W.; Wang, C.	CoMFA	<i>Acta Pharmacologica Sinica</i> 2004 , 25, 950-958	Inibição da reação 3'P
Yuan, H.; Parril, A.	MFA (<i>molecular field analysis</i>)	<i>Journal of Molecular Graphics and Modelling</i> 2005 , 23, 317-328	Inibição da reação 3'P
Nunthaboot, N.; Tonmunphean, S.; Parasuk, V.; Wolschann, P.; Kokpol, S.	CoMFA e CoMSIA	<i>European Journal of Medicinal Chemistry</i> 2006 , 41, 1359-1372	Inibição da reação 3'P
Leonard, J.T.; Roy, K.	MSA (<i>molecular shape analysis</i>)	<i>European Journal of Medicinal Chemistry</i> 2008 , 43, 81-92.	Inibição da reação 3'P
Dhaked, D.K.; Verma, J.; Saran, A.; Coutinho, E.C.	CoRIA (<i>comparative residue interaction analysis</i>)	<i>Journal of Molecular Modeling</i> 2009 , 15, 233-245	Inibição da reação 3'P

^(a) utilizam dados previamente obtidos para o planejamento *de novo* e síntese de novos derivados.

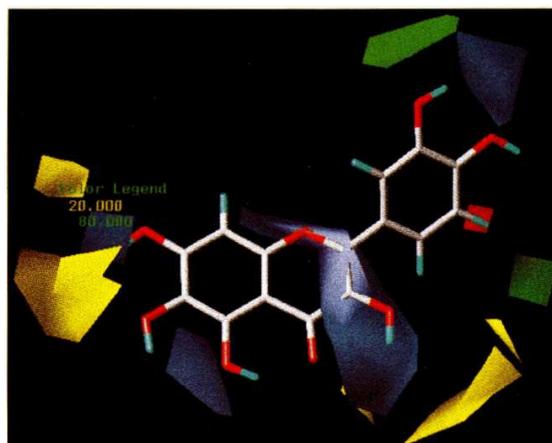


Figura 66. Mapas de contorno do modelo CoMFA relativos à inibição da reação ST, propostos por Raghavan e colaboradores (1995). A imagem corresponde à Figura 7 do artigo original. Reproduzida com permissão da American Chemical Society® licença nº 2224490959169 obtida *on-line* via <http://www.copyright.com>).

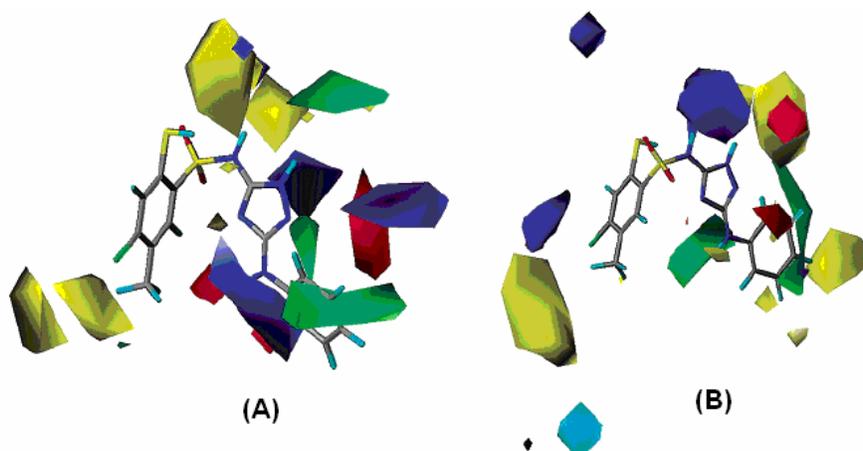


Figura 67. Mapas de contorno (para duas conformações, A e B) dos modelos CoMFA relativos à inibição da reação 3'P (porém similares aos mapas do modelo ST) propostos por Kuo e colaboradores (2004). A imagem corresponde a uma parte da Figura 4 do artigo original. Parcialmente adaptada e reproduzida com permissão da American Chemical Society® licença nº 2224491199544 obtida *on-line* via <http://www.copyright.com>).

Outro estudo, realizado por Makhija e Kulkarni (2002b), utilizou CoMFA para estudar a capacidade de inibição da reação ST de uma série de tiazolotiazepinas. Segundo os resultados apresentados na Figura 68, os campos eletrostáticos (azul e vermelho) e estéricos (verdes) correspondem a regiões ocupadas pelos aminoácidos Asp64 e Asp116 do complexo cristalográfico 1QS4, resíduos aos quais o íon Mg^{+2} encontra-se coordenado.

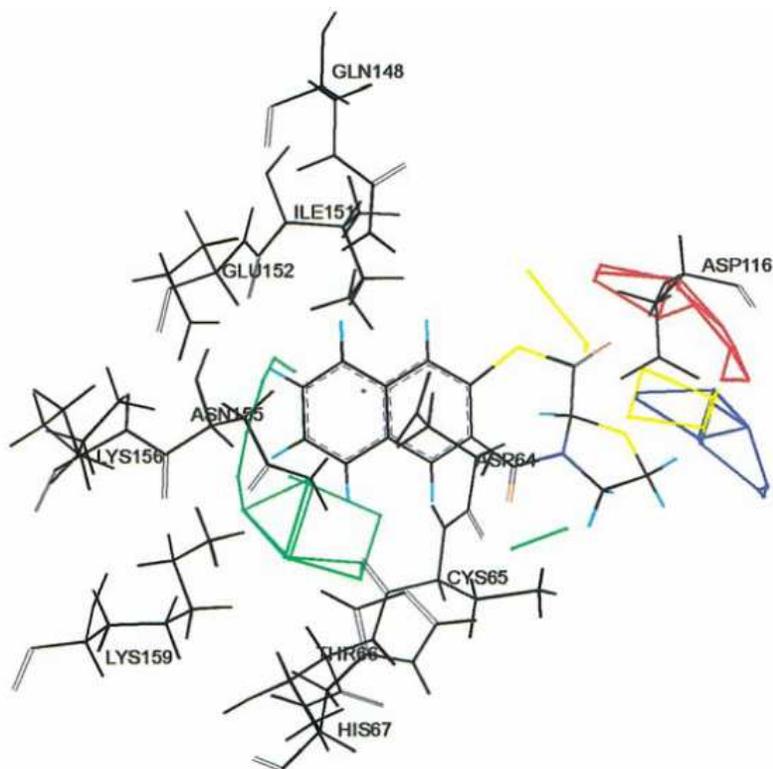


Figura 68. Mapas de contorno do modelos CoMFA relativos à inibição da reação ST propostos por Makhija e Kulkarni (2002b). A imagem corresponde a uma parte da Figura 11 do artigo original. Parcialmente adaptada e reproduzida com permissão da SpringerLink® (licença nº 2233250553421 obtida *on-line* via <http://www.copyright.com>).

Os três resultados apresentados em trabalhos diferentes ajudam a fortalecer as hipóteses levantadas com o Modelo XI. Ambos descreveram regiões de alta densidade eletrônica, próximas à região de ligação com os co-fatores metálicos com tendência a favorecerem a potência dos inibidores. Os resultados podem ser facilmente comparados

com o significado proposto para os descritores 15.15.18.Ar(C-H).LJ, 20.17.19.Zn²⁺.C e 21.15.16.Ar(C-H).LJ. Já as regiões próximas ao anel aromático do modelo de Kuo e colaboradores (2004) (região esquerda das partes A e B da Figura 68) podem ser comparadas aos descritores 26.18.14.Ar(C-H).LJ, 26.20.19.Ar(C-H).LJ e 24.13.12 Ar(C-H).LJ, já que os mesmos também são positivos, formam uma superfície razoavelmente ampla (Figura 69) e posicionam-se próximos à cadeia lateral aromática.

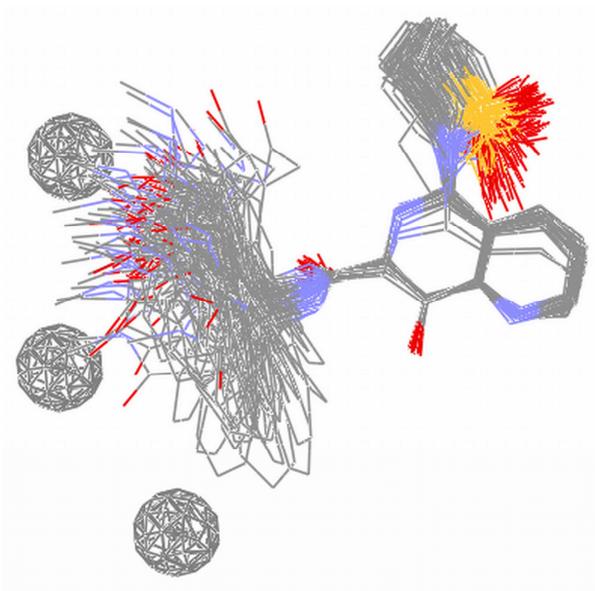


Figura 69. Descritores 26.18.14.Ar(C-H).LJ, 26.20.19.Ar(C-H).LJ e 24.13.12 Ar(C-H).LJ no espaço tridimensional em torno do composto **E5b**. Figura construída e interpretada no programa ViewerLite 4.2 (ACCELRY, 2001).

4.3 QSAR-2D/4D

Após a obtenção dos modelos QSAR-2D e 4D, os descritores de ambos foram combinados, formando uma matriz 85x21, na tentativa de gerar um modelo misto. Este estudo baseou-se na hipótese de que o modelo híbrido poderia originar LVs que codificariam tanto informações relacionadas diretamente com as interações fármaco-receptor quanto outras características importantes para a reação ST. Outros autores

(MAKHIIJA e KULKARNI, 2002b; HOPFINGER *et al.*, 1997; IYER e HOPFINGER, 2007; RAVI *et al.*, 2001) já procederam desta forma, em especial com o descritor LogP, sempre visando melhorar a qualidade dos modelos.

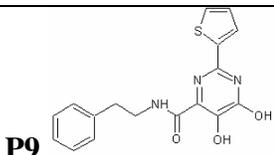
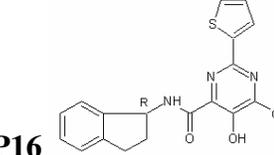
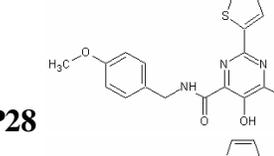
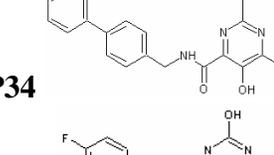
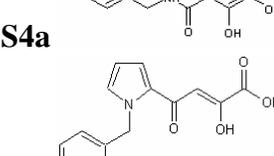
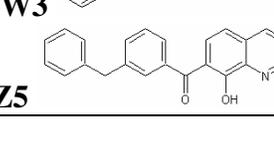
A matriz híbrida 85x21 foi analisada no Pirouette 4. A partir dela, foi obtido um modelo formado por dez LVs que acumularam 70,796% da informação, com qualidade estatística interna equivalente ao modelo QSAR-2D completo ($R^2 = 0,673$; $SEC = 0,550$; $Q^2_{LOO} = 0,468$; $SEV = 0,655$; $F_{(10,74)} = 15,267$). Esses resultados estatisticamente inferiores ao modelo QSAR-4D não eram inesperados, visto que todos os *outliers* de ambos os modelos IV e XI foram recolocados na matriz. Assim, na etapa seguinte, foi realizado o refinamento do modelo com a retirada de *outliers* e a eliminação de descritores que pudessem estar contribuindo pouco para a qualidade do modelo.

O melhor modelo (XIII) foi obtido após a retirada de sete *outliers* (**P16**, **P28**, **P34**, **P9**, **S4a**, **W3** e **Z5**, 8,23% do conjunto de treinamento), sendo que três (**P9**, **P34** e **W3**) foram classificados da mesma forma no estudo QSAR-4D e dois (**P34** e **S4a**) no estudo QSAR-2D completo (Tabela 19). Já compostos **P16**, **P28** e **Z5** são específicos do modelo XIII. O composto **P34** foi classificado novamente como *outlier*. Assim, o comportamento dos *outliers* apresentou alterações consideráveis em relação aos demais modelos, em especial ao modelo 4D. Novamente ocorreu predomínio de 4,5-diidroxipirimidinas carboxamidas. Mesmo assim, o número de *outliers* do Modelo XIII foi o menor entre todos os três modelos finais de cada etapa, e assim pode ser considerado o mais representativo da variabilidade estrutural e da faixa de atividade biológica dos inibidores em estudo.

$$\begin{aligned}
 \text{pIC}_{50} = & +3.115(\text{BELe8}) - 1.641(\text{G2e}) - 6.583(\text{R1u+}) \\
 & + 0.078[24.13.12.\text{Ar}(\text{C-H}).\text{LJ}] - 0.159[21.15.16.\text{Ar}(\text{C-H}).\text{LJ}] \\
 & + 0.198[26.18.14.\text{Ar}(\text{C-H}).\text{LJ}] + 0.239[19.21.24.\text{Ar}(\text{C-H}).\text{LJ}] \\
 & - 0.144[14.23.23.\text{Ar}(\text{C-H}).\text{LJ}] + 0.210[18.24.13.\text{Ar}(\text{C-H}).\text{LJ}] \\
 & - 0.189[20.17.19.\text{Zn}2+.\text{C}] + 0.052[22.16.9.\text{Ar}(\text{C-H}).\text{LJ}] \\
 & + 0.218[26.20.19.\text{Ar}(\text{C-H}).\text{LJ}] + 0.119[14.15.21.\text{Zn}2+.\text{LJ}] - 3.745
 \end{aligned}
 \tag{XIII}$$

$n = 78$; $R^2 = 0,844$; $SEC = 0,360$; $PRESS_{cal} = 9,200$; $F_{(6,71)} = 64,180$ ($cF = 2,229$);
 $Q^2_{LOO} = 0,764$; $SEV = 0,423$; $PRESS_{val} = 13,941$ ($SS_y = 59,101$).

Tabela 19. Outliers retirados do modelo XIII.

Outliers	pIC ₅₀
 <p>P9</p>	7,700
 <p>P16</p>	6,210
 <p>P28</p>	6,260
 <p>P34</p>	4,300
 <p>S4a</p>	7,070
 <p>W3</p>	6,660
 <p>Z5</p>	6,430

Quanto aos descritores, foram eliminados cinco, sendo um 2D (R_{Te}) e quatro 4D (18.18.19, 20.18.17, 20.13.16 e 15.15.18.Ar(C-H).LJ). Mesmo assim, os descritores mais importantes do modelo QSAR-4D foram mantidos. Os 13 descritores restantes originaram um modelo formado por seis VLs que acumularam 56,337% da informação, 12,912% a menos que o modelo XI. O Modelo XIII é capaz de explicar 84,40% da informação acumulada, 5,30% a menos que o modelo 4D. Seus *SEC* e *PRESS_{cal}* apresentaram qualidades um pouco inferiores, mas que também podem ser considerados bons resultados devido aos seus baixos valores. O resultado do teste *F* a 95% de confiança ($\alpha = 0,05$) também ficou abaixo do *F* do Modelo XI (8,647 unidades), porém muito acima de seu valor crítico.

O modelo XIII é capaz de prever 76,40% da variância, o que pode ser considerado apenas um pouco abaixo do modelo 4D (7,80%). Comportamento semelhante apresentou o seu *SEV*, sendo apenas 0,109 unidades maior que o *SEV* do Modelo XI. A diferença entre seus R^2 e Q^2_{LOO} também foi de apenas 0,08 unidades. Isto mostra que a probabilidade de ocorrência de sobreajuste de dados para esse modelo praticamente não existe. O resultado de seu *PRESS_{val}* ficou muito abaixo de seu valor SS_y de referência, sendo a diferença observada (45,160) a maior entre todos os modelos. Assim, pode-se considerar que, apesar dos resultados inferiores aos resultados do modelo 4D, o modelo híbrido também apresenta uma qualidade estatística satisfatória quanto às variâncias explicadas e preditas, e apresenta significância aceitável.

A Figura 70 apresenta os resultados do teste de randomização de *y* e da validação LNO para o Modelo XIII. É possível observar que o modelo XIII também não apresenta correlação ao acaso (intercepto de 0,123 para o teste com os R^2 e de -0,309 para o teste com os Q^2_{LOO}). A validação LNO também apresentou um comportamento homogêneo,

com baixos desvios padrões que apresentaram pequenas oscilações a partir de L100. O Q^2_{LNO} médio foi praticamente igual ao Q^2_{LOO} (0,760 e 0,764, respectivamente), e assim o Modelo XIII também pode ser considerado robusto.

O Modelo Real XIV (58,384% de informação acumulada) é muito parecido com o Modelo Auxiliar XIII. Os valores de seus respectivos R^2 apresentam diferença de apenas 0,006 unidades. Seu SEC é inferior em apenas superior 0,009 unidades. Seu teste F , por outro lado, é inferior em 10,079 unidades. Mesmo assim, é muito superior a seu valor crítico a 95% de confiança ($\alpha = 0,05$). Sua capacidade de predição interna também é equi-

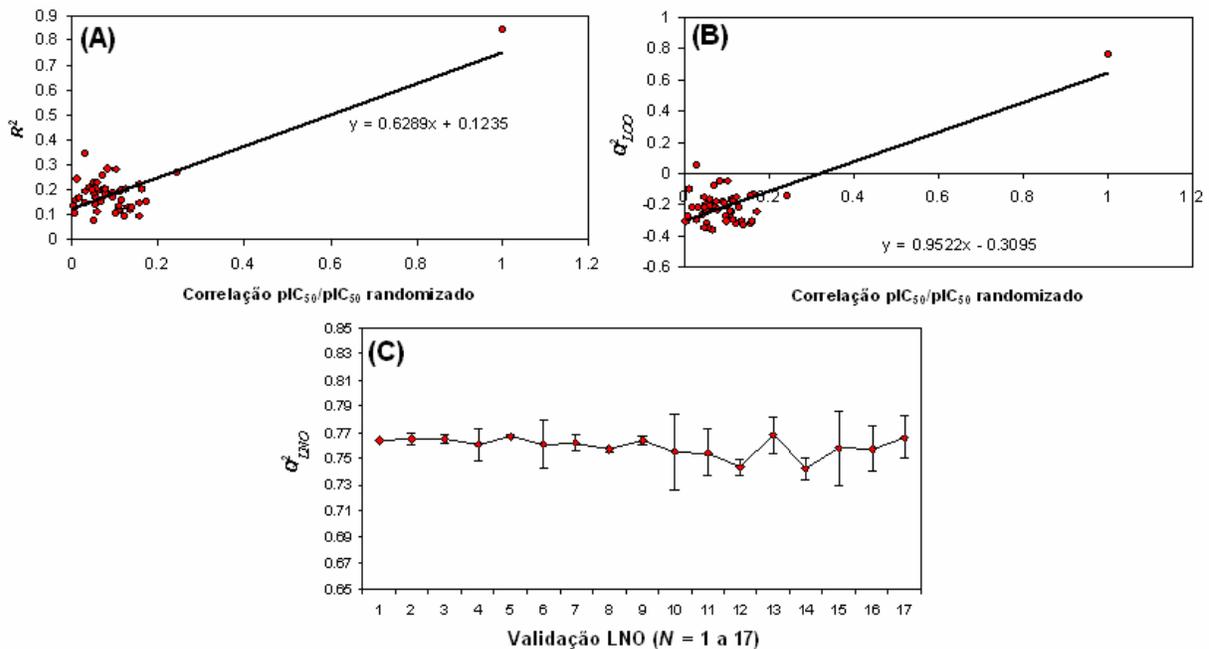


Figura 70. Resultados do teste de randomização de y (A e B) e da validação LNO (C) para o modelo XIII.

$$\begin{aligned}
& pIC_{50} = 3,016[BELe8] - 8,666[G2e] - 5,743[R1u+] \\
& + 0,084[24.13.12.Ar(C-H).LJ] - 0,169[21.15.16.Ar(C-H).LJ] \\
& + 0,205[26.18.14.Ar(C-H).LJ] + 0,196[19.21.24.Ar(C-H).LJ] \\
& + 0,013[20.13.16.Ar(C-H).LJ] - 0,161[14.23.23.Ar(C-H).LJ] \quad (XIV) \\
& + 0,173[18.24.13.Ar(C-H).LJ] - 0,185[20.17.19.Zn2+.C] \\
& + 0,039[22.16.9.Ar(C-H).LJ] + 0,231[26.20.19.Ar(C-H).LJ] \\
& + 0,096[14.15.21.Zn2+.LJ] - 2,321
\end{aligned}$$

$$n = 64; R^2 = 0,851; SEC = 0,350; PRESS_{cal} = 6,991; Q^2_{LOO} = 0,745; SEV = 0,433; PRESS_{val} = 11,985 (SSy = 46,990).$$

valente a de XIII (Q^2_{LOO} apenas 0,024 unidades menor e SEV apenas 0,014 unidades maior). Esses resultados mostram um modelo que sofreu pouca variação em relação ao seu auxiliar após a retirada do conjunto teste, podendo assim ser considerado o mesmo modelo. Além disso, ajuda a confirmar que o Modelo XIII é altamente robusto.

Na Figura 71 é apresentado o dendograma construído com o conjunto de treinamento completo, que mostra que o conjunto teste selecionado para o Modelo XIV também é tão bem distribuído quanto o modelo QSAR-4D. Os resultados da validação externa são apresentados na Tabela 20. Seus R^2_{pred} e SEP (0,779 e 0,445) também foram pouco inferiores aos obtidos para o modelo XI (0,839 e 0,384). Já o resultado do ARE_{pred} (6,102%) apresenta diferença de apenas 1,160 unidades. Esses resultados indicam que os valores preditos de novos compostos também tenderão a serem bem próximos dos reais, com uma probabilidade de erro pouco superior ao modelo XI. Finalmente, os resultados de k e k' localizaram-se dentro do intervalo proposto e o parâmetro $\left| R^2_0 - R'^2_0 \right|$ ficou abaixo de 0,3, confirmando a capacidade preditiva do modelo.

