

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FARMACÊUTICAS

LUANA JANAÍNA DE CAMPOS

**ESTUDOS QSAR DE DERIVADOS DE PIRIMIDONAS, PIRIMIDINAS E
PIRIDOPIRAZINAS CARBOXAMIDAS INIBIDORAS DA HIV 1 INTEGRASE**

CASCADEL - PARANÁ

2015

LUANA JANAÍNA DE CAMPOS

**ESTUDOS QSAR DE DERIVADOS DE PIRIMIDONAS, PIRIMIDINAS E
PIRIDOPIRAZINAS CARBOXAMIDAS INIBIDORAS DA HIV 1 INTEGRASE**

Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Ciências Farmacêuticas, da Universidade Estadual do Oeste do Paraná, como pré-requisito para obtenção do título de Mestre em Ciências Farmacêuticas. Área de concentração: Fármacos e Medicamentos

Orientador: Prof. Dr. Eduardo Borges de Melo

CASCADEL - PARANÁ

2015

Dados Internacionais de Catalogação-na-Publicação (CIP)

C214e	Campos, Luana Janaína de Estudos QSAR de derivados de pirimidonas, pirimidinas e piridopirazinas carboxamidas inibidoras da HIV 1 integrase./ Luana Janaína de Campos. — Cascavel, 2015. 134 p. Orientador: Prof. Dr. Eduardo Borges de Melo Dissertação (Mestrado) – Universidade Estadual do Oeste do Paraná, Campus de Cascavel, 2015 Programa de Pós-Graduação Stricto Sensu em Ciências Farmacêuticas 1. AIDS. 2. GRIND. 3. Volsurf. 4. OPS. 5. PLS. I. Melo, Eduardo Borges de . II. Universidade Estadual do Oeste do Paraná. III. Título. CDD 21.ed. 615.1 CIP – NBR 12899
-------	--

Ficha catalográfica elaborada por Helena Soterio Beijo – CRB 9ª/965

LUANA JANAÍNA DE CAMPOS

**ESTUDOS QSAR DE DERIVADOS DE PIRIMIDONAS, PIRIMIDINAS E
PIRIDOPIRAZINAS CARBOXAMIDAS INIBIDORAS DA HIV 1 INTEGRASE**

Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Ciências Farmacêuticas, da Universidade Estadual do Oeste do Paraná, como pré-requisito para obtenção do título de Mestre em Ciências Farmacêuticas. Área de concentração: Fármacos e Medicamentos

Orientador: Prof. Dr. Eduardo Borges de Melo

BANCA EXAMINADORA:

Prof. Dr. Eduardo Borges de Melo –
Universidade Estadual do Oeste do Paraná
UNIOESTE
Orientador

Profa. Dr^a. Lilian Sibelle Campos
Bernardes – *Universidade Federal de
Santa Catarina - UFSC*
Banca

Prof. Dr. Mauricio Ferreira da Rosa
Universidade Estadual do Oeste do Paraná
- UNIOESTE
Banca

CASCADEL – PR

2015

BIOGRAFIA

Nascida em 11 de Outubro de 1991, em Sete Quedas-MS, filha de Silvana Capeletto de Campos e Sérgio Alexandre de Campos, formou-se no Ensino Médio no ano de 2008 na cidade de Cascavel-PR. No ano seguinte, ingressou como acadêmica do curso de Farmácia da Universidade Estadual do Oeste do Paraná (UNIOESTE). Durante os cinco anos de graduação realizou atividades em diversas áreas, como Anatomia Humana, Química Analítica Qualitativa e Quantitativa, Modelagem Molecular, *Docking* e QSAR, Controle de Qualidade Físico-Químico e Análise Instrumentais em indústria farmacêutica. No ano de 2013 obteve o grau de Bacharel em Farmácia pela mesma universidade. Em 2014, iniciou suas atividades no Programa de Mestrado em Ciências Farmacêuticas da UNIOESTE, onde trabalhou com modelagem molecular e estudos QSAR.

Mesmo que eu tivesse o dom de profecia, e conhecesse todos os mistérios e toda a ciência, e ainda que tivesse toda a fé, a ponto de transpor montanhas, se não tiver caridade, nada seria.

1 Coríntios 13:2

“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar. Mas o mar seria menor se lhe faltasse uma gota”.

Madre Teresa de Calcutá

“... Ao transferir bens materiais, o doador perde a sua posse. Há, porém, alguns atributos intrínsecos que não podem ser transferidos de um indivíduo para o outro, como a beleza e a coragem. O conhecimento, por outro lado, é algo tão importante que Deus decidiu que o doador pode retê-lo mesmo que o tenha transmitido...”

Atribuído a Pitágoras de Samos, cerca de 2.500 anos atrás.

*À Deus, por sua infinita graça e misericórdia.
À minha família e amigos, pelo apoio, força,
incentivo e amizade. Sem vocês nenhuma
conquista valeria a pena.*

AGRADECIMENTOS

A Deus por me dar o dom da vida, por me amar incondicionalmente, ser minha força nos momentos de fraqueza e tornar meus sonhos possíveis.

Aos meus pais, Sérgio Alexandre e Silvana pelo suporte, pelos ensinamentos, pela torcida e oração, minha vitória é também de vocês. Aos meus irmãos Jhenyffer e João Alexandre, por partilharem da minha vida e me incentivarem. A todos os meus familiares, os quais mesmo na distância torceram por mim sempre.

Ao meu namorado Aramís, que esteve ao meu lado em todos os momentos, por me escutar nos momentos difíceis e me aconselhar.

As minhas amigas Ana Flávia, Iara e Jakeline, por todo o companheirismo, por terem feito estes anos mais felizes e por me ensinarem o valor da amizade. Aos colegas de mestrado, ao Programa de Pós-Graduação *Stricto Sensu* em Ciências Farmacêuticas da Unioeste e todos os professores pelo empenho e dedicação a este curso de pós-graduação. Ainda, um especial agradecimento à colega de mestrado e amiga Jessica Balbinot pelos preciosos momentos de cafezinho e de desabafo, obrigado por me escutar!

Ao meu orientador professor Eduardo Borges de Melo, pelos preciosos ensinamentos, tornando este trabalho realidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) pelo suporte financeiro ao trabalho através dos financiamentos 2010/7354 e 458355/2014-3 e a Fundação Araucária pelo apoio a este trabalho de pesquisa.

Ao professor Favero Reisdorfer Paula por gentilmente ter cedido seu tempo, conhecimento e recursos a mim e a este trabalho. À professora Kerly Fernanda Mesquita Pasqualoto pelas sugestões de grande valia para o melhoramento desta pesquisa.

RESUMO

ESTUDOS QSAR DE DERIVADOS DE PIRIMIDONAS, PIRIMIDINAS E PIRIDOPIRAZINAS CARBOXAMIDAS INIBIDORAS DA HIV 1 INTEGRASE

O campo do planejamento de fármacos auxiliado por computador tem atraído atenção no que se refere à descoberta de novos antirretrovirais para uso na farmacoterapia da AIDS. Com isso, o objetivo deste trabalho foi desenvolver modelos matemáticos que possam prever a atividade anti HIV1-IN de análogos ainda não sintetizados de pirimidonas, pirimidinas e piridopirazinas carboxamidas, agilizando assim a obtenção de novos compostos protótipos de fármacos. Desta maneira, foi selecionado da literatura um conjunto de 199 compostos descritos como inibidores da reação de transferência de fita e desenvolvido estudos de QSAR 2D e 3D. Para o primeiro utilizou-se a metodologia de seleção de variáveis *Ordered Predictors Selection* (OPS) e para o segundo Planejamento Fatorial Fracionário (*Fractional Factorial Design*, FFD) e OPS. Os modelos em ambos os estudos foram construídos através da regressão por quadrados mínimos parciais (PLS). Em relação aos estudos de QSAR 3D, foram empregadas duas diferentes metodologias para o cálculo dos campos de interação molecular (*Molecular Interaction Fields*, MIFs): GRID e GRIND (*GRId-INdependent descriptors*). Ademais, foram também utilizados descritores de propriedades farmacocinéticas provenientes do *software* Volsurf+ (extraídos de MIFs gerados com GRID) para construção de modelos em associação com descritores 2D. Os modelos com boa qualidade estatística, foram empregados para predição das atividades de um segundo conjunto de dados (conjunto de predição, N=145). Os resultados para as previsões foram satisfatórios. Os domínios de aplicabilidade de Willians e Euclidiano em relação as amostras do conjunto de predição revelam que as predições não ocorreram por extrapolação. Além disto, o descritor E_T (energia total da molécula), selecionado para o modelo QSAR 2D é consistente com achados bibliográficos prévios. Em relação ao estudo 3D, as características mais relevantes (interações N1-N1 e DRY-O) se mostram coerentes com o farmacóforo da respectiva classe sob estudo.

Palavras Chaves: AIDS; GRIND, VOLSURF; OPS; PLS.

ABSTRACT

QSAR STUDIES OF PYRIMIDONES, PYRIMIDINES AND PYRIDOPYRAZINES CARBOXAMIDES DERIVATIVES AS INHIBITORS OF HIV 1 INTEGRASE

The computer-aided drug design field have attracted attention with regard to the discovery of new antiretroviral for use in AIDS pharmacotherapy. Thus, the aim of this study was to develop mathematical models that can predict the anti HIV1-IN activity of analogues not synthesized from pyrimidones, pyrimidines and pyridopyrazine carboxamides, reducing the time to obtain new prototypes of drugs. Therefore, it was selected from the literature, a set of 199 compounds described as strand transfer reaction inhibitors and developed 2D and 3D QSAR studies. For the 2D QSAR was used the Ordered Predictors Selection (OPS) methodology in order to develop the variable selection and Fractional Factorial Design (FFD) and OPS for the 3D QSAR. The models in both studies were constructed using partial least squares (PLS). Regarding the 3D QSAR studies, were employed two different methodologies for the calculation of the molecular interaction fields (MIFs): GRID and GRIND (GRid-INdependent descriptors). In addition, they were also used descriptors of pharmacokinetic properties from the Volsurf+ software (extracted from MIFs generated with GRID) for building models in combination with 2D descriptors. The statistical models with good quality were used to predict the activities of a second data set (prediction set, $N=145$). The resultants predictions were satisfactory. The Williams and Euclidean applicability domains for the set of samples reveal that the predictions did not occur by extrapolation. Moreover, the descriptor E_T (total energy of the molecule), selected by the QSAR 2D model is consistent with earlier findings. Further, for the 3D study, the most relevant characteristics (N1-N1 interactions and DRY-O) are coherent with the pharmacophore of the class under study.

KEYWORDS: AIDS; GRIND, VOLSURF; OPS; PLS.

SUMÁRIO

1.	INTRODUÇÃO	1
2.	OBJETIVOS	2
3.	REVISÃO BIBLIOGRÁFICA	3
3.1	HIV/AIDS	3
3.2	Histórico	4
3.3	Morfologia e ciclo replicativo do HIV	5
3.4	Epidemiologia	8
3.5	Profilaxia	11
3.6	Terapia antirretroviral.....	11
3.6.1	Fármacos antirretrovirais	13
3.6.2	Inibidores da transcriptase reversa nucleosídeo-nucleotídeos (ITRN)	14
3.6.3	Inibidores da transcriptase reversa não-nucleosídeos (ITRNN)....	16
3.6.4	Inibidores de protease (IP).....	17
3.6.5	Inibidores da fusão (IF)	18
3.6.6	Antagonistas do co-receptor CCR5 (ACCR5).....	19
3.6.7	Inibidores da integrase	20
3.7	Integração e HIV-integrase como alvo para fármacos	21
3.8	Mecanismo de inibição da HIV-IN pelos INSTI	25
3.9	Pirimidonas/Pirimidinas/Piridopirazinas	29
3.10	Planejamento de fármacos auxiliado por computador	31
3.10.1	QSAR	32
3.10.1.1	Métodos de seleção de variáveis/construção dos modelos QSAR	
3.10.1.2	Validação dos modelos QSAR.....	35
4.	METODOLOGIA.....	40
4.1	Conjunto de dados	40
4.2	Modelagem Molecular	41
4.3	Estudo QSAR-2D	42
4.3.1	Obtenção dos descritores	42
4.3.2	Seleção de variáveis e construção dos modelos	44
4.3.3	Validação dos modelos.....	45
4.4	Estudo QSAR-3D	46
4.4.1	GRIND	47
4.4.1.1	Cálculo dos MIFs e descritores GRIND	48
4.4.1.2	Seleção de variáveis e construção dos modelos.....	49
4.4.1.3	Validação dos Modelos	50

4.4.2	Modelos VolSurf+	50
4.4.2.1	Geração dos MIFs/descriptores Volsurf e 2D	51
4.4.2.2	Seleção de variáveis e construção dos modelos.....	52
4.5	Avaliação da predição	52
5	RESULTADOS E DISCUSSÃO	52
5.1	QSAR 2D.....	52
5.2	QSAR 3D.....	56
5.2.1	GRIND	56
5.2.2	Modelos Volsurf/Descritores 2D.....	58
5.3	Avaliação da predição	62
5.4	Descritores selecionados	64
6.	CONCLUSÃO.....	69
7.	REFERÊNCIAS BIBLIOGRÁFICAS	70
8.	Anexos.....	82
8.1	Anexo 1.....	82
8.2	Anexo 2.....	87
8.3	Anexo 3.....	112

LISTA DE ILUSTRAÇÕES

Dissertação

Figura 1 Morfologia do HIV.....	6
Figura 2 Ciclo replicativo resumido do HIV.....	8
Figura 3 Distribuição global da prevalência de HIV em indivíduos com idade entre 15-49 anos.....	9
Figura 4 Estrutura química dos ITRN's em uso clínico.....	15
Figura 5 Fosforilação intracelular do AZT.....	16
Figura 6 Estrutura química dos ITRNN's em uso clínico.....	17
Figura 7 Estrutura química dos inibidores de protease em uso clínico.....	18
Figura 8 Estrutura química do enfuvirtide.....	19
Figura 9 Estrutura química do maraviroc.....	20
Figura 10 Estruturas químicas dos inibidores da HIV-IN em uso clínico.....	21
Figura 11 Representação da estrutura da integrase de HIV e representação tridimensional do complexo IN-DNA de PFV.....	22
Figura 12 Etapas da integração.....	23
Figura 13 Representação da atividade dos átomos de Mg^{2+} durante a etapa de transferência de fita.....	24
Figura 14 Demonstração do processo da integração realizado pela integrase...	25
Figura 15 Estrutura química de compostos detentores de atividade anti HIV-IN de diferentes classes químicas.....	25
Figura 16 Estrutura do ácido benzopirúvico e representação do equilíbrio ceto-enólico deslocado a favor da forma enólica.....	26
Figura 17 Farmacóforo dos DCA's.....	27
Figura 18 Esquema 2D da ligação do raltegravir ao sítio de ligação da PFV-IN..	27
Figura 19 Esquema 2D da ligação do elvitegravir ao sítio de ligação da PFV-IN.....	28
Figura 20 Esquema 2D da ligação do dolutegravir ao sítio de ligação da PFV-IN.....	29
Figura 21 A) Núcleo base das pirimidonas; B) pirimidinas; C) dihidroxipirimidinas; D) <i>N</i> -metilpirimidona.....	30

Figura 22 Composto pirimidínico baseado em β -dicetoácido.....	30
Figura 23 Estrutura com inserção do grupamento carboxamidas à pirimidona..	31
Figura 24 Estrutura básica das piridopirazinas em A; dihidroxipiridopirazinas em B.....	31
Figura 25 Regressões para os resultados obtidos em um conjunto arbitrário de dados com e sem interceptação.....	38
Figura 26 Estrutura base cristalográfica utilizada para construção do conjunto de compostos.....	42
Figura 27 Resultados da validação cruzada <i>Leave -14- out</i> do modelo A.....	55
Figura 28 Valor dos interceptos para Q^2_{LOO} e R^2 no processo de randomização do y do modelo A.....	56
Figura 29 Representação tridimensional do modelo GRIND.....	57
Figura 30 Resultados da validação cruzada <i>Leave -14- out</i> do modelo C.....	60
Figura 31 Valor dos interceptos para Q^2_{LOO} e R^2 no processo de randomização do y do modelo C.....	60
Figura 32 Resultados dos testes de domínio de aplicabilidade.....	64
Figura 33 Coeficientes PLS mostrando os descritores diretamente (valores positivos) e inversamente (valores negativos) correlacionados com a atividade anti HIV-1 Integrase no modelo B.....	66
Figura 34 Detalhamento do descritor 148.....	67
Figura 35 Detalhamento do descritor 261.....	68

Anexos

Anexo 2: Figura 1 Estruturas químicas dos inibidores da HIV-IN em uso clínico (A: raltegravir; B: elvitegravir; C: dolutegravir).....	88
Anexo 2: Figura 2 Núcleo básico das piridopirazinas em A; pirimidonas em B; pirimidinas em C; inserção do grupamento carboxamidas à pirimidona).....	89
Anexo 2: Figura 3 Estrutura básica. Representação 2D.....	90

Anexo 2: Figura 4 Resultados da validação cruzada LNO (A) e do teste de randomização do y (B).....	103
Anexo 2: Figura 5 Resultados dos testes de domínio de aplicabilidade.....	105

LISTA DE GRÁFICOS E TABELAS

Dissertação

Gráfico 1 Distribuição percentual de casos de AIDS por região no Brasil, dados coletados no período de 1980 a dezembro de 2013.....	10
Gráfico 2 Coeficiente de mortalidade por AIDS segundo Unidade Federada.....	10
Tabela 1 Fármacos antirretrovirais distribuídos gratuitamente para tratamento da AIDS no Brasil.....	12
Tabela 2 Classificação dos descritores moleculares.....	34
Tabela 3 Parâmetros estatísticos empregados na validação interna.....	36
Tabela 4 Parâmetros estatísticos empregados na validação externa.....	39
Tabela 5 Comparação dos parâmetros estatísticos dos modelos resultantes....	61
Tabela 6 Descritores selecionados no modelo A e seus respectivos coeficientes padronizados (abordagem de autoescalamento).....	65
Tabela 7 Descritores GRIND e suas correspondentes distâncias apresentados segundo a ordem de importância no modelo B. Tais variáveis foram identificadas como as mais correlacionadas com atividade biológica em estudo (segundo o coeficiente PLS).....	67

Anexos

Anexo 1: Tabela 1 Detalhamento dos valores preditos pelo modelo B e observados para a variável independente para o conjunto de predição (N=145).....	82
Anexo 2: Artigo Tabela 1 Parâmetros estatísticos empregados na validação interna/externa.....	95
Anexo 2: Artigo Tabela 2 Valores dos descritores selecionados para as amostras dos conjunto de dados e resultados dos processos de predição (validação cruzada e externa).....	99

Anexo 2: Artigo Tabela 3 Resultados dos parâmetros estatísticos para o modelo 1.....	102
--	-----

Anexo 2: Artigo Tabela 4 Descritores selecionados no modelo 1 e seus respectivos coeficientes/coeficientes padronizados (abordagem de autoescalamento).....	106
---	-----

Material suplementar artigo

Anexo 3: Tabela 1 Compostos descritos por Ferrara <i>et al.</i> (2010), pertencentes à classe das 2-pirrolidinil- <i>N</i> -metil pirimidonas carboxamidas.....	112
---	-----

Anexo 3: Tabela 2 Compostos descritos por Gardelli <i>et al.</i> (2007), pertencentes à classe das <i>N</i> -metil pirimidonas carboxamidas e compostos descritos por Nizi <i>et al.</i> (2009), pertencentes à classe das dihidroxipirimidinas carboxamidas e <i>N</i> -metil pirimidonas carboxamidas.....	113
--	-----

Anexo 3: Tabela 3 Compostos descritos por Petrochi <i>et al.</i> (2009), pertencentes à classe das pirazino pirimidonas carboxamidas.....	114
---	-----

Anexo 3: Tabela 4 Compostos descritos por Di Francesco <i>et al.</i> (2008), pertencentes à classe das <i>N</i> -metil pirimidonas carboxamidas.....	116
--	-----

Anexo 3: Tabela 5 Compostos descritos por Muraglia <i>et al.</i> (2008), pertencentes à classe das pirimidonas bicíclicas carboxamidas.....	117
---	-----

Anexo 3: Tabela 6 Compostos descritos por Donghi <i>et al.</i> (2009), pertencentes à classe das pirimidinas carboxamidas.....	118
--	-----

Anexo 3: Tabela 7 Compostos descritos por Pace <i>et al.</i> (2007), e compostos descritos por Summa <i>et al.</i> (2006), pertencentes à classe das dihidroxipirimidinas carboxamidas.....	120
---	-----

Anexo 3: Tabela 8 Compostos descritos por Petrochi <i>et al.</i> (2007), pertencentes à classe das dihidroxipirimidinas carboxamidas.....	121
---	-----

Anexo 3: Tabela 9 Compostos descritos por Wai <i>et al.</i> (2007), pertencentes à classe das dihidroxipiridopirazinas carboxamidas.....	123
--	-----

Anexo 3: Tabela 10 Detalhamento dos valores dos descritores presentes no modelo (equação 1) para os compostos do conjunto de predição ($N=145$), valores preditos pelo modelo e observados para a variável independente.....124

LISTA DE ABREVIATURAS

HIV	<i>Human Immunodeficiency Virus</i>
SIDA	Síndrome da Imunodeficiência Adquirida
AIDS	<i>Acquired Immunodeficiency Syndrome</i>
CV	Carga Viral
HTLV	<i>Human T Lymphotropic Virus</i>
TR	Transcriptase Reversa
SIV	<i>Simian Immunodeficiency Virus</i>
RNA	Ácido Ribonucleico
GLP's	Glicoproteínas
HIV-TR	HIV-Transcriptase Reversa
HIV-PR	HIV-Protease
HIV-IN	HIV-Integrase
cDNA	Ácido Desoxirribonucleico Complementar
RNA_m	RNA Mensageiro
PIC	Complexo de Pré-integração
Tat	<i>Trans-acting Transcription Transactivator</i>
OMS	Organização mundial da saúde
SINAN	Sistema de Informação de Agravos de Notificação
DST	Doenças Sexualmente Transmissíveis
TAR	Terapia antirretroviral
HBV	Vírus da hepatite B
ARV's	Antirretrovirais
DM	Diabetes Mellitus
ITRN	Inibidores da Transcriptase Reversa Nucleosídeo-Nucleotídeos
ITRNN	Inibidores da Transcriptase Reversa não-Nucleosídeos
IP	Inibidores de Protease
IF	Inibidores da Fusão
A CCR5	Antagonistas do Co-receptor CCR5
IIN	Inibidores da Integrase
AZT	Azidovudina
FDA	<i>Food and Drug Administration</i>
INSTI	<i>Intgrase Strand Transfer Inhibitors</i>
3'P	Processamento 3'
ST	<i>Strand Transfer</i>
NTD	Domínio N-terminal
CCD	Domínio Central Catalítico
PFV	<i>Prototype Foamy Virus</i>
LTR	<i>Long Terminal Repeat</i>
DCA's	Dicetoácidos
MCA	Monocetoácido
PDB	<i>Protein Data Bank</i>
CADD	<i>Computer-Aided Drug Design</i>
QSAR	<i>Quantitative Structure-Activity Relationship</i>

IUPAC	União Internacional de Química Pura e Aplicada
DFT	<i>Density Functional Theory</i>
MM	Mecânica Molecular
DM	Dinâmica Molecular
SCF	<i>Self-Consistent Field</i>
AM1	<i>Austin Model 1</i>
PM3	<i>Parameter Model 3</i>
B3LYP	<i>Becke's three-parameter hybrid exchange functional and the Lee-Yang-Parr correlation functional</i>
VI	Variável Independente
VD	Variável Dependente
AB	Atividade Biológica
MLR	<i>Multiple Linear Regression</i>
PLS	<i>Partial Least-Squares</i>
OPS	<i>Ordered Predictors Selection</i>
VL	Variável Latente
PCR	<i>Principal Component Regression</i>
FFD	<i>Fractional Factorial Design</i>
MIF	<i>Molecular Interaction Field</i>
GRIND	<i>GRInd-INdependent descriptors</i>
SRV	Sítio Receptor Virtual
R²	Coeficiente de Determinação Múltipla
RMSEC	Raiz Quadrada do Erro da Calibração
F	Teste de Fischer
LOO	<i>leave-One-out</i>
LNO	<i>leave-N-out</i>
Q²_{LoO}/Q²_{LNO}	Coeficiente de correlação entre as atividades biológicas observadas e preditas durante o processo de validação cruzada <i>leave-one-out/leave-n-out</i> .
RMSECV	Raiz Quadrada do Erro da Validação Cruzada
SDEP	Desvio Padrão do erro de Predição
RMSD	Raiz Quadrada do Desvio Médio
R²₀	Coeficiente de Determinação Centrado na Origem
R²_{pred}	Coeficiente de Correlação de Predição
K e K'	Inclinações das Linhas das Regressões Lineares de Predição
r_m²	Coeficiente de Correlação Modificado
Δr_m²	Variação de r _m ²
CSD	<i>Cambridge Structural Database</i>

1. INTRODUÇÃO

Uma das abordagens mais utilizadas no campo do planejamento de fármacos auxiliado por computador são os estudos de relação estrutura atividade quantitativa (QSAR: *quantitative structure-activity relationships*). Esta abordagem encontra ampla aplicação para a avaliação do impacto em potencial de produtos químicos na saúde humana, em sistemas ecológicos e em processos tecnológicos.