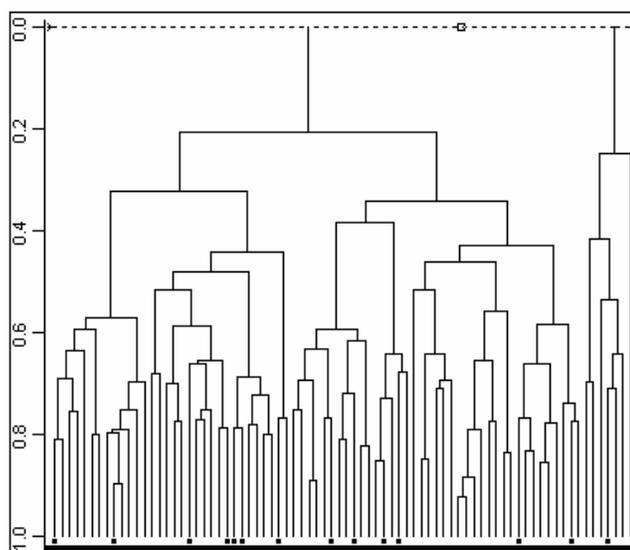


Figura 71. Dendrograma (dados autoescalados) do conjunto de treinamento completo para o Modelo XIV. O conjunto teste encontra-se destacado (pontos pretos).

Tabela 20. Resultados da etapa de validação externa realizada com o Modelo Real XIV a partir do conjunto teste selecionado do Modelo Auxiliar XIII.

Composto	$pIC_{50\text{ obs}}$	$pIC_{50\text{ prev}}$	Resíduos
E7	8,000	8,091	-0,091
E8b	7,350	7,783	-0,433
G12	7,400	7,198	0,202
L731927	6,300	6,595	-0,295
P10	6,000	6,310	-0,310
P13	6,210	7,002	-0,792
P19	4,770	5,796	-1,026
P32	6,890	7,136	-0,246
S12	6,700	7,023	-0,323
S9	7,300	7,142	0,158
W11	6,150	6,443	-0,293
W7	6,260	5,796	0,464
Z2	6,220	6,546	-0,326
Z7	8,000	8,285	-0,285
R^2_{pred}			0,779
SEP			0,445
ARE_{pred}			6,102%
k			0,964
k'			1,034
$ R_0^2 - R'_0 ^2$			0,0806

Na Tabela 21, são apresentadas as contribuições de cada descritor em cada uma das seis LVs, a contribuição total (em porcentagem) de cada LV para o modelo e os coeficientes autoescalados de cada descritor. A partir dos valores absolutos desses coeficientes, é possível observar que os descritores mais importantes apresentam seus valores no intervalo de 0,219-0,264. Dentre os descritores do modelo 2D, o mais importante é BELe8, o que fortalece a possibilidade da eletronegatividade atômica ser uma característica importante para a inibição da reação ST. Já os descritores G2e e R1u+ são menos importantes, além de serem desfavoráveis a atividade, uma indicação de que a forma das moléculas pode não ser um fator primordial para o aumento da potência quando é levada em consideração a flexibilidade das moléculas, situação não contemplada no estudo pelo modelo XI. Apesar disso, é possível observar que os sinais dos coeficientes são os mesmos do Modelo IV, e assim os significados dos descritores no modelo híbrido são os mesmos.

Quanto aos descritores 4D, o mais importante foi 19.21.24.Ar(C-H).LJ, enquanto o menos foi 20.17.19.Ar(C-H).LJ. O primeiro também é o descritor mais importante de todo o modelo híbrido, situação diferente do que ocorreu com o Modelo XI (Tabela 15), onde ele é apenas o décimo em importância. Já os descritores referentes à sonda Zn^{+2} apresentaram pouca alteração em seus coeficientes. É possível observar o descritor 21.15.16.Ar(C-H).LJ, relacionada a interação com o metal, teve seu coeficiente alterado de -0,323 para -0,169, enquanto o descritor 15.15.18.Ar(C-H).LJ foi eliminado na etapa de refinamento. Isto pode ser considerado a principal desvantagem do modelo XIII, pois foi perdida a informação provavelmente relacionada diretamente com a interação dos inibidores com os co-fatores metálicos expressa pela distância entre os descritores 21.15.16.Ar(C-H).LJ e 15.15.18.Ar(C-H).LJ.

Tabela 21. Contribuição dos descritores do Modelo XIII em cada LV, informação de cada LV e coeficientes autoescalados do modelo.

Descritor	LV1	LV2	LV3	LV4	LV5	LV6	<i>r</i>	Coefficientes autoescalados
BELe8	0,371	0,128	-0,001	0,117	-0,090	-0,342	0,513	0,264
G2e	-0,391	0,463	0,367	0,366	-0,062	0,254	-0,540	-0,015
R1u+	-0,440	0,311	-0,127	-0,193	-0,374	-0,553	-0,608	-0,236
24.13.12. Ar(C-H).LJ	0,177	-0,057	-0,008	0,195	-0,438	0,054	0,245	0,089
21.15.16. Ar(C-H).LJ	-0,212	-0,378	0,298	0,212	0,150	0,053	-0,293	-0,169
26.18.14. Ar(C-H).LJ	0,157	0,410	-0,128	0,225	0,227	0,050	0,217	0,223
19.21.24. Ar(C-H).LJ	0,243	0,241	0,501	-0,330	0,085	-0,221	0,336	0,271
20.13.16. Ar(C-H).LJ	-0,182	-0,158	-0,062	-0,102	0,147	0,268	-0,252	-0,165
14.23.23. Ar(C-H).LJ	0,228	0,254	0,382	-0,403	-0,147	0,420	0,316	0,241
18.24.13. Ar(C-H).LJ	-0,200	-0,053	-0,223	-0,630	0,070	0,127	-0,277	-0,219
20.17.19. Zn2+.C	0,226	-0,118	-0,150	0,028	-0,655	0,286	0,313	0,059
22.16.9. Ar(C-H).LJ	0,302	0,320	-0,451	0,034	0,293	0,119	0,418	0,223
26.20.19. Ar(C-H).LJ	0,299	-0,301	0,261	0,003	0,106	-0,309	0,413	0,136
Quantidade de informação	19,212%	9,229%	6,561%	8,113%	7,396%	5,827%		

Com os resultados apresentados, o modelo híbrido XIII pode ser considerado estatisticamente pouco inferior ao modelo QSAR-4D XI. Seu menor número de *ouliers* faz com que esse modelo seja mais representativo da faixa de variação da atividade biológica e da variabilidade estrutural entre os inibidores da reação ST que apresentam a subestrutura DCA. Assim, é possível propor que os descritores utilizados na construção do modelo 2D não lograram sucesso na obtenção de um modelo híbrido realmente superior, inclusive ocorrendo a perda de informações importantes relacionadas ao mecanismo de ação.

Todavia, quando o modelo híbrido é avaliado de modo individual, os resultados obtidos para todos os parâmetros estatísticos foram muito satisfatórios, estando muito acima dos limites mínimos aceitáveis e apresentando uma capacidade de predição interna muito próxima do Modelo XI. Esse modelo também mostrou ser autoconsistente (FERREIRA e KIRALJ, 2009), como pode ser visto na Tabela 21. Assim, também é possível propor que o Modelo XIII pode ser utilizado em conjunto com o modelo QSAR-4D em uma possível utilização em estudos de planejamento racional de inibidores da reação de transferência de fita, a fim de ajudar a fortalecer ou descartar possíveis previsões.

5. CONCLUSÕES

O trabalho realizado visou desenvolver modelos QSAR-2D, 4D e híbrido capazes de prever, com boa qualidade estatística, a potência de inibidores da reação de transferência de fita (ST: *strand transfer*) catalisada pela enzima viral HIV-1 integrase (HIV-IN) com base em 85 inibidores estruturalmente diversos, mas que possuem uma subestrutura em comum, o farmacóforo β -dicetoácido (DCA). Várias conclusões podem ser tiradas dos resultados obtidos:

(i) é possível, ao se valer apenas de descritores de simples interpretação, a obtenção de modelos relacionados diretamente com o mecanismo de ação mais aceito para o processo de inibição da HIV-IN: a ligação com um ou dois átomos de metal (Mg^{+2} ou Mn^{+2}) e a interação com o sítio hidrofóbico localizado na região da alça formada pelos aminoácidos Gly140, Ile141, Pro142, Tyr143, Asn144, Pro145, Gln146, Ser147, Gln148 e Gly149. Esta relação pode ser observada especialmente pela presença dos descritores E_{HOMO} e $SeaC2C2aa$, respectivamente, no modelo. Apesar do modelo possuir apenas razoável qualidade estatística interna, ele é robusto, apresenta boa capacidade de predição externa. Assim, o estudo preliminar permitiu a elaboração de um protocolo adequado para o estudo QSAR-2D com o conjunto de treinamento completo.

(ii) foi possível obter um modelo QSAR-2D utilizando o conjunto de treinamento selecionado, apesar do razoável grau de variabilidade estrutural que este apresentava. Porém, exatamente tal variabilidade pode ter sido a responsável pela seleção de apenas descritores de interpretação mais complexa que, segundo a literatura, são muito úteis exatamente em situações como esta. O modelo apresenta boa qualidade estatística interna, é robusto e possui boa capacidade de predição externa. Um fator importante quanto à interpretação perante o mecanismo de ação é que três dos descritores selecionados são ponderados pela eletronegatividade. Outros trabalhos disponíveis na literatura propõem que a presença deste tipo de descritor em um modelo relacionando a inibição da HIV-IN pode estar relacionada a interações com os co-fatores metálicos. Estes resultados indicam que o modelo apresenta um bom potencial para utilização no planejamento de novos inibidores da reação ST.

(iii) apesar dos descritores utilizados para a obtenção do modelo QSAR-4D para o conjunto de treinamento completo serem de origem completamente diferente daqueles utilizados no estudo QSAR-2D completo, o primeiro mostrou ser superior em todos os aspectos analisados: qualidade interna, externa, robustez e interpretabilidade perante o mecanismo de ação. Este resultado pode indicar que metodologias que utilizam descritores relacionados às energias de interações ligante-receptor, obtidas a partir das estruturas tridimensionais dos inibidores, podem ser de uso preferencial quando o conjunto de treinamento apresenta grau relevante de variabilidade estrutural.

(iv) o modelo híbrido QSAR-2D/4D obtido pode ser considerado estatisticamente inferior ao modelo QSAR-4D. Este resultado contrasta com os obtidos por outros autores, em que a utilização de descritores clássicos (como LogP) junto de descritores específicos de estudos QSAR-3D ou 4D não necessariamente levou à obtenção de melhores modelos. Apesar disto, o modelo é robusto e apresenta ótima qualidade de previsão interna e externa. Desta forma, também pode ser considerado uma opção quando da síntese de novos inibidores, separadamente ou em conjunto com o modelo QSAR-4D, o que permitiria comparações entre os resultados que aumentariam as chances de sucesso na predição da potência de um novo composto.

Assim, pode-se concluir que o objetivo principal deste trabalho foi alcançado: a obtenção de modelos QSAR para potencial utilização no planejamento de novos inibidores da HIV-IN relacionados direta ou indiretamente com o mecanismo de ação e que apresentassem qualidade estatística que atendessem aos requisitos recomendados pela literatura. O modelo QSAR-4D pode ser considerado o melhor modelo dentre os obtidos, pois, além de ser estatisticamente superior aos demais, sua interpretação mecanística apresentou uma relação clara e direta com o mecanismo de ação dos compostos em estudo. Sua utilização para o auxílio no direcionamento da síntese de novos derivados provavelmente permitiria a obtenção de descritores de energias de interações intermoleculares que descreveriam melhor as interações entre diferentes grupamentos hidrofóbicos que poderiam ser utilizados para originar a cadeia lateral aromática, podendo assim gerar inibidores com maior afinidade pela alça flexível. A região do grupamento ceto-enol apresenta menor potencial para modificações estruturais, já que a coordenação

com os co-fatores metálicos, e consequentemente a atividade, depende da presença dos grupamentos com pares de elétrons livres. Mesmo assim, as energias das interações formadas entre o grupamento ceto-enol e o sítio de ligação podem sofrer influências de modificações em outras partes dos derivados, o que pode permitir alterações estruturais que favoreçam as mesmas.

A grande limitação desse trabalho provavelmente deveu-se a heterogeneidade do conjunto de treinamento. A continuação desse projeto utilizando-se de um conjunto que apresente menor variabilidade estrutural, mas que cubra a maior faixa de atividade biológica possível, poder render modelos QSAR-2D (em especial) e 4D com melhores resultados estatísticos e interpretativos. Um bom ponto de partida a se considerar são as 42 4,5-diidroxipirimidinas carboxamidas utilizadas neste trabalho, exatamente devido a sua menor variabilidade estrutural combinada a sua ampla faixa de atividade biológica. Além disso, esses compostos são relacionados com o único inibidor da reação ST disponível para uso clínico, o raltegravir, o que aumenta as chances de sucesso na obtenção de derivados clinicamente úteis.

6. REFERÊNCIAS

Accelrys Inc. ViewerLite 4.2 for Windows. 2001.

Adesokan, A. A.; Roberts, V. A.; Lee, K. W.; Lins, R. D.; Briggs, J. M. **Prediction of HIV-1 integrase/viral DNA interactions in the catalytic domain by fast molecular docking.** *Journal of Medicinal Chemistry* **2004**, *47*, 821-828.

Agatonovic-Kustrin, S.; Turner, J.V.; Glass, B.D. **Molecular structural characteristics as determinants of estrogen receptor selectivity.** *Journal of Pharmaceutical and Biomedical Analysis* **2008**, *48*, 369–375.

Aids tem contaminação recorde: 40,3 milhões em 2005. In *Revista Riopharma*. 2005, *69*, 6-6.

Alencar, T.M.D.; Nemes, M.I.B.; Velloso, M.A. **Transformações da "Aids aguda" para a "Aids crônica": percepção corporal e intervenções cirúrgicas entre pessoas vivendo com HIV e Aids.** *Ciência e Saúde Coletiva* **2008**, *6*, 1841-1849.

Aptula, A.O.; Jeliaskova, N.G.; Schultz, T.W.; Cronin, M.T.D. **The better predictive model: high q^2 for the training set or low root mean square error of prediction for the test set?** *QSAR & Combinatorial Science* **2005**, *24*, 385-396.

Araújo, V.L.B.; Brito, D.M.S.; Gimenez, M.T.; Queiroz, T.A.; Tavares, C.M. **Características da Aids na terceira idade em um hospital de referência do Estado do Ceará, Brasil.** *Revista Brasileira de Epidemiologia* **2007**, *10*, 544-554.

Atkins, P.; Paula, J. **Teoria quântica: introdução e princípios.** In: *Atkins: Físico-Química*. V. 2, 7ª ed. Rio de Janeiro: LTC, **2004a**. p. 3-35.

Atkins, P.; Paula, J. **Estrutura molecular.** In: *Atkins: Físico-Química*. V. 2, 7ª ed. Rio de Janeiro: LTC, **2004b**. p. 122-163.

Avert – Averting HIV and AIDS. Origins of Aids and HIV. Disponível em <http://www.avert.org/origins.htm>. Acesso: 02 de abril de 2009a.

Avert – Averting HIV and AIDS. The Structure of HIV. Disponível em <http://www.avert.org/virus.htm>. Acesso: 02 de abril de 2009b.

Ayres, P.W.; Yang, W. **Density-functional theory.** In: Bultinck, P.; De Winter, H.; Langennaeker, W.; Tollenaere, J.P. (Eds.), *Computational Medicinal Chemistry for Drug Discovery*. Marcel Dekker, USA, **2004**, cap 4.

Barreca, M.L.; De Luca, L.; Ferro, S.; Rao, A.; Monforte, A.; Chimirri, A. **Computational and synthetic approaches for the discovery of HIV-1 integrase inhibitors.** *Arkivoc* **2006**, *7*, 224-244.

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.** *Química Nova* **1997**, *20*, 300-310.

Barreiro, E.J. **Estratégia de simplificação molecular no planejamento racional de fármacos: a descoberta de novo agente cardioativo.** *Química Nova* **2002**, *25*, 1172-1180.

Barreiro, E.J.; Fraga, C.A.M. *Química medicinal: as bases moleculares da ação dos fármacos.* Porto Alegre: Artmed, **2001**. 243 pp.

Barreiro, E.J. **A Química Medicinal e o paradigma do composto-protótipo.** *Revista Virtual de Química* **2009**, *1*, 26-34.

Barreiro, E.J.; Fraga, C.A.M.; Rodrigues, C.R.; Miranda, A.L.P. **Estratégias em química medicinal para o planejamento de fármacos.** *Revista Brasileira de Ciências Farmacêuticas*. **2001**, *37*, 269-290.

Barton, S. **Emptying reservoirs of latent HIV-1.** *Nature Reviews Drug Discovery* **2005**, *4*, 886-886.

[Basak, S.C.](#); [Balasubramanian, K.](#); [Gute, B.D.](#); [Mills, D.](#); [Gorzynska, A.](#); [Roszak, S.](#) **Prediction of cellular toxicity of halocarbons from computed chemodcriptors: A hierarchical QSAR approach.** *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 1103-1109.

Becke, A.D. **Density-functional thermochemistry. III. The role of exact exchange.** *Journal of Chemical Physics* **1993**, *98*, 5648-5652.

Bello-Ramírez, A.M.; Carreón-Garabito, B.Y.; Nava-Ocampo, A.A. **Do structural properties explain the anticonvulsant activity of valproate metabolites? A QSAR Analysis.** *Epilepsia* **2002**, *43*, 475-481.

Berkhout, B.; de Ronde, A.; van der Hoek, L. **Aggressive HIV-1?** *Retrovirology* **2005**, *2*. Disponível em: <http://www.retrovirology.com/content/2/1/13>.

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 Research* **2000**, *28*, 235-242.

Billings, E. **Molecular modeling and drug design.** In: Willians, D.A.; Lemke, T.L. (orgs.) *Foye's Principles of Medicinal Chemistry*. 5^a ed. Philadelphia: Lippincott Williams & Wilkins, **2002**. p. 68-85.

Billich, A. **S-1360 Shionogi-GlaxoSmithKline.** *Current Opinion in Investigative Drugs* **2003**, *4*, 206-209.

Brasil-MS-SVS-Programa Nacional de DST e Aids *Boletim epidemiológico AIDS/DST*. 2005, 1. Disponível em: <http://www.aids.gov.br/data/Pages/LUMIS624DE984PTBRIE.htm>.

Brasil-MS-SVS-Secretaria de Ciência, Tecnologia e Insumos Estratégicos. *Plano brasileiro de vacinas anti-HIV 2008-2012: pesquisa, desenvolvimento e inovação*. Brasília: Ministério da Saúde, **2008**. 72 p.

Bredow, T. **Semiempirical methods**. In: Bultinck, P.; De Winter, H.; Langenaeker, W.; Tollenaere, J.P. *Computational medicinal chemistry for drug design*. New York: Marcel Dekker Inc, **2004**. cap. 2.

Brito, A.M.; Castilho, E.A.; Szwarcwald, C.L. **AIDS e infecção pelo HIV no Brasil: uma epidemia multifacetada**. *Revista da Sociedade Brasileira de Medicina Tropical* **2000**, *34*, 207-217.

Buolamwini, J. K.; Assefa, H. **CoMFA and CoMSIA 3D-QSAR and docking studies on conformationally-restrained cinnamoyl HIV-1 integrase inhibitors: exploration of a binding mode at the active site**. *Journal of Medicinal Chemistry* **2002**, *45*, 841-852.

Burden, F.R. **Molecular identification number for substructure searches**. *Journal of Chemical Information and Computer Science* **1989**, *29*, 225-227.

Cambridge Crystallographic Data Centre Inc.; The Cambridge Structural Database. 2009.

Cambridge Structural Database software version 5.29-2007 (November 2007) + 1 update. By Cambridge Crystallographic Data Centre, England.

Carbó-Dorca, R.; Gironés, X. **Quantum similarity and quantitative structure-activity relationships**. In: Bultinck, P.; De Winter, H.; Langenaeker, W.; Tollenaere, J.P. *Computational medicinal chemistry for drug design*. New York: Marcel Dekker Inc, **2004**. cap. 14.

Carvalho, I.; Pupo, M.T.; Borges, A.D.L.; Bernardes, L.S.C. **Introdução a modelagem molecular de fármacos no curso experimental de Química Farmacêutica**. *Química Nova* **2003**, *26*, 428-438.

Castro, M.A.; Canuto, S. **O método Hartree-Fock**. In: Morgon, N.H.; Coutinho, K. (orgs.) *Métodos de química teórica e modelagem molecular*. São Paulo: Livraria da Física, **2007**. p. 1-28.

Challand, R.; Young, R. J. *Antiviral Chemotherapy*. Spectrum Press: Oxford. **1997**. 128 pp.

Chemaxon. *Marvin User's Guide*, **2008**. Disponível em <http://www.chemaxon.com/marvin/help/calculations/chargegroup.html>. Acessado em 03 de setembro de 2008.

Chen, X.; Tsiang, M.; Yu, F.; Hung, M.; Jones, G.S.; Zeynalzadegan, A.; Qi, X.; Jin, H.; Kim, C.U.; Swaminathan, S.; Chen, J.M. **Modeling, analysis, and validation of a novel HIV integrase structure provide insight into the binding modes of potent integrase inhibitors**. *Journal of Molecular Biology* **2008**, *380*, 504-519.

Chung, M.C.; Silva, A.T.A.; Castro, L.F.; Guido, V.C.; Nassute, J.C.; Ferreira, E.I. **Latenciação e formas avançadas de transporte de fármacos.** *Revista Brasileira de Ciências Farmacêuticas* **2005**, *41*, 2005.

Consonni, V.; Todeschini, R.; Pavan, M. **Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors.** *Journal of Chemical Information and Computer Science* **2002**, *42*, 682-692.

Coovadia, H.M.; Hadingham, J. **HIV/AIDS: global trends, global funds and delivery bottlenecks.** *Globalization and Health* **2005**, *1*. Disponível em: <http://www.globalizationandhealth.com/content/1/1/13>.

Costa, P.; Pilli, R.; Pinheiro, S.; Vasconcelos, M. *Substâncias carboniladas e derivados*. Porto Alegre: Bookman, **2003**. 411 pp.

Craigie, R. **HIV integrase, a brief overview from chemistry to therapeutics.** *Journal of Biological Chemistry* **2001**, *276*, 23213-23216.

Cramer, R.D.; Patterson, D.E.; Bunce, J.D. **Comparative molecular-field analysis (CoMFA) .1. Effect of shape on binding of steroids to carrier proteins.** *Journal of the American Chemical Society* **1988**, *110*, 5959-5967.

Cronin, M.T.D.; Schultz, T.W. **Pitfalls in QSAR.** *Journal of Molecular Structure (Theochem)* **2003**, *622*, 39-51.

Cotelle, P. **Patented HIV-1 Integrase Inhibitors (1998-2005).** *Current Medicinal Chemistry* **2006**, *1*, 1-15.

Cunico, W.; Gomes, C.R.B.; Vellasco Jr, W.T. **HIV - recentes avanços na pesquisa de fármacos.** *Química Nova* **2008**, *31*, 2111-2117.

Davidson, E.; Feller, D. **Basis set selection for molecular calculations.** *Chemical Reviews* **1986**, *86*, 681-696.

Dayam, R.; Neamati, N. **Small-molecule HIV-1 integrase inhibitors: the 2001-2002 update.** *Current Pharmaceutical Design* **2003**, *9*, 1789-1802.

Dayam, R.; Sanchez, T.; Clement, O.; Shoemaker, R.; Sei, S.; Neamati, N. **β -Diketo Acid Pharmacophore Hypothesis. 1. Discovery of a Novel Class of HIV-1 Integrase Inhibitors.** *Journal of Medicinal Chemistry*. **2005**, *48*, 111-120.

Dayam, R.; Al-Mawsawi, L.Q.; Neamati, N. **HIV-1 integrase inhibitors: an emerging clinical reality.** *Drugs in R & D* **2007**, *8*, 155-168.

Dayam, R.; Al-Mawsawi, L.Q.; Zawahir, Z.; Witvrouw, M.; Debyser, Z.; Neamati, N. **Quinolone 3-Carboxylic Acid Pharmacophore: Design of Second Generation HIV-1 Integrase Inhibitors.** *Journal of Medicinal Chemistry*. **2008**, *51*, 1136-1144.

Daeyaert, E.F.D.; Vinkers, H.M.; Jonge, M.R.; Heeres, J.; Koymans, L.M.H.; Lewi, P.J.; Janssen, P.A.J. **Ligand-based Computation of HIV-1 Integrase Inhibition Strength within a Series of β -ketoamide Derivatives.** *Internet Electronic Journal of Molecular Design* **2004**, *3*, 528-543.

Debnath, A.K. **Application of 3D-QSAR techniques in anti-HIV drug design – an overview.** *Current Pharmaceutical Design* **2005**, *11*, 3091-3110.

Dhaked, D.K.; Verma, J.; Saran, A.; Coutinho, E.C. **Exploring the binding of HIV-1 integrase inhibitors by comparative residue interaction analysis (CoRIA).** *Journal of Molecular Modeling* **2009**, *15*, 233-245.

Doweyko, A.M. **QSAR: dead or alive?** *Journal of Computer-Aided Molecular Design* **2008**, *22*, 81-89.

van Drie, J.H. **Monty Kier and the Origin of the Pharmacophore Concept.** *Internet Electronic Journal of Molecular Design* **2007**, *6*, 271-279.

Duarte, H.A.; Rocha, W.R. **Teoria do funcional de densidade.** In: Morgon, N.H.; Coutinho, K. (orgs.) *Métodos de química teórica e modelagem molecular*. São Paulo: Livraria da Física, **2007**. p. 73-112.

Dunn III, W.J. **Quantitative structure-activity relationships (QSAR).** *Chemometrics and Intelligent Laboratory Systems* **1989**, *6*, 181-190.

Eich, E.; Pertz, H.; Kaloga, M.; Schulz, J.; Fesen, M.R.; Mazumder, A.; Pommier, Y. **(-)-Arctigenin as a lead structure for inhibitors of human immunodeficiency virus type-1 integrase.** *Journal of Medicinal Chemistry* **1996**, *39*, 86-95.

Egbertson, M.S.; Moritz, M.; Melamed, J.Y.; Han, W.; Perlow, D.S.; Kuo, M.S.; Embrey, M.; Vacca, J.P.; Zrada, M.M.; Cortes, A.R.; Wallace, A.; Leonard, Y.; Hazuda, D.J.; Miller, M.D.; Felock, P.J.; Stillmock, K.A.; Witmer, M.V.; Schleif, W.; Gabryelski, L.J.; Moyer, G.; Ellis, J.D.; Jin, L.; Xu, W.; Braun, M.P.; Kassahun, K.; Tsou, N.N.; Young, S.D. **A potent and orally active HIV-1 integrase term inhibitor.** *Bioorganic & Medicinal Chemistry Letters* **2007**, *17*, 1392-1398.

Eriksson, L.; Jaworska, J.; Worth, A.P.; Cronin, M.T.D.; McDowell, R.M.; Gramática, P. **Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs.** *Environmental Health Perspectives* **2003**, *111*, 1361-1374.

Farhi, L.; Lima, D.B.; Cunha, C.B. Dislipidemia em pacientes HIV/AIDS em uso de anti-retrovirais num hospital universitário, Rio de Janeiro, Brasil. *Jornal Brasileiro de Patologia e Medicina Laboratorial* **2008**, *44*, 175-184.

- Ferreira, M.M.C.; Kiralj, R. **Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Applications.** *Journal of Brazilian Chemical Society* **2009**, *20*, 770-787.
- Ferreira, M.M.C.; Kiralj, R. **Métodos Quimiométricos em Relações Quantitativas Estrutura-Atividade (QSAR).** In: Carlos A. Montanari. (Org.). *Química medicinal: métodos e fundamentos em planejamento de fármacos.* 1ª ed. Viçosa: Editora UFV, **2008.** p. 532-633.
- Ferreira, M.M.C.; Montanari, C.A.; Gaudio, A.C. **Seleção de variáveis em QSAR.** *Química Nova* **2002**, *25*, 439-448.
- Flexner, C.; Bate, G.; Kirkpatrick, P. **Tipranavir.** *Nature Reviews: Drug Discovery.* **2005**, *4*, 955-956.
- Foresman, J.B.; Frisch, A. *Exploring chemistry with electronic structure methods.* 2ª ed. Pittsburg: Gaussian Inc, **1996.** 302 pp.
- Freitas, M.P. **Multivariate image analysis applied to QSAR: evaluation to a series of potential anxiolytic agents.** *Chemometrics and Intelligent Laboratory Systems* **2008**, *91*, 173-176.
- Furusjö, E.; Svenson, A.; Rahmberg, M.; Andersson, M. **The importance of outlier detection and training set selection for reliable environmental QSAR predictions.** *Chemosphere* **2006**, *63*, 99-108.
- Gallo, R. **The Early Years of HIV/AIDS.** *Science* **2002**, *298*, 1728-1730.
- Gallo, R.; Montagneir, L. **Enhanced: Prospects for the Future.** *Science* **2002**, *298*, 1730-1731.
- Garcia, G.C.; Ruiz, I.L.; Gómez-Nieto, M.A.; Domel, J.A.C.; Plaza, A.G. **From Wiener Index to molecules.** *Journal of Chemical Information and Modelling* **2005**, *45*, 231-238.
- Garvey, E.P.; Johns, B.A.; Gartland, M.J.; Foster, S.A.; Miller, W.H.; Ferris, R.G.; Hazen, R.J.; Underwood, M.R.; Boros, E.E.; Thompson, J.B.; Weatherhead, J.G.; Koble, C.S.; Allen, S.H.; Schaller, L.T.; Sherrill, R.G.; Yoshinaga, T.; Kobayashi, M.; Wakasa-Morimoto, C.; Miki, S.; Nakahara, K.; Noshi, T.; Sato, A.; Fujiwara, T. **The Naphthyridinone GSK364735 Is a Novel, Potent Human Immunodeficiency Virus Type 1 Integrase Inhibitor and Antiretroviral.** *Antimicrobial Agents and Chemotherapy* **2008**, *52*, 901-908.
- Gasteiger, G.; Marsili, M.; **Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges.** *Tetrahedron* **1980**, *36*, 3219-3288.
- Gaudio, A.C.; Zandonade, E. **Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica.** *Química Nova* **2001**, *24*, 658-671.

- Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.; Lee, K.; Tropsha, A. **Rational selection of training and test sets for the development of validated QSAR models.** *Journal of Computer-Aided Molecular Design* **2003**, *17*, 241-253.
- Golbraikh, A.; Tropsha, A. **Beware of q^2 !** *Journal of Molecular Graphics and Modelling* **2002**, *20*, 269-276.
- Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. **Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design.** *Proceedings of the National Academy of Science of USA* **1999**, *96*, 13040-13043.
- Gonçalves, M.M.; Fraceto, L.F.; Vila, M.M.D.C.; Oliveira, R.V.M. **Polarizabilidade atômica efetiva α_d pode ditar a ionização de radioligantes $^{99}\text{T}^{\text{cm}}$ - diaminoditiol alquilamínicos?** *Química Nova* **2006**, *29*, 1072-1077.
- González, M.P.; C. Terán, C.; Teijeira, M.; González-Moa, M.J. **GETAWAY descriptors to predicting A2A adenosine receptors agonists.** *European Journal of Medicinal Chemistry* **2005**, *40*, 1080-1086.
- Gramatica, P. **WHIM descriptors of shape.** *QSAR & Combinatorial Chemistry* **2006**, *25*, 327-332.
- Gramatica, P. **Principles of QSAR models validation: internal and external.** *QSAR & Combinatorial Chemistry* **2007**, *26*, 694-701.
- Gregoy, M.C.; Tolman, K.G. **Doenças: manifestações e fisiopatologia.** In: Gennaro, A.R (org.). *Remington: a ciência e a prática da Farmácia*. 20^a ed. Rio de Janeiro: Guanabara Koogan, **2004**. p. 1092-1140.
- Guare, J.P.; Wai, J.S.; Gomez, R.P.; Anthony, N.J.; Jolly, S.M.; Cortes, A.R.; Vacca, J.P.; Felock, P.J.; Stillmock, K.A.; Schleif, W.A.; Moyer, G.; Gabryelski, L.J.; Jin, L. Chen, T.; Hazuda, D.J.; Young, S.D. **A series of 5-aminosubstituted 4-fluorobenzyl-8-hydroxy-[1,6]naphthyridine-7-carboxamide HIV-1 integrase inhibitors.** *Bioorganic & Medicinal Chemistry Letters* **2006**, *16*, 2900-2904.
- Gupta, S. P.; Nagappa, A. N. **Design and development of integrase inhibitors as anti-HIV agents.** *Current Medicinal Chemistry* **2003**, *10*, 1779-1794.
- Hashemianzadeh, M.; Safarpour, M.A.; Gholamjani-Moghaddam, K.; Mehdipour, A.R. **DFT-based QSAR study of valproic acid and its derivatives.** *QSAR & Combinatorial Chemistry* **2008**, *27*, 469-474.
- Hazuda, D.J.; Felock, P.J.; Hastings, J.C.; Pramanik, B.; Wolfe, A.L. **Differential divalent cation requirements uncouple the assembly and catalytic reactions of human immunodeficiency virus type 1 integrase.** *Journal of Virology* **1997**, *71*, 7005-7011.

Hazuda, D.J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J.A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M.D. **Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells.** *Science* **2000**, *287*, 646-650.

Hazuda, D.J.; Young, S.D.; Guare, J.P.; Anthony, N.J.; Gomez, R.P.; Wai, J.S.; Vacca, J.P.; Handt, L.; Motzel, S.L.; Klein, H.J.; Dornadula, G.; Danovich, R.M.; Witmer, M.V.; Wilson, K.A.; Tussey, L.; Schleif, W.A.; Gabryelski, L.S.; Jin, L.; Miller, M.D.; Casimiro, D.R.; Emini, E.A.; Shiver, J.W. **Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques.** *Science* **2004**, *305*, 528-532

Healy, E.F.; Sanders, J.; King, P.J.; Robinson Jr, W.E. **A docking study of L-chicoric acid with HIV-1 integrase.** *Journal of Molecular Graphics and Modelling* **2009**, *27*, 584-589.

Helguera, A.M.; Cordeiro, N.D.S.; Pérez, M.A.C.; Combes, R.D.; González, M.P. **QSAR modeling of the rodent carcinogenicity of nitrocompounds.** *Bioorganic & Medicinal Chemistry* **2008**, *16*, 3395-3407.

M. C. Hemmer, V. Steinhauer, J. Gasteiger. **The Prediction of the 3D Structure of Organic Molecules from Their Infrared Spectra.** *Vibrational Spectroscopy* **1999**, *19*, 151-164.

Heralth, K.B.; Jayasuriya, H.; Bills, G.F.; Polishook, J.D.; Dombrowski, A.W.; Guan, Z.; Felock, P.J.; Hazuda, D.J.; Singh, S.B. **Isolation, structure, absolute stereochemistry, and HIV-1 integrase inhibitory activity of integrasone, a novel fungal polyketide.** *Journal of Natural Products* **2004**, *67*, 872-874.

Herr, R.J. **5-Substituted-1H-tetrazoles as carboxylic acid isosteres: medicinal chemistry and synthetic methods.** *Bioorganic and Medicinal Chemistry* **2002**, *10*, 3379-3393.

Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M. C.; Mazumder, A.; Zhao, H.; Burke Jr., T.R.; Pommier, Y.; Milne, G.W.A. **Discovery of HIV-1 Integrase Inhibitors by Pharmacophore Searching.** *Journal of Medicinal Chemistry* **1997**, *40*, 930-936.

Hopfinger, A.J.; Wang, S.; Tokarski, J.S.; Jin, B.; Albuquerque, M.; Madhav, P.J.; Duraiswami, C. **Construction of 3D-QSAR models using the 4D-QSAR analysis formalism.** *Journal of American Chemistry Society* **1997**, *119*, 10509-10524.

Hypercube Inc. HyperChem 7.1 for Windows. 2002.

Hypercube Inc. *HyperChem Computational Chemistry*. Gainesville: Hypercube Inc, 2002. 860 pp.

Informetrix Inc. Pirouette 4 for Windows. 2007.

Iyer, M.; Hopfinger, A.J. **Treating chemical diversity in QSAR analysis: modeling diverse HIV-1 integrase inhibitors using 4D fingerprints.** *Journal of Chemical Information and Modeling* **2007**, *47*, 1945-1960.

Jayasuriya, H.; Guan, Z.; Polishook, J.D.; Dombrowski, A.W.; Felock, P.J.; Hazuda, D.J.; Singh, S.B. **Isolation, structure, and HIV-1 integrase inhibitory activity of cytosporic acid, a fungal metabolite produced by a *Cytospora* sp.** *Journal of Natural Products* **2003**, *66*, 551-553.

Jornada, D.H.; Pizzolato, M. **Sistemática para Avaliação da Robustez de Métodos de Ensaio Através de Projetos de Experimentos.** In: *ENQUALAB-2007 – Congresso da Qualidade em Metrologia Rede Metrológica do Estado de São Paulo – REMESP*. 11 a 14 de junho de 2007, São Paulo, Brasil. Disponível em <http://www.portalcertificar.com.br>.

Jurs, P.C.; Dixon, S.L.; Ego, L.M. **Molecular Concepts.** Em: Watebeend, H. (Ed.) *Chemometrics Methods in Molecular Design*. Weinheim: VCH, **1994**, 15-38

Karelson, M.; V.S. Lobanov, A.R. Katritzky. **Quantum-Chemical Descriptors in QSAR/QSPR Studies.** *Chemical Reviews* **1996**, *96*, 1027-1043.

Katritzky, A.R.; Gordeeva, E.V. **Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research.** *Journal of Chemical Information and Computer Sciences* **1993**, *33*, 835-857.

Kawasuji, T.; Fuji, M.; Yoshinaga, T.; Sato, A.; Fujiwara, T.; Kiyama, R. **A platform for designing HIV integrase inhibitors. Part 2: A two-metal binding model as a potential mechanism of HIV integrase inhibitors.** *Bioorganic and Medicinal Chemistry* **2006**, *14*, 8420-8429.

Keserû, G.M.; Kolossvary, I. **Fully flexible low-mode docking: application to the induced fit in HIV-integrase.** *Journal of American Chemical Society* **2001**, *123*, 12708-12709.

Kier, L.B.; Hall, L.H. In: Devillers, J.; Balaban, A.T.(Eds.) *Topological Indices and Related Descriptors in QSAR and QSPR*. London: Gordon and Breach, **1999**, pp. 491-562.

Kim, K.H. **Outliers in SAR and QSAR: is unusual binding mode a possible source of outliers?** *Journal of Computer-Aided Molecular Design* **2007a**, *21*, 63-86.

Kim, K.H. **Outliers in SAR and QSAR: 2. Is a flexible binding site a possible source of outliers?** *Journal of Computer-Aided Molecular Design* **2007b**, *21*, 421-435.

Kim, H.J.; Woo, E.; Shin, C.; Park, H. **A new flavonol glycoside gallate ester from acer okamotoanum and its inhibitory activity against human immunodeficiency virus-1 (HIV-1) integrase.** *Journal of Natural Products* **1998**, *61*, 145-148.

Knekta, E.; Andersson, P.L.; Johansson, M.; Tysklind, M. **An overview of OSPAR priority compounds and selection of a representative training set.** *Chemosphere* **2004**, *57*, 1495-1503.

Kobayashi, M.; Nakahara, K.; Seki, T.; Miki, S.; Kawauchi, S.; Suyama, A.; Wakasa-Morimoto, C.; Kodama, M.; Endoh, T.; Oosugi, E.; Matsushita, Y.; Murai, H.; Fujishita, T.; Yoshinaga, T.; Garvey, E.; Foster, S.; Underwood, M.; Johns, B.; Sato, A.; Fujiwara, T.

Selection of diverse and clinically relevant integrase inhibitor-resistant human immunodeficiency virus type 1 mutants. *Antiviral Research* **2008**, *80*, 213-222.

Kubinyi, H. *QSAR: Hanch Analysis and Related Approaches*. Weinheim: VCH, **1993**. 240 pp.

Kubinyi, H. **Variable Selection in QSAR Studies. II. A Highly Efficient Combination of Systematic Search and Evolution.** *Quantitative structure-activity relationships* **1994**, *13*, 393-401.

Kubinyi, H. **QSAR and 3D QSAR in drug design. Part 1: methodology.** *Drug Discovery Today* **1997**, *2*, 457-467.

Kuo, C.L.; Assefa, H.; Kamath, S.; Brzozowski, Z.; Slawinski, J.; Saczewski, F.; Buolamwini, J.K.; Neamati, N. **Application of CoMFA and CoMSIA 3D-QSAR and docking studies in optimization of mercaptobenzenesulfonamides as HIV-1 integrase inhibitors.** *Journal of Medicinal Chemistry* **2004**, *47*, 385-399.

Kuritzkes, D.; Kar, S.; Kirkpatrick, P.; **Maraviroc.** *Nature Reviews Drug Discovery* **2008**, *7*, 15-16.

LaBonte, J.; Lebbos, J.; Kirkpatrick, P. **Enfuvirtide.** *Nature Reviews Drug Discovery* **2003**, *2*, 345-346.

Lehrman, G.; Hogue, I.B.; Palmer, S.; Jennings, C.; Spina, C.A.; Wiegand, A.; Landay, A.L.; Coombs, R.W.; Richman, D.D.; Mellors, J.W.; Coffin, J.M.; Bosch, R.J.; Margolis, D.M. **Depletion of latent HIV-1 infection in vivo: a proof-of-concept study.** *Lancet* **2005**, *366*, 549-555.

Levine, I.N. *Quantum chemistry*. 4^a ed. Englewood Cliffs: Prentice Hall, **1991**. 629 pp.

Lewis, D.S.; Castelo Filho, A.; Accentturi, C.A.; Granato, C.F.H. **Síndrome da Imunodeficiência Adquirida (AIDS).** Em *Atualização Terapêutica: Manual Prático de Diagnóstico e Tratamento*; Ramos, O.L.; Rothschild, H.A., ed.; Artes Médicas: São Paulo, **1997**, 18 ed., Seção I, p.55-63.

Lignani Jr, L.; Greco, D.B.; **Carneiro, M. Avaliação da aderência aos anti-retrovirais em pacientes com infecção ple HIV/Aids.** *Revista de Saúde Pública* **2005**, *35*, 495-501.

Liljefors, T.; Gundertofte, K.; Norrby, P.; Pettersson, I. **Molecular mechanics and comparison of force fields.** Em: *Computational Medicinal Chemistry for Drug Discovery*. New York: Marcel Dekker, **2004**. Cap. 1.

Livingstone, D.J. **Quantitative structure-activity relationships.** In: King, F.D. (org.) *Medicinal Chemistry: Principles and Practice*. 2^a ed. Cambridge: Royal Society of Chemistry, **2003**. p. 215-242.

Lohray, B.B.; Gandhi, N.; Srivastava, B.K.; Lohray, V.B. **3D QSAR studies of N-4-arylacryloylpiperazin-1-yl-phenyl-oxazolidinones: A novel class of antibacterial agents.**

Bioorganic and Medicinal Chemistry Letters **2006**, *16*, 3817-3823.

Lombardino, J.G.; Lowe III, J.A. **The role of the medicinal chemistry in drug discovery – then and now.** *Nature Reviews Drug Discovery* **2004**, *3*, 853-862.

de Luca, L.; Vistoli, G.; Pedretti, A.; Barreca, M. L.; Chimirri, A. **Molecular dynamics studies of the full-length integrase–DNA complex.** *Biochemical and Biophysical Research Communications* **2005**, *336*, 1010-1016.

Makhija, M.T.; Kulkarni, V.M. **Eigen value analysis of HIV-1 integrase inhibitors.** *Journal of Chemical Information and Computer Science* **2001**, *41*, 1569-1577.

Makhija, M.T.; Kulkarni, V.M. **QSAR of HIV-1 integrase inhibitors by genetic function approximation method.** *Bioorganic and Medicinal Chemistry* **2002a**, *10*, 1483-1497.

Makhija, M.T.; Kulkarni, V.M. **3D-QSAR and molecular modeling of HIV-1 integrase inhibitors.** *Journal Computer-Aided Molecular Design* **2002b**, *16*, 181-200.

Martins, J.P.A.; Barbosa, E.G.; Pasqualoto, K.F.M.; Ferreira, M.M.C. **LQTA-QSAR: a new 4D-QSAR methodology.** *Journal of Chemical Information and Modeling* **2009**, *49*, 1428–1436.

Martins, J.P.A.; Ferreira, M.M.C. QSAR Modeling. Disponível em <http://lqta.iqm.unicamp.br>. 2009.

Masunari, A.; Tavares, L.C. **A new class of nifuroxazide analogs: synthesis of 5-nitrothiophene derivatives with antimicrobial activity against multidrug-resistant *Staphylococcus aureus*.** *Bioorganic and Medicinal Chemistry* **2007**, *15*, 4229-4236.