Considerando a necessidade do desenvolvimento de novos fármacos antirretrovirais, uma classe de compostos detentores de atividade contra o vírus da imunodeficiência humana (HIV: *human immunodeficiency virus*) amplamente estudada são os derivados de dicetoácidos (DCAs). Estes compostos são conhecidos por inibirem a enzima HIV-integrase (HIV-IN), pois podem quelar o cofator Mg^{2+} e inibir a reação de transferência de fita (ST: *strand transfer*) catalizada por esta enzima.

Pirimidinas, pirimidonas e piridopirazinas são compostos heterocíclicos, os quais na presença do grupamento carboxamida apresentam o farmacóforo dos DCAs. Um dos fatos que torna esta classe terapêutica atrativa para o desenvolvimento de novos antirretrovirais é que as células de mamíferos não apresentam enzimas homólogas a HIV-IN, o que aumenta a especificidade destes fármacos, apresentando baixo potencial de toxicidade.

As metodologias de QSAR baseiam-se na hipótese de que, utilizando modelos matemáticos, é possível prever a atividade biológica de novos análogos, auxiliando na seleção daqueles de maior interesse em sua preparação. Assim, este trabalho foi realizado com intuito de obter um modelo QSAR estatisticamente validado, com alta capacidade de predição de novos inibidores da HIV-IN derivados de pirimidinas, pirimidonas e piridopirazinas carboxamidas, potencialmente útil como ferramenta de apoio ao planejamento de novos derivados.

2. OBJETIVOS

Este trabalho teve por objetivo a obtenção de modelos QSAR estatisticamente validados, com alta capacidade preditiva, potencialmente úteis para a predição da atividade inibitória sobre a HIV1-IN, ou seja, úteis como ferramenta de apoio ao planejamento de novos derivados. Para cumprir este objetivo, foram selecionados da literatura 199 derivados de pirimidonas, pirimidinas e piridopirazinas previamente sintetizados que apresentaram atividade quando ensaiados por um mesmo protocolo. Além disto, utilizar os modelos obtidos como ferramenta para uma melhor compreensão do mecanismo de ação dos referidos compostos.

3. REVISÃO BIBLIOGRÁFICA

3.1 HIV/AIDS

A infecção pelo vírus HIV ocorre através de relações sexuais sem proteção (anal ou vaginal), transfusão de sangue contaminado, compartilhamento de agulhas contaminadas, transmissão vertical (da mãe para feto ou recém-nascido) durante a gravidez, parto ou amamentação (GALLO e MONTAGNIER, 2002; WHO, 2014a). Um amplo espectro de apresentações clínicas pode ocorrer, desde a fase aguda (que pode ser assintomática, oligossintomática ou se manifestar como síndrome retroviral aguda) até a fase avançada, com as manifestações que definem a síndrome da imunodeficiência adquirida (AIDS: *acquired immunodeficiency syndrome*) (BRASIL, Ministério da Saúde, 2008).

A AIDS caracteriza-se por ser uma patologia de longa duração, com um grande espaço de tempo entre a exposição ao agente etiológico e um profundo estado de imunossupressão (o estágio mais avançado da infecção), e pode levar de 10 a 15 anos para um portador desenvolver a doença. A imunossupressão ocorre devido a uma drástica redução no número de linfócitos T CD4+ (LT-CD4+), células responsáveis pela ativação de outros linfócitos que realizam a defesa do organismo, sendo esta a principal característica desta patologia. Entretanto, macrófagos, monócitos e células de Langerhans, entre outras, são também susceptíveis à infecção (PEÇANHA *et al.*, 2002; GALLO, 2002; REQUEJO, 2006; WHO, 2014a).

À medida que a infecção progride, o sistema imunológico se torna mais debilitado e o organismo mais suscetível a diversas infecções oportunistas (WHO, 2014a), em especial a tuberculose e a pneumonia, os dois quadros patológicos que mais causam óbitos em pacientes (FERREIRA, 1996). A infecção pelo HIV predispõe o surgimento de várias neoplasias, especialmente linfoma não-Hodgkin e sarcoma de Kaposi (SCHULZ *et al.*, 1996; FRÖHLICH *et al.*, 2000), provavelmente devido a uma interferência com o controle tumoral mediado pelo sistema imune (FRÖHLICH *et al.*, 2000). O aparecimento de infecções oportunistas e neoplasias é definidor da AIDS (BRASIL, Ministério da Saúde, 2008). Também já foram descritas doenças por dano direto a certos

órgãos a processos inflamatórios, como miocardiopatia, nefropatia e neuropatias, que podem estar presentes durante toda a evolução da infecção pelo HIV (BRASIL, Ministério da Saúde, 2008).

Para estimar o prognóstico e avaliar a indicação de início da terapia antirretroviral, monitora-se a evolução da contagem de LT-CD4+, cuja contagem normal varia de 800 a 1200 células/mm³, e a quantificação plasmática da carga viral do HIV. Os indivíduos que apresentam contagem média de 500 células/mm³ são definidos como infectados pelo agente etiológico (soropositivos) ou portadores, e indivíduos com contagem inferior a 350 células/mm³ (limite brasileiro) são considerados doentes de AIDS. A contagem de LT-CD4+ é utilizada internacionalmente como marcador do estado imunológico dos indivíduos, enquanto que a quantificação da carga viral (CV) serve como marcador do risco de queda subsequente nas contagens LT-CD4+ (BRASIL, Ministério da Saúde, 2008).

3.2 Histórico

Apesar das evidências de mortes causadas pela AIDS há diversas décadas, os primeiros relatos oficiais datam de junho de 1981 em Los Angeles e Nova Iorque (BARRE-SINOSSI *et al.*, 1983; GALLO, 2002). A possibilidade de que o agente etiológico fosse um vírus, levou os pesquisadores Robert C. Gallo e Luc Montagnier à hipótese de que o HTLV-1 (*human T lymphotropic virus type 1*) poderia ser o agente, pois as possíveis vias de transmissão eram as mesmas. Além disto, o mesmo marcador biológico, uma diminuição nos níveis de LT CD4+, era observado. No início de 1983 o grupo de Gallo encontrou sequências de DNA relacionadas ao HTLV somente em dois indivíduos de um grupo de 33 pessoas com AIDS. Já o grupo de Luc Montagnier encontrou traços de transcriptase reversa (TR), enzima já conhecida e essencial para o HIV, no sobrenadante da cultura de linfócitos de pacientes, sem que ocorresse a precipitação imune de anticorpos para HTLV. Em 1984, com o crescimento do vírus isolados de pacientes em culturas de LT CD4+, confirmou-se que o HIV, da subfamília dos *lentiretrovirus*, era o agente etiológico da doença (GALLO, 2002; REQUEJO, 2006).

Atualmente, acredita-se que a infecção tenha surgido por volta dos anos 1900 na região sudeste da república do Camarões (FARIA *et al.*, 2014), sendo que o mesmo passou de primatas para o homem, sendo provavelmente um descendente resultante de mutações do vírus da imunodeficiência símia (SIV: *simian immunodeficiency virus*) (FORATTINI, 1993; REQUEJO, 2006). Cerca de 85% dos isolados de HIV a partir de seres humanos são agrupados em dois tipos, o HIV-1 e HIV-2, sendo o primeiro o principal agente etiológico, enquanto o segundo é restrito a algumas regiões da África Ocidental e Central (REQUEJO, 2006).

3.3 Morfologia e ciclo replicativo do HIV

A estrutura morfológica dos HIV-1 e 2 inclui proteínas estruturais, funcionais, um genoma de ácido ribonucleico (RNA) protegido por um capsídeo localizado dentro de um envelope viral constituído por uma bicamada lipídica de origem celular, e por glicoproteínas (GLP) virais complexas, denominadas *env*, incorporados neste envelope (Figura 1). A *env* constituída pelas GLP's gp41, transmembrana, e gp120, que fica exposta ao meio externo, possui a função de fixação e fusão do vírus com as células hospedeiras, e condução da entrada na célula alvo (GREENE *et al.*, 2012; HUGHES e ADMIRAAL, 2012).

O capsídeo é formado pela proteína viral p24, e é envolvido pela p17, conhecida como matriz. Na parte mais interna encontram-se os elementos mais importantes: dois filamentos simples de RNA protegidos por um nucleocapsídeo formado pela proteína p7 e as três enzimas virais essenciais, transcriptase reversa (HIV-TR), protease (HIV-PR) e integrase (HIV-IN) (Figura 1) (DOMS e MOORE, 2000; PEÇANHA *et al.*, 2002; FANALES-BELASIO *et al.*, 2010). No citoplasma, o RNA de cadeia simples é convertido em ácido desoxirribonucleico complementar (cDNA) pela HIV-TR. Este novo material genético é integrado ao genoma da célula hospedeira pela HIV-IN, formando o pró-vírus maduro, que é capaz de sintetizar novas fitas de RNA genômico viral e de RNA mensageiro (RNAm viral) (PEÇANHA *et al.*, 2002; GREENE *et al.*, 2012). Como o processo de integração causa danos ao material genético, este também é relacionado como uma das causas dos cânceres causados por retrovírus (SCHULZ *et al.*, 1996).

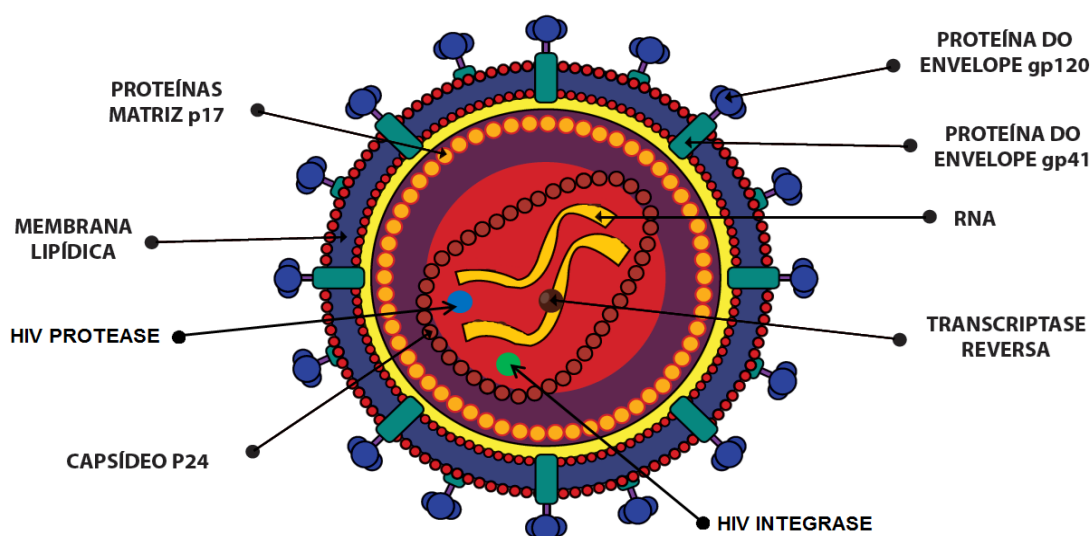


Figura 1 Morfologia do HIV.

Fonte: Adaptado de manual técnico para o diagnóstico da infecção pelo HIV (BRASIL, Ministério da Saúde, 2013a).

O ciclo replicativo do HIV é constituído de várias etapas, e cada uma apresenta potencial para o desenvolvimento de estratégias terapêuticas contra a AIDS (Figura 2). As etapas do ciclo são denominadas:

- (1) Adsorção/fusão: envolve a interação das glicoproteínas do envelope com os receptores CD4 e de quimiocinas (CCR5 ou CXCR4) localizados na superfície da célula hospedeira (DOMS e MOORE, 2000; GREENE *et al.*, 2012);
- (2) Desencapsulamento: após a fusão ocorre o desencapsulamento. Estudos sugerem o envolvimento da fosforilação de proteínas de matriz virais e alterações locais de pH na promoção do desencapsulamento. Este processo leva a liberação do genoma e enzimas virais no citoplasma da célula hospedeira (GREENE *et al.*, 2012);
- (3) Transcrição reversa: no citoplasma a HIV-TR promove a síntese de cDNA, catalisando reações de polimerização de DNA dependente de RNA. Ela também realiza a clivagem da porção de RNA do híbrido RNA-DNA formado durante o processo (atividade ribonuclease) (PEÇANHA *et al.*, 2002; GREENE *et al.*, 2012);

(4) Formação do complexo cDNA viral/integrase: o cDNA e a HIV-IN combinam-se no citoplasma, formando o complexo de pré-integração (PIC), sendo em seguida rapidamente transportado para o núcleo da célula hospedeira (PEÇANHA *et al.*, 2002; FANALES-BELASIO *et al.*, 2010; GREENE *et al.*, 2012);

(5) Integração: a HIV-IN também promove a integração estável do cDNA no DNA hospedeiro. Esta enzima cliva nucleotídeos de cada extremidade 3' do cDNA criando duas extremidades coesivas, realizando em seguida o processo conhecido como transferência de fita, formando assim um pró-vírus maduro (PEÇANHA *et al.*, 2002; FANALES-BELASIO *et al.*, 2010; GREENE *et al.*, 2012);

(6) Transcrição: a transcrição inicial do cDNA em RNAm resulta na síntese de proteínas reguladoras, tais como Tat (*trans-acting transcription transactivator*). Na ausência de Tat, as RNA polimerases são geralmente incapazes de transcrever sequências maiores que algumas centenas de nucleotídeos, de modo que esta proteína estimula a transcrição e a formação de partículas de RNA mais longos (PEÇANHA *et al.*, 2002; FANALES-BELASIO *et al.*, 2010);

(7) Tradução das proteínas virais: o RNAm transcrito é transportado para o citoplasma, onde serão traduzidos (PEÇANHA *et al.*, 2002). As proteínas maiores formadas são clivadas pela HIV-PR, formando proteínas menores e funcionais. Aquelas codificadas pelos genes *gag* e *pol* formarão o núcleo, enquanto aqueles codificados pelo gene *env* formam as GLP (FANALES-BELASIO *et al.*, 2010; GREENE *et al.*, 2012);

(8) Montagem: as partículas virais são inicialmente montadas próximo à membrana celular na forma de partículas imaturas formadas por um glicoproteico, RNA genômico e poliproteínas virais (PEÇANHA *et al.*, 2002);

(9) Brotamento e maturação: as partículas imaturas montadas migram para a superfície da célula. As grandes moléculas proteicas precursoras são então clivadas pela HIV-PR resultando em novas partículas virais infecciosas. Durante o brotamento, em que ocorre a fusão da partícula viral com membrana da célula

hospedeira, o HIV obtém o envelope viral. Nos linfócitos-T este processo ocorre por gemulação na superfície das células, enquanto nos monócitos e macrófagos ocorre acúmulo de partículas virais nos vacúolos intracelulares, que são então libertados (PEÇANHA *et al.*, 2002; FANALES-BELASIO *et al.*, 2010; GREENE *et al.*, 2012).

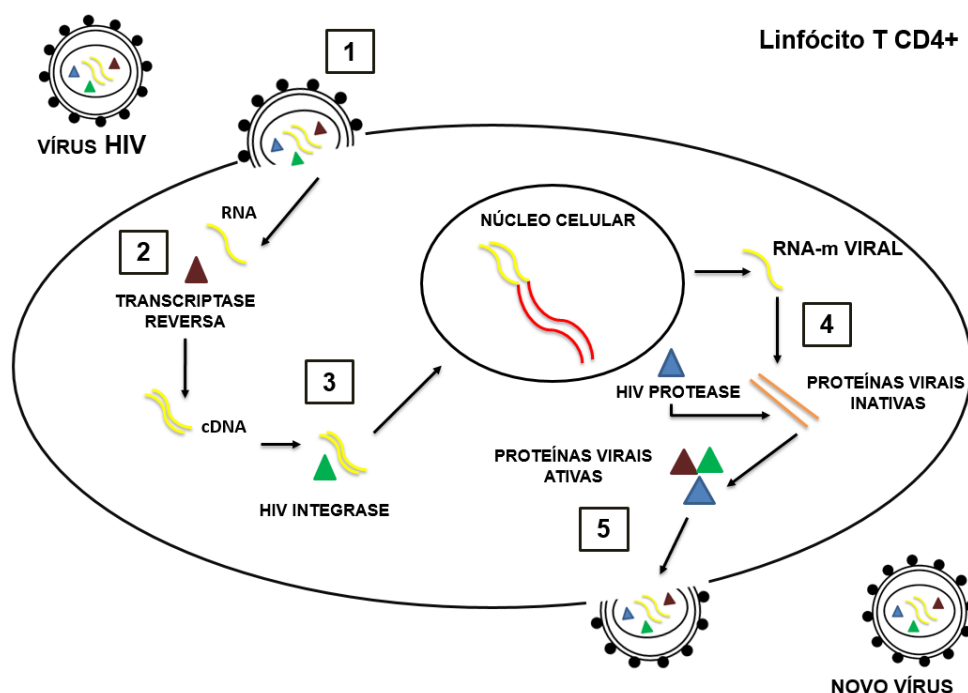


Figura 2 Ciclo replicativo resumido do HIV. 1) Adsorção/fusão; 2) Transcrição reversa; 3) Integração; 4) Tradução; 5) Montagem e liberação, baseado em PEÇANHA *et al.*, 2002; FANALES-BELASIO *et al.*, 2010.

3.4 Epidemiologia

O HIV continua sendo um grande desafio para a saúde global. Segundo dados da Organização Mundial de Saúde (OMS) (*World Health Organization*, WHO), até o final de 2013, trinta e cinco milhões de pessoas viviam com o HIV no mundo (WHO, 2014a). Desde o início da epidemia, quase 75 milhões de pessoas foram infectadas com o vírus HIV e cerca de 36 milhões de pessoas foram a óbito. Globalmente, estima-se que 0,8% dos adultos com idades entre 15-49 anos estão vivendo com o HIV em todo o mundo (Figura 3). A África Subsaariana continua sendo a região mais afetada, com quase 1 em cada 20 adultos infectados, respondendo por 71% das pessoas infectadas em todo o mundo (WHO, 2014b).

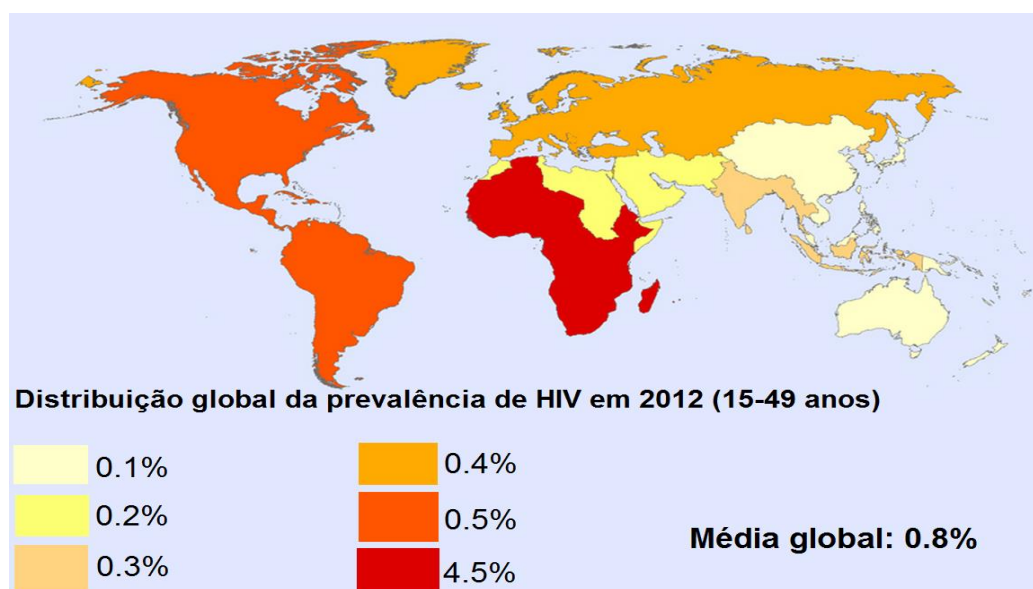


Figura 3 Distribuição global da prevalência de HIV em indivíduos com idade entre 15-49 anos.

Fonte: Adaptado de Organização Mundial da Saúde (WHO, 2014b).

Estima-se que 1,6 milhão de pessoas vivam com HIV na América Latina. A maioria dos casos (75%) se concentra em cinco países – Argentina, Brasil, Colômbia, México e Venezuela. A região teve queda de 3% em novas infecções entre 2005 e 2013, mas os índices variam de país para país. No Brasil, as infecções por HIV aumentaram 11% entre 2005 e 2013. No ano de 2012, o país registrou 47% de todos os novos casos contabilizados na América Latina (BRASIL, Ministério da Saúde, 2014b).

Desde os anos 1980, a vigilância epidemiológica do HIV/AIDS Brasil é baseada na notificação compulsória de casos de AIDS por meio do Sistema de Informação de Agravos de Notificação (SINAN) da Secretaria de Vigilância em Saúde do Ministério da Saúde. Segundo dados do último boletim epidemiológico sobre HIV/AIDS, em dezembro de 2013 aproximadamente 718 mil pessoas viviam com HIV/AIDS no Brasil. Em relação à distribuição da infecção nas diferentes regiões brasileiras, o maior número de casos se concentra na região sudeste (Gráfico 1) (BRASIL, Ministério da Saúde, 2013b).

A prevalência na população em geral é muito baixa quando comparada a dos países da África. Porém, ocorre uma concentração em determinados grupos mais vulneráveis, como homens que fazem sexo com homens, travestis, transexuais, usuários de drogas, profissionais do sexo, nos quais a prevalência pode ser maior que 10%. Entre 2003 e 2012 as maiores taxas de detecção de AIDS no Brasil foram observadas entre indivíduos com 30 a 49 anos. Entretanto,

observa-se uma tendência de queda na taxa nesta faixa, além de uma leve estabilização entre os indivíduos com 40 a 49 anos. Também se observa uma tendência de aumento entre os jovens de 15 a 24 anos e entre os adultos com 50 anos ou mais (BRASIL, Ministério da Saúde, 2013b).

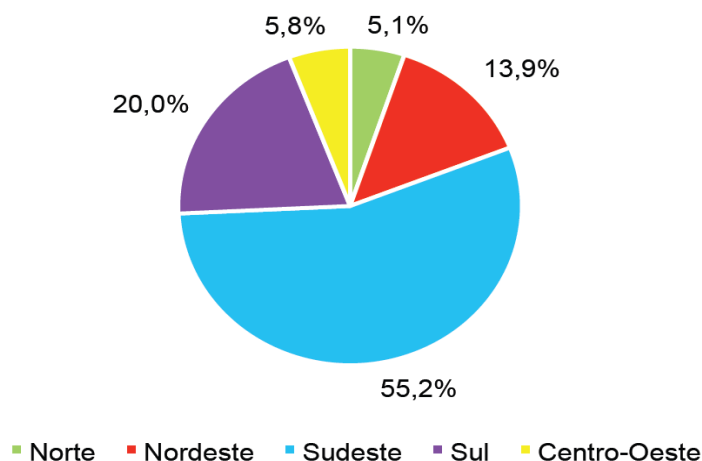


Gráfico 1 Distribuição percentual de casos de AIDS por região no Brasil, dados coletados no período de 1980 a dezembro de 2013.

Fonte: Disponível em Boletim Epidemiológico - AIDS e DST, Ministério da Saúde, Brasil, 2013b.

Em relação à mortalidade, desde a descoberta dos primeiros casos (1980) até o ano de 2012, foram declarados 265.698 óbitos classificados como causa básica “doenças pelo HIV”. Mais da metade ocorreram na Região Sudeste (62,6%). A Região Sul representa 17,1%, o Nordeste 11,6%, o Centro-Oeste 4,9% e o Norte 3,8% (Gráfico 2) (BRASIL, Ministério da Saúde, 2013b).