Mathworks Inc. Matlab 7 for Windows. 2006.

Matos, K.T.F.; Santos, M.C.M.; Muccioli, C. **Manifestações oculares do paciente infectado pelo HIV atendido no Departamento de Oftalmologia da Universidade Federal de São Paulo.** *Revista da Associação Médica Brasileira* **1999**, *45*, 323-326.

Maurin, C.; Bailly, F.; Buisine, E.; Vezin, H.; Mbemba, G.; Mouscadet, J.F.; Cotelle, P. **Spectroscopic Studies of Diketoacids–Metal Interactions. A Probing Tool for the Pharmacophoric Intermetallic Distance in the HIV-1 Integrase Active Site.** *Journal of Medicinal Chemistry* **2004**, *47*, 5583–5586.

Melagraki, G.; Afantitis, A.; Sarimveis, H.; Koutentis, P.A.; Markopolus, J.; Igglessi-Markopoulou, O. **Optimization of biaryl piperidine and 4-amino-2-biarylurea MCH1 receptor antagonists using QSAR modeling, classification techniques and virtual screening.** *Journal of Computer-Aided Molecular Design* **2007**, *21*, 251-267.

Melo, E.B.; Gomes, A.S.; Carvalho, I. **alfa- and beta-Glucosidase inhibitors: chemical structure and biological activity.** *Tetrahedron* **2006**, *62*, 10277-10302.

Menesia, E.O.; Passos, A.D.C.; Monteiro, M.E.; Dal-Fabbro, A.L.; Laprega, M.R. **Sobrevivência de pacientes com AIDS em uma cidade do sudeste brasileiro.** *Revista Panamericana de Salud Pública* **2001**, *10*, 29-36.

Miller, K.J.; Savchik, J.A. **A new empirical method to calculate average molecular polarizabilities.** *Journal of American Chemical Society* **1979**, *101*, 7206-7213.

Milne, G.W.A.; Nicklaus, M.C.; Driscoll, J.S.; Wang, S. **National Cancer Institute Drug Information System 3D Database.** *Journal of Chemical Information and Computer Science* **1994**, *34*, 1219-1224.

Ministério da Saúde. *Recomendações para terapia anti-retroviral em adultos e adolescentes infectados pelo HIV – 2002/2003.* **2002a.** Disponível em <http://www.aids.gov.br/final/tratamento/adulto.pdf>. Acessada em 15 de abril de 2009.

Ministério da Saúde. *Guia de tratamento clínico da infecção pelo HIV em crianças.* **2002b.** Disponível em <http://www.aids.gov.br/final/tratamento/crianca.pdf>. Acessada em 15 de abril de 2009.

Ministério da Saúde. *Boletim Epidemiológico AIDS.* v.1, ano 5, **2008.** 36 pp. Disponível em <http://www.aids.gov.br/data/Pages/LUMIS9A49113DPTBRIE.htm>. Acessado em 15 de abril de 2009.

Mitchell, S.S.; Rhodes, D.; Bushman, F.D.; Faulkner D.J. **Cyclodidemniserinol trisulfate, a sulfated serinolipid from the palauan *Ascidian didemnum guttatum* that inhibits HIV-1 integrase.** *Organic Letters* **2000**, *2*, 1605-1607.

Mitscher, L.A. **Antibiotics and antimicrobial agents.** In: Willians, D.A.; Lemke, T.L. (orgs.) *Foye's Principles of Medicinal Chemistry.* 5^a ed. Philadelphia: Lippincott Williams & Wilkins, **2002.** p. 819-866.

Mitsuya, H.; Shirasaka, T.; Broder, S. Em *Design of Anti-AIDS Drugs.* De Clerq, E. (org.). Amsterdam: Elsevier Press, 1990, cap. 2.

Molteni, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F.D.; Siegel, J.S.; Choe, S. **Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain.** *Acta Crystallographica Section D: Biological Crystallography* **2001**, *57*, 536-544 .

Montagneir, L. **A history of HIV discovery.** *Science* **2002**, *298*, 1728-1730.

Morgon, N.H.; Custódio, R. **Teoria do Funcional de Densidade.** *Química Nova* **1995**, *18*, 44-55.

Morley, C. Open BabelGUI 2.1.1 for Windows, 2006. Disponível em http://openbabel.sourceforge.net/wiki/Windows_GUI.

Morley, J.O.; Matthews, T.P. **Structure–activity relationships in nitrothiophenes.**

Bioorganic and Medicinal Chemistry **2006**, *14*, 8099-8108.

Nadal, S.R.; Horta, S.H.C.; Calore, E.E.; Manzione, C.R. **Outcome of treatment of anal squamous cell carcinoma and its precursor in HIV-infected patients.** *Revista da Associação Médica Brasileira* **2007**, *53*, 365-369.

National Aids Drug Policy. **2001.** Disponível em <http://www.aids.gov.br/final/biblioteca/drug/drug2.htm>. Acessada em 15 de abril de 2009.

Neamati, N.; Sunder, S.; Pommier, Y. **Design and discovery of HIV-1 integrase inhibitors** *Drug Discoveru Today* **1997**, *2*, 487-498.

Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittikul, V. **A practical overview of quantitative structure-activity relationship.** *Excli Journal* **2009**, *8*, 74-88.

Netzeva, T.I.; Aptula, A.O.; Benfenati, E.; Cronin, M.T.D.; Gini, G.; Lessigiarska, I.; Maran, U.; Vracko, M.; Schüürmann, G. **Description of the electronic structure of organic chemicals using semiempirical and ab initio methods for development of toxicological QSAR.** *Journal of Chemical Information and Modelling* **2005**, *45*, 106-114.

Neves, P.J.; Costa, J.B.M.; Ndujae, P.M. **TOP – um programa de cálculo de descritores topológicos para uso em correlações entre estrutura e atividade.** *Química Nova* **1998**, *21*, 709-713.

Ni, H.; Sottriffer, C.A.; McCammon, J.A.; **Ordered water and ligand mobility in the HIV-1 integrase-5CITEP complex: a molecular dynamics study.** *Journal of Medicinal Chemistry* **2001**, *44*, 3043-3047.

Nikolic, K.M. **QSAR Study of Aromatic Organochalcogens with Glutathione Peroxidase - like Antioxidant Activity.** *QSAR & Combintorial Science* **2007**, *26*, 358-367.

OECD-Organization for Economic Co-Operation and Development. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models.* Paris: OECD, **2007**. Disponível em <http://www.oecd.org/ehs>.

Oliveira, B.G.; Araújo, R.C.M.U. **Relação entre transferência de carga e as interações intermoleculares em complexos de hidrogênio heterocíclicos.** *Química Nova* **2007**, *30*, 791-796.

Ondeyka, J.G.; Zink, D.L.; Dombrowski, A.W.; Polishook, J.D.; Felock, P.J.; Hazuda, D.J.; Singh, S.B. **Isolation, Structure and HIV-1 Integrase Inhibitory Activity of Exophillic Acid, a Novel Fungal Metabolite from *Exophiala pisciphila*.** *Journal of Antibiotics* **2003**, *56*, 1018-1023.

Opar, A. **Raltegravir.** *Nature Reviews Drug Discovery* **2007**, *6*, 258-259.

Ortiz, A.R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R.C. **Reliability of comparative molecular force field analysis models: effects of data sacaling and variable**

- selection using a set of human synovial fluid phospholipase A2 inhibitors. *Journal of Medicinal Chemistry* **1997**, *40*, 1136-1148.
- Padrón, J.A.; Carrasco, R.; Pellón, R.F. **Molecular descriptor base on molar refractivity partition using Randic-type graph-theoretical invariant.** *Journal of Pharmacy and Pharmaceutical Sciences* **2002**, *5*, 258-266.
- Pais, G.C.G.; Zhang, X.; Marchand, C.; Neamati, N.; Cowansage, K.; Svarovskaia, E.S.; Pathak, V.K.; Tang, Y.; Nicklaus, M.; Pommier, Y.; Burke Jr, T.R. **3-Aryl-1,3-diketo-Containing Compounds** *Journal of Medicinal Chemistry* **2002**, *45*, 3184-3194.
- Patani, G.A.; LaVoie, E.J. **Bioisoterism: a rational approach in drug design.** *Chemical Reviews* **1996**, *96*, 277-188.
- Parril, A.L. **HIV-1 Integrase Inhibition: Binding Sites, Structure Activity Relationships and Future Perspectives.** *Current Medicinal Chemistry* **2003**, *10*, 1875-1888.
- Peçanha, E.P.; Antunes, O.A.C.; Tanuri, A. **Estratégias farmacológicas para a terapia anti-AIDS.** *Quimica Nova* **2002**, *25*, 1108-1116.
- Petrocchi, A.; Koch, U.; Matassa, V.G.; Pacini, B.; Stillmock, K.A.; Summa, V. **From dihydroxypyrimidine carboxylic acids to carboxamide HIV-1 integrase inhibitors: SAR around the amide moiety.** *Bioorganic & Medicinal Chemistry Letters* **2007**, *17*, 350-353.
- Philips, O.A.; Udo, E.E.; Samuel, S.M. **Synthesis and structure-antibacterial activity of triazolyl oxazolidinones containing long chain acyl moiety.** *European Journal of Medicinal Chemistry* **2008**, *43*, 1095-1104.
- Pirard, B.; Pickett, S.D. **Classification of kinase inhibitors using BCUT descriptors.** *Journal of Chemical Information and Computer Science* **2000**, *40*, 1431-1440.
- Pluymers, W.; Pais, G.; Van Maele, B.; Pannecouque, C.; Fikkert, V.; Burke Jr, T. R.; De Clercq, E.; Witvrouw, M.; Neamati, N.; Debyser, Z. **Inhibition of human immunodeficiency virus type 1 integration by diketo derivatives.** *Antimicrobial Agents and Chemotherapy* **2002**, *46*, 3292-3297.
- Pommier, Y.; Marchand, C.; Neamati, N. **Retroviral integrase inhibitors year 2000: update and perspectives.** *Antiviral Research* **2000**, *47*, 139-148.
- Prusiner, S.B. **Discovering the cause of AIDS.** *Science*. **2002**, *298*, 1726-1726
- Puzyn, T.; Suzuki, N.; Haranczyk, M.; Rak, J. **Calculation of quantum-mechanical descriptors for QSPR at the DFT level: is it necessary?** *Journal of Chemical Information and Modelling* **2008**, *48*, 1174-1180.
- Raghavan, K.; Buolamwini, J. K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W.; Weinstein, J. N. **Three-dimensional quantitative structure-activity relationship (QSAR) of HIV Integrase**

- Inhibitors: A Comparative Molecular Field Analysis (CoMFA) study.** *Journal of Medicinal Chemistry* **1995**, *38*, 890-897.
- Ramalho, E.D. **Variabilidade genética de nef em amostras de HIV-1 circulantes.** *Dissertação de Mestrado*. Instituto de Ciências Biológicas: Universidade de Brasília, **2006**. 101 pp.
- Ravi, M.; Hopfinger, A.J.; Hormann, R.E.; Dinan, L. **4D-QSAR analysis of a set of ecdysteroids and a comparison to CoMFA modeling.** *Journal of Chemical Information and Computer Science* **2001**, *41*, 1587-1604.
- Rocha, C.; Gouvêa, A.T.B; Machado, D.M.; Hörnke, L.; Succi, R.C.M. **Infarto cerebral em duas crianças infectadas pelo HIV-1.** *Arquivos de Neuropsiquiatria* **2003**, *61*, 1015-1017.
- Rodrigues, C.R. **Modelagem molecular.** *Química Nova na Escola* **2001**, *3*, 43-49
- Roy, P.P.; Leonard, J.T.; Roy, K. **Exploring the impact of size of training sets for the development of predictive QSAR models.** *Chemometrics and Intelligent Laboratory Systems* **2008**, *90*, 31-42.
- Roy, P.P.; Roy, K. **On some aspects of variable selection for partial least squares regression models.** *QSAR & Combinatorial Science* **2008**, *27*, 302-313.
- Rücker, C.; Rücker, G.; Meringer, M. **y-Randomization and its variants in QSAR/QSPR.** *Journal of Chemical Information and Modelling* **2007**, *47*, 2345-2357.
- Rustici, V.C.F.; Caramori, G.F.; Galembeck, S.E. **Ffeitos de substituintes na ligação de hidrogênio do 3-hidroxioprenal.** *Química Nova* **2006**, *29*, 1187-1192.
- Sadowski, J.; Schwab, C.H.; Gasteiger, J. **3D Structure Generation and Conformational Searching.** Em: *Computational Medicinal Chemistry for Drug Discovery*. New York: Marcel Dekker, **2004**. Cap. 7.
- Saíz-Urra, L.; González, M.P.; Fall, Y.; Gómez, G. **Quantitative structure-activity relationship studies of HIV-1 integrase inhibition. 1. GETAWAY descriptors.** *European Journal of Medicinal Chemistry* **2007**, *42*, 64-70.
- Sanderson, R.T. **Electronegativities in inorganic chemistry.** *Journal of Chemical Education* **1955**, *23*, 539-544.
- Sanderson, R.T. **Electronegativity and bond energy.** *Journal of American Chemical Society* **1983**, *105*, 2259-2261.
- Sanner, M.F. **Python: A Programming Language for Software Integration and Development.** *Journal of Molecular Graphics and Modeling*, **1999**, *17*, 57-61.
- Sant'Anna, C.M.R. **Glossário de termos usados no planejamento de fármacos (recomendações da IUPAC para 1997).** *Química Nova* **2002**, *25*, 505-512.

Sant'Anna, C.M.R. **Métodos de modelagem molecular para estudo e planejamento de compostos bioativos: uma introdução.** *Revista Virtual de Química* **2009**, *1*, 49-57.

Santos Filho, O.A.; Alencastro, R.B. **Modelagem de proteínas por homologia.** *Química Nova* **2003**, *26*, 253-259.

Schuettelkopf, A.W.; van Aalten, D.M.F.; **PRODRG: a tool for high-throughput crystallography of protein-ligand complexes.** *Acta Crystallographica Section D: Biological Crystallography*. **2004**, *D60*, 1355-1363.

Sethi, M.L. **Antiviral agents and protease inhibitors.** In: Williams, D.A.; Lemke, T.L. (Org.) *Foye's Principles of Medicinal Chemistry*. Philadelphia: Lippincott Williams & Wilkins, 5^a ed. **2002**. p. 952-980.

Serrao, E.; Odde, S.; Ramkumar, K.; Neamati, N. **Raltegravir, elvitegravir, and metoogravir: the birth of "me-too" HIV-1 integrase inhibitors.** *Retrovirology* **2009**, *6*. Disponível em: <http://www.retrovirology.com/content/6/1/25>.

Sheo B. Singh, Deborah L. Zink, Brian Heimbach, Olga Genilloud, Ana Teran, Keith C. Silverman, Russell B. Lingham, Peter Felock, Daria J. Hazuda. **Structure, stereochemistry, and biological activity of integramycin, a novel hexacyclic natural product produced by *Actinoplanes sp.* that inhibits HIV-1 integrase.** *Organic Letters* **2002**, *4*, 1123-1126.

Sheo B. Singh, Kithsiri Herath, Ziqiang Guan, Deborah L. Zink, Anne W. Dombrowski, Jon D. Polishook, Keith C. Silverman, Russell B. Lingham, Peter J. Felock, Daria J. Hazuda. **Integramides A and B, two novel non-ribosomal linear peptides containing nine C-methyl amino acids produced by fungal fermentations that are inhibitors of hiv-1 integrase.** *Organic Letters* **2002**, *4*, 1431-1434.

Shiomi, K.; Matsui, R.; Isozaki, M.; Chiba, H.; Sugai, T.; Yamaguchi, Y.; Masuma, R.; Tomoda, H.; Chiba, T.; Yan, H.; Kitamura, Y.; Sugiura, W.; Omura, S.; Tanaka, H.; **Fungal phenalenones inhibit HIV-1 integrase.** *Journal of Antibiotics* **2005**, *58*, 654-658.

Shodor Education Foundation, Inc. *Background Reading for Basis Sets*, **2000**. Disponível em <http://www.shodor.org/chemviz/basis/students/background.html>. Acessado em 05 de abril de 2009.

Silakari, P.; Shrivastava, S.D.; Silakari, G.; Kohli, D.V.; Rambabu, G.; Srivastava, S.; Shrivastava, S.K.; Silakari, O. **QSAR analysis of 1,3-diaryl-4,5,6,7-tetrahydro-2H-isoindole derivatives as selective COX-2 inhibitors.** *European Journal of Medicinal Chemistry* **2008**, *43*, 1559-1569.

Silva, T. H. A. **Modelagem molecular.** Em: Andrei, C.C.; Ferreira, D.T.; Faccione, M.; Faria, T.J. (orgs.) *Da química medicinal à química combinatória e modelagem molecular – um curso prático*. Barueri: Manole, **2003**. p. 111-139

Silva, V.B.; Silva, C.H.T.P. **Modelagem molecular de proteínas-alvo por homologia estrutural.** *Revista Eletrônica de Farmácia* **2007**, *4*, 15-26.

Simas, A.M.; Rocha, G.B.; **Métodos semi-empíricos de estrutura eletrônica em química quântica**. In: Morgon, N.H.; Coutinho, K. (orgs.) *Métodos de química teórica e modelagem molecular*. São Paulo: Livraria da Física, **2007**. p. 29-72.

Singh, S.B.; Felock, P.; Hazuda, D.J. **Chemical and enzymatic modifications of integrase acid and HIV-1 integrase inhibitory activity**. *Bioorganic and Medicinal Chemistry Letters* **2000**, *10*, 235-238.

Singh, S.B.; Zink, D.; Polishook, J.; Valentino, D.; Shaflee, A.; Silverman, K.; Felock, P.; Teran, A.; Vilella, D.; Hazuda, D.J.; Lingham, R.B. **Structure and absolute stereochemistry of HIV-1 integrase inhibitor integrase acid. A novel eremophilane sesquiterpenoid produced by a *Xylaria* sp.** *Tetrahedron Letters* **1999**, *40*, 8775-8779.

Singh, S.B.; Jayasuriya, H.; Salituro, G.M.; Zink, D.L.; Shafiee, A.; Heimbuch, B.; Silverman, K.C.; Lingham, R.B.; Genilloud, O.; Teran, A.; Vilella, D.; Felock, P.; Hazuda, D. **The complestatins as HIV-1 integrase inhibitors. Efficient isolation, structure elucidation, and inhibitory activities of isocomplestatin, chloropectin I, new complestatins, A and B, and acid-hydrolysis products of chloropectin I**. *Journal of Natural Products* **2001**, *64*, 874-882.

Singh, S.B.; Zink, D.L.; Bills, G.F.; Pelaez, F.; Teran, A.; Collado, J.; Silverman, K.C.; Lingham, R.B.; Felock, P.J.; Hazuda, D.J. **Discovery, structure and HIV-1 integrase inhibitory activities of integracins, novel dimeric alkyl aromatics from *Cytonaema* sp.** *Tetrahedron Letters* **2002**, *43*, 1617-1620.

Singh, S.B.; Pelaez, F.; Hazuda, D.J.; Lingham, R.B.; **Discovery of natural products inhibitors of HIV-1 integrase at Merck**. *Drugs of the Future* **2005**, *30*, 277-299.

Smith, S.M. **New York City HIV superbug: fear or fear not**. *Retrovirology* **2005**, *2*. Disponível em: <http://www.retrovirology.com/content/2/1/14>.

Sotriffer, C.A.; Ni, H.; McCammon, J.A. **HIV-1 integrase inhibitors interactions at the active site: prediction of binding modes unaffected by crystal packing**. *Journal of American Chemical Society* **2000**, *122*, 6136-6137.

Souza, H.F. **Introdução à química computacional**. Apostila. Campinas: Centro Nacional de Processamento de Alto Desempenho de São Paulo – CENAPAD, **2000**. 25 pp.

Souza, M.V.N.; Almeida, M.V. **Drogas anti-VIH: passado, presente e perspectivas futuras**. *Química Nova* **2003**, *26*, 366-372.

Souza, S.L.S.; Feitoza, P.V.S.; Araújo, J.R.; Andrade, R.V.; Ferreira, L.C.L. **Causas de óbito em pacientes com síndrome da imunodeficiência adquirida, necropsiados na Fundação de Medicina Tropical do Amazonas**. *Revista da Sociedade Brasileira de Medicina Tropical* **2008**, *41*, 247-251.

van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. **GROMACS: Fast, flexible, and free.** *Journal of Computational Chemistry* **2005**, *26*, 1701-1718.

Stanton, D.T. **Evaluation and use of BCUT descriptors in QSAR and QSPR studies.** *Journal of Chemical Information and Computer Science* **1999**, *39*, 11-20.

Summa, V.; Petrocchi, A.; Matassa, V.G.; Gardelli, C.; Muraglia, E.; Rowley, M.; Paz, O.G.; Laufer, R.; Monteagudo, E.; Pace, P. **4,5-Dihydroxypyrimidine Carboxamides and N-Alkyl-5-hydroxypyrimidinone Carboxamides Are Potent, Selective HIV Integrase Inhibitors with Good Pharmacokinetic Profiles in Preclinical Species.** *Journal of Medicinal Chemistry* **2006**, *49*, 6646–6649.

Taktakishvili, M.; Neamati, N.; Pommier, Y.; Pal, S.; Nair, V. **Recognition and inhibition of HIV-integrase by novel dinucleotides.** *Journal of American Chemical Society* **2000**, *122*, 5671-5677.

Tavares, L.C. **QSAR: a abordagem de Hansch.** *Química Nova* **2004**, *27*, 631-639.

Teixeira, P.R. *Boletim epidemiológico Aids* 2002. **2003**. Disponível em http://www.aids.gov.br/final/biblioteca/bol_dezembro_2002/boletim.asp. Acessada em 15 de abril de 2009.

Teófilo, R.F., Martins, J.P.A., Ferreira M.M.C. **Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression.** *Journal of Chemometrics* **2009**, *23*, 32-48.

Thomas, M.; Brady, L. **HIV-integrase: a target for Aids therapeutics.** *Trends in Biotechnology* **1997**, *15*, 167-172.

Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, Weinheim: Wiley-VCH, 2000. 667 pp.

Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. Dragon 3.0 Web Version, 2003.

Tropsha, A.; Gramatica, P.; Gombar, V.K. **The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models.** *QSAR & Combinatorial Science*. **2003**, *22*, 69-77.

Toit, K.; Elgorashi, E.E.; Malan, S.F.; Drewes, S.E. van Staden, J.; Croueh, N.R.; Mulholland, D.A. **Anti-inflammatory activity and QSAR studies of compounds isolated from Hyacinthaceae species and Tachiadenus longiflorus Griseb. (Gentianaceae).** *Bioorganic and Medicinal Chemistry* **2005**, *13*, 2561-2568.

Unaids. *Unaids Report on the Global HIV/AIDS epidemics*. Geneva: Unaids, **2002**. 229 pp.

Unaids. *Unaids Report on the Global HIV/AIDS epidemics*. Geneva: Unaids, **2005**. 98 pp.

- Unaids. *Unaid Report on the Global HIV/AIDS epidemics*. Geneva: Unaid, **2008**. 357 pp.
- Vacca J.; Wai J.; Fisher T.; Embrey M.; Hazuda D.; Miller M.; Felock P.; Witmer M.; Gabryelski L.; Lyle T. **Discovery of MK-2048 – subtle changes confer unique resistance properties to a series of tricyclic hydroxypyrrole integrase strand transfer inhibitors**. In: *4th IAS Conference on HIV Pathogenesis, Treatment and Prevention*. Sydney, Australia. **2007**. Disponível em <http://www.ias2007.org/pag/Abstracts.aspx?AID=3234>. Acessado em 21 de julho de 2009.
- Vedani, A; Dobler, M. **Multidimensional QSAR: Moving from three- to five-dimensional concepts**. *Quantitative Structure-Activity Relationships* **2002**, *21*, 382-390.
- Vedani, A.; Dobler, M.; Dollinger, H.; Hasselbach, K.M.; Birke, F.; Lill, M.A. **Novel ligands for the chemokine receptor-3 (CCR3): A receptor-modeling study based on 5D-QSAR**. *Journal of Medicinal Chemistry* **2005a**, *48*, 1515-1527.
- Vedani, A; Dobler, M; Lill, MA. **Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor**. *Journal of Medicinal Chemistry* **2005b**, *48*, 3700-3703.
- Vedani, A.; Zumstein, M.; Lill, M.A.; Ernst, B. **Simulating alpha/beta selectivity at the human thyroid hormone receptor: Consensus scoring using multidimensional QSAR**. *ChemMedChem* **2007**, *2*, 78-87.
- Verma, R.P.; Hansch, C. **An approach toward the problem of outliers in QSAR**. *Bioorganic & Medicinal Chemistry* **2005**, *13*, 4597–4621.
- Voisset, C.; Andrawiss, M. **Retroviruses at a glance**. *Genome Biology* **2000**, *1*. Disponível em <http://genomebiology.com/2000/1/3/reports/4015/>
- Votano, J.R.; M. Parham, L.H. Hall, L.B. Kier, S. Oloff, A. Tropsha, Q. Xie, W. Tong, **Three new consensus QSAR models for the prediction of Ames genotoxicity**. *Mutagenesis* **2004**, *19*, 365-377.
- Wai, J.S.; Egbertson, M.S.; Payne, L.S.; Fisher, T.E.; Embrey, M.W.; Tran, L.O.; Melamed, J.Y.; Langford, H.M.; Guare, Jr., J.P.; Zhuang, L.; Grey, V.E.; Vacca, J.P.; Holloway, M.K.; Naylor-Olsen, A.M.; Hazuda, D.J.; Felock, P.J.; Wolfe, A.L.; Stillmock, K.A.; Schleif, W.A.; Gabryelski, L.J.; Young, S.D. **4-Aryl-2,4-dioxobutanoic Acid Inhibitors of HIV-1 Integrase and Viral Replication in Cells**. *Journal of Medicinal Chemistry* **2000**, *43*, 4923–4926.
- Waltz, E. **Intermittent AIDS therapy is dangerous, trial shows**. *Nature Medicine* **2006**, *12*, 157-157.
- Wan, J.; Zhang, L.; Yang, G.; Zhan, C. **Quantitative Structure–Activity Relationship for Cyclic Imide Derivatives of Protoporphyrinogen Oxidase Inhibitors: A Study of Quantum Chemical Descriptors from Density Functional Theory**. *Journal of Chemical Information and Computer Sciences* **2004**, *44*, 2099-2105.

- Waterbeemd, H. **Quantitative approaches to structure-activity relationships**. In: Warmuth, C.G. (Ed.), *The practice of Medicinal Chemistry*, Academic Press, San Diego, **1996**, pp. 367-389.
- Waterbeemd, H. **Introduction**. Em Waterbeemd, H.(Edt.) *Chemometrics Methods in Molecular Design*. Weinheim: VCH, **1994**, 1-13.
- Wold, S. **PLS for multivariate linear modeling**. In: H. van de Waterbeemd (Ed.), *Chemometric Methods in Molecular Design*, VCH, Weinheim, **1998**, pp. 195-218.
- Wold, S.; Eriksson, L.; **Statistical validation of QSAR results**. In: H. van de Waterbeemd (Org.) *Chemometric Methods in Molecular Design*. VCH: Weinheim, **1998**. pp. 309-318.
- Wold, S.; Johansson, E.; Cocchi, M. **PLS – Partial-least squares projections to latent structure**. In: H. Kubinyi (Org.) *3D QSAR in Drug Design*. Dordrecht: Kluwer/Escom, **2000**. pp. 523-550.
- Wold, S.; Sjöström, M.; Eriksson, L. **PLS-regression: a basic tool of chemometrics**. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58*, 109-130.
- Xinliang, Y.; Bing, Y.; Xueye, W. **Quantitative structure–property relationships for the reactivity parameters of acrylate monomers**. *European Polymer Journal* **2008**, *44*, 3997–4001.
- Yakasai, B. A. **Neuropsychiatric complications of HIV/AIDS**. *Annals of African Medicine* **2004**, *3*, 63-65.
- Yan, X.; Xiao, H.; Gong, X.; Ju, X. **A comparison of semiempirical and first principle methods for establishing toxicological QSARs of nitroaromatics**. *Journal of Molecular Structure (Theochem)* **2006**, *764*, 141-148,
- Yaun, H.; Parril, A.L. **QSAR development to describe HIV-1 integrase inhibition**. *Journal of Molecular Structure - Theochem* **2000**, *529*, 273-282.
- Yaun, H.; Parril, A.L. **QSAR studies of HIV-1 integrase inhibition**. *Bioorganic and Medicinal Chemistry* **2002**, *10*, 4169-4183.
- Yaun, H.; Parril, A.L. **Cluster analysis and three-dimensional QSAR studies of HIV-1 integrase inhibitors**. *Journal of Molecular Graphics and Modelling* **2005**, *23*, 317-328.
- Young, D.C. *Computational chemistry: a practical guide for applying techniques to real-world problems*. New York: Wiley-Interscience, **2001**. 369 pp.
- Young, S.D. **Inhibition of HIV-1 integrase by small molecules: the potential for a new class of Aids chemotherapeutics** *Current Opinion in Drug Discovery & Development*. **2001**, *4*, 402-410.

- [Yu, X.L.](#); [Yi, B.](#); [Liu, F.](#); [Wang, X.Y.](#) **Prediction of the dielectric dissipation factor $\tan \delta$ of polymers with an ANN model based on the DFT calculation.** *Reactive and Functional Polymers* **2008**, 68, 1557-1562.
- Zeinalipour-Loizidou, E.; Nicolaou, C.; Nicolaidis, A.; Kostrikis, L.G. **HIV-1 integrase: from biology from chemotherapeutics.** *Current HIV Research* **2007**, 5, 365-388.
- Zhang, S.; Wei, L.; Bastow, K.; Zheng, W.; Brossi, A.; Lee, K.; Tropsha, A. **Antitumor agents 252. Application of validated QSAR models to database mining: discovery of novel tylophorine derivatives as potential anticancer agents.** *Journal of Computer-Aided Molecular Design* **2007**, 21, 97-112.
- Zhang, J.; Xiao, J.J. **Theoretical studies on heats of formation for cubyl nitrates using density functional theory B3LYP method and semiempirical MO methods.** *International Journal of Quantum Chemistry* **2002**, 86, 305-312.
- Zheng, R.; Jenkins, T.M.; Craigie, R. **Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity.** *Proceedings of the National Academy of Sciences USA* **1996**, 93, 13659-13664.
- Zhuang, L.; Wai, J.S.; Embrey, M.W.; Fisher, T.E.; Egbertson, M.S.; Payne, L.S.; Guare, Jr., J.O.; Vacca, J.P.; Hazuda, D.J.; Felock, P.J.; Wolfe, A.L.; Stillmock, K.A.; Witmer, M.V.; Moyer, G.; Schleif, W.A.; Gabryelski, L.J.; Leonard, Y.M.; Lynch Jr., J.J.; Michelson, S.R.; Young, S.D. **Design and Synthesis of 8-Hydroxy-[1,6]Naphthyridines as Novel Inhibitors of HIV-1 Integrase in Vitro and in Infected Cells.** *Journal of Medicinal Chemistry* **2003**, 46, 453-456.

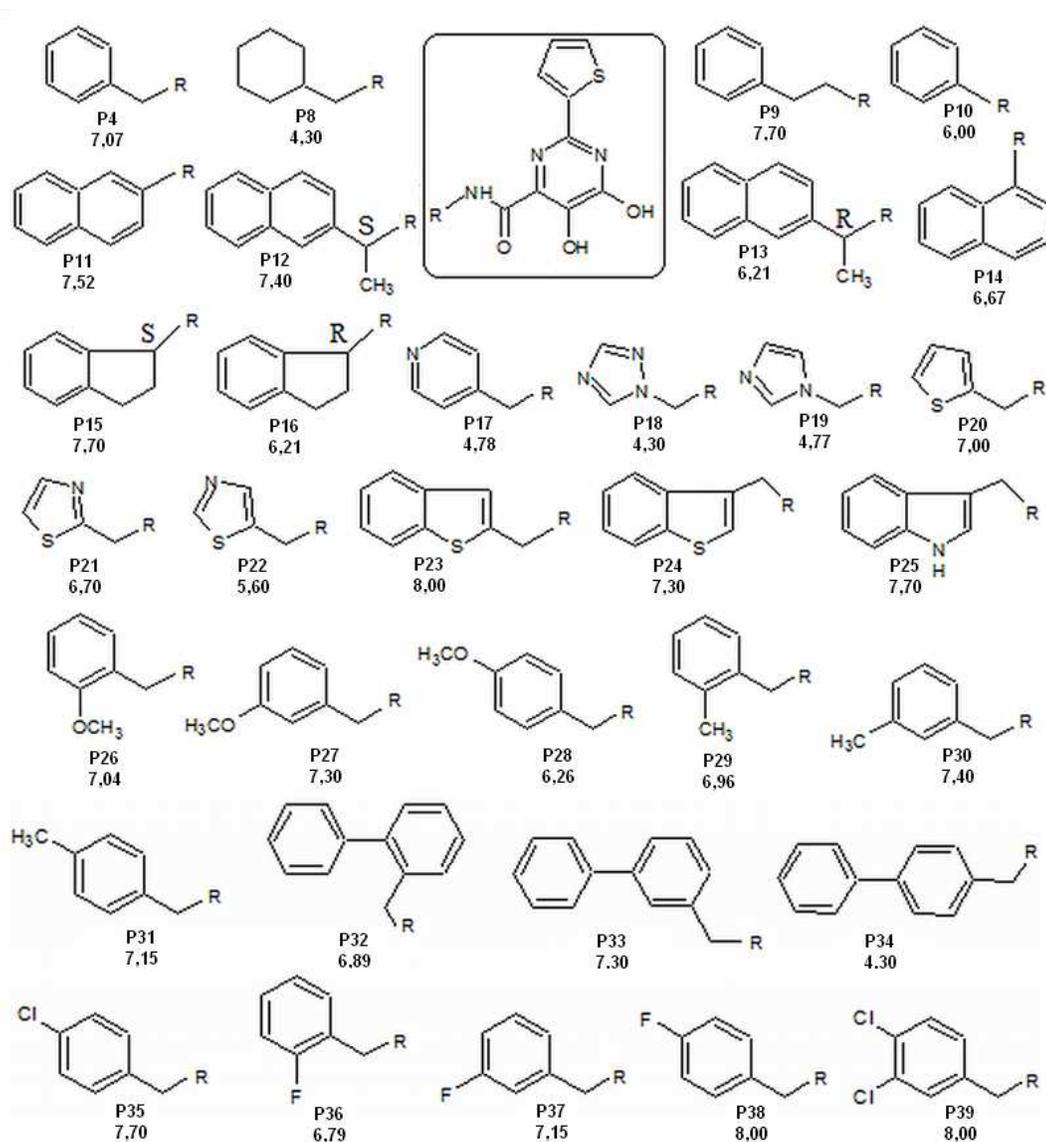
ANEXO 1. ESTRUTURAS E VALORES DE pIC_{50} DOS COMPOSTOS DO CONJUNTO DE TREINAMENTO.


Figura 1.1. Compostos descritos por Petrocchi e colaboradores (2007). Pertencem a classe das 4,5-diidroxipirimidinas carboxamidas.

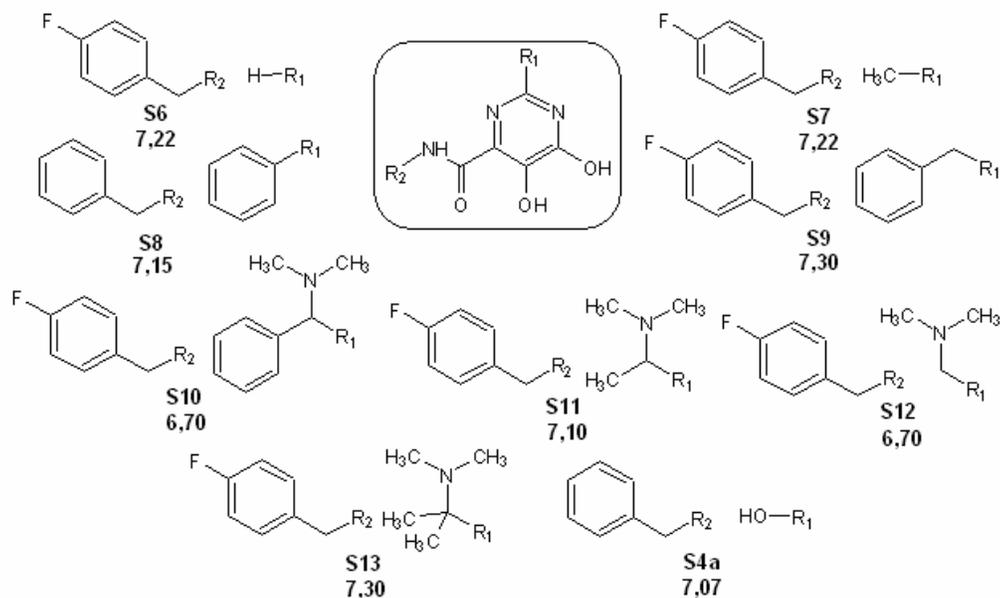


Figura 1.2. Compostos descritos por Summa e colaboradores (2006). Pertencem a classe das 4,5-dihidropirimidinas carboxamidas.

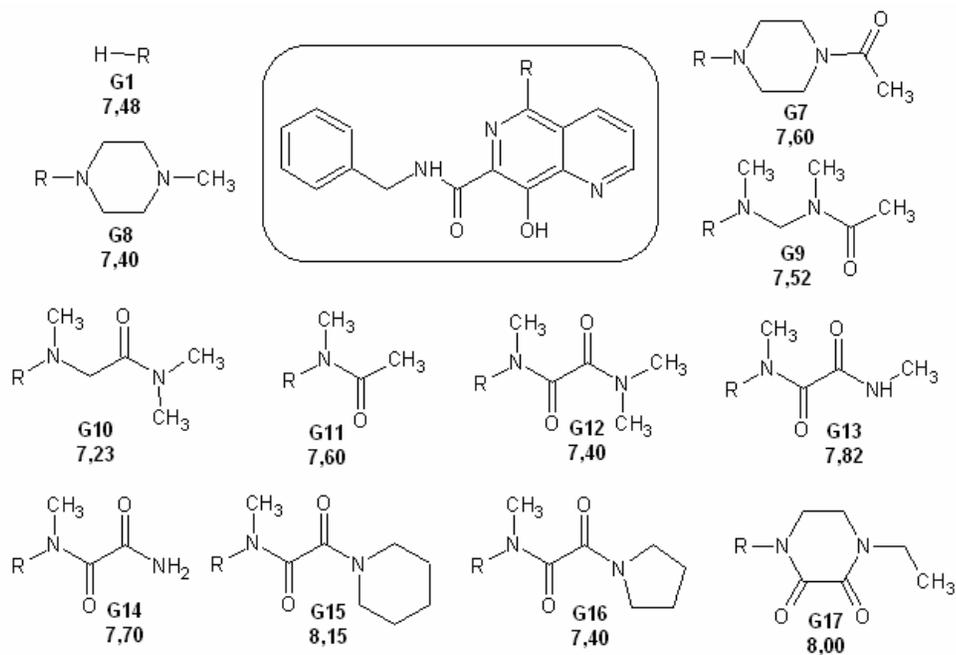


Figura 1.3. Compostos descritos por Guare e colaboradores (2006). Pertencem a classe das 8-hidroxi-1,6-naftiridinas.

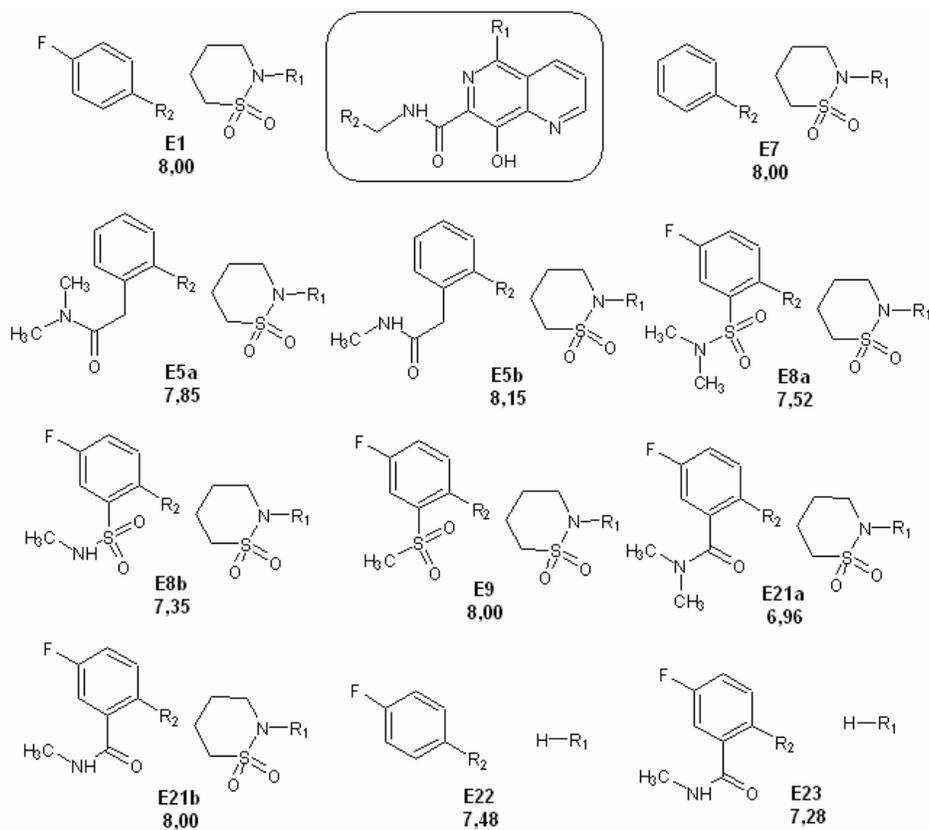


Figura 1.4. Compostos descritos por Egbertson e colaboradores (2007). Pertencem a classe das 8-hidroxi-1,6-naftiridinas.

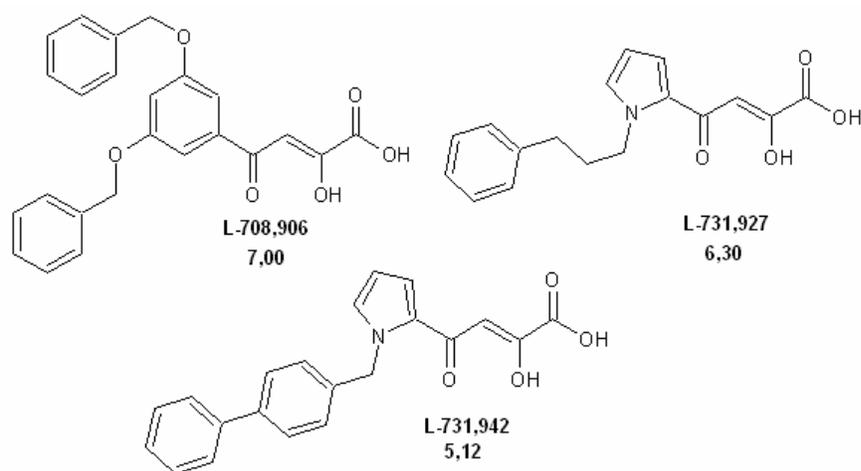


Figura 1.5. Compostos descritos por Hazuda e colaboradores (2000). Pertencem a classe dos β -dicetoácidos.