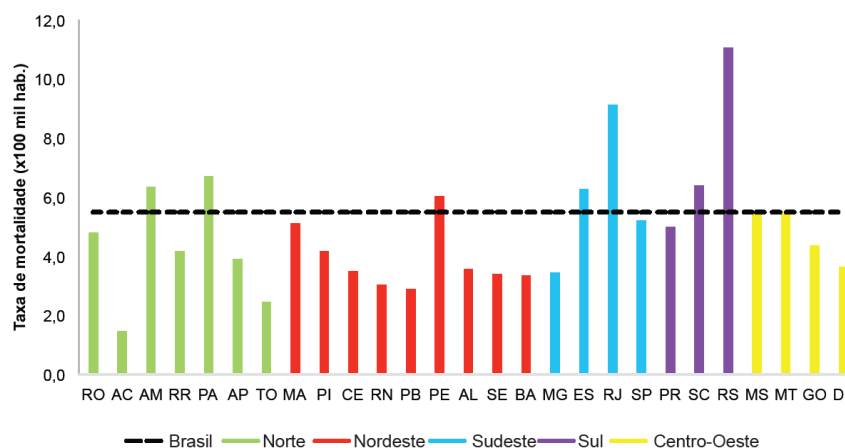


Gráfico 2 Coeficiente de mortalidade por AIDS segundo Unidade Federada. Brasil, 2012.

Fonte: Disponível em Boletim Epidemiológico - AIDS e DST, Ministério da Saúde, Brasil, 2013b.

3.5 Profilaxia

A descoberta de uma vacina eficaz que previna ou cure a infecção por HIV permanece em aberto (ESTE e CIHLAR, 2010). Poucas vacinas entraram em ensaios clínicos de fase III e apenas uma mostrou proteção parcial contra a infecção pelo HIV-1 (SMITH *et al.*, 2014). No respectivo estudo, voluntários saudáveis com baixo risco de infecção receberam doses de uma vacina com vetor viral recombinante expressando os genes *gag*, *pro* e *env*, juntamente com a gp120. Os voluntários do grupo vacina adquiriram 31,2% menos infecções pelo HIV-1 do que os no grupo placebo, após 3 anos de estudo. Além disso, anticorpos anti-gp120 encontravam-se presentes em 90% dos pacientes após 20 semanas. Citotoxicidade celular dependente de anticorpos para alvos revestidos com gp120 foi detectada em 75% dos pacientes, porém esse efeito sofria redução em 20 semanas (RERKS-NGARM *et al.*, 2009). A maior dificuldade na geração de uma vacina eficaz reside no fato de que o inóculo de HIV é composto por uma população geneticamente diversa, pois o vírus passa por uma alta incidência de mutações em seu material genético, permitindo que durante o curso da infecção consiga escapar de uma resposta imune apropriada (ESTE e CIHLAR, 2010; SMITH *et al.*, 2014).

Assim, no que tange às abordagens visando à prevenção, as medidas mais efetivas disponíveis são as estratégias do uso de preservativos, educação sexual, controle de sangue e derivados sanguíneos, entre outras medidas preventivas, as quais ainda constituem as melhores abordagens para evitar a contaminação pelo HIV, enquanto a terapia antirretroviral continua sendo a melhor abordagem para o tratamento dos pacientes (BRODER, 2010).

3.6 Terapia antirretroviral

Em 2013 a OMS lançou as novas recomendações clínicas para tratamento das pessoas soropositivas, levando em consideração a estimativa de que 26 milhões de pessoas em todo o mundo, portadores do HIV em países de baixa e média renda, sejam elegíveis para a terapia com medicamentos antirretrovirais (ARVs). Dentre estas recomendações clínicas está o tratamento dos adultos, adolescentes e crianças o mais cedo possível, iniciando a terapia antirretroviral (TAR) em todos os indivíduos com uma contagem de células CD4

igual ou inferior a 500 células/mm³, dando prioridade aos indivíduos com infecção grave ou avançada pelo HIV e aos que possuam uma contagem de células CD4 igual ou inferior a 350 células/mm³. Além disto, a TAR deve ser iniciada com qualquer contagem de células CD4 em pessoas co-infectadas com o vírus da hepatite B (HBV), com doença hepática crônica grave, com parceiros soropositivos em casais soro discordantes, mulheres grávidas ou lactantes e crianças menores de cinco anos de idade (WHO, 2013).

No Brasil, a Lei 9113/96 dispõe sobre a obrigação do estado de distribuir, de forma universal e gratuita, os medicamentos para o tratamento dos portadores do HIV. Segundo dados de dezembro de 2012, 313 mil pessoas recebem regularmente os remédios para tratar a doença. Atualmente, 21 medicamentos antirretrovirais estão disponíveis para distribuição gratuita pelo governo brasileiro (Tabela 1) (BRASIL, Ministério da Saúde, 2014a e 2015).

Tabela 1 Fármacos antirretrovirais distribuídos gratuitamente para tratamento da AIDS no Brasil.

Classe	Fármacos
Inibidor de transcriptase reversa nucleosídeo-nucleotídeo	Abacavir, Didanosina, Estavudina, Lamivudina, Tenofovir, Zidovudina, Lamivudina/Zidovudina
Inibidor da transcriptase reversa não-nucleosídeo	Efavirenz, Nevirapina e Etravirina.
Inibidor de protease	Atazanavir, Darunavir, Fosamprenavir, Indinavir, Lopinavir, Nelfinavir, Ritonavir, Saquinavir e Tipranavir.
Inibidor da fusão	Enfuvirtida
Inibidor da integrase	Raltegravir, Dolutegravir*

Fonte: Ministério da Saúde, Brasil, 2014a e 2015.

* O medicamento Dolutegravir foi incorporado à lista de antirretrovirais distribuídos pelo ministério da saúde em Novembro de 2015 e estará disponível para a população a partir de 2016.

O prognóstico da infecção pelo HIV melhorou drasticamente graças aos medicamentos antirretrovirais, porém ainda há grandes desafios para a terapia medicamentosa. Resistência, toxicidade, complexidade de certos regimes e a adesão do paciente ao tratamento permanecem como grandes desafios. A baixa adesão é a principal causa de falha terapêutica, sendo um fator determinante

para os resultados de longo prazo. A perda de controle virológico como consequência da não-adesão pode levar ao surgimento de resistência total ou parcial aos antirretrovirais devido a mutações que levam a variações no genoma viral (SOSA, 2007; AFANI S e GALLARDO O, 2011; INFO, 2014). 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, constitui outro fator limitante da terapia (PEÇANHA *et al.*, 2002).

As falhas na aderência dos pacientes a TAR também ocorre devido a alta taxa de efeitos colaterais causados pelo uso dos antirretrovirais, sendo os mais comuns efeitos gastrointestinais como vômitos, náuseas, diarreias, dores abdominais, efeitos colaterais gerais como dispepsias, adinamias (redução da força muscular), astenias, além de efeitos neurológicos como cefaleias, insônia, parestesias periorais, e efeitos dermatológicos como pruridos, exantemas, modificações da cor da pele (LIGNANI JÚNIOR *et al.*, 2001; GIR *et al.*, 2005).

As complicações metabólicas também se tornaram uma das principais desvantagens da TAR. O acúmulo anormal de gordura visceral e subcutânea como lipoatrofia (diminuição da gordura nos membros e face) são duas complicações que são vistas com frequência elevada em pacientes recebendo ARVs. Esta distribuição irregular de gordura no corpo decorrente do uso prolongado dos medicamentos foi caracterizada como síndrome lipodistrófica do HIV. Do mesmo modo, o aumento de triglicérides e de colesterol nestes pacientes, e o aumento da resistência periférica à insulina, levando a diabetes mellitus (DM), leva ao aumento do risco de doença cardiovascular. Outras desordens metabólicas observadas foram o aumento da pressão diastólica, colesterol total, diminuição do HDL, aumento do PAI-1 (inibidor do ativador do plasminogênio) e tPA (ativador do plasminogênio tecidual) (SOSA, 2007; DE ALENCAR *et al.*, 2008).

Pelos fatores apresentados, fica evidente a necessidade da descoberta e/ou desenvolvimento de novos e eficientes fármacos ARVs que sejam efetivos para compor regimes terapêuticos adequados, menos tóxicos e com atividade contra vírus resistentes (SOSA, 2007; ESTE e CIHLAR, 2010).

3.6.1 Fármacos antirretrovirais

O tratamento da infecção pelo HIV tem evoluído continuamente. A disponibilidade de diferentes fármacos e seu uso combinado transformou o tratamento dos pacientes de tal modo que a morbidade e mortalidade declinaram drasticamente (BRODER, 2010). Tradicionalmente, a pesquisa para o desenvolvimento de novos antirretrovirais contra o HIV tem se concentrado na inibição dos sítios ativos enzimáticos virais (SERRAO *et al.*, 2013), porém o ciclo de replicação complexo deste vírus oferece muitas oportunidades para intervenção farmacológica (PEÇANHA *et al.*, 2002; ADAMSON e FREED, 2010; ESTE e CIHLAR, 2010).

Até o fim de 2014, vinte e oito medicamentos antirretrovirais pertencentes a seis classes farmacológicas diferentes foram aprovados para o tratamento do HIV/AIDS (PAU e GEORGE, 2014): inibidores da transcriptase reversa nucleosídeo-nucleotídeos (ITRN), inibidores da transcriptase reversa não-nucleosídeos (ITRNN), inibidores de protease (IP), inibidores da fusão (IF), antagonistas do co-receptor CCR5 (ACCR5) e inibidores da integrase (IIN) (BRODER, 2010; ESTE e CIHLAR, 2010; MEHELLOU e DE CLERCQ, 2010; BRASIL, Ministério da Saúde, 2014a). Dependendo do estado e da idade do paciente, a TAR é composta por no mínimo três antirretrovirais combinados, sendo dois medicamentos de classes diferentes (que poderão ser combinados em um só comprimido) (BRASIL, Ministério da Saúde, 2014a) de maneira a inibir duas etapas da replicação viral, podendo diminuir em até 100 vezes o ritmo de produção do vírus (MELO *et al.*, 2006; PORTELA e LOTROWSKA, 2006). Geralmente, o esquema terapêutico de primeira linha é composto por dois ITRNs e um terceiro agente, este último, geralmente pertence à classe dos ITRNNs ou um dos vários inibidores de protease. Medicamentos de classes farmacológicas mais recentes, como o raltegravir ou maraviroc, ou novos fármacos de classes existentes anteriormente, com perfis de resistência melhorados (por exemplo, darunavir, tipranavir e etravirina) são opções para pacientes que exibem resistência aos esquemas de tratamento de primeira linha (ESTE e CIHLAR, 2010).

3.6.2 Inibidores da transcriptase reversa nucleosídeo-nucleotídeos (ITRN)

O primeiro antirretroviral para uso clínico, o Retrovir® (Azidotimidina ou Azidovudina, AZT) (Figura 4 A), um ITRN, foi aprovado pelo *Food and Drug Administration* (FDA) americano em 1987 (CIHLAR e RAY, 2010; ESTE e CIHLAR, 2010). Após 27 anos, estão atualmente disponíveis para uso na terapêutica anti-HIV os seguintes medicamentos desta classe: didanosina, zalcitabina, estavudina, lamivudina, abacavir tenofovir disoproxil fumarato (pró-fármaco para a administração oral do tenofovir) e a emtricitabina (Figura 4) (PAU e GEORGE, 2014).

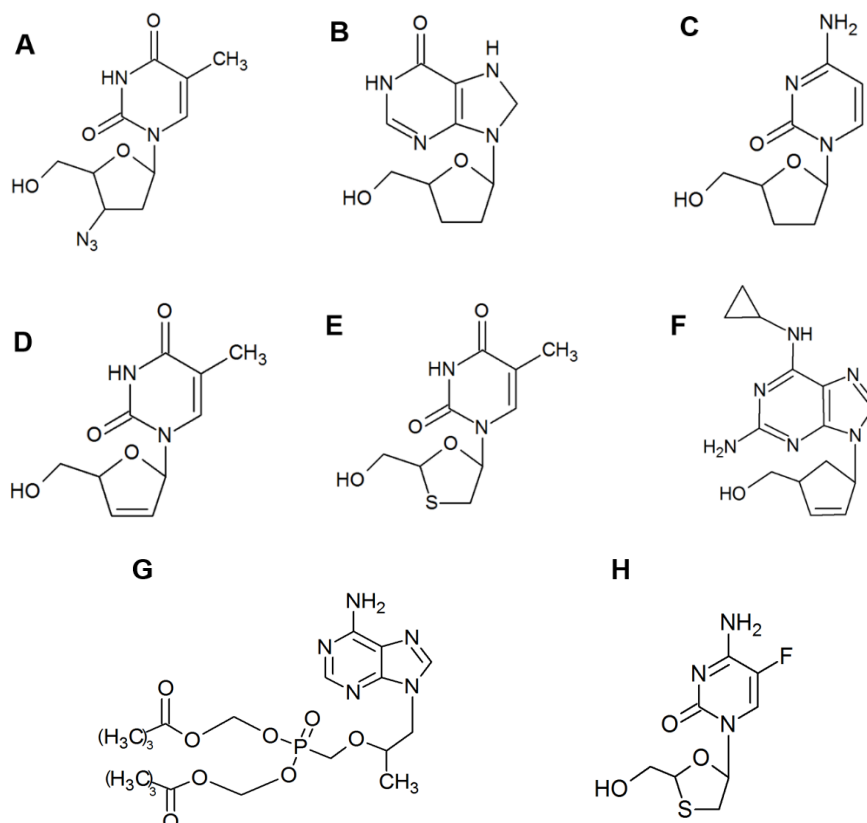


Figura 4 Estrutura química dos ITRN's em uso clínico. A) AZT; B) didanosina; C) zalcitabina; D) estavudina; E) lamivudina; F) abacavir; G) tenofovir disoproxil; H) emtricitabina.

O domínio catalítico da transcriptase reversa, heterodímero p66/p51, chamado de domínio “palma”, apresenta dois sítios de ligação que podem ser considerados alvos para intervenção quimioterápica. Os ITRNs agem no sítio de ligação do substrato, porém precisam sofrer fosforilação intracelular que os converte em sua forma 5'-trifosfato (Figura 5). Assim, eles competem com seus correspondentes endógenos para incorporação pela TR. Se o inibidor for incorporado à cadeia de DNA em formação, ele atua como um terminador de

cadeia. Já os ITRNNs agem em um sítio alostérico (PEÇANHA *et al.*, 2002; CIHLAR e RAY, 2010; MEHELLOU e DE CLERCQ, 2010).

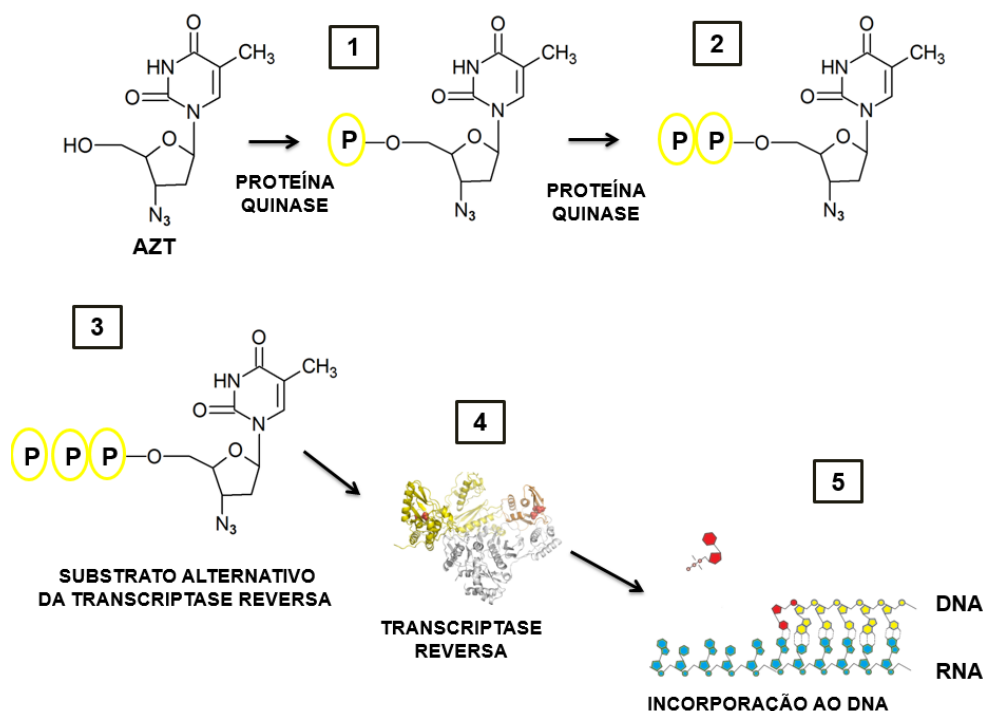


Figura 5 Fosforilação intracelular do AZT. Etapas 1 a 4: fosforilação do AZT por enzimas quinases intracelulares; 4 associação como substrato alternativo da TR; 5 incorporação ao material genético viral e interrupção da cadeia. Adaptado de PEÇANHA *et al.*, 2002; CIHLAR e RAY, 2010.

3.6.3 Inibidores da transcriptase reversa não-nucleosídeos (ITRNN)

Os ITRNNs (Figura 6) não necessitam de fosforilação intracelular para exercer sua ação farmacológica, uma vez que atuam em um sítio alostérico, sua principal vantagem em relação aos ITRNs. Esta resulta em uma mudança conformacional que leva a inativação enzimática. Porém, em virtude destes fármacos apresentarem praticamente o mesmo modo de ligação, mutações que levam à resistência a fármacos desta classe podem levar a resistência cruzada (PEÇANHA *et al.*, 2002; ESTE e CIHLAR, 2010; PAU e GEORGE, 2014).

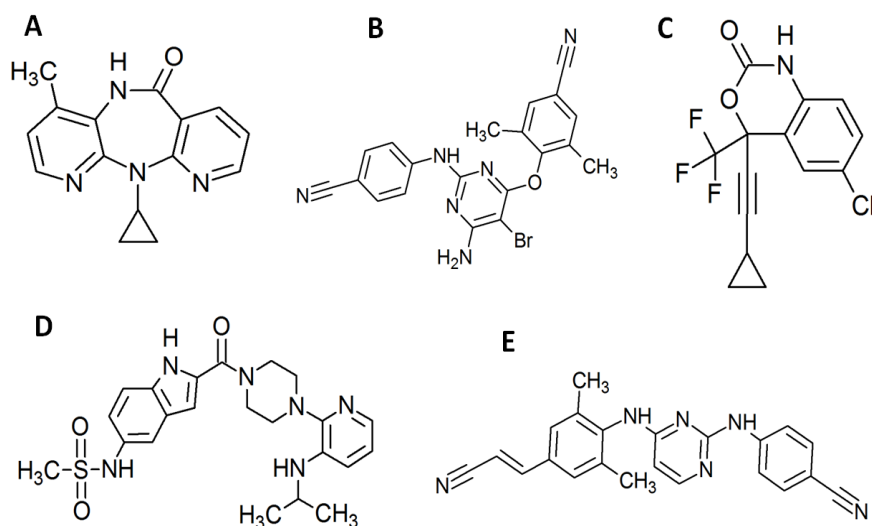


Figura 6 Estrutura química dos ITRNN's em uso clínico. A) nevirapina; B) etravirina; C) efavirenz; D) delavirdina, E) rilpivirina.

3.6.4 Inibidores de protease (IP)

Os IPs (Figura 7) exibem sua ação através da ligação com o sítio catalítico da protease, conduzindo ao bloqueio das atividades proteolíticas desta enzima. Assim, as partículas virais infecciosas maduras não são formadas, uma vez que a HIV protease é a responsável pelo processamento das poliproteínas virais, levando à formação das proteínas estruturais e funcionais do vírus (PEÇANHA *et al.*, 2002; MEHELLOU e DE CLERCQ, 2010; PAU e GEORGE, 2014).

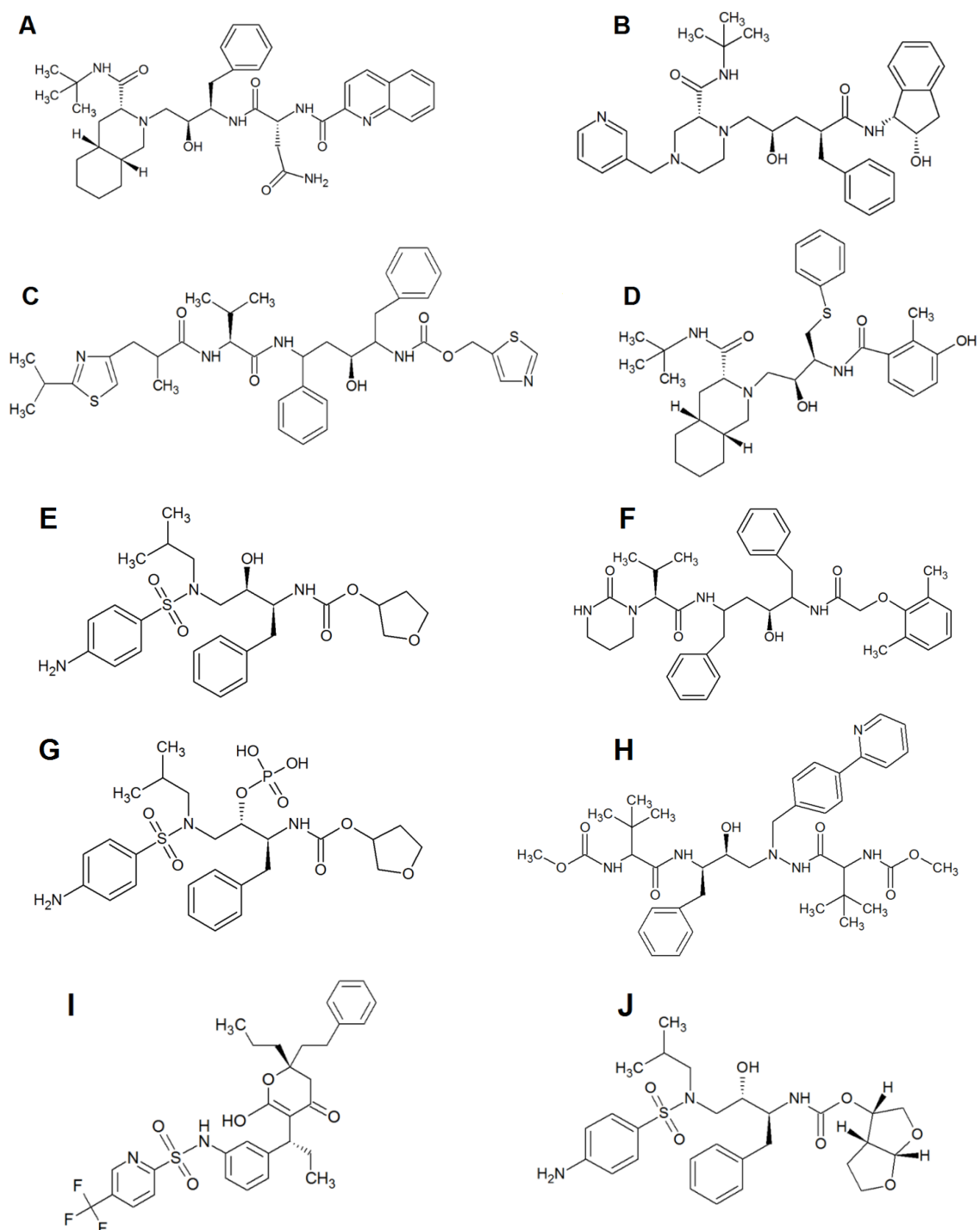


Figura 7 Estrutura química dos inibidores de protease em uso clínico. A) saquinavir; B) indinavir; C) ritonavir; D) nelfinavir; E) amprenavir; F) lopinavir; G) fosamprenavir; H) atazanavir; I) tipranavir; J) darunavir.