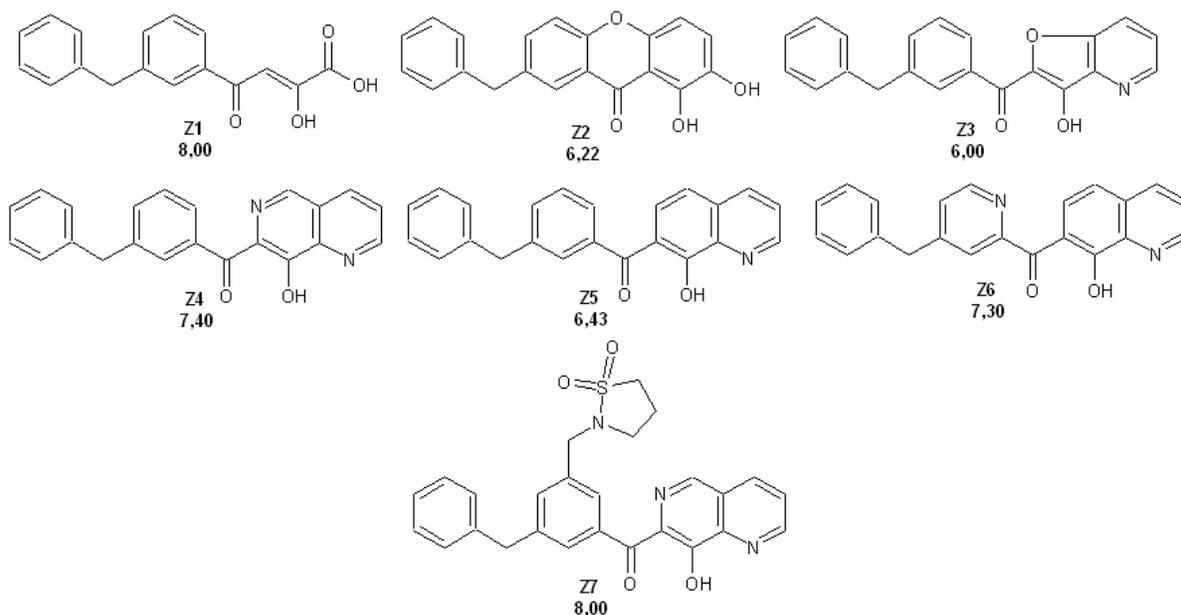


Figura 1.6. Compostos descritos por Zhuang e colaboradores (2003). O conjunto é formado por um dicetoácido, duas 8-hidroxi-1,6-naftiridinas e quatro derivados relacionados.

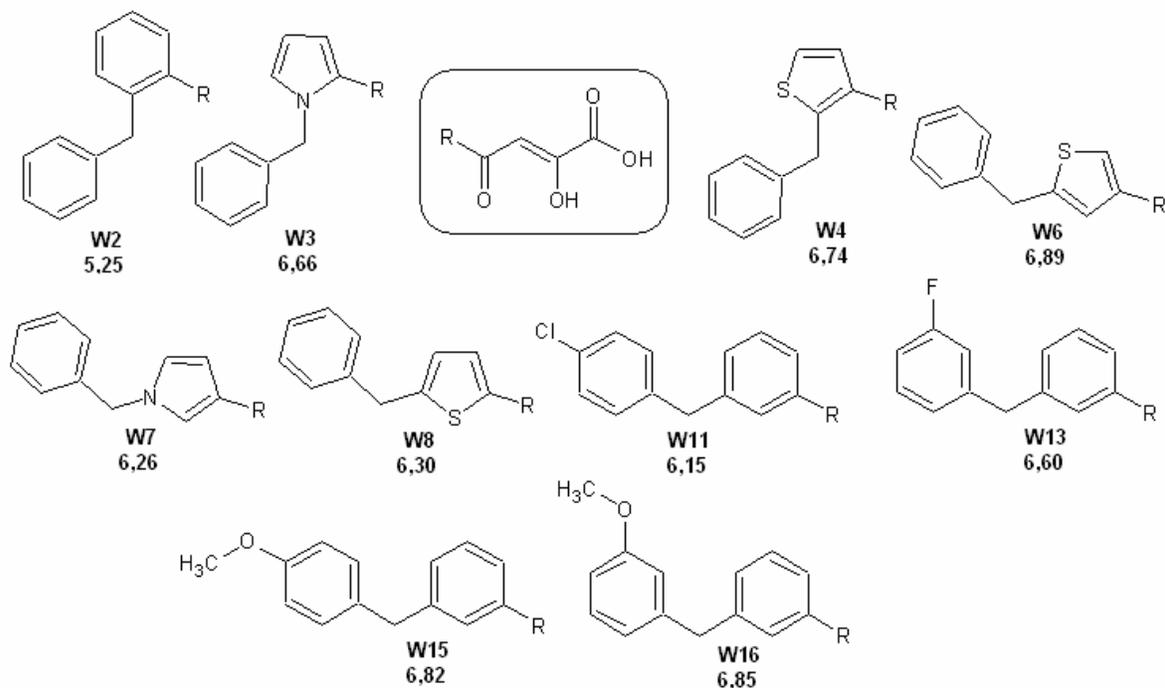


Figura 1.7. Compostos descritos por Wai e colaboradores (2000). Pertencem a classe dos β -dicetoácidos.

ANEXO 2. ALGORITMOS DESNVOLVIDOS POR MEMBROS DO GRUPO LQTA UTILIZADOS NESSE TRABALHO.

2.1. Algoritmo para reduzir a quantidade de descritores de Lennard-Jones antes do pré-processamento *blockscaling*, de autoria de Euzébio Guimarães Barobosa (grupo LQTA).

```
%  
% cutoff  
%  
X=LJ;  
[m,n]=size(X);  
for f=1:m  
    for j=1:n  
        if (X(f,j) >= 125.52)  
            X(f,j) = 125.52+log10(X(f,j)-125.52);  
        else  
            X(f,j) = X(f,j);  
        end  
    end  
end  
LJ=X;
```

Para utilizar:

> cutoff

2.2. Algoritmo para seleção de variáveis independentes com tendência linear e distribuição normal em relação a variável dependente, de autoria de Euzébio Guimarães Barobosa (grupo LQTA).

```
% X descritores
clear jan
[m,n]=size(X);
% janela de 10
faixa = 10;
for i=1:n,
    jan(:,i) = (max(X(:,i))-min(X(:,i)))/faixa;
end
% frequencias nos quadrantes
clear FREQ
%
for g=1:faixa
for f=1:m
    for j=1:n
        if (X(f,j) >= ((g-1) * jan(:,j)+ min((X(:,j)))) & (X(f,j) <= (g *
jan(:,j)+ min((X(:,j))))))
            FREQ(j,g,f) = 1;
        else
            FREQ(j,g,f) = 0;
        end
    end
end
end
end
F=sum(FREQ,3);
F=F';
for k=1:n
    param(1,k) = max(F(:,k))./m;
end
```

Para utilizar:

```
> seldescri3
```

Após o processo, deve ser selecionado a frequência máxima desejada:

```
> X(:,param > 0.4) = [ ]
> names(:,param > 0.4) = [ ]
```

Onde X é a matriz de dados selecionada, names é a matriz com a identificação desses dados, e 0,4 é o f_{\max} desejado.

2.3. Algoritmo para realização do teste de aleatorização do vetor y de um modelo PLS no Matlab, de autoria de João Paulo Ataíde Martins (grupo LQTA).

```
function parametros = yrandomization(X,y,nVL,preproc,nmax,nret);
yr = [];
m = length(y);
for i=1:nmax
    a = rand(m,1);
    [h,ind] = sort(a);
    y1 = y(ind);
    PLSmodel = val_cruzada(X,y1,nVL,preproc,preproc,nret);
    R2(i) = PLSmodel.r2(nVL);
    Q2(i) = PLSmodel.q2(nVL);
    corr(i) = correlograma(y1,y);
    yr = [yr y1];
end
PLSmodel = val_cruzada(X,y,nVL,preproc,preproc,nret);
R2(i+1) = PLSmodel.r2(nVL);
Q2(i+1) = PLSmodel.q2(nVL);
corr(i+1) = 1;
plot(R2,Q2,'o');
title('y-randomization');
xlabel('R^2');
ylabel('Q^2');
yr = [yr y];
parametros.R2 = R2;
parametros.Q2 = Q2;
parametros.corr = corr;
parametros.yr = yr;
```

Para utilizar:

```
>Yrand= yrandomization(X,y,nVL,preproc,nmax,nret);
```

Onde: X = matriz com as variáveis dependentes selecionadas; $y1$ = matriz com a variável dependente em estudo; nVL = número de variáveis latentes do modelo; $preproc$ = tipo de pré-processamento de dados desejado, sendo 1 para dados autoescalados, 2 para dados centrados na média; $nmax$ = número máximo de aleatorizações desejadas; $nret$ = número de amostras que devem ser retiradas para a validação cruzada.

O algoritmo devolve uma matriz composta de três dados diferentes: $R2$: os novos valores de R^2 para os modelos reconstruídos após a aleatorização do vetor y ; $Q2$: os novos valores de Q^2_{LOO} para os modelos reconstruídos após a aleatorização do vetor y ; yr : as correlações entre os valores do vetor y não aleatorizado e cada uma das aleatorizações.

2.4. Algoritmo para realização da validação leave- N -out (LNO) no Matlab, de autoria de João Paulo Ataíde Martins (grupo LQTA).

```
function parametros = leavenout(X,y,nVL,preproc,nmax,nrepet);
for i=1:nmax
    for j=1:nrepet
        PLSmodel = val_cruzada(X,y,nVL,preproc,preproc,i);
        rmsecv(i,j) = PLSmodel.RMSECV(nVL);
        Q2(i,j) = PLSmodel.q2(nVL);
    end
end
figure
hold
for i=1:nmax
    plot(i*ones(1,nrepet),rmsecv(i,:), 'o');
    title('Leave-N-out');
    xlabel('Amostras retiradas');
    ylabel('RMSECV');
end
figure
hold
for i=1:nmax
    plot(i*ones(1,nrepet),Q2(i,:), 'o');
    title('Leave-N-out');
    xlabel('Amostras retiradas');
    ylabel('Q2');
end
parametros.rmsecv = rmsecv;
parametros.Q2 = Q2;
```

Para utilizar:

```
>LNO=leavenout(X,y,nVL,preproc,nmax,nrepet);
```

Onde: X = matriz com as variáveis independentes selecionadas; y = matriz com a variável dependente em estudo; nVL = número de variáveis latentes do modelo; $preproc$ = tipo de pré-processamento de dados desejado, sendo 1 para dados autoescalados, 2 para dados centrados na média; $nmax$ = número máximo “ N ” a ser realizado; $nrepet$ = número de replicatas para cada “ N ”.

2.5. Algoritmo para construção de modelos PLS a partir de uma matriz específica (obs: necessário para utilização dos algoritmos dos testes de aleatorização do vetor y e da validação LNO). Autor: João Paulo Ataíde Martins, grupo LQTA.

```
function PLSmodel = val_cruzada(X1,y,npcs,preprocx,preprocy,nret)
% VAL_CRUZADA Realiza a validação cruzada leave-one-out de um modelo de
regressão
%
% Entrada: X (m,n): matriz com as variáveis independentes
% y (m,1): vetor com a variável dependente
% npcs: número máximo de componentes principais ou variáveis latentes a ser
% utilizado
% preprocx (1, 2 ou 3): método de préprocessamento a ser utilizado para as
% variáveis independentes
% preprocy (1, 2 ou 3): método de préprocessamento a ser utilizado para a
% variável dependente
% nret número de amostras a serem retiradas
% Saída: PLSmodel: struct com o modelo PLS obtido
%
% PLSmodel = val_cruzada(X1,y,npcs,preprocx,preprocy)
%
% See also: PREPROCESSA, PROCESSA, REESCALA, CORRELOGRAMA

% Copyright © LQTA-Unicamp OPS 1.0, Inc. 2006-2007
% Qualquer mudança, recompilação, tradução ou conversão de "M-files"
% não poderão ser realizadas sem a permissão escrita de
% LQTA-Unicamp OPS 1.0, Inc.
%
% Modified 02/07
% Checked on MATLAB 7 by JPAM 06/02/07
% -standard IO mods

[m,n] = size(X1);
[X1,y,ind] = aleatoriza(X1,y);
nconj = floor(m/nret);
X = X1((nret+1):m,:);
[X,media,desvio] = preprocessa(X,preprocx);
yaux = y((nret+1):m,:);
[yaux,mediay,desvioy] = preprocessa(yaux,preprocy);
b = opsimpls(X,yaux,npcs);
vet = processa(X1(1:nret,:),media,desvio,preprocx);
ycv(1:nret,:) = vet*b;
ycv(1:nret,:) = reescala(ycv(1:nret,:),mediay,desvioy,preprocy);
for i=2:nconj
    X = [X1(1:((i-1)*nret),:);X1((nret*i+1):m,:)];
    [X,media,desvio] = preprocessa(X,preprocx);
    yaux = [y(1:((i-1)*nret),:);y((nret*i+1):m,:)];
    [yaux,mediay,desvioy] = preprocessa(yaux,preprocy);
    b = opsimpls(X,yaux,npcs);
    vet = processa(X1((nret*(i-1)+1):(nret*i),:),media,desvio,preprocx);
    ycv((nret*(i-1)+1):(nret*i),:) = vet*b;
    ycv((nret*(i-1)+1):(nret*i),:) = reescala(ycv((nret*(i-
1)+1):(nret*i),:),mediay,desvioy,preprocy);
end
rest = m-(nconj*nret);
```

```

if rest~=0
    X = X1(1:m-rest,:);
    [X,media,desvio] = preprocessa(X,preprocx);
    yaux = y(1:m-rest,:);
    [yaux,mediay,desvioy] = preprocessa(yaux,preprocy);
    b = opsimpls(X,yaux,npcs);
    vet = processa(X1((m-rest)+1:m,:),media,desvio,preprocx);
    ycv((m-rest)+1:m,:) = vet*b;
    ycv((m-rest)+1:m,:) = reescala(ycv((m-
rest)+1:m,:),mediay,desvioy,preprocy);
end
X1 = desaleatoriza(X1,ind);
y = desaleatoriza(y,ind);
ycv = desaleatoriza(ycv,ind);
mediay = mean(y);
ssy = sum((y-mediay).^2);
%[m,n] = size(b);
[m,n] = size(X1);
[X,media,desvio] = preprocessa(X1,preprocx);
[yaux,mediay,desvioy] = preprocessa(y,preprocy);
b = opsimpls(X,yaux,npcs);
ycal = X*b;
ycal = reescala(ycal,mediay,desvioy,preprocy);
for j=1:npcs
    presscv = sum((y-ycv(:,j)).^2);
    rmsecv(j) = sqrt(presscv/m);
    param(j) = sqrt(presscv/(m-j+1));
    rcv(j) = correlograma(ycv(:,j),y);
    q2(j) = 1-(presscv/ssy);
    presscal = sum((y-ycal(:,j)).^2);
    rmsec(j) = sqrt(presscal/m);
    rcal(j) = correlograma(ycal(:,j),y);
    R2(j) = 1-(presscal/ssy);
    [breal(:,j),tind(j)] =
reprocessavetreg(b(:,j),preprocx,media,mediay,desvio,desvioy);
end
PLSmodel.regproc = b;
PLSmodel.reg = breal;
PLSmodel.tind = tind;
PLSmodel.RMSECV = rmsecv;
PLSmodel.rcv = rcv;
PLSmodel.Ypred = ycv;
PLSmodel.q2 = q2;
PLSmodel.RMSEC = rmsec;
PLSmodel.rcal = rcal;
PLSmodel.Ycal = ycal;
PLSmodel.r2 = R2;
PLSmodel.param = param;

```

ANEXO 3. TABELAS COM OS VALORES DOS DESCRITORES OBTIDOS PARA OS MODELOS APRESENTADOS NESTE TRABALHO.

Tabela 3.1. Descritores utilizados para obtenção do Modelo III, atividades observadas e valores preditos na validação cruzada LOO (*outliers* já excluídos).

	VAR	BEHp2	RDF025m	Mor16v	G2v	HGM	HATS1e	R1e+	pIC ₅₀ obs	pIC ₅₀ pred	Resíduos
P4	81	3,800	19,597	0,272	0,162	6,669	0,240	0,164	7,070	6,805	0,265
P8	81	3,690	21,357	0,320	0,156	5,935	0,207	0,167	4,300	5,796	-1,496
P9	92	3,808	20,624	0,331	0,159	6,353	0,196	0,161	7,700	6,982	0,718
P10	70	3,779	18,237	0,418	0,165	4,313	0,138	0,153	6,000	6,423	-0,423
P11	78	3,906	25,102	0,294	0,159	3,416	0,121	0,121	7,520	7,084	0,436
P12	104	3,908	28,213	0,387	0,154	5,300	0,216	0,172	7,400	6,750	0,650
P13	104	3,908	28,545	0,394	0,154	5,288	0,212	0,156	6,210	7,063	-0,853
P14	82	3,906	28,477	0,216	0,156	5,203	0,235	0,193	6,670	6,542	0,128
P15	79	3,891	21,888	0,273	0,158	5,446	0,232	0,147	7,700	6,624	1,076
P16	79	3,891	21,888	0,275	0,158	5,446	0,232	0,147	6,210	6,859	-0,649
P17	81	3,739	18,518	0,382	0,163	6,777	0,263	0,165	4,780	6,219	-1,439
P18	68	3,615	15,911	0,400	0,167	7,111	0,311	0,194	4,300	4,941	-0,641
P19	68	3,658	15,879	0,475	0,165	7,017	0,292	0,189	4,770	5,025	-0,255
P20	68	3,869	19,694	0,351	0,165	7,035	0,277	0,191	7,000	6,794	0,206
P21	68	3,792	19,538	0,328	0,167	5,426	0,176	0,131	6,700	7,105	-0,405
P22	68	3,800	19,809	0,370	0,167	6,718	0,312	0,245	5,600	5,481	0,119
P23	86	3,910	22,139	0,356	0,173	6,180	0,226	0,143	8,000	8,290	-0,290
P24	88	3,910	22,776	0,476	0,159	5,810	0,233	0,155	7,300	6,873	0,427
P25	88	3,908	20,577	0,492	0,158	6,240	0,218	0,158	7,700	7,045	0,655
P26	81	3,809	21,567	0,336	0,158	6,164	0,215	0,157	7,040	6,618	0,422
P27	96	3,808	21,451	0,226	0,158	6,247	0,209	0,145	7,300	7,136	0,164
P28	111	3,808	21,439	0,070	0,158	6,121	0,235	0,167	6,260	7,076	-0,816
P29	78	3,834	21,461	0,463	0,159	6,253	0,221	0,151	6,960	6,652	0,308
P30	80	3,831	21,578	0,320	0,159	6,329	0,212	0,141	7,400	7,078	0,322
P31	96	3,830	21,594	0,258	0,159	6,301	0,225	0,165	7,150	7,014	0,136
P32	96	3,812	26,840	0,636	0,164	5,287	0,197	0,138	6,890	6,664	0,226
P35	96	3,799	19,590	0,248	0,162	6,608	0,245	0,166	7,700	6,866	0,834
P36	92	3,825	19,883	0,473	0,162	6,699	0,245	0,186	6,790	6,595	0,195
P37	106	3,899	20,033	0,225	0,162	6,670	0,243	0,163	7,150	7,867	-0,717
P38	117	3,898	19,873	0,245	0,162	6,656	0,251	0,173	8,000	7,659	0,341
P39	128	3,897	19,734	0,358	0,162	6,584	0,246	0,164	8,000	7,658	0,342
S6	59	3,655	29,114	-0,248	0,149	4,874	0,159	0,085	7,220	5,889	1,331
S7	66	3,683	15,745	-0,033	0,167	7,636	0,259	0,174	7,220	6,713	0,507
S8	88	3,793	20,259	0,165	0,159	6,156	0,224	0,161	7,150	6,733	0,417
S9	112	3,792	21,634	0,116	0,156	6,199	0,204	0,156	7,300	7,184	0,116
S10	136	3,794	25,625	-0,005	0,151	4,951	0,186	0,156	6,700	6,995	-0,295
S11	96	3,709	21,710	-0,239	0,156	5,717	0,196	0,169	7,100	6,663	0,437
S12	88	3,693	19,024	-0,081	0,158	6,292	0,214	0,169	6,700	6,439	0,261
S13	104	3,731	25,241	-0,181	0,153	5,430	0,186	0,154	7,300	6,811	0,489
S4A	54	3,659	13,641	0,155	0,188	8,203	0,292	0,275	7,070	6,127	0,943
G1	67	3,797	19,848	-0,099	0,164	7,298	0,219	0,140	7,480	8,186	-0,706
G7	130	3,797	32,090	-0,407	0,149	4,487	0,170	0,103	7,600	7,948	-0,348
G8	116	3,797	29,084	-0,250	0,150	4,703	0,172	0,107	7,400	7,571	-0,171

G9	123	3,797	23,674	-0,162	0,149	4,866	0,157	0,094	7,520	7,645	-0,125
G10	123	3,797	28,983	0,037	0,149	4,812	0,157	0,107	7,230	7,374	-0,144
G11	102	3,797	25,233	-0,230	0,155	5,490	0,191	0,115	7,600	7,748	-0,148
G12	130	3,797	28,982	-0,247	0,150	4,684	0,171	0,100	7,400	7,783	-0,383
G13	123	3,797	27,818	-0,102	0,152	4,914	0,183	0,101	7,820	7,530	0,290
G14	116	3,797	26,931	-0,030	0,154	5,251	0,198	0,141	7,700	7,166	0,534
G15	157	3,798	32,839	-0,153	0,146	4,223	0,158	0,094	8,150	7,667	0,483
G16	148	3,797	30,213	-0,376	0,147	4,419	0,162	0,093	7,400	8,074	-0,674
G17	139	3,797	32,955	-0,246	0,149	4,635	0,171	0,113	8,000	7,762	0,238
Z1	63	3,737	19,107	0,193	0,163	6,887	0,216	0,153	8,000	6,717	1,283
Z2	84	3,801	24,850	0,073	0,160	6,383	0,206	0,154	6,220	7,549	-1,329
Z3	91	3,856	22,845	0,349	0,158	5,869	0,205	0,159	6,000	7,033	-1,033
Z4	98	3,864	25,252	0,005	0,156	5,347	0,184	0,149	7,400	7,595	-0,195
Z5	98	3,870	22,653	0,137	0,156	4,873	0,168	0,125	6,430	7,528	-1,098
Z6	98	3,852	25,587	0,216	0,156	5,248	0,183	0,151	7,300	7,139	0,161
Z7	138	3,917	34,928	0,212	0,146	4,108	0,165	0,094	8,000	7,658	0,342
W2	62	3,732	18,361	0,337	0,163	6,553	0,239	0,178	5,250	5,961	-0,711
W3	57	3,718	16,440	0,247	0,165	7,549	0,243	0,202	6,660	6,249	0,411
W4	57	3,745	19,201	0,317	0,167	7,218	0,257	0,184	6,740	6,360	0,380
W6	58	3,748	18,912	0,249	0,184	6,345	0,253	0,200	6,895	6,979	-0,084
W7	58	3,725	16,028	0,396	0,165	7,263	0,246	0,205	6,260	5,874	0,386
W8	58	3,750	17,444	0,238	0,167	6,384	0,235	0,165	6,300	6,411	-0,111
W11	76	3,747	18,258	0,216	0,163	6,829	0,221	0,163	6,150	6,829	-0,679
W13	64	3,736	18,825	0,183	0,163	6,877	0,222	0,164	6,600	6,675	-0,075
W15	91	3,744	19,985	0,073	0,159	6,414	0,199	0,165	6,820	6,948	-0,128
W16	78	3,743	20,203	0,067	0,159	6,016	0,207	0,154	6,850	6,589	0,261
E1	144	3,824	28,973	-0,016	0,151	5,132	0,179	0,131	8,000	7,705	0,295
E7	123	3,824	28,567	0,089	0,151	5,107	0,168	0,104	8,000	7,627	0,373
E5B	164	3,849	33,786	0,290	0,145	4,268	0,140	0,074	8,150	7,705	0,445
E21A	134	3,845	33,170	0,027	0,147	4,641	0,153	0,097	6,960	7,941	-0,981
E21B	140	3,842	34,719	-0,155	0,145	4,087	0,157	0,118	8,000	7,527	0,473
E22	80	3,796	20,187	-0,135	0,164	7,282	0,230	0,172	7,480	7,939	-0,459
E23	82	3,838	24,195	0,090	0,157	6,071	0,202	0,127	7,280	7,552	-0,272
L708906	120	3,780	24,795	0,148	0,151	3,278	0,122	0,096	7,000	6,533	0,467
L731927	76	3,777	18,127	0,205	0,159	5,862	0,221	0,191	6,300	5,999	0,301
L731942	100	3,799	23,852	0,410	0,156	5,395	0,219	0,187	5,120	6,011	-0,891

Tabela 3.2. Descritores utilizados para obtenção do Modelo IV, atividades observadas e valores preditos na validação cruzada LOO (*outliers* já excluídos).

	BELe8	RTe	G2e	R1u+	pIC₅₀ obs	pIC₅₀ pred	Resíduos
P4	3,800	0,164	0,162	0,143	7,070	6,781	0,289
P9	3,808	0,161	0,159	0,129	7,700	7,005	0,695
P10	3,779	0,153	0,165	0,121	6,000	6,836	-0,836
P11	3,906	0,121	0,159	0,105	7,520	7,798	-0,278
P12	3,908	0,172	0,154	0,137	7,400	7,534	-0,134
P14	3,906	0,193	0,156	0,154	6,670	7,329	-0,659
P15	3,891	0,147	0,158	0,142	7,700	7,370	0,330
P16	3,891	0,147	0,158	0,142	6,210	7,423	-1,213
P17	3,739	0,165	0,179	0,172	4,780	5,962	-1,182
P18	3,615	0,194	0,184	0,205	4,300	4,935	-0,635
P19	3,658	0,189	0,165	0,200	4,770	5,620	-0,850
P10	3,869	0,191	0,165	0,167	7,000	6,761	0,239
P21	3,792	0,131	0,167	0,127	6,700	6,846	-0,146
P22	3,800	0,245	0,167	0,259	5,600	5,457	0,143
P23	3,910	0,143	0,159	0,137	8,000	7,474	0,526
P24	3,910	0,155	0,173	0,133	7,300	7,087	0,213
P25	3,908	0,158	0,158	0,126	7,700	7,542	0,158
P26	3,809	0,157	0,158	0,141	7,040	6,968	0,072
P27	3,808	0,145	0,158	0,115	7,300	7,196	0,104
P28	3,808	0,167	0,158	0,134	6,260	7,001	-0,741
P29	3,834	0,151	0,159	0,149	6,960	7,020	-0,060
P30	3,831	0,141	0,159	0,126	7,400	7,206	0,194
P31	3,830	0,165	0,173	0,132	7,150	6,683	0,467
P32	3,812	0,138	0,153	0,146	6,890	7,139	-0,249
P33	3,799	0,133	0,153	0,141	7,300	7,120	0,180
P35	3,799	0,166	0,162	0,132	7,700	6,841	0,859
P36	3,825	0,186	0,162	0,148	6,790	6,800	-0,010
P37	3,899	0,163	0,162	0,144	7,150	7,257	-0,107
P38	3,898	0,173	0,162	0,132	8,000	7,266	0,734
P39	3,897	0,164	0,162	0,131	8,000	7,304	0,696
S6	3,655	0,085	0,149	0,086	7,220	7,071	0,149
S8	3,793	0,161	0,159	0,147	7,150	6,803	0,347
S9	3,792	0,156	0,156	0,124	7,300	7,067	0,233
S10	3,794	0,156	0,151	0,125	6,700	7,221	-0,521
S11	3,709	0,169	0,156	0,135	7,100	6,530	0,570
S12	3,693	0,169	0,158	0,134	6,700	6,420	0,280
S13	3,731	0,154	0,153	0,122	7,300	6,865	0,435
G1	3,797	0,140	0,164	0,142	7,480	6,783	0,697
G7	3,797	0,103	0,149	0,109	7,600	7,549	0,051
G8	3,797	0,107	0,150	0,111	7,400	7,502	-0,102
G9	3,797	0,094	0,149	0,086	7,520	7,766	-0,246
G10	3,797	0,107	0,149	0,113	7,230	7,520	-0,290
G11	3,797	0,115	0,155	0,122	7,600	7,252	0,348
G12	3,797	0,100	0,150	0,106	7,400	7,563	-0,163
G13	3,797	0,101	0,152	0,107	7,820	7,481	0,339

G14	3,797	0,141	0,154	0,129	7,700	7,145	0,555
G15	3,798	0,094	0,146	0,089	8,150	7,793	0,357
G16	3,797	0,093	0,147	0,098	7,400	7,734	-0,334
G17	3,797	0,113	0,149	0,120	8,000	7,417	0,583
Z2	3,801	0,154	0,160	0,163	6,220	6,732	-0,512
Z3	3,856	0,159	0,158	0,167	6,000	7,024	-1,024
Z4	3,864	0,149	0,156	0,157	7,400	7,175	0,225
Z5	3,870	0,125	0,156	0,115	6,430	7,637	-1,207
Z6	3,852	0,151	0,156	0,160	7,300	7,090	0,210
Z7	3,917	0,094	0,146	0,099	8,000	8,325	-0,325
W2	3,732	0,178	0,163	0,160	5,250	6,300	-1,050
W3	3,718	0,202	0,165	0,161	6,660	6,035	0,625
W4	3,745	0,184	0,167	0,177	6,740	6,046	0,694
W6	3,748	0,200	0,167	0,159	6,890	6,146	0,744
W7	3,725	0,205	0,165	0,163	6,260	6,066	0,194
W8	3,750	0,165	0,167	0,156	6,300	6,317	-0,017
W11	3,747	0,163	0,163	0,154	6,150	6,435	-0,285
W13	3,736	0,164	0,163	0,161	6,600	6,309	0,291
W15	3,744	0,165	0,159	0,132	6,820	6,674	0,146
W16	3,743	0,154	0,159	0,161	6,850	6,474	0,376
E1	3,824	0,131	0,151	0,101	8,000	7,592	0,408
E7	3,824	0,104	0,151	0,110	8,000	7,604	0,396
E5a	3,850	0,067	0,144	0,071	7,850	8,374	-0,524
E5b	3,849	0,074	0,153	0,079	8,150	7,999	0,151
E9	3,991	0,093	0,156	0,099	8,000	8,445	-0,445
E21b	3,842	0,118	0,145	0,092	8,000	7,952	0,048
E22	3,796	0,172	0,164	0,133	7,480	6,748	0,732
E23	3,838	0,127	0,170	0,124	7,280	6,990	0,290
L708906	3,780	0,096	0,151	0,080	7,000	7,702	-0,702
L731927	3,777	0,191	0,159	0,152	6,300	6,615	-0,315
L731942	3,799	0,187	0,168	0,149	5,120	6,563	-1,443

Tabela 3.3. Descritores utilizados para obtenção do Modelo V, atividades observadas e valores preditos na validação cruzada LOO (*outliers* já excluídos).

	BELm7	BEHy2	BEHe5	RDF060m	RDF030e	Mor16p	G2e	H4v	pIC ₅₀ obs	pIC ₅₀ pred	Resíduos
P4	0,972	3,786	2,938	11,202	6,203	0,363	0,162	0,313	7,070	6,791	0,279
P8	1,100	3,674	3,026	10,603	14,898	0,414	0,156	0,375	4,300	5,649	-1,349
P9	1,078	3,795	3,065	12,188	8,084	0,402	0,159	0,400	7,700	6,737	0,963
P10	0,868	3,764	2,842	8,152	6,436	0,505	0,165	0,295	6,000	6,201	-0,201
P11	1,014	3,833	3,216	8,337	6,658	0,376	0,159	0,472	7,520	7,247	0,273
P12	1,160	3,837	3,252	19,723	9,343	0,478	0,154	0,580	7,400	7,678	-0,278
P13	1,160	3,837	3,252	6,872	11,179	0,468	0,154	0,570	6,210	6,574	-0,364
P14	1,126	3,834	3,279	7,737	11,488	0,296	0,156	0,410	6,670	7,085	-0,415
P15	1,010	3,829	3,159	15,872	16,659	0,360	0,158	0,359	7,700	7,228	0,472
P16	1,010	3,829	3,159	15,872	16,659	0,362	0,158	0,359	6,210	7,431	-1,221
P17	0,920	3,734	2,903	12,280	6,445	0,454	0,179	0,302	4,780	6,155	-1,375
P18	0,826	3,602	2,844	9,230	6,784	0,441	0,184	0,209	4,300	4,932	-0,632
P19	0,846	3,665	2,844	8,538	5,865	0,509	0,165	0,232	4,770	5,782	-1,012
P10	0,794	3,764	2,851	8,709	7,088	0,416	0,165	0,224	7,000	6,563	0,437
P21	0,658	3,681	2,851	11,151	4,235	0,411	0,167	0,300	6,700	6,568	0,132
P22	0,794	3,691	2,843	9,800	5,235	0,430	0,167	0,252	5,600	6,191	-0,591
P23	1,008	3,836	3,183	13,345	6,757	0,444	0,159	0,476	8,000	7,497	0,503
P24	0,989	3,836	3,196	11,499	9,836	0,589	0,173	0,488	7,300	6,417	0,883
P25	1,048	3,836	3,238	11,124	8,867	0,586	0,158	0,453	7,700	7,055	0,645
P26	1,023	3,795	3,135	11,451	6,164	0,406	0,158	0,415	7,040	7,146	-0,106
P27	1,008	3,795	3,127	10,361	6,901	0,301	0,158	0,492	7,300	7,076	0,224
P28	1,082	3,794	3,119	12,499	5,557	0,126	0,158	0,478	6,260	7,439	-1,179
P29	1,001	3,812	3,052	10,777	9,583	0,555	0,159	0,405	6,960	6,631	0,329
P30	0,985	3,810	3,057	10,929	6,894	0,388	0,159	0,412	7,400	7,073	0,327
P31	1,082	3,810	3,062	10,402	7,058	0,334	0,173	0,421	7,150	6,264	0,886
P32	0,933	3,794	2,949	15,476	9,112	0,744	0,153	0,555	6,890	6,609	0,281
P33	0,946	3,787	2,940	14,421	8,619	0,470	0,153	0,606	7,300	6,735	0,565
P35	0,939	3,787	2,957	13,612	6,371	0,321	0,162	0,326	7,700	7,117	0,583
P36	0,907	3,801	2,952	11,393	7,024	0,567	0,162	0,314	6,790	6,796	-0,006
P37	1,110	3,833	3,318	11,355	6,187	0,316	0,162	0,320	7,150	7,752	-0,602
P38	1,108	3,833	3,315	11,825	5,717	0,338	0,162	0,311	8,000	7,649	0,351
P39	1,082	3,833	3,313	14,520	6,359	0,421	0,162	0,359	8,000	7,731	0,269
S6	0,710	3,669	2,851	19,614	13,673	-0,201	0,149	0,688	7,220	7,698	-0,478
S7	0,711	3,697	2,952	7,989	4,137	-0,008	0,167	0,202	7,220	7,206	0,014
S8	1,071	3,789	2,960	12,614	4,732	0,208	0,159	0,380	7,150	7,056	0,094
S9	1,082	3,789	3,123	10,095	4,041	0,173	0,156	0,393	7,300	7,359	-0,059
S10	1,186	3,791	3,317	12,802	10,576	0,048	0,151	0,593	6,700	7,466	-0,766
S11	1,082	3,722	3,057	10,656	11,208	-0,236	0,156	0,393	7,100	6,872	0,228
S12	1,082	3,707	2,964	11,256	6,545	-0,070	0,158	0,366	6,700	6,672	0,028
S13	1,249	3,741	3,113	13,595	14,155	-0,119	0,153	0,528	7,300	6,462	0,838
S4a	0,775	3,674	2,929	6,331	4,791	0,166	0,169	0,190	7,070	6,307	0,763
G1	0,932	3,792	2,978	6,438	3,583	-0,065	0,164	0,327	7,480	7,310	0,170
G7	1,365	3,793	3,332	13,865	10,216	-0,368	0,149	0,742	7,600	7,587	0,013
G8	1,349	3,793	3,227	10,968	10,372	-0,208	0,150	0,643	7,400	6,981	0,419
G9	1,352	3,793	3,319	18,480	17,896	-0,143	0,149	0,722	7,520	7,169	0,351
G10	1,249	3,793	3,328	23,581	15,668	0,098	0,149	0,735	7,230	7,738	-0,508
G11	1,082	3,793	3,195	12,080	10,731	-0,191	0,155	0,480	7,600	7,708	-0,108

G12	1,249	3,793	3,333	13,198	14,812	-0,228	0,150	0,685	7,400	7,384	0,016
G13	1,099	3,793	3,273	10,970	11,097	-0,079	0,152	0,619	7,820	7,478	0,342
G14	1,082	3,793	3,221	11,415	9,793	-0,014	0,154	0,554	7,700	7,462	0,238
G15	1,355	3,793	3,486	22,139	20,348	-0,069	0,146	0,839	8,150	7,458	0,692
G16	1,353	3,793	3,471	13,917	18,016	-0,364	0,147	0,726	7,400	7,522	-0,122
G17	1,119	3,793	3,309	15,320	13,691	-0,184	0,149	0,716	8,000	7,821	0,179
Z2	0,980	3,798	3,052	5,121	3,329	0,128	0,160	0,478	6,220	7,061	-0,841
Z3	1,080	3,855	3,104	4,485	6,212	0,393	0,158	0,449	6,000	6,860	-0,860
Z4	1,082	3,862	3,114	6,534	5,054	0,058	0,156	0,494	7,400	7,555	-0,155
Z5	1,082	3,867	3,121	6,757	4,159	0,183	0,156	0,467	6,430	7,638	-1,208
Z6	1,082	3,849	3,105	4,324	5,203	0,269	0,156	0,501	7,300	6,913	0,387
Z7	1,300	3,881	3,437	15,897	10,800	0,298	0,146	0,788	8,000	7,841	0,159
W2	0,943	3,727	3,078	2,831	5,839	0,373	0,163	0,278	5,250	6,209	-0,959
W3	0,856	3,716	3,002	3,302	5,506	0,266	0,165	0,227	6,660	6,259	0,401
W4	0,835	3,719	2,990	4,138	4,263	0,374	0,167	0,259	6,740	6,202	0,538
W6	0,850	3,723	2,906	4,838	5,090	0,338	0,167	0,274	6,890	6,048	0,842
W7	0,863	3,723	2,922	5,318	7,166	0,417	0,165	0,263	6,260	5,948	0,312
W8	0,822	3,721	2,921	2,168	6,394	0,289	0,167	0,248	6,300	5,942	0,358
W11	0,877	3,740	3,020	4,658	4,118	0,262	0,163	0,365	6,150	6,576	-0,426
W13	0,913	3,733	3,018	5,030	3,566	0,240	0,163	0,334	6,600	6,538	0,062
W15	1,081	3,741	3,095	2,925	3,726	0,113	0,159	0,496	6,820	6,216	0,604
W16	1,000	3,741	3,106	5,186	3,843	0,100	0,159	0,367	6,850	6,843	0,007
E1	1,082	3,794	3,252	16,862	16,381	0,022	0,151	0,557	8,000	7,588	0,412
E7	1,082	3,793	3,252	21,638	16,429	0,173	0,151	0,592	8,000	7,713	0,287
E5a	1,364	3,845	3,500	19,742	20,076	0,039	0,144	0,790	7,850	7,709	0,141
E5b	1,348	3,843	3,460	18,573	19,369	0,357	0,153	0,672	8,150	6,874	1,276
E8a	1,294	3,899	3,502	24,568	24,240	-0,055	0,145	0,781	7,520	8,659	-1,139
E8b	1,294	3,894	3,498	21,185	24,747	-0,045	0,146	0,731	7,350	8,312	-0,962
E9	1,294	3,903	3,496	25,027	23,200	-0,243	0,156	0,768	8,000	8,735	-0,735
E21b	1,294	3,839	3,363	17,945	25,241	-0,132	0,145	0,645	8,000	7,349	0,651
E22	0,894	3,793	2,980	6,655	2,980	-0,099	0,164	0,320	7,480	7,555	-0,075
E23	1,081	3,838	3,215	9,361	8,053	0,127	0,170	0,558	7,280	6,945	0,335
L708906	1,082	3,775	3,217	9,156	1,569	0,208	0,151	0,651	7,000	7,428	-0,428
L731927	1,071	3,776	3,047	5,277	7,458	0,228	0,159	0,338	6,300	6,435	-0,135
L731942	1,082	3,803	3,071	6,902	6,659	0,436	0,168	0,503	5,120	6,187	-1,067

Tabela 3.4. Descritores utilizados para obtenção do Modelo VI, atividade observadas e valores preditos na validação cruzada LOO (*outliers* já excluídos).

	RDF065u	VAR	HGM	BEHe5	BELm5	G1e	Mor16v	R1e+	BEHv2	qXMull_I	pIC ₅₀ obs	pIC ₅₀ pred	Resíduos
P4	10,531	81	6,669	2,938	1,131	0,162	0,272	0,164	3,786	-0,548	7,070	7,133	-0,063
P8	17,284	81	5,935	3,026	1,372	0,156	0,320	0,167	3,674	-0,549	4,300	5,949	-1,649
P9	13,506	92	6,353	3,065	1,164	0,159	0,331	0,161	3,795	-0,548	7,700	7,335	0,365
P12	17,298	104	5,300	3,252	1,447	0,154	0,387	0,172	3,837	-0,548	7,400	6,910	0,490
P13	22,880	104	5,288	3,252	1,447	0,154	0,394	0,156	3,837	-0,547	6,210	6,882	-0,672
P14	20,903	82	5,203	3,279	1,457	0,156	0,216	0,193	3,834	-0,549	6,670	5,981	0,689
P15	10,627	79	5,446	3,159	1,279	0,158	0,273	0,147	3,829	-0,548	7,700	6,988	0,712
P16	10,627	79	5,446	3,159	1,279	0,158	0,275	0,147	3,829	-0,548	6,210	7,129	-0,919
P17	9,727	81	6,777	2,903	1,125	0,179	0,382	0,165	3,734	-0,546	4,780	5,425	-0,645
P18	9,378	68	7,111	2,844	1,153	0,167	0,400	0,194	3,602	-0,545	4,300	5,042	-0,742
P19	9,045	68	7,017	2,844	1,188	0,165	0,475	0,189	3,665	-0,545	4,770	5,444	-0,674
P10	10,475	68	7,035	2,851	1,135	0,165	0,351	0,191	3,764	-0,548	7,000	6,199	0,801
P22	11,056	68	6,718	2,843	1,130	0,167	0,370	0,245	3,691	-0,547	5,600	4,674	0,926
P23	11,386	86	6,180	3,183	1,222	0,159	0,356	0,143	3,836	-0,547	8,000	7,525	0,475
P24	9,020	88	5,810	3,196	1,179	0,159	0,476	0,155	3,836	-0,547	7,300	7,402	-0,102
P25	14,391	88	6,240	3,238	1,315	0,158	0,492	0,158	3,836	-0,549	7,700	7,103	0,597
P26	18,677	81	6,164	3,135	1,418	0,158	0,336	0,157	3,795	-0,550	7,040	6,330	0,710
P27	13,185	96	6,247	3,127	1,378	0,158	0,226	0,145	3,795	-0,548	7,300	7,253	0,047
P28	14,823	111	6,121	3,119	1,350	0,171	0,070	0,167	3,794	-0,548	6,260	6,260	0,000
P29	13,620	78	6,253	3,052	1,410	0,159	0,463	0,151	3,812	-0,548	6,960	6,405	0,555
P30	10,592	80	6,329	3,057	1,332	0,159	0,320	0,141	3,810	-0,548	7,400	7,071	0,329
P31	11,826	96	6,301	3,062	1,263	0,159	0,258	0,165	3,810	-0,548	7,150	7,414	-0,264
P32	22,086	96	5,287	2,949	1,127	0,153	0,636	0,138	3,794	-0,548	6,890	6,979	-0,089
P33	17,484	78	5,106	2,940	1,130	0,153	0,398	0,133	3,787	-0,548	7,300	6,882	0,418
P35	12,339	96	6,608	2,957	1,127	0,162	0,248	0,166	3,787	-0,547	7,700	7,303	0,397
P36	12,256	92	6,699	2,952	1,126	0,162	0,473	0,186	3,801	-0,549	6,790	7,059	-0,269
P37	10,194	106	6,670	3,318	1,425	0,162	0,225	0,163	3,833	-0,547	7,150	7,584	-0,434
P38	11,338	117	6,656	3,315	1,416	0,162	0,245	0,173	3,833	-0,547	8,000	7,547	0,453
P39	12,024	128	6,584	3,313	1,407	0,162	0,358	0,164	3,833	-0,546	8,000	7,652	0,348
S6	26,491	59	4,874	2,851	0,986	0,149	-0,248	0,085	3,669	-0,549	7,220	6,886	0,334
S7	6,872	66	7,636	2,952	1,082	0,167	-0,033	0,174	3,697	-0,512	7,220	6,426	0,794
S8	9,062	88	6,156	2,960	1,082	0,159	0,165	0,161	3,789	-0,515	7,150	7,479	-0,329
S9	13,950	112	6,199	3,123	1,306	0,156	0,116	0,156	3,789	-0,512	7,300	7,681	-0,381
S10	31,761	136	4,951	3,317	1,405	0,151	-0,005	0,156	3,791	-0,548	6,700	7,584	-0,884
S11	17,922	96	5,717	3,057	1,385	0,156	-0,239	0,169	3,722	-0,549	7,100	6,680	0,420
S12	19,799	88	6,292	2,964	1,249	0,158	-0,081	0,169	3,707	-0,549	6,700	6,587	0,113
S13	22,057	104	5,430	3,113	1,391	0,153	-0,181	0,154	3,741	-0,550	7,300	7,045	0,255
G1	4,353	67	7,298	2,978	1,087	0,164	-0,099	0,140	3,792	-0,463	7,480	7,527	-0,047
G7	20,954	130	4,487	3,332	1,512	0,149	-0,407	0,103	3,793	-0,481	7,600	8,020	-0,420
G8	24,395	116	4,703	3,227	1,410	0,150	-0,250	0,107	3,793	-0,466	7,400	7,468	-0,068
G9	28,646	123	4,866	3,319	1,505	0,149	-0,162	0,094	3,793	-0,547	7,520	7,954	-0,434
G10	29,472	123	4,812	3,328	1,520	0,149	0,037	0,107	3,793	-0,482	7,230	7,101	0,129
G11	14,790	102	5,490	3,195	1,397	0,155	-0,230	0,115	3,793	-0,465	7,600	7,421	0,179
G12	23,948	130	4,684	3,333	1,504	0,150	-0,247	0,100	3,793	-0,456	7,400	7,583	-0,183
G13	20,305	123	4,914	3,273	1,485	0,152	-0,102	0,101	3,793	-0,463	7,820	7,357	0,463
G14	18,740	116	5,251	3,221	1,356	0,154	-0,030	0,141	3,793	-0,464	7,700	7,165	0,535
G15	32,398	157	4,223	3,486	1,520	0,146	-0,153	0,094	3,793	-0,459	8,150	7,839	0,311