3.6.5 Inibidores da fusão (IF)

Uma das principais etapas da infecção por HIV é a interação entre a gp120 e o receptor na superfície celular CD4. Isto resulta em mudanças conformacionais que levam ao desmascaramento de um segundo sítio de

ligação na gp120 para co-receptores. O estabelecimento da ligação produz uma conformação ativa da proteína transmembrana gp41, que permite a fusão do vírus com a membrana celular (MEHELLOU e DE CLERCQ, 2010). O único IF aprovado é o enfuvirtide (Figura 8), o qual interfere com este processo através da ligação à gp41, de maneira a impedir que ocorram as mudanças conformacionais. Devido ao seu mecanismo de ação único, não há resistência cruzada com outros antirretrovirais. Entretanto, deve ser administrado por via subcutânea duas vezes ao dia, produzindo na maioria dos pacientes algum grau de reação no local da injeção, que se manifesta como dor, eritema, nódulos, equimoses, além de reações de hipersensibilidade (PAU e GEORGE, 2014).

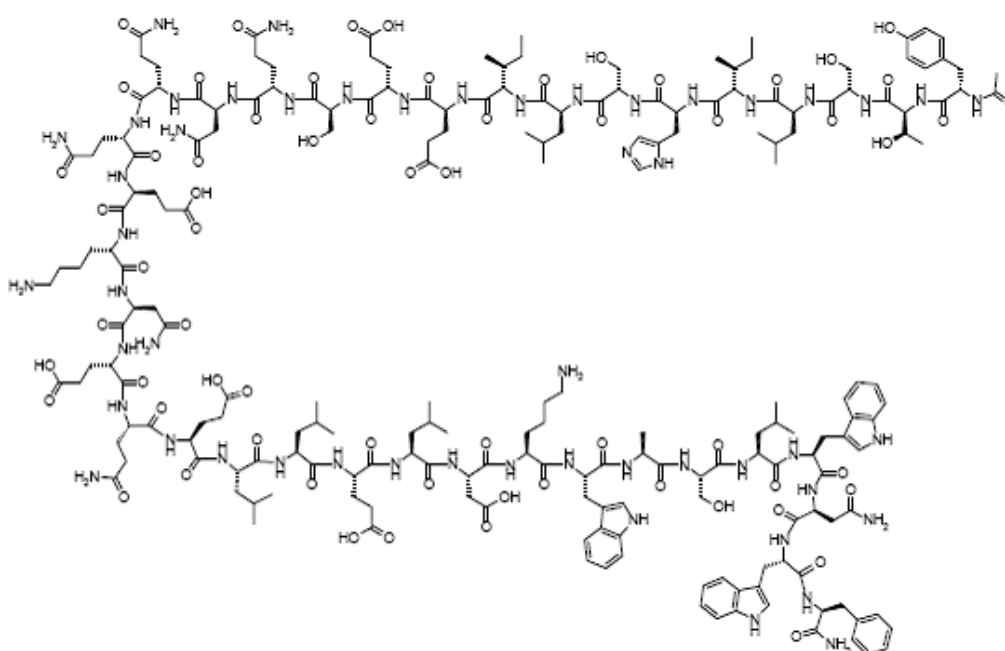


Figura 8 Estrutura química do enfuvirtide.

3.6.6 Antagonistas do co-receptor CCR5 (ACCR5)

A fusão também depende da ligação da gp120 a receptores de quimiocinas, ou co-receptores, sendo os mais importantes o CCR5 e CXCR4 (MEHELLOU e DE CLERCQ, 2010). O maraviroc (Figura 9), introduzido em 2007, é o único antagonista de CCR5 aprovado para utilização em pacientes sujeitos a cepas virais que utilizam como co-receptor o CCR5. Este fármaco liga-se seletivamente ao CCR5, e assim bloqueia a interação do HIV. No entanto, ele não bloqueia a entrada de vírus do HIV CXCR4-trópico ou cepas que utilizem tanto o CCR5 quanto o CXCR4 para entrada na célula hospedeira. Assim, deve

ser realizado um teste de tropismo antes de sua prescrição (PAU e GEORGE, 2014). Atualmente, é o único agente terapêutico que tem como alvo uma estrutura da célula hospedeira.

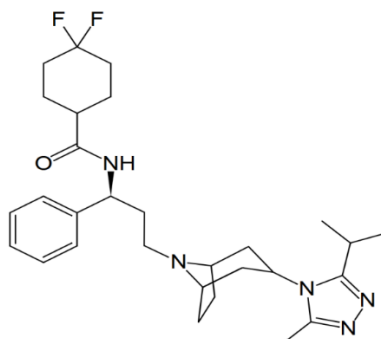


Figura 9 Estrutura química do maraviroc.

3.6.7 Inibidores da integrase

Os inibidores da integrase compõem a mais recente classe de fármacos anti-HIV aprovada para uso clínico no tratamento da AIDS, sendo a HIV-IN um dos alvos mais promissores para pesquisas de novos antirretrovirais (DI SANTO, 2014).

A incorporação do DNA proviral no DNA cromossômico da célula hospedeira é um elemento essencial no ciclo de replicação do HIV, este processo, conhecido como integração, é realizado por intermédio da HIV-IN. Os IINs bloqueiam esta enzima, ligando-se à interface do complexo IN-DNA, impedindo que seja realizada a formação de ligações covalentes entre o DNA hospedeiro e o DNA proviral, evitando a incorporação deste último (MEHELLOU e DE CLERCQ, 2010; PAU e GEORGE, 2014). Atualmente, existem apenas três inibidores da integrase licenciados pelo FDA: raltegravir (2007), elvitegravir (2012) e dolutegravir (2013) (Figura 10). Estes agentes são classificados como inibidores de transferência de fita da integrase (INSTI: *integrase strand transfer inhibitors*), os quais atuam quelando átomos de Mg^{2+} que a enzima utiliza como cofator (VOET *et al.*, 2011; CARCELLI *et al.*, 2014; PAU e GEORGE, 2014).

Os INSTI são bem tolerados e potentes, podendo diminuir rapidamente o conteúdo de RNA viral. Todos os fármacos desta classe, devido ao seu mecanismo de ação, podem ligar-se a cátions polivalentes e, assim, têm o

potencial de interação com medicamentos como antiácidos de magnésio, alumínio e cálcio (PAU e GEORGE, 2014).

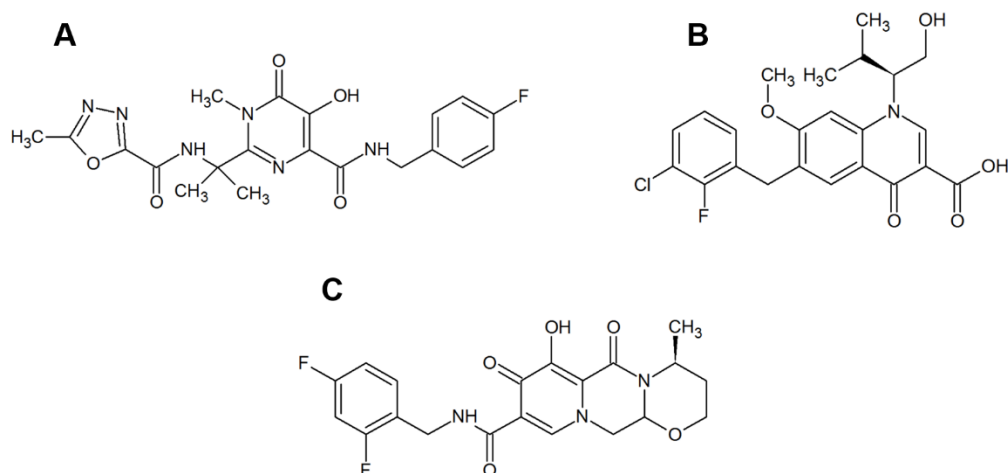


Figura 10 Estruturas químicas dos inibidores da HIV-IN em uso clínico. Em A raltegravir; B elvitegravir; C dolutegravir.

3.7 Integração e HIV-integrase como alvo para fármacos

O processo de integração é dividido em duas fases, ambas realizadas pela HIV-IN: a reação de processamento 3'(3'P), que corresponde à clivagem endonucleotídica de 2 nucleotídeos do 3'-terminal do cDNA viral, e a reação ST, etapa em que o cDNA é incorporado ao DNA da célula hospedeira. A integração ocorre dentro do complexo de nucleoproteínas, chamado de complexo de pré-integração (PIC), formado no citoplasma, após a transcrição reversa, pelo cDNA e proteínas virais como a HIV-IN e posteriormente transportado para o núcleo celular (VOET *et al.*, 2011; DI SANTO, 2014).

A HIV1-IN (Figura 11) é uma proteína de 32 kDa, composta por três domínios estruturais: N-terminal (NTD, resíduos 1 a 50), onde ocorre a ligação com Zn^{2+} , central catalítico (CCD, resíduos 51 a 212), que contém a tríade catalítica formada por três aminoácidos ácidos altamente conservados (ácidos aspárticos 64 e 116, e ácido glutâmico 152) responsáveis pelas reações 3'P e ST, e o C-terminal (CTD, resíduos 213 a 288), onde ocorre a ligação ao DNA (MELO *et al.*, 2006; VOET *et al.*, 2011; DI SANTO, 2014). As estruturas dos três domínios foram determinadas individualmente, além das estruturas do domínio CCD ligado aos domínios NTD e CTD. As interações entre estas subunidades

são altamente dinâmicas, uma propriedade que é essencial para a função biológica da HIV1-IN (DI SANTO, 2014).

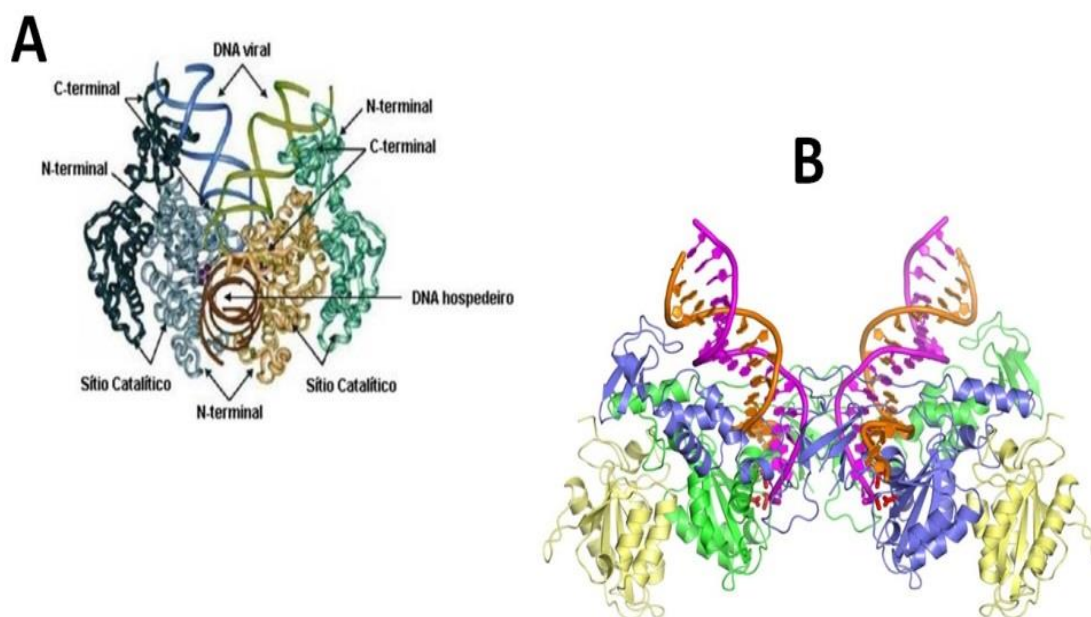


Figura 11 Em A: representação da estrutura da integrase de HIV (WIELENS *et al.*, 2005). Em B: representação tridimensional do complexo IN-DNA de PFV (*prototype foamy vírus*), código no Protein Data Bank (PDB) 3OS0 (MAERTENS *et al.*, 2010).

No domínio NTD estão dois resíduos de histidina e dois de cisteína, os responsáveis pela coordenação e ligação a átomos de Zn^{2+} (esta região é conhecida como “HHCC *zinc-binding motif*”), que seriam importantes, porém não essenciais, para a função biológica. Apesar de sua função biológica ainda não ter sido determinada com exatidão, o mais aceito é que o NTD interaja com o domínio catalítico, formando um multímero essencial para as reações 3’P e ST, além de facilitar a multimerização da integrase (ERIKETI *et al.*, 2007). Presente no domínio central catalítico, a tríade catalítica Asp 64, Asp 116 e Glu 152 apresenta grupos carboxilato carregados negativamente coordenados com dois íons metálicos Mg^{2+} , os quais atuam como cofatores enzimáticos. No domínio CTD ocorre a ligação ao DNA do hospedeiro de forma inespecífica. Embora a atividade catalítica ocorra primariamente devido a ação do domínio principal, tanto o CTD quanto o NTD são necessários para ambas as reações de 3’P e ST (VOET *et al.*, 2011; BLANCO e MARTINEZ-PICADO, 2012).

O processo de integração inicia-se no citoplasma, após a conversão do RNA viral em DNA. O material genético do HIV reúne sequências específicas de

quatro nucleotídeos (CAGT) dentro de regiões de repetições terminais longas (LTR: *long terminal repeat*,) em cada porção final da fita de DNA viral. A enzima reconhece estas sequências e catalisa a clivagem dos nucleotídeos GT das extremidades 3', originando um DNA processado com o final CAOH-3' (Figura 12). Esta reação ocorre através da hidrólise “*one-step*” da ligação fosfodiéster do dinucleotídeo 3'-CA. Para que este procedimento ocorra são necessários íons metálicos Mg^{2+} e um nucleófilo. Os átomos do metal ativam moléculas de água, as quais atuam como nucleófilo desta reação. Posteriormente, o grupo OH ativado ataca a ligação fosfodiéster no nucleotídeo adenosina e como produto desta reação têm-se CAOH -3'e 5'-GT-3' (ERIKETI *et al.*, 2007; DI SANTO, 2014).

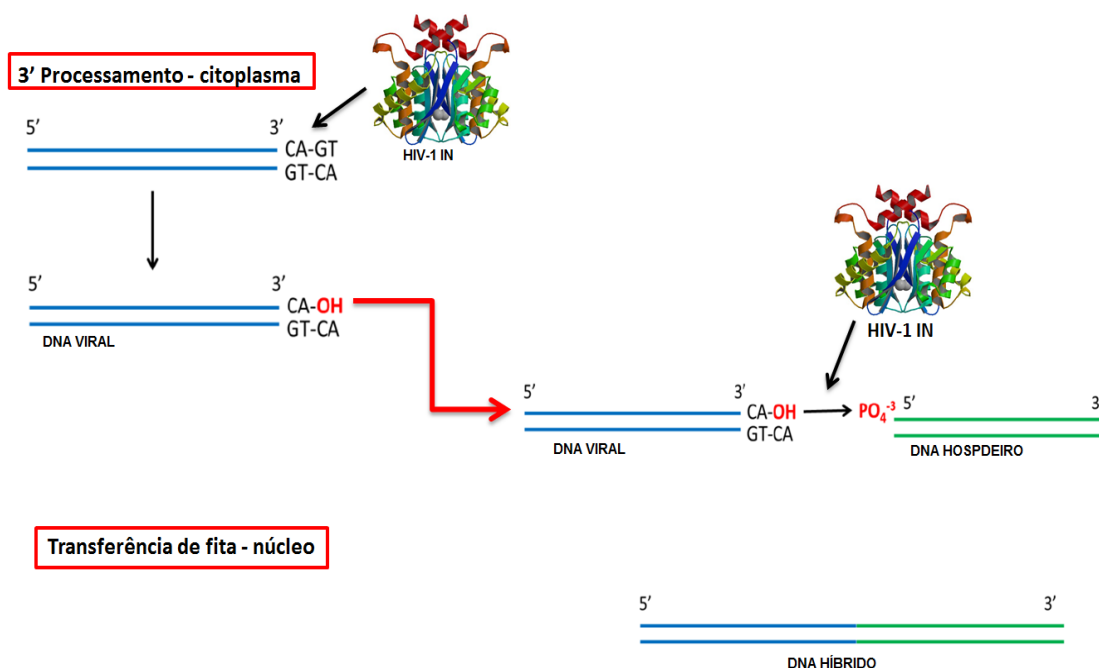


Figura 12 Etapas da integração.

O grupo hidroxila do terminal 3'-OH formado será o nucleófilo que a integrase necessita para atacar as ligações fosfodiéster em ambas a fitas do DNA hospedeiro, abrindo a cadeia e permitindo a incorporação do material genético viral, em uma reação de transferência de cadeia, ou transferência de fita (Figura 12), que ocorre no núcleo. Esta reação também necessita de átomos de Mg^{2+} para ocorrer. Neste processo, a HIV-IN catalisa a reação de junção das terminações 3'-OH do DNA viral com 5'-fosfato do DNA do hospedeiro, a qual é seguida de etapas de reparação e ligação (ERIKETI *et al.*, 2007; VOET *et al.*, 2011; DI SANTO, 2014). Os íons metálicos desempenham um duplo papel

durante a etapa ST: eles ajudam a estabilizar o complexo DNA-enzima e facilitam o fluxo de carga do grupamento viral 3'-OH para o grupo 5'-OH do DNA hospedeiro (DI SANTO, 2014) (Figura 13).

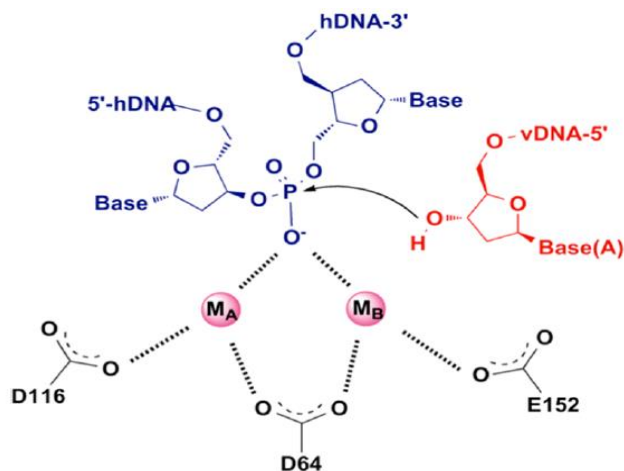


Figura 13 Representação da atividade dos átomos de Mg^{2+} durante a etapa de transferência de fita. Ataque da extremidades 3'-OH do DNA viral (em vermelho) à ligação fosfodiéster do DNA do hospedeiro (em azul) sendo coordenado por íons metálicos (DI SANTO, 2014).

O processo de obtenção de complexos cristalográficos da HIV-IN completa e ligada ao DNA permanece muito complexo, fazendo com que os estudos cristalográficos que auxiliam no entendimento da ação desta enzima têm se focado no uso de retrovírus relacionados. O mais empregado para este fim é o PFV. Apesar de ser inofensivo, a estrutura de sua integrase (Figura 14) é muito semelhante à do HIV, e é um excelente modelo para estudos de integração retroviral (MAERTENS *et al.*, 2010).

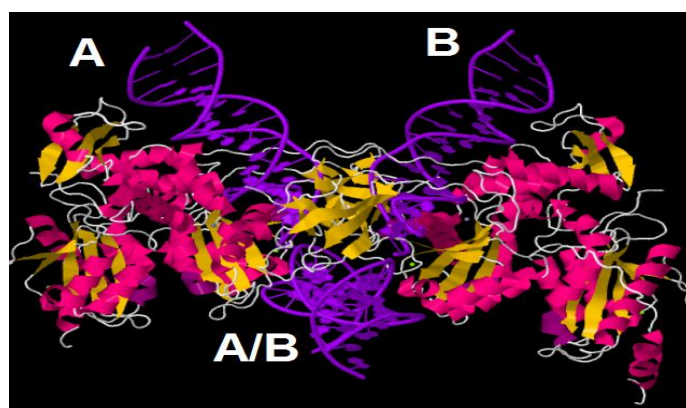


Figura 14 Demonstração do processo da integração realizado pela integrase. Em A o DNA do hospedeiro e em B DNA viral. Representação tridimensional do complexo IN-DNA de PFV (PDB 3OS0).

Os INSTI podem ser considerados uma classe de fármacos promissores, pois atuam em um alvo novo comparado aos primeiros medicamentos introduzidos na prática clínica, possuem atividade contra cepas de HIV-1 resistentes aos inibidores da HIV-TR e HIV-PR, além do fato de que células de mamíferos não apresentam enzimas integrase, o que torna essa classe altamente específica contra o vírus e, portanto, de baixo potencial para toxicidade (BLANCO e MARTINEZ-PICADO, 2012).

3.8 Mecanismo de inibição da HIV-IN pelos INSTI

A quelação do cofator Mg^{2+} , mecanismo pelo qual os INSTI exercem sua ação, provou ser uma estratégia de sucesso na concepção de inibidores desta enzima (CARCELLI *et al.*, 2014), de maneira que um grande número de compostos detentores desta atividade têm sido estudados como inibidores da HIV-IN, como os dicetoácidos (GOLDGUR *et al.*, 1999) (DCA's), e análogos como as naftiridonas (GARVEY *et al.*, 2008), além dos ácidos 4-quinolona-3-carboxílicos (DAYAM *et al.*, 2008), pirimidonas carboxamidas (ROGOLINO *et al.*, 2012) (Figura 15) e outras classes químicas relacionadas aos DCA's (VOET *et al.*, 2011)

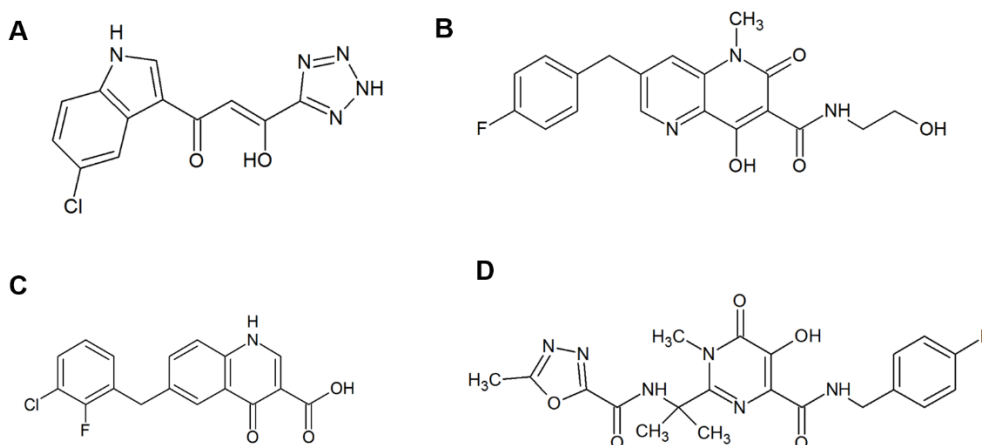


Figura 15 Estrutura química de compostos detentores de atividade anti HIV-IN de diferentes classes químicas. A) DCA; B) naftiridona; C) ácido 4-quinolona-3-carboxílico; D) pirimidona carboxamidas.

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, sendo que em algumas situações o equilíbrio pode ser deslocado a favor da forma enólica. Nas

β -dicetonas este equilíbrio tautomérico é deslocado a favor do enol (no caso, ceto-enol), pois ocorre a formação de uma ligação hidrogênio intramolecular entre o oxigênio carbonílico e o hidrogênio enólico (Figura 16) (RUSTICI *et al.*, 2006). Este grupamento, quando na presença de íons metálicos e de pH básico, pode ser facilmente desprotonado, gerando um diânion que pode interagir com dois íons divalentes (MAURIN *et al.*, 2004). Isto faz com que os DCAs e compostos relacionados consigam quelar os co-fatores da HIV-IN (CARCELLI *et al.*, 2014).

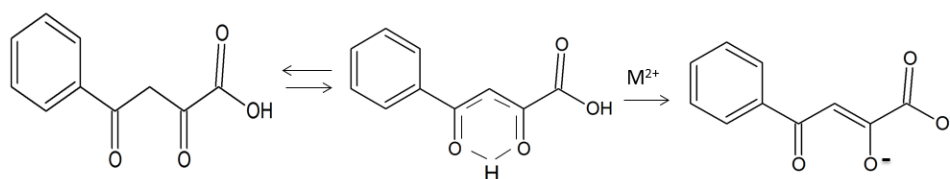


Figura 16 Estrutura do ácido benzopirúvico e representação do equilíbrio ceto-enólico deslocado a favor da forma enólica.

Outra característica estrutural importante para a inibição da reação ST pelos DCAs e análogos é a presença de uma cadeia lateral aromática ligado ao grupamento dicetona por uma cadeia flexível. A estrutura mais comum entre os inibidores mais potentes é um grupamento benzila ($-\text{CH}_2\text{-Ph}$) 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) e aos fármacos inibidores da HIV-IN já em uso (Figura 10) (CARCELLI *et al.*, 2014). Esta importância reside no fato de que esta subestrutura liga-se a HIV1-IN através da interação com uma alça desordenada na enzima formada pelos resíduos de aminoácidos 140 a 149. Essa ligação ocorre primariamente por uma interação hidrofóbica aromática com o resíduo Tyr143, sendo que a presença de mutações neste aminoácido em cepas virais leva à resistência ao Raltegravir. Na integrase de PFV este resíduo corresponde a Tyr 212, o qual auxilia a formação da alça correspondente (resíduos 209 a 218) neste vírus, assim como a Pro 214, conforme pode ser visualizado na Figura 18 (HARE *et al.*, 2010; METIFIOT *et al.*, 2011).

Além dessas características estruturais, é possível que outras interações também sejam necessárias para que ocorra a inibição da reação ST pelos DCA's e análogos, como ligações hidrogênio, que foram observadas em estudos de

docagem e dinâmica molecular (GOLDGUR *et al.*, 1999; HEALY *et al.*, 2009; ROGOLINO *et al.*, 2012). O farmacóforo dos DCA's é representado na Figura 17.

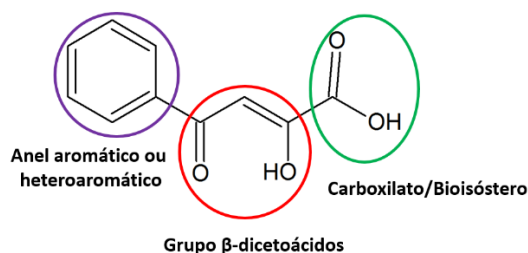


Figura 17 Farmacóforo dos DCA's, adaptado de ROGOLINO *et al.*, 2012.

A partir da estrutura dos DCAs, muitos estudos foram desenvolvidos, com diferentes classes de estruturas químicas análogas, até que o primeiro sucesso útil na prática clínica veio apenas em 2007, com a aprovação pelo FDA e lançamento do raltegravir (*N*-[(4-flúor-fenil)-metil]-1,6-di-hidro-5-hidróxi-1-metil 2-[1-metil-1-[[[(5-metil-1,3,4-oxadiazol-2-il)-carbonil]-amino]-etil]-6-oxo-4 pirimidino-carboxamida) (VOET *et al.*, 2011; ROGOLINO *et al.*, 2012; PAU e GEORGE, 2014), ou Isentress®, pela Merck. O esquema 2D simplificado da ligação do raltegravir ao sítio de ligação da PFV-IN é apresentado na Figura 18. As Figuras 18 a 20 apresentam os três fármacos da classe dos IN em uso clínico, destaca-se a ligação ao metal Mg^{2+} e a interação com o aminoácido Pro 214, as quais são essenciais para a inibição da reação ST pelos DCAs e análogos.

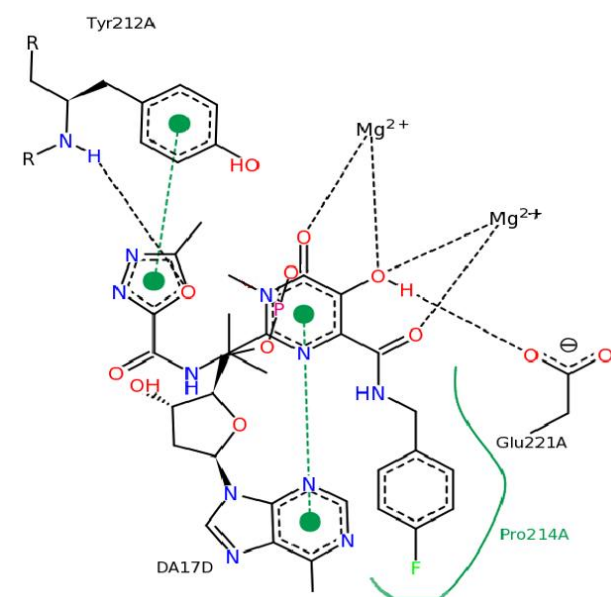


Figura 18 Esquema 2D da ligação do raltegravir ao sítio de ligação da PFV-IN. Figura gerada com o software *PoseView* (STIERAND *et al.*, 2006) a partir de PDB 3OYA.

Após o sucesso da pesquisa com o raltegravir, muitas outras tentativas de desenvolvimento de fármacos que atuassem pelo mesmo mecanismo foram realizadas utilizando estruturas bioisósteras dos DCAs e através da estratégia de simplificação molecular. Estes estudos levaram a descoberta do elvitegravir, segundo fármaco da primeira geração dos inibidores da HIV-IN, (6-[(3-cloro 2-flúor-fenil)metil]-1-[(2S)-1-hidróxi-3-metil-butan-2-il]-7-metóxi-4oxo-1,4 dihidroquinolina- 3 ácido carboxílico), o qual é um monocetoácido (MCA), resultante de modificações no núcleo DCA (DI SANTO, 2014). O esquema 2D simplificado da ligação do elvitegravir ao sítio de ligação da PFV-IN é apresentado na Figura 19.

Posteriormente, foi lançado no mercado o primeiro fármaco da segunda geração de inibidores da HIV-IN, o dolutegravir (3S,7R)-N-[(2,4-difluórfenil)-metil]-11-hidróxi-7-metil-12-oxo-4-oxa-1,8-diazatriciclo[8.4.0.0]tetradeca-10,13-dieno-13-carboxamida) (DI SANTO, 2014). A Figura 20 apresenta o esquema 2D simplificado da ligação do dolutegravir ao sítio de ligação da PFV-IN.

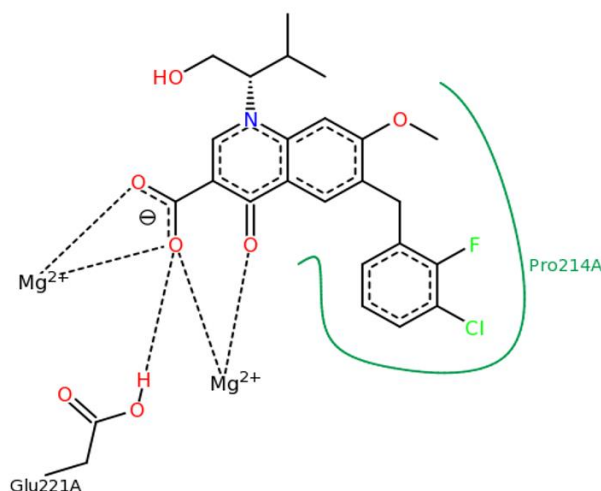


Figura 19 Esquema 2D da ligação do elvitegravir ao sítio de ligação da PFV-IN. Figura gerada com o software *PoseView* (STIERAND *et al.*, 2006) a partir de PDB 3L2U.

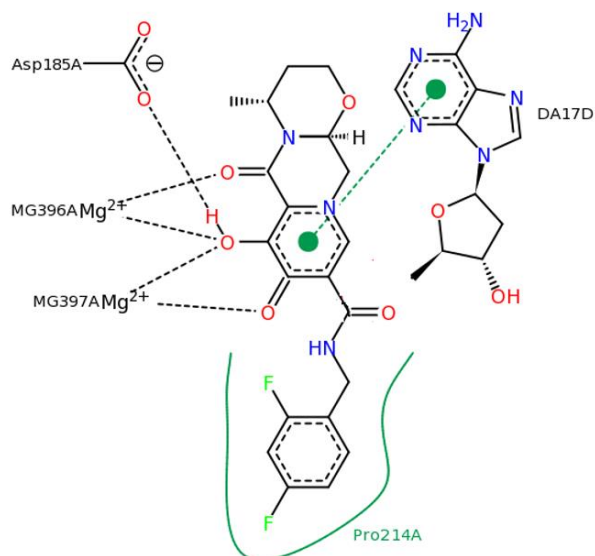


Figura 20 Esquema 2D da ligação do dolutegravir ao sítio de ligação da PFV-IN. Figura gerada com o software *PoseView* (STIERAND *et al.*, 2006) a partir de PDB 3L2U.

Além da inibição da etapa de ST, outras estratégias têm sido pesquisadas com intuito de promover inibição do processo de integração: (i) inibição da interação do cDNA ao sítio ativo da enzima, através da ligação à extremidade do cDNA (antagonistas da ligação da HIN-IN ou do PIC ao DNA viral) (PANNECOUQUE *et al.*, 2002; VOET *et al.*, 2011); (ii) inibição da multimerização da HIV-IN, impedindo assim que a enzima desenvolva sua atividade biológica (VOET *et al.*, 2011); (iii) inibição da interação da HIV-IN com cofatores celulares, como a LEDGF/75, uma proteína que atua como um determinante do sítio de integração no genoma viral (ERIKETI *et al.*, 2007; VOET *et al.*, 2011; BHATT *et al.*, 2014).

3.9 Pirimidonas/Pirimidinas/Piridopirazinas

Pirimidonas são compostos heterocíclicos contendo o núcleo base representado na Figura 21, sendo provenientes de estruturas chamadas de pirimidinas. Este núcleo é uma das estruturas heterocíclicas de maior relevância biológica, aparecendo com destaque nas estruturas dos ácidos nucleicos e em muitos compostos farmacologicamente ativos (DHUGURU *et al.*, 2011; PATEL *et al.*, 2014). As pirimidinas compõem uma classe de compostos de enorme interesse para as pesquisas na área do desenvolvimento de novos fármacos com

ação anti-HIV. Vários antirretrovirais já em uso são derivados destes compostos através de inserções de diferentes substituintes em seu núcleo básico, como por exemplo, o AZT e o raltegravir (PATEL *et al.*, 2014).

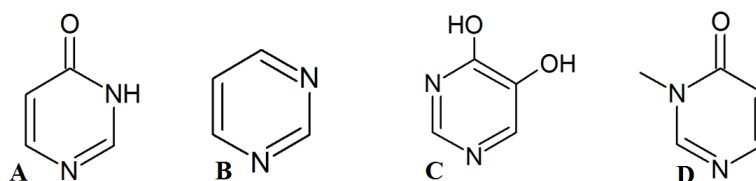


Figura 21 A) Núcleo base das pirimidonas; B) pirimidinas; C) dihidroxipirimidinas; D) *N*-metilpirimidona.

Focando-se na inibição do processo de transferência de fita, após a descoberta dos DCAs foi descoberto que um derivado de pirimidona contendo o farmacóforo dos DCA's (Figura 22) apresentava ação inibitória sobre a HIV-IN (PATEL *et al.*, 2014). Posteriormente, algumas pirimidinas (dihidroxipirimidinas) carboxamidas e *N*-metilpirimidonas (Figura 23) foram testadas para avaliação da capacidade inibitória sobre a HIV-IN, mostrando-se potentes. A presença do grupamento carboxamidas (Figura 23) provou ser de grande valia para atividade anti-HIV das estruturas contendo o farmacóforo DCA. A partir desses núcleos bases supracitados, muitas variações com diferentes substituições foram realizadas, gerando diversas estruturas como as pirimidonas bicíclicas, pirimidonas com anéis tiazólicos, variações de dihidroxipirimidinas, entre outros compostos com atividade anti HIV-IN, levando em última análise a descoberta do raltegravir (MAERTENS *et al.*, 2010; PATEL *et al.*, 2014).

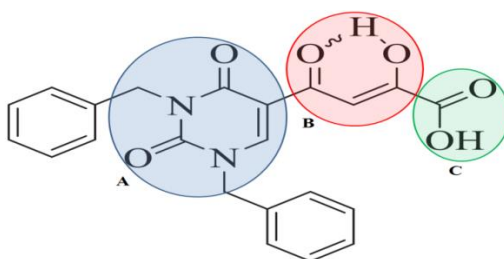


Figura 22 Composto pirimidínico baseado em β -dicetoácido. Em A porção pirimidínica, em B porção β -dicetoácida, em C carboxila como um bioisóstero presente no farmacóforo dos DCA's. Adaptada de Patel e colaboradores (PATEL *et al.*, 2014).

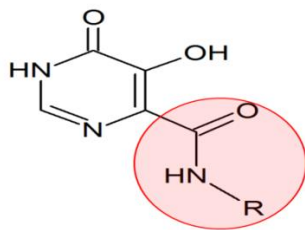


Figura 23 Estrutura com inserção do grupamento carboxamidas à pirimidona.

Piridopirazinas são estruturas hetero-bicíclicas que possuem o núcleo básico demonstrado na Figura 24. Dihidroxipiridopirazinas (Figura 24) são estruturas derivadas das piridopirazinas, e apresentam atividade anti HIV-IN pelo mecanismo de inibição da transferência de fita, sendo compostos relacionados aos DCAs (WAI *et al.*, 2007).

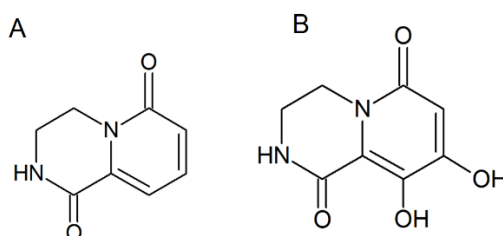


Figura 24 Estrutura básica das piridopirazinas em A; dihidroxipiridopirazinas em B.

3.10 Planejamento de fármacos auxiliado por computador

O desenvolvimento de um novo fármaco envolve estudos altamente complexos e interdisciplinares. Por muito tempo, este processo era realizado apenas com ensaios biológicos e modificações de estruturas químicas utilizando síntese orgânica ou a partir de produtos naturais. Atualmente, além das técnicas clássicas de síntese e ensaios biológicos, métodos como Cristalografia de Proteínas, Ressonância Magnética Nuclear (RMN), Ensaios em Larga Escala (*High-Throughput Screening*), Química Combinatória, e vários outros, são de larga utilização no desenvolvimento de novos compostos químicos. Entre eles encontra-se os métodos de Planejamento de Fármacos Auxiliado por Computador (CADD: *Computer-Aided Drug Design*), área que envolve uma série de técnicas computacionais úteis para descobrir, planejar e aperfeiçoar *in silico* compostos biologicamente ativos, com a finalidade de serem utilizados como novos fármacos (SANTOS FILHO e ALENCASTRO, 2003; BARREIRO *et al.*, 1997; SANT'ANNA, 2002; LIMA, 2007).

As pesquisas de CADD sobre inibidores de HIV-IN evoluíram após cristalização e depósito no PDB do complexo 1QS4 (GOLDGUR *et al.*, 1999). Além de ter sido muito útil para confirmar a região onde se localiza a tríade catalítica, este complexo ajudou a aumentar a compreensão sobre a forma como esta enzima age e sobre o mecanismo de ação dos inibidores. Também foi possível a realização dos primeiros estudos de docagem molecular, uma vez que o arquivo cristalográfico se tornou disponível publicamente (BARRECA e MONFORTE, 2006; ROGOLINO *et al.*, 2012; BHATT *et al.*, 2014). Abordagens de CADD vêm sendo utilizadas há anos para pesquisar os vários aspectos necessários para o planejamento racional de inibidores da HIV-IN.

Além dos trabalhos baseados na estrutura do alvo molecular supracitados, vários estudos QSAR (*Quantitative Structure-Activity Relationship*) foram realizados com diversas classes de compostos, 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 (MELO *et al.*, 2006), como os estudos de Raghavan e colaboradores (1995), Mahindra e Kulkarni (2002), Saíz-Urra e colaboradores (2007), Magalhães e colaboradores (2013) e Bhatt e colaboradores (2014).

A realização dos estudos *in silico* voltados para o planejamento de novos fármacos vem despertando o interesse de autoridades regulatórias e indústrias. O estímulo para a realização destes estudos visa, basicamente, a redução de custos gerais e aumento da velocidade na obtenção de bons resultados, a diminuição da utilização de animais de laboratório, e um melhor gerenciamento ambiental dos sistemas produtivos, graças à geração de uma menor quantidade de resíduos químicos e biológicos decorrentes dos processos de síntese e ensaios biológicos (ERIKSSON, L. *et al.*, 2003).

3.10.1 QSAR

Uma das abordagens mais utilizadas em CADD para o planejamento de novas moléculas candidatas a fármacos e para predição do potencial de impacto ambiental de substâncias químicas são os estudos QSAR. Esta abordagem 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 (FERREIRA *et al.*, 2002; TAVARES, 2004).

Estes modelos, quando obtidos com sucesso, são capazes de explicar as complexas relações entre as variáveis independentes (VI) e dependentes (VD) em estudo. No caso em questão, as VIs são descritores moleculares que codificam aspectos estruturais e propriedades físico-químicas de interesse. Já as VD correspondem a uma ou mais atividades biológicas. Ferramentas matemáticas são empregadas com intuito de obter equações que descrevam as variações na propriedade de interesse em função dos descritores, e desta maneira, possam ser utilizadas para realizar a predição de uma atividade biológica, ou mesmo outra propriedade físico-química de interesse. Posteriormente à obtenção das equações (modelos), estas podem ser empregadas para orientar a síntese de novas moléculas com propriedades otimizadas, pois de antemão ter-se-á a provável potência terapêutica ou toxicidade do mesmo, o que indicará se pode haver interesse ou não em sua preparação. Como os estudos QSAR consideram diferentes moléculas com diferentes potências perante um mesmo sistema biológico, os modelos de predição que utilizam os descritores consideram o quanto a atividade está variando em função destes descritores, que por sua vez dependem basicamente das variações nas estruturas químicas em estudo (FERREIRA *et al.*, 2002; KUBINYI, 2008; MONTANARI, 2011).

Atualmente, em estudos QSAR, estão disponíveis para uso diferentes tipos de descritores moleculares. A Tabela 2 apresenta as principais classes de descritores moleculares, de acordo com a classificação e segundo a metodologia utilizada para obtê-los (MONTANARI, 2011).

Tabela 2 Classificação dos descritores moleculares.

Classe	Exemplo de descritores
Constitucionais	Massa molar, número de elétrons de valência
Físico-químicos	Ponto de fusão, solubilidade
Topológicos	Índices de Randic, índices de kier
Geométricos (estruturais)	Comprimentos de ângulos de ligação, volume de Van der Waals
Eletrônicos	Cargas atômicas, momento dipolar
Combinados	Densidade de carga superficial, índices eletrotológicos

Fonte: MONTANARI, 2011.

3.10.1.1 Métodos de seleção de variáveis/construção dos modelos QSAR

Em QSAR, normalmente são obtidos descritores em número muito elevado. Assim, é necessário utilizar metodologias que selecionem aqueles que mais contribuem para a atividade biológica, os quais farão parte do modelo (MARTINS, 2010). O processo de seleção de variáveis baseia-se em encontrar, dentre aquelas selecionadas para estudo, aqueles sub-conjuntos capazes de produzir modelos matemáticos que descrevam adequadamente os valores observados da atividade biológica (FERREIRA *et al.*, 2002).

Existem vários métodos disponíveis na literatura para a realização da seleção de variáveis, estes são empregados em conjunto com as metodologias de obtenção dos modelos matemáticos (FERREIRA *et al.*, 2002; MONTANARI, 2011). Para a seleção, as ferramentas mais empregadas são os métodos de busca sistemática, algoritmo genético e quimiométricos.

Em relação à obtenção dos modelos, os métodos mais empregados são os baseados em regressão linear, como exemplo tem-se a regressão linear múltipla (*multiple linear regression*, MLR), regressão por quadrados mínimos parciais (*partial least-squares*, PLS), regressão por componentes principais (*principal component regression*, PCR), e métodos não-lineares como os baseados em redes neurais artificiais (MARTINS e FERREIRA, 2013).

Historicamente, a regressão multivariada era feita usando MLR, o qual tem bom desempenho quando o número de descritores é menor do que o

número de compostos. Porém, programas modernos de modelagem usados em estudos de QSAR geram um elevado número de descritores que frequentemente são altamente correlacionados entre si, especialmente em análises de QSAR-3D a 6D. Levando isto em consideração, o método MLR não pode ser usado nesses casos, a menos que se faça uma seleção de variáveis criteriosa. Para evitar esses problemas, são utilizados métodos de projeção, também conhecidos como métodos de regressão bilineares, como o PCR ou PLS. Quando esses métodos são aplicados, o número de descritores e as correlações entre eles deixam de ser um problema (MARTINS e FERREIRA, 2013).

3.10.1.2 Validação dos modelos QSAR

Para que uma equação de regressão seja usada como um modelo matemático de previsão é preciso que ela seja validada através de diversos testes estatísticos. Esta validação deve ser realizada antes de sua interpretação e utilização, para que atividades biológicas ou outras características de interesse sejam previstas de maneira confiável. Os testes estatísticos devem ser usados para avaliar a capacidade de previsão dos modelos em relação aos compostos que auxiliaram na construção destes, processo denominado validação interna, e também a compostos que não foram empregados em sua obtenção, ou validação externa (KIRALJ e FERREIRA, 2009).

Em relação à validação interna (Tabela 3), os testes estatísticos avaliam o grau de ajuste, significância e previsibilidade. O grau de ajuste do modelo é avaliado através dos parâmetros R^2 (coeficiente de determinação múltipla) e do $RMSEC$ (raíz quadrada do erro da calibração). Para avaliação da significância é aplicado o teste F (Teste de Fischer). O grau de previsibilidade do modelo, na validação interna, é testado pela validação cruzada através da exclusão individual de uma amostra, reconstrução do modelo sem esta, e finalmente o uso do modelo para calcular o valor da amostra excluída. Ao final deste processo é possível obter o Q^2_{LOO} (coeficiente de determinação entre as atividades biológicas observadas e previstas durante o processo de validação cruzada *leave-one-out*, LOO), o qual representa a quantidade de informação ou variabilidade que o modelo pode prever. Na validação cruzada é também

calculado o RMSECV (raiz quadrada do erro da validação cruzada), que deve apresentar o menor valor possível (FERREIRA *et al.*, 2002).

Na validação interna também são realizados os testes de validação cruzada *leave-N-out* (LNO), visando medir a robustez do modelo. Nesse teste 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, podendo-se, então calcular o Q^2_{LNO} (coeficiente de correlação entre as atividades biológicas observadas e previstas durante o processo de validação cruzada *leave-N-out*). Além disto, faz-se a análise da possibilidade de que as variabilidades explicadas e previstas pelo modelo devem-se a ocorrências ao acaso (KIRALJ e FERREIRA, 2009; MARTINS, 2010b; MARTINS e FERREIRA, 2013). Com este intuito, emprega-se o teste de randomização do y (*y-randomização*), em que se constroem modelos paralelos com os valores de atividade biológica trocados, enquanto que os descritores originais são mantidos inalterados. Desta maneira, espera-se que os modelos paralelos construídos nestas condições sejam de qualidade ruim e com valores de Q^2 bem menores do que o valor obtido para o modelo original (MARTINS e FERREIRA, 2013).

Tabela 3 Parâmetros estatísticos empregados na validação interna.

PARÂMETRO	SIGNIFICADO	EQUAÇÃO
R^2	Coeficiente de determinação ^a	$R^2 = 1 - \frac{\sum_i (y_{obsi} - y_{ci})^2}{\sum_i (y_{obsi} - \bar{y}_{ci})^2}$
$RMSEC$	Raiz quadrada do erro da calibração ^a	$RMSEC = \sqrt{\frac{\sum_i (y_{obsi} - y_{ci})^2}{n}}$
F	Teste F (com 95% de confiança, $\alpha = 0,05$) ^b	$F = \frac{\frac{\sum_i (y_{obsi} - y_{ci})^2}{k}}{\frac{\sum_i (y_{obsi} - \bar{y}_{obs})^2}{n - p - 1}}$
Q^2_{LNO}	Coeficiente de determinação da validação cruzada ^{c,d}	$Q^2_{LNO} = 1 - \frac{\sum_i (y_{obsi} - y_{vi})^2}{\sum_i (y_{obsi} - \bar{y}_{obs})^2}$
$RMSECV$	Raiz quadrada do erro da validação cruzada ^c	$RMSECV = \sqrt{\frac{\sum_i (y_{obsi} - y_{vi})^2}{n}}$

^a qualidade do ajuste; ^b significância; ^c validação cruzada, se $n = 1$, e teste de robustez *leave-N-out*, se $n > 1$; ^d qualidade da predição interna; y: atividade biológica; $\bar{y}$: atividade biológica média; obs: valores experimentais; ci: atividade estimada no modelo de regressão para o conjunto de treinamento completo; vi: valor obtido na validação cruzada LOO; 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.

A validação externa avalia a capacidade de predição das atividades biológicas de compostos que não foram utilizados para construção do modelo. Desta maneira, pode-se comparar as atividades previstas com o valor real. Com este intuito, escolhe-se uma alíquota de amostras que não serão incluídas no processo de construção do modelo, ou conjunto de teste. Enquanto que, as amostras que farão parte da construção do modelo são nomeadas conjunto de treinamento. Assim, constrói-se um modelo com as moléculas do conjunto de treinamento e a atividade biológica das amostras do conjunto de teste é calculada pelo modelo construído (também pode ser referido como modelo real). Como a atividade biológica real das amostras do conjunto de teste é conhecida, pode-se fazer uma comparação entre o valor previsto pelo modelo e o valor real utilizando-se parâmetros estatísticos (GOLBRAIKH *et al.*, 2003; MARTINS, 2010b). Desta maneira, a validação externa é muito mais confiável para assegurar a capacidade preditiva do modelo quando comparado com a validação cruzada, justamente pelo fato de que os compostos não são usados para a construção do modelo (GOLBRAIKH e TROPSHA, 2002; TROPSHA, 2010).

Para avaliar a capacidade de previsão do modelo na validação externa, um dos parâmetros mais empregados é o coeficiente de determinação da validação externa (R^2_{pred}). 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. Gobraikh e Tropha (2003) sugeriram 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 (dados experimentais) e preditos pelo modelo (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.

Além dos testes citados acima, Roy e colaboradores (2008) sugeriram uma nova métrica de validação externa semelhante ao conceito de Golbraikh e Tropsha (CHIRICO e GRAMATICA, 2011), o r_m^2 (coeficiente de correlação modificado). Os autores afirmam que mesmo com um R^2_{pred} com valor adequado

pode ocorrer que os valores preditos e observados não sejam próximos, pois existe a possibilidade de haver considerável diferença numérica entre estes, entretanto, manter uma boa correlação no geral (ROY *et al.*, 2013).

O r_m^2 é calculado com base na correlação entre os dados de resposta observados e previstos do conjunto teste com intercepto (r^2) e sem a intercepção (r_0^2) (ROY *et al.*, 2013) (Figura 25). A partir de r_m^2 , através da troca dos eixos das atividades observadas e previstas calcula-se $r_m'^2$. Em seguida, faz-se a análise da capacidade de predição externa utilizando a média de r^2 modificado ($Average\ r_m^2$) e variação de r_m^2 (Δr_m^2), propostos por Roy *et al.* (2012).

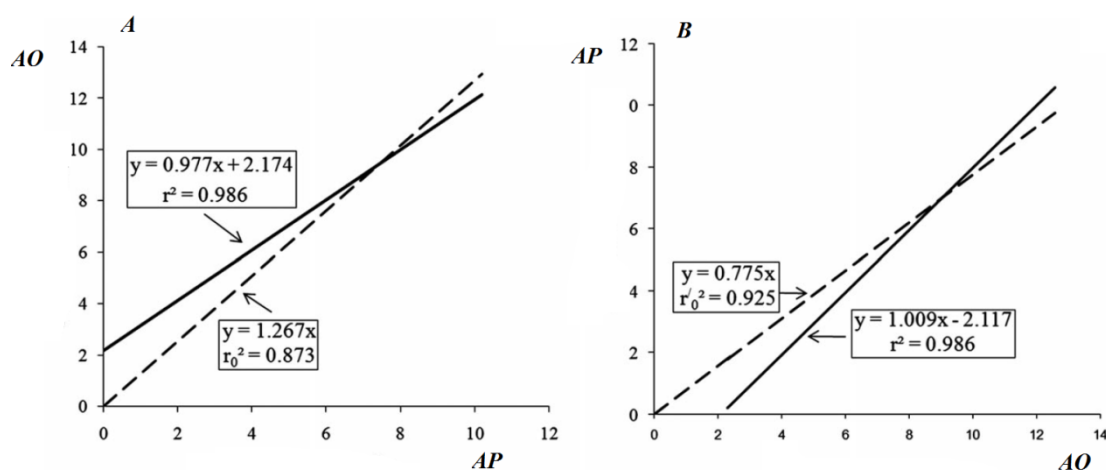


Figura 25 Regressões para os resultados obtidos em um conjunto arbitrário de dados com e sem intercepção: (A) dados previstos são plotados ao longo do eixo X, enquanto os dados observados são plotados ao longo do eixo Y; (B) eixos são trocados. *AO: atividade observada; AP: atividade prevista. Adaptado de ROY *et al.*, 2013.

A média destes dois últimos parâmetros resulta no $average\ r_m^2$, enquanto que a diferença entre estes resulta na variação de r_m^2 (Δr_m^2). Estes parâmetros avaliam de maneira bastante rigorosa a capacidade preditiva dos modelos QSAR, determinando a proximidade entre a atividade observada experimentalmente e a prevista. O uso destes critérios visa garantir uma mínima confiabilidade, qualidade e eficácia dos modelos de regressão para fins práticos (KIRALJ e FERREIRA, 2009; DE MELO, 2012; ROY *et al.*, 2013). Os parâmetros mais empregados para avaliar a capacidade preditiva dos modelos, usados na validação externa estão listados na tabela 4.

Tabela 4 Parâmetros estatísticos empregados na validação externa.

PARÂMETRO	SIGNIFICADO	EQUAÇÃO
R^2_{pred}	Coeficiente de determinação da validação externa ^e	$R^2_{pred} = 1 - \frac{\sum_i (y_{obsi} - y_{evi})^2}{\sum_i (y_{obsi} - \bar{y}_{obs})^2}$
k e k'	Inclinações das linhas das regressões lineares de predição ^e	$k = \frac{\sum_i (y_{obsi} - y_{evi})}{\sum_i y_{evi}}$ $k' = \frac{\sum_i (y_{obsi} - y_{evi})}{\sum_i y_{obsi}}$
$ R^2_0 - R'^2_0 $	Diferença absoluta entre os coeficientes de determinação múltipla ^e	$R_0^2 = \left(1 - \frac{\sum (y_{obsi} - k y_{evi})^2}{\sum (y_{obsi} - \bar{y}_{evi})^2} \right)$ $R'^2_0 = \left(1 - \frac{\sum (y_{obs} - k y_{obs})^2}{\sum (y_{obs} - \bar{y}_{obs})^2} \right)$
Average r_m^2	Média de r^2 modificado ^e	$Média\ r_m^2 = \frac{(r_m^2 + r'^2_m)}{2}$
Δr_m^2	Variação de r^2 modificado ^e	$\Delta r_m^2 = r_m^2 - r'^2_m $

^e validação externa; obs: valores experimentais; $\bar{y}_{obsi}$: atividade média observada para o conjunto de treinamento completo; evi: valor obtido na validação externa; r_m^2 : coeficiente de correlação modificado obtido a partir da regressão com dados observados e preditos; r'^2_m : coeficiente de correlação modificado obtido da regressão com a troca de eixos.

4. METODOLOGIA

4.1 Conjunto de dados

Para este estudo, foram utilizados 199 inibidores de HIV-IN previamente disponibilizados na literatura, sendo este conjunto formado por: (i) derivados de pirimidonas previamente sintetizados e testados por FERRARA *et al.*, 2010, MURAGLIA *et al.*, 2008, GARDELLI *et al.*, 2007, DI FRANCESCO *et al.*, 2008, NIZI *et al.*, 2009, PETROCCHI *et al.*, 2009 e DONGHI *et al.*, 2009; (ii) derivados de pirimidinas descritos por PACE *et al.*, 2007, SUMMA *et al.*, 2006 e PETROCCHI *et al.*, 2007; (iii) derivados de piridopirazinas descritos por WAI *et al.*, 2007. Todos estes compostos apresentaram a capacidade de inibir o processo de transferência de fita (ST) realizado pela HIV-1 integrase. Moléculas que se apresentavam como a forma de misturas racêmicas foram excluídas. Todos os compostos foram ensaiados pela metodologia descrita por HAZUDA *et al.*, 1997. Através de uma análise minuciosa das estruturas químicas presentes em cada uma das referências citadas constatou-se que todas apresentavam o mesmo farmacóforo dos DCAs apresentada na Figura 17.

A atividade inibitória enzimática foi medida quanto à concentração necessária (em nanomolar, nM, ou micromolar, μ M, dependendo da referência) para reduzir em 50% a reação de ST (IC_{50}). Os valores observados de IC_{50} foram convertidos em seus correspondentes $-\log IC_{50}$ (ou pIC_{50}), e desta maneira as atividades ficaram distribuídas dentro do intervalo de 3.067 unidades logarítmicas (pIC_{50} 5.602 a 8.699).

O conjunto de dados foi dividido em dois subconjuntos. O primeiro foi destinado a modelagem dos dados (conjunto de treinamento), com 54 compostos (ver material suplementar artigo, destaques nas Tabelas 1 a 9) selecionados de modo a representar adequadamente a variabilidade estrutural e a faixa de atividade biológica do conjunto de dados. O segundo, com os demais compostos, foi utilizado como um conjunto de predição, visando avaliar o potencial dos modelos como ferramenta de apoio a síntese de novos derivados, levando em conta que este é o objetivo principal de estudos QSAR aplicados a planejamento de novos fármacos.

4.2 Modelagem Molecular

Estudos de modelagem molecular foram utilizados em vários momentos neste trabalho. Esta abordagem compreende um grande conjunto de métodos computacionais que auxiliam a tornar mais eficiente o processo de planejamento racional de compostos bioativos. 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 ou propriedade de interesse. Para isto, são realizados cálculos de energia de conformação, de propriedades termodinâmicas, de orbitais moleculares e estatísticos (BARREIRO *et al.*, 1997; JENS *et al.*, 2003). Outra é a visualização tridimensional de moléculas, isoladas ou na forma de complexos com macromoléculas de interesse biológico (especialmente enzimas), o que permite a obtenção de informações sobre os requisitos estruturais necessários para a interação de um fármaco com seu receptor (BARREIRO *et al.*, 1997).

Considerando o exposto, as estruturas dos compostos que formam o conjunto de dados (material suplementar artigo, ver destaques nas Tabelas 1 a 9) foram construídas no programa *HyperChem 7* (HYPERCUBE, 2002) com base na estrutura cristalográfica de código DOTRUZ (Figura 26) extraída do banco de dados *Cambridge Structural Database* (CSD) (CAMBRIDGE, 2007).

Todas as geometrias foram otimizadas, com intuito de se obter estruturas com mínimos de energias, utilizando a teoria de mecânica molecular (MM+) através do método *Polak-Ribiere*, e em seguida a nível semi-empírico em *Austin Model 1* (AM1). Estas otimizações também foram realizadas no *HyperChem 7* (HYPERCUBE, 2002). Posteriormente, os arquivos de saída foram convertidos para arquivos de entrada do programa *Gaussian 0.9* (GAUSSIAN INC., 2009) utilizando o software *Open Babel* (OPEN BABEL, 2011). Neste programa, as estruturas foram otimizadas ao nível *ab initio*, inicialmente em *Hartree-Fock* (HF)

e, em seguida, pela Teoria do Funcional de Densidade ou DFT (*Density Functional Theory*). Os cálculos DFT foram realizados usando o funcional B3LYP e as funções de base foram 6-311G++(d, p). As funções difusas (++) nos cálculos em DFT representam melhor sistemas aniônicos, uma vez que adição de orbitais s e p muito difusos melhora a descrição de pares de elétrons de alta energia (FORESMAN e FRISCH, 2002; SANT'ANNA, 2009). A utilização dessas funções deve-se ao mecanismo de quelação do metal magnésio pelos compostos que apresentam o farmacóforo dos DCA's, conforme descrito anteriormente, como é o caso das estruturas aqui estudadas.

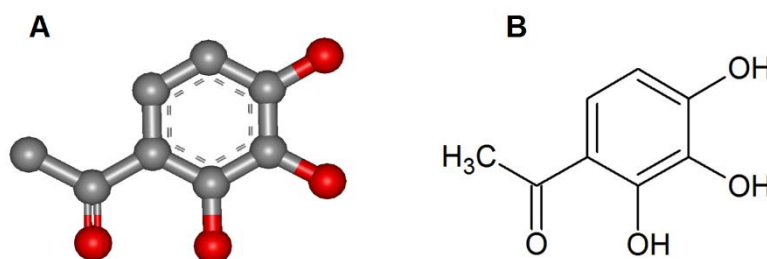


Figura 26 Estrutura base cristalográfica utilizada para construção do conjunto de compostos. Em A representação da geometria 3D e em B 2D.

4.3 Estudo QSAR-2D

QSAR bidimensional envolve principalmente a representação 2D gráfica de moléculas e a informação contida na mesma. Nessa análise, a informação química é derivada a partir das estruturas moleculares 2D. No entanto, o QSAR bidimensional também pode incluir descritores moleculares, assim como parâmetros físico-químicos, entre outros. A descrição molecular codificada na forma de descritores varia de contagens atômicas simples, medidas de peso molecular ou mesmo características espaciais ou geométricas complexas as quais podem ser associadas aos descritores bidimensionais com intuito de agregar maior quantidade de informação química a um modelo QSAR (ROY e NARAYAN DAS, 2014).

4.3.1 Obtenção dos descritores

Utilizando as estruturas tridimensionais otimizadas, foram obtidos os seguintes descritores eletrônicos: cargas parciais de Mulliken, NBO (*Natural*

Bond Orders) e *CHELPG* (*Charges from Electrostatic Potential Grid-based*), energia dos orbitais moleculares de valência ($E_{\text{HOMO}-1}$, E_{HOMO} , E_{LUMO} , $E_{\text{LUMO}+1}$), momento de dipolo total (D) e nos eixos x , y e z (D_x , D_y e D_z), e energia total (E_T). Além disso, utilizando as equações descritas por TODESCHINI e CONSONNI (2009) e os valores obtidos para os descritores de energia dos orbitais moleculares, foram derivados mais 10 descritores: GAP (diferença energética entre E_{HOMO} e E_{LUMO}), índice de energia de ativação (*activation energy index*, AEI), fração de energia HOMO/LUMO (*HOMO/LUMO energy fraction* $f(\text{H/L})$), dureza (*hardness*, n), potencial de ionização (*ionization potential*, IP), afinidade eletrônica (*electronic affinity*, EA), suavidade (*softness*, S), índice de eletrofilicidade no estado fundamental (*electrophilicity index in the ground state*, ω_{GS}), índice de eletrofilicidade (*eletrofilicity*, ω) eletronegatividade molecular (*molecular electronegativity*, χ). Deste modo, foram utilizados no estudo 52 descritores de caráter eletrônico, os quais foram obtidos através do *software* Gauss View 5 (Semichem Inc, 2009).

As geometrias de menor energia também foram utilizadas no programa *Dragon 6.0* (TALETE, 1997) para a obtenção dos descritores das seguintes classes: constitucionais (*constitutional descriptors*), descritores de contagem de grupos funcionais (*functional groups counts*), descritores de carga (*charge descriptors*), propriedades moleculares (*molecular properties*), contagem de circuitos e caminhos (*walk and path counts*), índices de informação (*information indices*), índices de adjacência de ponta (*edge adjacency indices*), índices de cargas topológicas (*topological charge indices*), descritores topológicos (*topological descriptors*), índices de conectividade (*connectivity indices*), descritores geométricos (*geometrical descriptors*), auto correlações 2D (*2D autocorrelations*), auto correlações 3D (*3D autocorrelations*), pares de átomos 2D e 3D (*2D e 3D atom pairs*), autovalores de Burden (*Burden eigenvalues*), descritores 3D de Morse (*3D-MoRSE descriptors*), autovalores baseados em índices (*eigenvalues-based indices*), descritores GETAWAY (*GETAWAY descriptors*). De modo geral, estas classes podem ser descritas como descritores constitucionais, topológicos, geométricos ou moleculares. Ao final foram gerados 4855 descritores.

Após a derivação dos descritores, foram aplicados os filtros para redução de variáveis, também disponíveis no programa *Dragon 6.0*. Assim, foram

reduzidos descritores através da eliminação de um membro dentre pares com alta correlação entre si (maior do que 0,9). Além disto, foram eliminados descritores invariantes e quase-invariantes, além daqueles que apresentavam desvio padrão menor que 0,001, visando à exclusão de dados invariantes. Finalmente, foi realizada uma redução manual, com intuito de excluir descritores com baixa variação, os quais não foram previamente eliminados pelo programa, já que estes não agregam informação relevante para o modelo. Como resultado, obteve-se uma matriz com 863 descritores. Esta última matriz foi submetida a mais uma etapa de redução de descritores no programa *QSAR Modeling* (MARTINS e FERREIRA, 2013): foram excluídos aqueles que apresentavam correlação absoluta com a atividade biológica ($|r|$) abaixo de 0,3, visando eliminar descritores que apresentassem variação numérica, porém pouco correlacionados com a variável dependente. Assim, a matriz utilizada para a etapa de seleção de variáveis era constituída por 319 descritores.

4.3.2 Seleção de variáveis e construção dos modelos

Para realização da seleção de variáveis e construção dos modelos foi utilizada a associação entre o algoritmo de seleção por preditores ordenados (*Ordered Predictors Selection*, OPS) e a regressão por quadrados mínimos parciais (PLS). O OPS é um método de seleção que atribui importância a cada descritor de acordo com um vetor informativo, para que em seguida a matriz de descritores seja rearranjada de modo que os mais importantes sejam representados pelas primeiras colunas da matriz, sendo então escolhidos dentre estes os mais relevantes e diversos modelos PLS são construídos aumentando-se a quantidade de descritores (MARTINS, 2010; MARTINS e FERREIRA, 2013).

No método de regressão PLS a matriz de descritores é correlacionada com o vetor atividade biológica (ou vetor Y). Assim, os dados são otimizados para estimar os valores de Y, originando novas variáveis X', ou variáveis latentes (VL). Estas novas variáveis têm a vantagem de serem ortogonais entre si. Assim, o problema da correlação entre as variáveis, uma limitação em RLM, não impede sua aplicabilidade (FERREIRA *et al.*, 2002; MARTINS, 2010).

A associação entre o OPS e PLS empregada neste estudo está implantada no programa *QSAR Modeling* (MARTINS, 2010). O método OPS foi utilizado considerando inicialmente o valor de $RMSECV$ (raiz quadrada do erro da validação cruzada), seguido dos valores de Q^2_{LOO} , com intuito de obter modelos com o menor erro e com boa capacidade de previsão. Os modelos foram construídos utilizando dados autoescalados, pré-processamento que consiste em subtrair de cada elemento de uma coluna da matriz de dados o valor médio da respectiva coluna e dividir o resultado pelo seu desvio padrão. Com este procedimento, a influência de uma variável dominante é minimizada em cálculos posteriores. Quando as variáveis têm diferentes unidades ou quando a faixa de variação dos dados é grande, ocorrência comum em QSAR faz-se necessário o autoescalamamento das variáveis (MARTINS e FERREIRA, 2013).

4.3.3 Validação dos modelos

O uso de um modelo de predição é condicionado a sua prévia aprovação em uma série de testes estatísticos. Isto visa garantir que as variáveis dependentes de interesse sejam previstas de maneira mais confiável possível. O método mais empregado de validação de modelos QSAR consiste em duas etapas: (i) testes usados para avaliar a capacidade de previsão dos modelos em relação aos compostos que auxiliaram em sua construção, ou validação interna; (ii) e testes de avaliação da previsão para compostos que não foram empregados na obtenção, ou validação externa (KIRALJ e FERREIRA, 2009; TROPSHA, 2010).

Na etapa de validação interna foram utilizados os parâmetros descritos na Tabela 3, permitindo assim a avaliação do grau de ajuste dos dados pelo modelo, a significância do modelo e a sua capacidade de predição interna. Nesta etapa deve-se avaliar a robustez dos modelos, através da validação cruzada *leave-N-out* (LNO), com uso do coeficiente de determinação entre as atividades biológicas observadas e preditas durante o processo de validação cruzada *leave-N-out* (Q^2_{LNO}). Este teste foi aplicado para $N= 1-14$ (26% do conjunto de dados) e repetido seis vezes para cada “ N ”. Também foi avaliado a presença de correlação ao acaso nas informações explicadas e preditas, através do teste de randomização do y (y -randomização) (KIRALJ e FERREIRA, 2009), onde

apenas este vetor é reandomizado. Foi utilizada a abordagem sugerida por Eriksson e colaboradores (2003) onde o $|r|$ entre o vetor y original e os vetores randomizados é utilizado para avaliar a presença de correlações espúrias (ERIKSSON *et al.*, 2003). Estes testes foram realizados nos programa *QSAR Modeling* (MARTINS, 2010). Também foram utilizados os testes de Roy aplicados a validação cruzada (*Average $r_{m^2_{LOO}}$ -scaled* e *$\Delta r_{m^2_{LOO}}$ -scaled*) (ROY e MITRA, 2012; ROY *et al.*, 2013), calculados com o software gratuito *Xternal Validation Metric Calculator* 1.0 (<http://dtclab.webs.com/software-tools>).

Para a validação externa (Tabela 4), um conjunto-teste de 13 compostos (**35, 45, 66, 72, 86, 97, 98, 121, 157, 164, 171, 195 e 196**), totalizando 24 % do conjunto de treinamento, os quais são representativos da faixa de variação da atividade biológica e da variabilidade estrutural das amostras (MARTINS e FERREIRA, 2013). Como a atividade biológica das amostras do *test set* é conhecida, pode-se fazer uma comparação entre o valor previsto pelo modelo e o valor real, utilizando parâmetros estatísticos (GOLBRAIKH *et al.*, 2003). Assim, a validação externa é muito mais confiável para assegurar a capacidade preditiva dos modelos QSAR, pois as amostras não foram utilizadas na construção dos modelos (GOLBRAIKH e TROPSHA, 2002; TROPSHA, 2010). A qualidade da predição foi avaliada através dos parâmetros descritores na Tabela 6. Para o cálculo dos parâmetros avaliados na validação externa, também foi empregado o software *Xternal Validation Metric Calculator* 1.0 (<http://dtclab.webs.com/software-tools>).

Além dos testes já citados, foi também realizada a detecção de *outliers*, os quais são valores atípicos, apresentando um grande afastamento dos demais da série para a previsão da atividade biológica em relação à experimental, não sendo capazes de se encaixar em um modelo QSAR (VERMA e HANSCH, 2005; MARTINS, 2010; MARTINS e FERREIRA, 2013). Para a identificação de compostos *outliers* levou-se em consideração a análise dos resíduos de Student pelos respectivos *leverages*.

4.4 Estudo QSAR-3D

Estudos de relação estrutura atividade quantitativa tridimensionais, ou QSAR-3D, têm como objetivo encontrar correlação estatística entre descritores

tridimensionais, cuja disposição no espaço dá origem a campos de interação moleculares (MIFs, *molecular interaction fields*), e a atividade biológica sob estudo (VERMA *et al.*, 2010). Os MIFs representam o processo de interação do ligante com o receptor, e podem ser calculados através de duas abordagens distintas: (i) baseados no receptor, onde os MIFs identificam regiões em que determinados grupos químicos podem interagir favoravelmente, sugerindo posições onde um ligante deva alocar grupos químicos os quais possam determinar interações químicas; e (ii) baseados nos ligantes, onde as regiões nas moléculas que mostram energia favorável de interação representam posições onde grupos de um potencial receptor poderiam interagir favoravelmente com estes ligantes. Para isso, usam-se diferentes sondas ou átomos de prova, através dos quais se podem obter, para um determinado conjunto de ligantes, posições em que interações específicas constituem um “sítio receptor virtual” (SRV). Esta entidade abstrata define um local ideal para complementar uma determinado composto químico e representaria a potencial capacidade de um ligante interagir com uma biomolécula (PASTOR *et al.*, 2000; VERMA *et al.*, 2010).

Com intuito de desenvolver os estudos de QSAR-3D, foram empregadas duas diferentes metodologias para o cálculo dos MIFs: descritores GRIND (*GRId-INdependent Descriptors*) (PASTOR *et al.*, 2000) e descritores de propriedades farmacocinéticas provenientes do programa VolSurf+ (CRUCIANI *et al.*, 2000), os quais são extraídos de MIFs gerados com GRID (GOODFORD, 1985) para construção de modelos em associação com descritores 2D.

4.4.1 GRIND

Os métodos tradicionais de QSAR-3D dependem da etapa de alinhamento, a qual muitas vezes consome longo tempo para sua realização e pode introduzir viés do usuário. Desta maneira, o modelo é sempre dependente do alinhamento. Existem vários métodos para superar este problema, mas, em geral, as transformações necessárias impedem uma interpretação simples dos modelos resultantes. O método GRIND apresenta descritores moleculares para compor modelos tridimensionais independentes do alinhamento. Estes descritores são derivados de tal forma a serem altamente relevantes para a

descrição das propriedades biológicas dos compostos em estudo, serem quimicamente interpretáveis e fáceis de calcular (PASTOR *et al.*, 2000; DAMALE *et al.*, 2014).

O funcionamento deste método também é baseado na construção do SRV, obtido através de átomos de prova como por exemplo, o N1 (nitrogênio amida), que é empregado para calcular energias de interação relacionadas à ligações de hidrogênio. Entretanto, os descritores GRIND constituem um pequeno grupo de variáveis que representam as relações geométricas entre regiões relevantes do SRV, e assim são independentes das coordenadas de onde os MIFs são calculados. De maneira geral, a obtenção destas variáveis envolve três passos principais: (i) cálculo dos MIFs: os compostos são posicionados no centro de um grade virtual (ou caixa) com intuito de calcular as energias de interação dos diferentes grupos nos ligantes com diferentes átomos de prova localizados nas intersecções do grid gerando os MIFs; (ii) filtragem dos MIFs, pois já que nem todos os pontos de interação presentes no MIFs descrevem informação de interação entre os átomos de prova e grupos químicos nos ligantes, é necessário filtrar a informação contida nos MIFS; (iii) codificação do SRV em descritores GRIND. Uma vez gerados os valores das interações que compõem os MIFs, estes são codificados na forma de descritores GRIND, de maneira que estes representem as relações geométricas entre as regiões do SRV. Para isto, o algoritmo empregado calcula o produto dos valores energéticos de interação entre todos os pares de interações possíveis. Deste modo, a informação não é mais dependente das posições 3D no espaço (PASTOR *et al.*, 2000).

4.4.1.1 Cálculo dos MIFs e descritores GRIND

Para desenvolver o método GRIND foi empregado o programa Pentacle (MOLECULAR DISCOVERY, 2015). Os MIFS foram calculados utilizando quatro átomos de prova diferentes: sonda DRY (representa interações hidrofóbicas), oxigênio carbonila (O sp^2 carbonila, representa grupos aceptores de ligação de Hidrogênio), N1 (Nitrogênio amida, representa grupos doadores de ligação de Hidrogênio), sonda TIP (representa a forma da molécula, em termos de pontos estéricos relevantes). As regiões com o MIF mais relevantes foram extraídas e

filtradas mediante a aplicação do algoritmo AMANDA (DURÁN *et al.*, 2008) implantado no programa Pentacle. Em cada intersecção do *grid* de 0.5 Å, a energia de interação (E_{xyz}) foi calculada como uma soma das energias de Lennard-Jones (estéricas), ligação de hidrogênio e interações eletrostáticas. Os limites de corte para os valores energéticos dos átomos de prova foram aplicados segundo o padrão do programa (-0,5 DRY; -2.6 O sp^2 ; -4.2 N1; -0,75 TIP). O algoritmo MACC2 (*Maximum Auto-and Cross-Correlation*) (PASTOR *et al.*, 2000) foi utilizado para codificar as interações pré-filtradas em descritores GRIND. Os valores obtidos a partir desta análise foram representados diretamente em correlogramas no programa, em que o produto dos valores das energias é representado versus a distância que separa essas interações. De maneira geral, o método GRIND possui vantagens sobre outros métodos 3D, uma vez que é independente do alinhamento molecular e possui adicionalmente as sondas para cálculo das ligações de Hidrogênio e interações estéricas (DAMALE *et al.*, 2014).

4.4.1.2 Seleção de variáveis e construção dos modelos

Para proceder a seleção de variáveis utilizou-se a seleção por Planejamento Fatorial Fracionário (FFD, *fractional factorial design*) (BARONI *et al.*, 1993) e a construção dos modelos foi feita a partir do método PLS. O procedimento de FFD tem como objetivo selecionar as variáveis que têm o maior efeito sobre a previsibilidade, para isso este método emprega o desvio padrão do erro de predição (*SDEP, standard deviation of error of predictions*,) para avaliar e decidir se um determinado grupo de variáveis será mantido ou descartado do modelo final, dependendo da mudança no desempenho na validação cruzada *leave-one-out*, relacionados com esses mesmos grupos (BARONI *et al.*, 1993; TOSCO e BALLE, 2011). Em relação à construção de modelos QSAR-3D, devido ao fato de que o número de descritores usados nas abordagens tridimensionais é muito elevado os métodos baseados em regressão linear múltipla não podem ser utilizados. Deste modo, o emprego do PLS permite a obtenção de modelos QSAR com variáveis ortogonais entre si dentro de uma gama de soluções possíveis (VERMA *et al.*, 2010). Os modelos QSAR GRIND foram construídos usando PLS com dados autoescalados. Os métodos de seleção de variáveis e construção dos modelos foram utilizados considerando a

qualidade estatística dos modelos, em relação aos parâmetros R^2 , Q^2_{LOO} e $SDEP$, com intuito de obter modelos com boa capacidade preditiva e menor erro.

4.4.1.3 Validação dos Modelos

A avaliação da qualidade estatística Tabela 5 dos modelos GRIND também foi realizada utilizando os parâmetros descritos na Tabela 3. Os testes citados acima foram aplicados nos modelos GRIND com o uso do software Pentacle (MOLECULAR DISCOVERY, 2015). Também foram utilizados os testes de Roy aplicados a validação cruzada (*Average $r_{m^2\text{LOO}}\text{-scaled}$ e $\Delta r_{m^2\text{LOO}}\text{-scaled}$*) (ROY e MITRA, 2012; ROY *et al.*, 2013). Todos os modelos foram construídos com dados autoescalados. A impossibilidade de extração dos dados dos modelos gerados neste programa impediu com que outros testes de validação fossem usados. Assim, somente o modelo auxiliar (contendo 54 compostos) é apresentado.

4.4.2 Modelos VolSurf+

Campos de interação moleculares tridimensionais em geral contêm uma quantidade grande de dados, os quais podem ser redundantes ou pouco correlacionados com a atividade biológica em questão. Levando isto em consideração, novas ferramentas são necessárias para a extração de descritores úteis de MIFs e correlacionar estas informações com as estruturas moleculares. Desta maneira, uma das metodologias disponíveis para extrair informações relevantes de campos tridimensionais e codificá-las em um número reduzido de descritores é o método VolSurf+ (CRUCIANI *et al.*, 2000).

VolSurf+ é um protocolo automático para conversão de campos moleculares 3D em descritores físico-químicos relevantes. Estes, de modo geral, são fáceis de entender e correlacionar com propriedades farmacocinéticas, principalmente relacionadas à permeação em membranas, uma vez que a interação de drogas com membranas biológicas é mediada por propriedades da superfície molecular como forma e tamanho da molécula, forças de van der Waals, eletrostáticas, ligações de hidrogênio e interações hidrofóbicas. Tais propriedades citadas podem ser extraídas de campos tridimensionais, os quais

calculam interações energéticas ao redor da superfície molecular. De modo geral, este método funciona em duas etapas: (i) cálculo dos MIFs utilizando o campo de força GRID (GOODFORD, 1985); (ii) cálculo dos descritores VolSurf+ a partir das informações tridimensionais, os quais basicamente se referem ao tamanho e forma da molécula e das regiões hidrofóbicas e hidrofílicas, além do balanço entre estas últimas (CRUCIANI *et al.*, 2000).

4.4.2.1 Geração dos MIFs/descriptores Volsurf e 2D

Os seguintes átomos de prova foram empregados: H₂O (usada para simular os processos de solvatação e dessovaltação); sonda Dry (representa interações hidrofóbicas); oxigênio carbonila (O sp² carbonila, representa grupos aceptores de ligação de Hidrogênio), N1 (Nitrogênio amida, representa grupos doadores de ligação de Hidrogênio). Posteriormente, foram geradas as seguintes classes de descritores VolSurf+: descritores de forma e tamanho molecular, de regiões hidrofílicas e hidrofóbicas, de momentos de energia de interação e mistos. Ao total foram gerados 128 descritores VolSurf+.

Com o intuito de ampliar a quantidade de informação química a ser investigada na construção dos modelos e associar as metodologias de QSAR 3D com QSAR 2D, os descritores Volsurf foram aliados com os descritores das classes constitucionais, topológicos, geométricos (obtidos no programa Dragon 6.0 (TALETE, 1997) e eletrônicos (obtidos no programa Gauss View 5 (GAUSSIAN INC., 2009) já anteriormente empregados no estudo 2D. Para isto, as variáveis obtidas no programa VolSurf+ (CRUCIANI *et al.*, 2000) foram exportadas e inseridas em uma matriz juntamente com os outros descritores citados. Após a derivação dos descritores, foram aplicados os filtros para redução de variáveis, disponíveis no programa Dragon 6.0 (TALETE, 1997): eliminação de um membro dentre pares com alta correlação entre si (maior do que 0,9), descritores invariantes e quase-invariantes, desvio padrão menor que 0,001, eliminação manual (exclusão de descritores com baixa variação, os quais não foram previamente eliminados pelo programa). Como resultado, obteve-se uma matriz com 717 descritores. Ainda, esta matriz foi submetida a mais uma etapa de redução de descritores no programa *QSAR Modeling* (MARTINS e FERREIRA, 2013): foram excluídos aqueles que apresentavam correlação

absoluta com a atividade biológica ($|r|$) abaixo de 0,3, visando eliminar descritores que apresentassem variação numérica, porém pouco correlacionados com a variável dependente. Assim, a matriz utilizada para a etapa de seleção de variáveis era constituída por 360 descritores.

4.4.2.2 Seleção de variáveis e construção dos modelos

Para realização da seleção de variáveis e construção dos modelos, assim como no estudo QSAR 2D, foi utilizada a associação entre o OPS e PLS implantada no programa *QSAR Modeling* (MARTINS, 2010). O método OPS foi utilizado considerando inicialmente o valor de *RMSECV*, seguido dos valores de Q^2_{Loo} , com intuito de obter modelos com o menor erro e com boa capacidade de previsão, além disto, as equações foram construídas utilizando dados autoescalados.

Os modelos gerados associando-se os descritores Volsurf com outros descritores 2D foram validados pelo mesmo protocolo de validação usado no estudo QSAR-2D, já descrito anteriormente.

4.5 Avaliação da predição

Para avaliar o potencial dos modelos QSAR como ferramenta de apoio a síntese de novos derivados, os modelos validados satisfatoriamente foram submetidos à predição de um segundo conjunto (conjunto de predição, $n=145$) (Material suplementar artigo, Tabela 1 a 9). Para isto, foi empregado o parâmetro de raiz quadrada do desvio médio (*RMSD*, *root mean square deviation*) para avaliar o erro das predições. Além disto, os compostos foram avaliados quanto à adequação no domínio de aplicabilidade do respectivo modelo utilizado para a predição. Foram empregados tanto os testes de domínio de aplicabilidade euclidiano quanto o de Willians (TROPSHA *et al.*, 2003; WOLD *et al.*, 2008).

5 RESULTADOS E DISCUSSÃO

5.1 QSAR 2D

O melhor modelo obtido é apresentado na equação 1. A análise dos resíduos de Student pelos respectivos *leverages* mostrou que o modelo não apresenta nenhum *outlier* (amostras anômalas) (VERMA e HANSCH, 2005; MARTINS, 2010; MARTINS e FERREIRA, 2013). Desta maneira, o modelo real obtido foi formado por 7 descritores, que deram origem a quatro VLs (informação acumulada: 77,438%; VL1: 34,142%; VL2: 6,744%; VL3: 12,657%; VL4: 23,894). Os resultados dos testes de validação interna e externa são disponibilizados na tabela 5.

$$\text{pIC}_{50} = 7,503 + 1,740 (\text{VE1_B(p)}) - 30,263 (\text{R5m+}) - 1,111 (\text{Eig09_EA(bo)}) + 0,827 (\text{H2m}) - 0,140 (\text{Mor28s}) - 0,001 (\text{E}_T) - 26,036 (\text{SpPosA_X})$$

Equação 1 Modelo de predição real A

A qualidade do ajuste dos dados, ou variância explicada pelo modelo, equivale a 75,6%, já que o mínimo recomendado é 60% ($R^2 > 0,6$), este apresenta valor satisfatório. O *RMSEC* representa a variabilidade na atividade não explicada pelo modelo e, conseqüentemente, deve ser o menor possível, também apresenta valor aceitável (KIRALJ e FERREIRA, 2009; DE MELO, 2012). A razão entre a variabilidade explicada pelo modelo e a variabilidade que permanece sem explicação é expressa pelo resultado do teste *F* (95% de confiança, $\alpha=0,05$). Considerando que o valor obtido é maior que seu valor crítico ($F_c=2,641$, para $p=4$, $n-p-1=35$), e que quanto maior o valor de *F*, mais significativo é a equação, o modelo A pode então ser classificado como significativo (GAUDIO e ZANDONADE, 2001; MARTINS e FERREIRA, 2013).

A previsibilidade interna foi testada pela validação cruzada LOO, que avalia a quantidade de informação ou variabilidade que um modelo pode prever. Se esta apresentar elevado grau de previsibilidade, seu Q^2_{LOO} será o mais próximo possível de 1, com mínimo recomendado é 0,5 (50% da informação ou variabilidade), enquanto o *RMSECV* deve apresentar o menor valor possível. Os resultados mostram que o modelo é capaz de prever 63,2% de informação, com um *RMSECV* associado de 0,456. Para os parâmetros *average $r_{m^2\text{LOO-scaled}}$* e *$\Delta r_{m^2\text{LOO-scaled}}$* , modelos com boa previsibilidade devem apresentar valores

superiores a 0,5 para o primeiro e menores que 0,2 para o segundo, resultados que foram aqui obtidos (ROY *et al.*, 2012). Finalmente, a diferença entre os valores de R^2 e Q^2_{LOO} é de apenas 0,124, o que indica que a probabilidade de estar ocorrendo sobreajuste dos dados é muito pequena (GAUDIO e ZANDONADE, 2001; BESALÚ e VERA, 2008; MARTINS e FERREIRA, 2013). Assim, considerando o exposto, o modelo A possui boa capacidade de predição, é significativo, e possui e baixa tendência de sofrer sobreajuste.

Na Figura 27 e 28 são apresentados os gráficos correspondentes aos resultados dos testes de validação cruzada LNO e randomização do y . A robustez do modelo foi avaliada através do procedimento de *leave-N-out*, outro processo de validação cruzada, onde um maior número de amostras são retiradas do conjunto de dados (geralmente, entre 20 a 30%). Como resultado, avalia-se a oscilação nos valores de cada Q^2_{LNO} e do Q^2_{LNO} médio em relação ao valor de Q^2_{LOO} , e do desvio padrão de cada replicata (KIRALJ e FERREIRA, 2009). Neste estudo, foi realizado um processo de *leave-14-out* (26% do conjunto de treinamento). Na Figura 27, pode-se observar que o modelo pode ser considerado robusto, pois mantém seu Q^2_{LNO} (0.610) em apenas 0.022 unidades menor que o Q^2_{LOO} , algumas variações são observadas nos resultados, entretanto o valor para a maior variação individual em relação ao Q^2_{LOO} foi de apenas 0,079 unidades ($Q^2_{L-14-O} = 0.553$). Assim, o resultado pode ser considerado aceitável.

Já o teste de randomização do y (Figura 28) é realizado através de modelos paralelos, os quais mantêm as colunas de descritores originais enquanto os valores do vetor y são aleatoriamente alterados. Os modelos resultantes devem apresentar baixa qualidade estatística. Segundo Eriksson *et al.* (ERIKSSON *et al.*, 2003), deve-se realizar uma regressão entre os valores absolutos das correlações de Pearson, $|r|$, entre o vetor y original e cada novo vetor obtido após uma aleatorização, com os valores de R^2 ou Q^2_{LOO} correspondentes. As intersecções das retas obtidas devem apresentar interceptos menores que 0,05 para Q^2_{LOO} e menores que 0,3 para R^2 para que o modelo possa ser considerado livre de correlações espúrias (KIRALJ e FERREIRA, 2009; MARTINS e FERREIRA, 2013). O resultado obtido mostra que o modelo obtido apresenta baixa probabilidade de sofrer de correlação ao acaso.

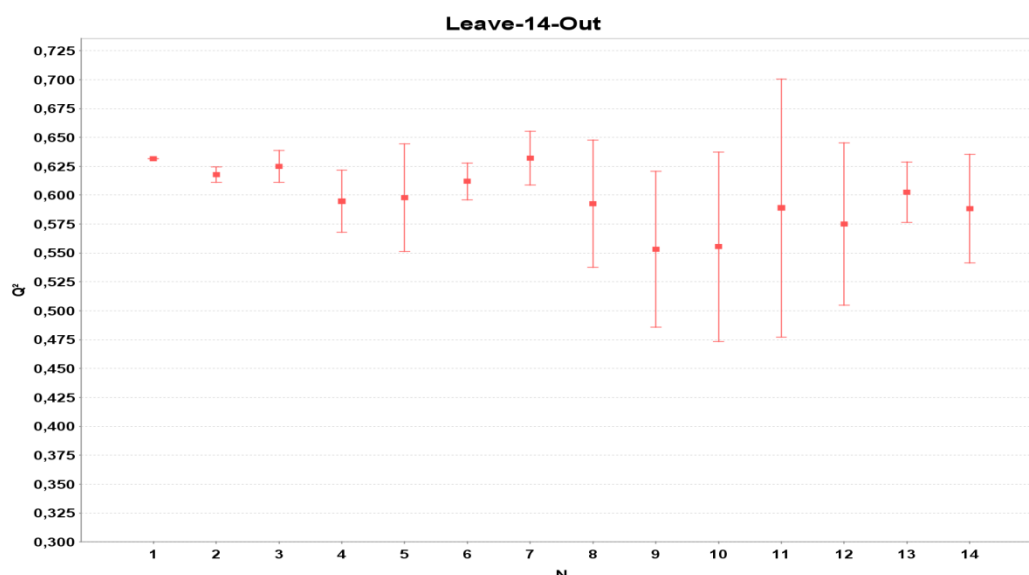


Figura 27 Resultados da validação cruzada *Leave -14- out* do modelo A.

Quanto a validação externa, os resultados apresentados na Tabela 5 mostram que o modelo real A foi aprovado em todos os testes adotados. Alguns autores sugerem que o valor limite de R^2_{pred} deve ser maior que 0,6 (GOLBRAIKH *et al.*, 2003), enquanto outros propõem valores maiores que 0,5 (PRATIM ROY *et al.*, 2009). Apesar de o último valor ser o mais comumente encontrado na literatura, o valor obtido neste estudo foi de 0,735, mostrando uma ótima capacidade preditiva.

Independente dos valores adotados, a maioria dos autores atualmente concorda que o uso somente do R^2_{pred} não é suficiente para avaliar a capacidade preditiva externa de um modelo. GOLBRAIKH *et al.*, 2003 e Roy *et al.*, 2012 propuseram alguns parâmetros para avaliação do poder de predição externa de modelos QSAR baseados nas inclinações de regressões centradas na origem (ou seja, o resultado de uma previsão ideal). Pode-se observar pelos resultados dos parâmetros k , k' (resultados esperados: dentro do intervalo de 0,85 a 1,15), $|R^2_0 - R^2_0|$ (resultado esperado: $< 0,3$) (GOLBRAIKH e TROPSHA, 2003), $average\ r_{m^2\ pred} - scaled$ (resultado esperado: $> 0,5$), e $\Delta r_{m^2\ pred} - scaled$ (resultado esperado: $< 0,2$) (ROY e MITRA, 2012; ROY *et al.*, 2013) que o modelo obtido possui boa capacidade de predição (ROY *et al.*, 2012).

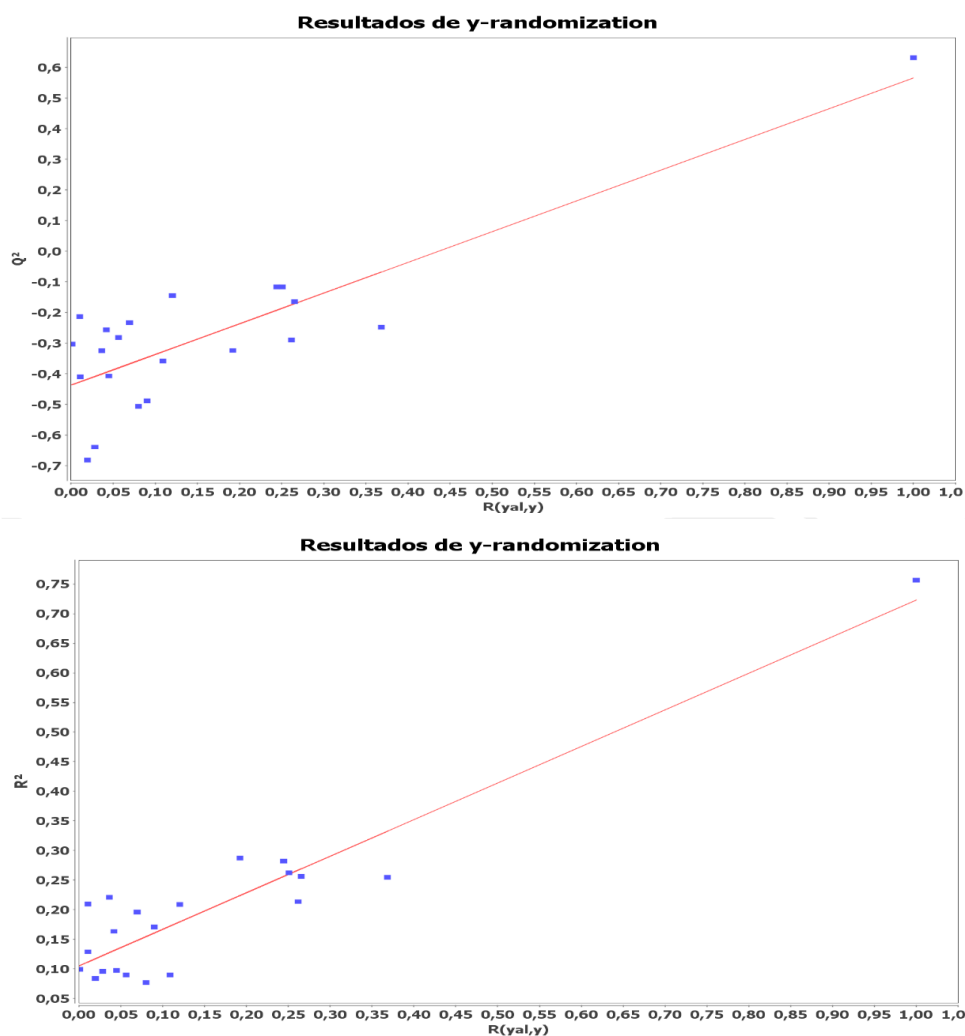


Figura 28 Valor dos interceptos para Q^2_{LOO} e R^2 no processo de randomização do y do modelo A.

5.2 QSAR 3D

5.2.1 GRIND

A fim de identificar bons modelos 3D um modelo PLS inicial foi construído, usando o conjunto completo de 550 variáveis disponíveis no *software* Pentacle (MOLECULAR DISCOVERY, 2015). Isto resultou em um modelo de 5 VLs com um valor de R^2 de 0,930, Q^2_{LOO} de 0,170, e RMSCV 0,900, resultado muito insatisfatório. Assim, a seleção de variáveis foi aplicada para reduzir o número de descritores com intuito de melhorar a qualidade estatística dos modelos. Isto resultou no modelo B representado na Figura 29 com 91 descritores e 5 VLs (informação acumulada: 45,180%; VL1: 12,460%; VL2: 13,340%; VL3: 8,570%; VL4: 5,360%; VL5: 5,450%), o qual mostra grande melhora nos parâmetros

avaliados. Detalhes sobre os principais descritores apresentados por este modelo estão identificados na Tabela 7. Os resultados dos testes de validação são disponibilizados na Tabela 5.

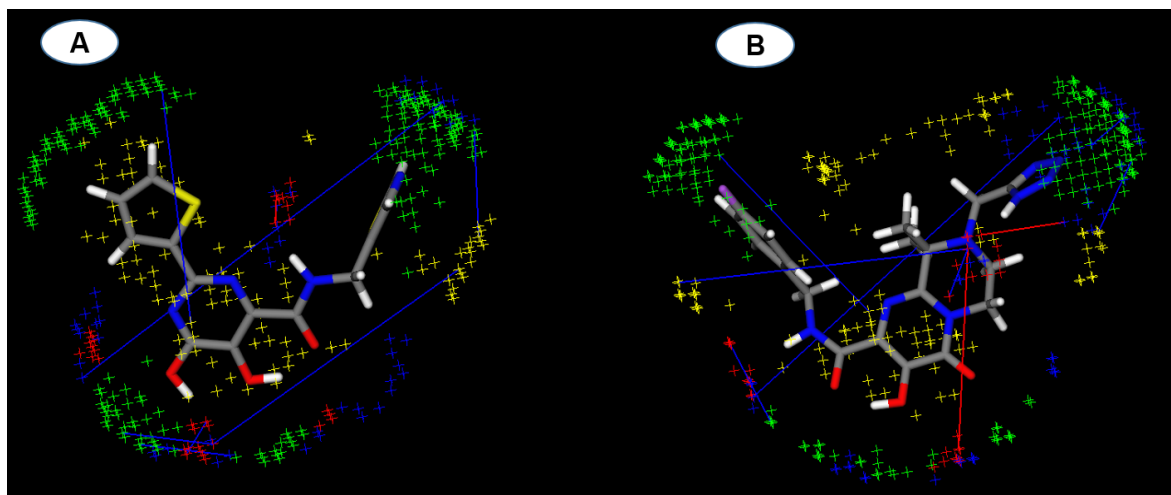


Figura 29 Representação tridimensional do modelo GRIND. A e B representam a sobreposição dos descritores sobre a molécula menos e mais ativa do conjunto de dados (**174** e **76** respectivamente). As linhas representam os intervalos de distância entre os descritores mais influentes sobre o modelo. As colorações dos pontos na figura indicam o tipo de interação que os descritores representam. Verde representa interações do tipo TIP, amarelo DRY, azul N1 (grupo doador de ligação Hidrogênio na molécula) e vermelho O (grupo aceptor).

A qualidade do ajuste dos dados do modelo B equivale a 94,0%, apresentando valor satisfatório. O *RMSEC* representa a variabilidade na atividade não explicada pelo modelo e, conseqüentemente, deve ser o menor possível, também apresenta valor aceitável (KIRALJ e FERREIRA, 2009). A significância do modelo foi avaliada pelo teste *F* (95% de confiança, $\alpha=0,05$), sendo o valor obtido maior que seu valor crítico ($F_c=2,40$, para $p=5$, $n-p-1=48$). Levando em consideração que quanto maior o valor de *F*, mais significativo é a equação, o modelo B pode então ser classificado como significativo, uma vez que apresenta um valor muito acima do valor de referência. Este valor de referência é dependente dos parâmetros apresentados pelo modelo a ser testado e varia para cada caso (GAUDIO e ZANDONADE, 2001; MARTINS e FERREIRA, 2013). A previsibilidade interna foi testada pelo Q^2_{LOO} e pelo *RMSECV*. Os resultados mostram que o modelo é capaz de prever 71,0% da informação, com um *erro* associado de 0,530, valores considerados adequados para um modelo de predição. Para os parâmetros *average $r_m^2_{LOO-scaled}$* e *$\Delta r_m^2_{LOO-scaled}$* , o modelo B novamente apresenta valores satisfatórios (0,618 e

0,091), o que reforça a qualidade preditiva do modelo. Finalmente, a diferença entre os valores de R^2 e Q^2_{LOO} é de 0,230, resultado dentro do limite aceitável, indicando que o modelo não sofre sobreajuste dos dados (GAUDIO e ZANDONADE, 2001; BESALÚ e VERA, 2008; MARTINS e FERREIRA, 2013). Desta maneira, é possível afirmar que o modelo B possui boa capacidade de predição, é significativo e possui e baixa tendência de sofrer sobreajuste.

5.2.2 Modelos Volsurf/Descritores 2D

Em relação a associação de descritores Volsurf 3D com os 2D, o melhor modelo obtido é apresentado na equação 2. A análise dos resíduos de Student pelos respectivos *leverages* mostrou que o modelo não apresenta nenhum *outlier*. O modelo real selecionado foi formado por 9 descritores, que deram origem a quatro VLs (informação acumulada: 71,340%; VL1: 35,255%; VL2: 8,056%; VL3: 18,494%; VL4: 9,535%). Dentre os descritores que compõe este modelo seis são das classes geométrica e topológica (Eig09_EA(bo), R5m+, H2m, VE1_B(p), SpPosA_X, Mor28s), dois são eletrônicos (E_T e CG.Chelpg6) e apenas um descritor Volsurf foi selecionado (SKIN). Quando comparado com o modelo A (2D) nota-se que o modelo C se mostra bastante semelhante em termos dos descritores selecionados (os mesmos descritores das classes topológica e geométrica estão presentes nos dois modelos), o que reforça a importância destes para a predição da atividade inibitória da HIV-1 Integrase. Os resultados dos testes de validação interna e externa são disponibilizados na Tabela 5.

$$\text{pIC}_{50} = 8,239 - 1,018 (\text{Eig09_EA(bo)}) - 25,032 (\text{R5m+}) + 0,635 (\text{H2m}) + 1,678 (\text{VE1_B(p)}) - 24,017 (\text{SpPosA_X}) - 0,718 (\text{CG.Chelpg6}) - 0,001 (E_T) + 0,117 (\text{SKIN}) - 0,107 (\text{Mor28s})$$

Equação 2 Modelo de predição real C

A variância explicada pela equação 2, equivale a 75,0%, enquanto o seu erro associado *RMSEC* apresenta valor de 0,376, o qual representa a variabilidade na atividade não explicada pelo modelo (KIRALJ e FERREIRA, 2009; DE MELO, 2012). Os resultados para o teste *F* (95% de confiança, $\alpha=0,05$)

do modelo C se apresentam satisfatórios, considerando que o valor obtido é maior que seu valor crítico ($F_c=2,641$, para $p=4$, $n-p-1=35$) o modelo C pode então ser classificado como significativo (GAUDIO e ZANDONADE, 2001; MARTINS e FERREIRA, 2013). Em relação à avaliação da previsibilidade interna, o Q^2_{LOO} corresponde à 0,553 enquanto o $RMSECV$ possui valor de 0,503, resultados dentro dos limites esperados. Ademais, a diferença entre os valores de R^2 e Q^2_{LOO} é de apenas 0,197, o que indica que a probabilidade de estar ocorrendo sobreajuste dos dados é pequena (GAUDIO e ZANDONADE, 2001; BESALÚ e VERA, 2008; MARTINS e FERREIRA, 2013).

Para os parâmetros *average $r_{m^2_{LOO}}-scaled$* e *$\Delta r_{m^2_{LOO}}-scaled$* , este modelo apresentou valores de 0,425 e 0,198, respectivamente. Segundo a literatura, modelos com boa previsibilidade devem apresentar valores superiores a 0,5 para o primeiro e menores que 0,2 para o segundo parâmetro (ROY *et al.*, 2012). Embora, para o *$\Delta r_{m^2_{LOO}}-scaled$* o valor apresentado se mostre dentro do intervalo adequado, para *average $r_{m^2_{LOO}}-scaled$* o mesmo não acontece. Isto indica que este modelo mesmo apresentando bons resultados para os demais testes, falha na capacidade preditiva interna.

Nas Figuras 30 e 31 são apresentados os gráficos correspondentes aos resultados dos testes de validação cruzada LNO e randomização do y . A robustez deste modelo foi também avaliada através do procedimento de *leave-14-out* (26% do conjunto de treinamento). Na figura 30 pode-se observar que o modelo C não pode ser considerado robusto, pois para algumas amostras o seu Q^2_{LNO} apresenta variações maiores que 0,100 unidades em relação ao Q^2_{LOO} (0,553), o que pode ser visto no gráfico com distribuição não estável nesta faixa de valores. Para o teste de randomização do y as intersecções das retas obtidas devem apresentar interceptos menores que 0,05 para Q^2_{LOO} e menores que 0,3 para R^2 para que o modelo possa ser considerado livre de correlações espúrias (KIRALJ e FERREIRA, 2009; MARTINS e FERREIRA, 2013). A Figura 31 mostra o resultado obtido neste teste para o modelo C, através da análise desta figura é possível perceber que o modelo obtido apresenta baixa probabilidade de sofrer de correlação ao acaso.

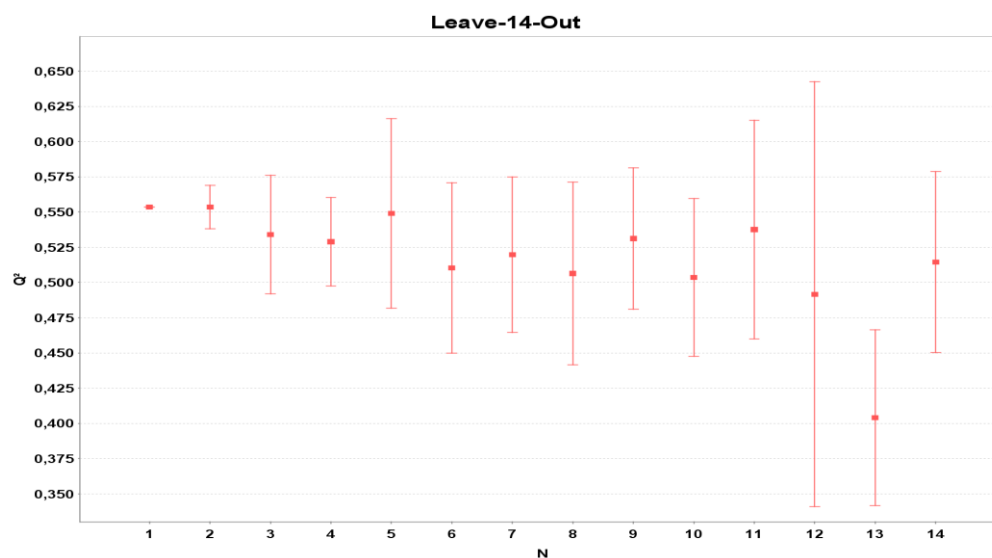


Figura 30 Resultados da validação cruzada *Leave -14- out* do modelo C.

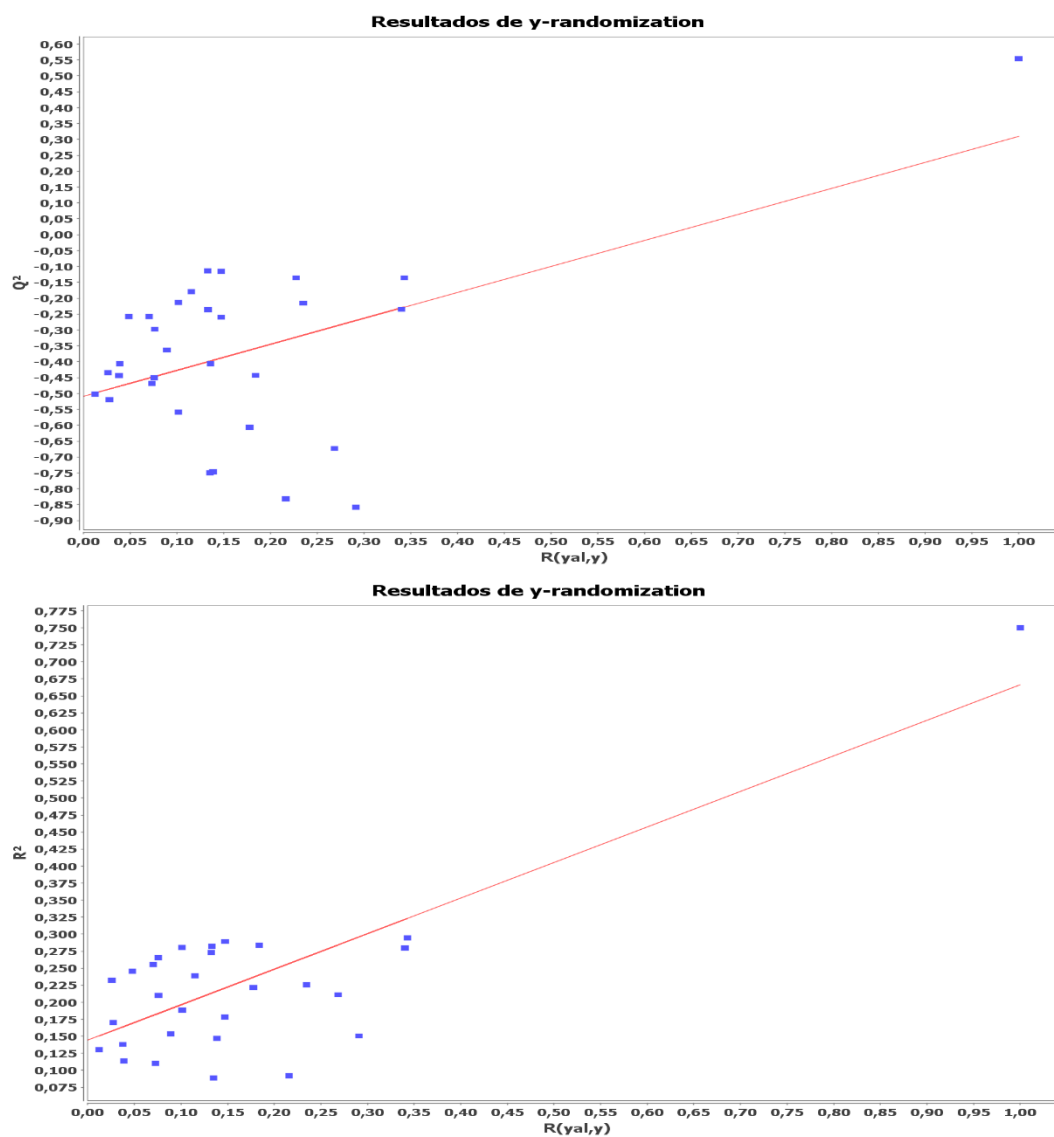


Figura 31 Valor dos interceptos para Q^2_{LOO} e R^2 no processo de randomização do y do modelo C.

Os resultados da validação externa são apresentados na Tabela 5. Pela análise desta tabela nota-se que o modelo real C não foi aprovado novamente em apenas um dos testes adotados. Segundo os parâmetros de R^2_{pred} (resultado esperado: $> 0,6$) (GOLBRAIKH *et al.*, 2003), k e k' (resultados esperados: dentro do intervalo de 0,85 a 1,15), $|R^2_0 - R^2_o|$ (resultado esperado: $< 0,3$) (GOLBRAIKH e TROPSHA, 2002) a equação 2 apresenta boa capacidade preditiva externa, uma vez que em todos estes testes valores desejados foram alcançados. Entretanto, segundo as métricas de *average $r_{m^2_{pred-scaled}}$* (resultado esperado: $> 0,5$) e *$\Delta r_{m^2_{pred-scaled}}$* (resultado esperado: $< 0,2$) (ROY e MITRA, 2012; ROY *et al.*, 2013) o modelo obtido não possui boa capacidade de predição, pois apresenta valor adequado para apenas a primeira métrica citada (ROY *et al.*, 2012). Desta maneira, segundo o exposto, o modelo C embora apresente bons valores para a variabilidade explicada, seja significativo, possua baixa tendência de sofrer sobreajuste e apresente resultados satisfatórios em alguns testes de previsibilidade, não pode ser considerado um modelo com boa capacidade preditiva e robusto.

Tabela 5 Comparação dos parâmetros estatísticos dos modelos resultantes.

Parâmetro	Modelo A	Modelo B	Modelo C	Valor referência
Nº de compostos	40	54	40	-
Nº de descritores	7	91	9	-
Nº de variáveis latentes	4	5	4	-
R^2	0,756	0,940	0,750	$> 0,6$
F	27,111	150,20	26,25	$>$ Valor referência
Valor referência teste F	2,641	2,408	2,641	-
$RMSEC$	0,371	0,180	0,376	$<$ Possível

<i>RMSECV</i>	0,456	0,530	0,503	< Possível
Q^2_{LOO}	0,632	0,710	0,553	> 0,5
$R^2 - Q^2_{\text{LOO}}$	0,124	0,230	0,197	< 0,3
<i>Average $r_{m^2\text{LOO}}$-scaled</i>	0,510	0,618	0,425	> 0,5
$\Delta r_{m^2\text{LOO-scaled}}$	0,199	0,091	0,198	< 0,2
Informação acumulada	77,438%	45,180%	71,340%	-
<i>Outlier</i>	0	0	0	-
R^2_{pred}	0,735	-	0,670	> 0,6
<i>Average $r_{m^2\text{pred}}$-scaled</i>	0,645	-	0,553	> 0,5
$\Delta r_{m^2\text{pred-scaled}}$	0,053	-	0,232	< 0,2
<i>k</i>	1,049	-	1,005	0,85<k<1,15
<i>k'</i>	0,950	-	0,991	0,85<k'<1,15
$ R^2_0 - R^2_0 $	0,126	-	0,183	< 0,3

5.3 Avaliação da predição

Posteriormente ao processo de validação estatística, o modelo A e B, os quais obtiveram resultados satisfatórios nos testes a que foram submetidos, tiveram seu poder preditivo também avaliado através de um segundo conjunto de moléculas inibidoras da enzima HIV-1 Integrase (conjunto de predição, $n=145$) (ver material suplementar artigo, Tabela 1 a 9). Estas, por sua vez, são provenientes das mesmas fontes dos compostos do conjunto de treinamento. Este teste de predição foi aplicado para avaliar o potencial dos modelos QSAR

como ferramenta de apoio a síntese de novos derivados. Para isto, estas moléculas foram avaliadas quanto à adequação no domínio de aplicabilidade do respectivo modelo utilizado para a predição. Foram usados tanto os testes de domínio de aplicabilidade (DA) euclidiano quanto o de Willians (TROPSHA *et al.*, 2003; WOLD *et al.*, 2008). A avaliação do DA dos compostos para o modelo B não pode ser realizada devido ao fato de que para a aplicação destes testes são necessários os dados de todos os descritores que compõem o modelo para todo o conjunto de compostos avaliado e que o software empregado para a construção dos modelos não permite a extração destes dados.

As predições da atividade biológica sob estudo para o conjunto de predição ($N=145$) foram realizadas com os modelos auxiliares A (equação 3) e B. Modelos auxiliares são construídos a partir da utilização de todas as amostras do conjunto de treinamento (neste caso 54 moléculas). Embora tanto o modelo de treinamento quanto real são formados pelos mesmos descritores, o primeiro citado possui quantidade maior de informação química em relação ao modelo real (40 moléculas). Para isso, as equações dos modelos auxiliares devem apresentar grande semelhança (sinal e grandeza dos coeficientes na equação) em relação aos modelos reais, com intuito de serem consideradas equivalentes.

Em relação ao *RMSD* entre os valores preditos e observados o valor para o modelo A foi de 0,698, enquanto que para o modelo C foi de 0,628. Apesar de estar acima dos valores obtidos para os parâmetros *RMSEC*, *RMSECV* e *RMSEP*, apenas 15 amostras (10,34% do conjunto avaliado) para o modelo A (Material suplementar artigo, Tabela 10) e 8 compostos (5,52%) para o modelo C (Anexo 1, Tabela 1) apresentaram resíduos acima de 1 unidade logarítmica, resultado que pode ser considerado aceitável quando analisado o grande número de compostos submetidos à predição.

Na Figura 32, são apresentados os domínios de aplicabilidade pelas abordagens de Willians (Fig. 32a) e pela distância euclidiana (Fig. 32b) para o modelo A. No primeiro, apenas 6 amostras (4,14% do segundo *test set*) ficaram localizados fora do domínio, apesar da amostra **103** (Material suplementar, Tabela 5) estar muito próxima da linha limite superior de resíduos. Já no domínio de aplicabilidade Euclidiano, nenhuma das amostras apresentou uma distância média normalizada (*normalized mean distance*) maior que a linha limite. Assim, é possível propor que as predições das atividades realizadas para as amostras

que compõem o segundo *test set* não são frutos de extrapolação, e que o modelo é representativo do espaço químico das moléculas em estudo.

$$\text{pIC}_{50} = 8,131 + 1,549 (\text{VE1_B(p)}) - 36,773 (\text{R5m+}) - 1,195 (\text{Eig09_EA(bo)}) + 1,251 (\text{H2m}) - 0,164 (\text{Mor28s}) - 0,001 (\text{E}_T) - 25,000 (\text{SpPosA_X})$$

Equação 3 Modelo de treinamento A

$n = 54$; outliers: 0; LVs= 4; informação acumulada: 79.755%; $R^2 = 0.759$; RMSEC = 0.372; $F(2.790) = 50.538$; $Q^2_{LOO} = 0.679$; RMSECV= 0.430

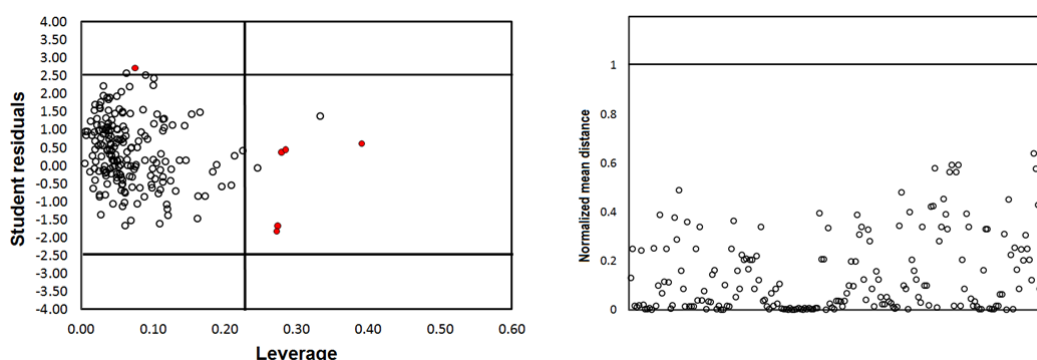


Figura 32 Resultados dos testes de domínio de aplicabilidade. Todas as amostras do estudo ($n = 199$) são apresentadas. A: resíduos de *Student* versus valores de *leverage* (DA de Willian). As amostras do segundo *test set* fora do espaço químico são representados como pontos vermelhos. A linha vertical é o valor limite (0,222). B: DA euclidiano. A linha horizontal é o valor limite de distância média normalizada.

5.4 Descritores selecionados

Embora os modelos A e B tenham-se mostrados adequados para fins de predição, quando possível deve-se realizar a interpretação mecanística do modelo em relação ao mecanismo de ação das moléculas em estudo, quando este for conhecido (OECD, 2007). Pirimidonas, pirimidinas e piridopirazinas carboxamidas são compostos relacionados aos DCA's, que inibem a HIV-1 IN por quelação dos cofatores Mg^{2+} presentes no sítio de ligação através de ligações com a tríade catalítica formada pelos resíduos Asp64, Asp116, and Glu152. Também estão presentes outras interações, em especial interações hidrofóbicas da cadeia lateral aromática com uma alça desordenada formada

pelos resíduos 140-145 (HARE *et al.*, 2010; METIFIOT *et al.*, 2011; CARCELLI *et al.*, 2014).

Para o modelo A os descritores selecionados pertencem a categorias diversas (Tabela 6), sendo dois topológicos, quatro geométricos e um eletrônico. Os coeficientes PLS autoescalados mostram que todos os descritores são relevantes para o modelo. Apesar das duas primeiras categorias serem de difícil interpretação, o descritor molecular E_T (e outros descritores termodinâmicos) pode ser relacionado com a estabilidade molecular (DU TOIT *et al.*, 2005; LOHRAY *et al.*, 2006; PHILLIPS *et al.*, 2008). De modo interessante, este descritor foi anteriormente selecionado no modelo previamente publicado por de Melo e Ferreira (2009) (DE MELO e FERREIRA, 2009) para um conjunto de 33 dihidroxipirimidinas carboxamidas, obtido pela mesma abordagem (OPS/PLS).

Tabela 6 Descritores selecionados no modelo A e seus respectivos coeficientes/coeficientes padronizados (abordagem de autoescalamamento).

Símbolo	Descritor	Classe	Coefficiente autoescalado
VE1_B(p)	Coefficiente da soma do último autovetor da matriz Burden/ponderado pela polarizabilidade	Topológico/ Descritores baseados na matrix 2D	0,552
R5m+	Autocorrelação R máxima entre átomos separados por 5 ligações químicas/ ponderada pela massa	Geométrico/ Descritores GETAWAY ^a	– 0,484
Eig09_EA(bo)	Autovalor n.9 da matriz de adjacência de ponta/ ponderado pela ordem de ligação	Topológico/ Índices de adjacência de ponta.	– 0,550
H2m	Autocorrelação H entre átomos separados por 2 ligações químicas/ ponderada pela massa	Geométrico/ Descritores GETAWAY ^a	0,288
Mor28s	Sinal 28/ ponderado pelo I-estado	Geométrico/ Descritores 3D-MorSE ^b	– 0,265

E_T	Energia total da molécula	Eletrônico/ Descritor termodinâmico	– 0,397
SpPosA_X	Soma positiva espectral normalizada a partir da matriz <i>