G17	13,581	139	4,635	3,309	1,531	0,149	-0,246	0,113	3,793	-0,464	8,000	8,124	-0,124
Z1	11,937	63	6,887	3,014	1,082	0,163	0,193	0,153	3,732	-0,656	8,000	7,164	0,836
Z2	13,005	84	6,383	3,052	1,175	0,160	0,073	0,154	3,798	-0,548	6,220	7,534	-1,314
Z3	10,537	91	5,869	3,104	1,325	0,158	0,349	0,159	3,855	-0,500	6,000	7,198	-1,198
Z4	12,651	98	5,347	3,114	1,340	0,156	0,005	0,149	3,862	-0,463	7,400	7,329	0,071
Z5	14,839	98	4,873	3,121	1,346	0,156	0,137	0,125	3,867	-0,498	6,430	7,393	-0,963
Z6	12,413	98	5,248	3,105	1,293	0,156	0,216	0,151	3,849	-0,482	7,300	7,153	0,147
Z7	26,262	138	4,108	3,437	1,498	0,146	0,212	0,094	3,881	-0,463	8,000	7,993	0,007
W2	18,812	62	6,553	3,078	1,082	0,163	0,337	0,178	3,727	-0,550	5,250	6,028	-0,778
W3	14,379	57	7,549	3,002	1,116	0,165	0,247	0,202	3,716	-0,660	6,660	6,613	0,047
W4	9,855	57	7,218	2,990	1,082	0,167	0,317	0,184	3,719	-0,653	6,740	6,538	0,202
W6	6,623	58	6,345	2,906	1,082	0,167	0,249	0,200	3,723	-0,657	6,890	6,116	0,775
W7	8,509	58	7,263	2,922	1,121	0,165	0,396	0,205	3,723	-0,663	6,260	6,596	-0,336
W8	6,215	58	6,384	2,921	1,082	0,167	0,238	0,165	3,721	-0,656	6,300	6,573	-0,273
W11	13,306	76	6,829	3,020	1,082	0,163	0,216	0,163	3,740	-0,558	6,150	6,906	-0,756
W13	11,966	64	6,877	3,018	1,081	0,163	0,183	0,164	3,733	-0,550	6,600	6,643	-0,043
W15	16,552	91	6,414	3,095	1,287	0,159	0,073	0,165	3,741	-0,550	6,820	6,848	-0,028
W16	11,453	78	6,016	3,106	1,318	0,159	0,067	0,154	3,741	-0,550	6,850	6,609	0,241
E1	27,131	144	5,132	3,252	1,359	0,151	-0,016	0,131	3,794	-0,467	8,000	7,676	0,324
E7	30,617	123	5,107	3,252	1,376	0,151	0,089	0,104	3,793	-0,467	8,000	7,145	0,855
E5a	47,582	157	4,210	3,500	1,538	0,144	0,004	0,067	3,845	-0,468	7,850	7,886	-0,036
E5b	45,640	164	4,268	3,460	1,536	0,145	0,290	0,074	3,843	-0,466	8,150	7,456	0,694
E8a	47,811	132	4,098	3,502	1,528	0,145	-0,096	0,096	3,899	-0,467	7,520	7,546	-0,026
E8b	42,117	136	4,441	3,498	1,499	0,146	-0,043	0,106	3,894	-0,467	7,350	7,899	-0,549
E9	37,472	136	4,585	3,496	1,423	0,147	-0,270	0,093	3,903	-0,465	8,000	8,691	-0,691
E21a	39,371	134	4,641	3,457	1,528	0,147	0,027	0,097	3,842	-0,461	6,960	7,438	-0,478
E21b	37,156	140	4,087	3,363	1,519	0,145	-0,155	0,118	3,839	-0,459	8,000	7,421	0,579
E22	6,420	80	7,282	2,980	1,082	0,164	-0,135	0,172	3,793	-0,463	7,480	7,492	-0,012
E23	17,451	82	6,071	3,215	1,359	0,157	0,090	0,127	3,838	-0,463	7,280	6,975	0,305
L708906	27,098	120	3,278	3,217	1,369	0,151	0,148	0,096	3,775	-0,549	7,000	6,564	0,436
L731927	13,001	76	5,862	3,047	1,370	0,159	0,205	0,191	3,776	-0,660	6,300	6,857	-0,557

Tabela 3.5. Descritores derivados da sonda Zn^{+2} utilizados para obtenção do Modelo XI e as atividade observadas dos compostos do modelo, organizados em ordem crescente (*outliers* já excluídos).

	20.17.19.Zn2+.C	14.15.21.Zn2+.LJ	pIC ₅₀
P8	-0,609	-0,809	4,300
P19	-0,336	-0,689	4,770
P17	-1,247	-0,828	4,780
L731942	4,704	-0,683	5,120
Z3	-0,281	-0,810	6,000
P10	-0,426	-0,760	6,000
W11	1,485	-0,729	6,150
P13	-0,935	0,379	6,210
P16	0,279	-0,706	6,210
Z2	1,255	-0,697	6,220
W7	2,568	-0,708	6,260
P28	-0,444	-0,745	6,260
W8	1,733	-0,690	6,300
L731927	0,962	1,406	6,300
Z5	1,127	1,397	6,430
W13	0,490	-0,699	6,600
P14	-0,035	-0,787	6,670
S10	-1,106	-0,836	6,700
S12	0,169	-0,725	6,700
P21	-1,084	-0,835	6,700
W4	1,074	0,292	6,740
P36	0,048	-0,739	6,790
W15	0,560	-0,746	6,820
W16	0,396	-0,688	6,850
P32	-0,062	-0,751	6,890
W6	-0,187	-0,583	6,895
E21A	0,189	1,405	6,960
P29	0,151	-0,723	6,960
P20	0,247	-0,723	7,000
S4A	1,020	-0,714	7,070
P4	0,032	-0,577	7,070
S11	-0,077	-0,717	7,100
S8	0,217	-0,839	7,150
P31	-0,414	-0,784	7,150
P37	-0,206	-0,756	7,150
S6	-0,956	-0,741	7,220
S7	1,141	-0,698	7,220
G10	0,294	-0,438	7,230
E23	-1,180	-0,317	7,280
S9	1,443	-0,700	7,300
S13	-1,031	1,413	7,300
Z6	0,855	-0,990	7,300
P24	-1,068	-0,716	7,300
P27	0,280	-0,631	7,300

P33	-0,714	-0,801	7,300
E8B	-0,327	-0,645	7,350
G8	1,535	1,388	7,400
G12	-1,398	1,392	7,400
G16	-0,121	1,402	7,400
Z4	-0,181	1,378	7,400
P12	-0,692	-0,794	7,400
P30	-1,209	-0,917	7,400
E22	-1,275	1,376	7,480
G1	-0,217	1,387	7,480
E8A	-1,015	1,330	7,520
G9	1,679	1,389	7,520
P11	1,446	-0,684	7,520
G7	1,466	1,395	7,600
G11	-0,217	1,394	7,600
G14	0,190	1,371	7,700
P25	-1,179	-0,819	7,700
P35	-0,532	1,411	7,700
G13	-1,011	1,381	7,820
E5A	0,668	1,376	7,850
E1	0,220	1,397	8,000
E21B	-0,881	1,378	8,000
E7	-0,050	1,401	8,000
E9	0,421	1,384	8,000
G17	-1,015	1,402	8,000
Z1	0,218	-0,711	8,000
Z7	-0,105	1,379	8,000
P23	-1,040	-0,723	8,000
P38	-1,086	-0,825	8,000
E5B	-0,830	1,378	8,150
G15	-0,154	1,390	8,150

Tabela 3.6. Descritores derivados da sonda Ar(C-H) utilizados para obtenção do Modelo XI e as atividade observadas dos compostos do modelo, organizados em ordem crescente (*outliers* já excluídos) – **primeira parte.**

	24.13.12	18.18.19	21.15.16	26.18.14	19.21.24	20.13.16	pIC ₅₀
	.Ar(C-H).LJ	.Ar(C-H).LJ	.Ar(C-H).LJ	.Ar(C-H).LJ	.Ar(C-H).LJ	.Ar(C-H).LJ	
P8	-0,446	0,621	1,345	-1,444	-1,208	0,899	4,300
P19	-1,122	0,534	-0,077	-1,474	-1,145	0,875	4,770
P17	-0,788	0,487	2,357	-1,101	-1,248	0,097	4,780
L731942	-1,183	-2,208	2,237	0,822	0,894	0,879	5,120
Z3	-1,106	-1,578	-0,470	-1,378	1,011	0,900	6,000
P10	-1,127	0,118	0,232	0,781	-1,314	0,877	6,000
W11	1,044	-2,598	0,982	0,918	-1,237	0,914	6,150
P13	0,979	-1,431	0,529	-1,085	-1,243	0,941	6,210
P16	-0,275	0,210	0,554	-0,478	0,830	0,487	6,210
Z2	-1,115	-0,267	-0,956	0,891	-1,058	-1,047	6,220
W7	0,997	-1,054	2,054	0,846	-0,304	0,895	6,260
P28	-1,107	0,440	-0,882	-1,411	0,830	-1,236	6,260
W8	0,977	-1,361	0,732	0,852	0,079	-0,244	6,300
L731927	-1,106	-1,804	-0,589	-1,396	0,928	-0,787	6,300
Z5	1,047	0,851	2,298	0,837	0,825	0,880	6,430
W13	1,085	0,031	2,735	0,778	0,932	0,927	6,600
P14	0,265	0,039	1,226	0,783	-1,307	0,905	6,670
S10	-0,760	0,639	-1,016	-0,863	-0,955	-1,265	6,700
S12	0,973	0,095	-0,341	0,538	0,932	0,912	6,700
P21	0,951	0,638	-0,525	0,777	-1,091	-1,207	6,700
W4	-0,820	-0,594	1,423	0,820	0,905	0,946	6,740
P36	-1,112	-0,240	-0,649	-0,739	-0,737	-0,339	6,790
W15	-1,106	-2,274	-1,254	0,872	0,978	-1,355	6,820
W16	-0,942	-1,611	-0,921	0,953	-1,267	-0,477	6,850
P32	0,966	-0,071	0,690	-0,909	0,853	0,925	6,890
W6	-1,095	-0,040	-0,884	0,834	0,932	0,901	6,895
E21A	0,995	0,079	0,450	-1,422	0,836	0,970	6,960
P29	0,966	0,440	-0,555	0,755	0,858	-1,229	6,960
P20	0,979	0,772	-0,528	0,251	-1,096	-0,382	7,000
S4A	0,951	0,891	0,910	0,748	-1,265	0,881	7,070
P4	-1,133	-0,522	1,145	-1,312	0,920	0,988	7,070
S11	-0,972	0,235	-0,409	0,818	0,853	-0,573	7,100
S8	-1,013	0,571	-1,100	0,754	-0,657	-1,269	7,150
P31	0,978	0,512	-0,360	-1,376	0,891	-0,575	7,150
P37	0,944	0,762	0,121	0,768	0,610	0,894	7,150
S6	0,986	0,756	1,270	0,764	-1,260	0,873	7,220
S7	-1,126	0,224	-0,289	0,768	-1,168	0,920	7,220
G10	-1,101	0,654	0,230	-0,621	0,873	-1,324	7,230
E23	0,978	1,176	-0,543	0,762	-1,271	-1,417	7,280
S9	0,977	-0,202	-0,966	-1,323	0,820	-1,114	7,300
S13	0,962	0,606	-0,454	-1,515	-1,304	0,685	7,300
Z6	-1,126	-1,884	-1,366	0,895	0,890	0,915	7,300
P24	0,993	-2,620	-1,197	0,866	-1,349	0,858	7,300
P27	0,985	0,133	1,055	0,773	0,878	-0,997	7,300
P33	-1,128	0,451	-0,195	0,769	0,921	-1,216	7,300
E8B	-1,149	-0,654	-0,444	-1,185	-1,268	0,913	7,350

G8	0,974	0,834	-0,104	0,687	0,864	0,334	7,400
G12	-0,977	0,764	-0,525	-1,604	0,815	-1,345	7,400
G16	0,986	-0,536	0,243	0,797	0,915	0,866	7,400
Z4	-1,109	-0,899	-0,587	0,835	-1,278	-1,114	7,400
P12	1,061	0,580	-0,455	0,772	-1,120	0,883	7,400
P30	0,980	0,978	1,008	0,793	-1,104	0,864	7,400
E22	0,953	1,006	-0,063	-1,119	-1,253	-1,355	7,480
G1	0,973	0,480	0,142	-1,256	0,850	0,878	7,480
E8A	-1,136	0,888	-1,120	-1,458	-1,373	-1,301	7,520
G9	-0,547	-0,104	-0,419	0,752	-0,541	-1,163	7,520
P11	0,971	0,588	-0,378	0,769	0,827	-1,271	7,520
G7	1,001	-0,703	-0,088	-0,946	0,892	0,865	7,600
G11	0,990	0,306	0,638	0,754	0,827	0,878	7,600
G14	-1,040	0,810	0,177	0,752	0,862	-0,521	7,700
P25	0,979	0,633	-0,570	0,776	-1,259	-1,312	7,700
P35	-1,113	-0,539	-1,279	-1,010	0,853	0,902	7,700
G13	0,939	0,985	0,236	-1,462	0,947	-1,327	7,820
E5A	1,044	1,417	0,490	0,869	0,847	0,924	7,850
E1	1,000	0,274	0,288	0,756	0,832	0,885	8,000
E21B	-1,125	1,006	-1,235	0,747	0,839	-1,374	8,000
E7	0,982	0,741	-0,387	0,795	0,832	-1,105	8,000
E9	0,984	0,432	-0,582	0,751	0,818	-1,267	8,000
G17	-0,947	1,088	-1,042	0,762	0,916	-1,386	8,000
Z1	0,992	0,158	-0,816	0,841	0,956	-0,156	8,000
Z7	0,080	2,051	0,935	0,791	1,002	0,893	8,000
P23	1,059	-2,604	-1,242	0,962	-1,362	-1,612	8,000
P38	0,393	0,819	-0,919	0,771	-1,189	-1,299	8,000
E5B	-1,137	0,896	-1,149	0,763	0,841	-1,383	8,150
G15	0,955	0,741	1,309	0,786	0,825	-0,684	8,150

Tabela 3.7. Descritores derivados da sonda Ar(C-H) utilizados para obtenção do Modelo XI e as atividade observadas dos compostos do modelo, organizados em ordem crescente (*outliers* já excluídos). – **segunda parte.**

	22.16.9	20.18.17	15.15.18	26.20.19	14.23.23	18.24.13	pIC ₅₀
	.Ar(C-H).LJ	.Ar(C-H).LJ	.Ar(C-H).LJ	.Ar(C-H).LJ	.Ar(C-H).LJ	.Ar(C-H).LJ	
P8	-0,883	1,043	1,002	-2,171	1,305	-1,191	4,300
P19	-0,897	0,136	0,705	-2,001	0,624	-1,402	4,770
P17	-0,809	-0,447	0,706	-2,630	1,631	-0,753	4,780
L731942	-1,156	-0,346	-0,743	-2,841	1,647	0,862	5,120
Z3	-1,074	-0,400	-0,342	-2,107	1,629	0,840	6,000
P10	-1,110	0,246	0,759	0,403	1,387	-1,376	6,000
W11	1,098	-0,097	0,796	0,406	-0,771	-0,112	6,150
P13	1,037	-0,227	0,377	0,370	-0,774	-1,299	6,210
P16	-1,083	0,548	-0,703	0,436	-0,294	-1,425	6,210
Z2	-1,080	0,081	0,506	0,630	-0,791	-1,377	6,220
W7	-1,113	1,043	0,515	0,394	1,636	0,886	6,260
P28	-1,078	-0,236	0,913	0,542	0,310	0,824	6,260
W8	1,124	0,108	0,103	0,546	-0,785	0,874	6,300
L731927	-1,083	0,207	0,687	0,447	1,736	0,758	6,300
Z5	-1,094	-0,315	-0,327	0,546	-0,799	-1,255	6,430
W13	1,090	0,540	-0,001	0,795	-0,766	0,858	6,600
P14	1,035	0,320	-0,664	0,385	-0,296	-1,377	6,670
S10	1,056	-0,371	0,646	0,385	1,642	0,830	6,700
S12	1,046	1,103	-0,324	0,374	-0,739	-1,314	6,700
P21	0,175	1,103	1,043	0,335	0,879	-1,252	6,700
W4	-1,119	0,469	0,252	0,347	-0,795	0,827	6,740
P36	-1,040	-0,002	0,381	0,484	-0,679	0,794	6,790
W15	-1,069	1,103	-0,519	0,500	-0,791	-1,319	6,820
W16	1,083	0,297	0,912	0,588	-0,775	-0,303	6,850
P32	-1,081	0,907	-0,321	0,380	-0,314	0,793	6,890
W6	1,046	3,112	-2,976	0,447	1,754	-0,791	6,895
E21A	1,059	0,247	-0,780	0,357	-0,303	-0,829	6,960
P29	-1,079	0,494	0,459	0,401	-0,363	-1,408	6,960
P20	-0,992	-0,186	-0,565	0,363	0,093	-1,451	7,000
S4A	-0,407	1,176	-0,237	0,487	-0,791	0,681	7,070
P4	-1,081	-2,828	0,393	0,374	-0,815	0,893	7,070
S11	-1,085	-0,208	-0,169	0,386	-0,671	-1,287	7,100
S8	0,179	-0,263	0,988	0,419	1,661	-0,637	7,150
P31	1,076	-0,366	0,457	0,453	0,263	0,870	7,150
P37	1,041	0,649	0,386	0,388	-0,344	-1,349	7,150
S6	1,055	1,211	0,931	0,370	-0,782	0,806	7,220
S7	1,019	0,055	0,721	0,447	-0,801	0,787	7,220
G10	1,029	-0,217	-0,265	0,354	-0,978	0,820	7,230
E23	1,058	-0,348	1,134	-2,518	-0,803	0,815	7,280
S9	-1,084	0,208	0,738	0,430	-0,526	0,829	7,300
S13	1,068	-0,309	0,798	0,418	-0,998	0,794	7,300
Z6	-1,083	0,005	-0,450	0,425	-0,823	0,885	7,300
P24	1,130	-1,769	-0,319	-1,271	-0,774	0,881	7,300
P27	1,029	0,247	-0,689	0,367	-0,642	-1,412	7,300
P33	1,040	-0,248	0,894	0,500	0,665	0,883	7,300
E8B	-1,126	-2,836	-3,003	0,364	-0,809	0,908	7,350

G8	1,038	-0,442	0,183	-1,797	-0,911	0,796	7,400
G12	0,615	-0,433	0,956	0,335	1,729	-0,448	7,400
G16	-0,668	0,582	-0,488	0,343	1,724	0,960	7,400
Z4	-1,077	-0,256	-0,345	0,538	-0,794	-1,321	7,400
P12	1,152	-0,170	0,860	0,523	1,009	0,809	7,400
P30	1,028	-0,377	0,634	0,380	1,036	0,798	7,400
E22	1,062	-0,355	1,124	0,445	-0,800	0,842	7,480
G1	1,065	0,473	-0,416	0,438	-0,787	-1,287	7,480
E8A	-1,118	-0,558	-0,795	0,379	0,606	0,849	7,520
G9	-1,102	0,201	-0,420	0,486	-0,343	0,808	7,520
P11	-1,087	0,209	-0,386	0,602	-0,779	-1,413	7,520
G7	-1,087	-0,005	-0,315	0,329	-0,825	0,888	7,600
G11	1,031	-0,280	-0,125	0,414	-0,792	-1,361	7,600
G14	-0,749	0,318	-0,109	0,393	0,455	-0,235	7,700
P25	1,059	-0,319	0,774	0,406	1,624	0,906	7,700
P35	1,043	-0,107	-2,885	0,406	-0,804	0,904	7,700
G13	1,000	-0,389	0,500	0,431	1,666	0,810	7,820
E5A	1,051	0,111	-0,145	0,471	-0,855	-0,873	7,850
E1	0,294	-0,288	-0,634	0,403	0,093	0,401	8,000
E21B	1,025	-0,297	-0,856	-1,395	-0,827	0,858	8,000
E7	1,030	-0,278	-0,598	0,371	-0,344	0,162	8,000
E9	0,415	0,372	-0,545	0,385	-0,676	-1,467	8,000
G17	0,254	-0,461	-0,251	0,376	1,731	0,816	8,000
Z1	-1,097	0,060	0,435	0,522	-0,597	0,913	8,000
Z7	-1,077	0,177	-0,392	0,507	-0,329	0,799	8,000
P23	0,195	-1,960	-0,313	0,439	-0,775	-1,348	8,000
P38	1,059	-0,443	0,709	0,370	1,622	0,797	8,000
E5B	0,591	-0,651	-0,037	0,458	-0,954	0,838	8,150
G15	-0,065	-0,228	-0,666	0,421	-0,984	0,784	8,150
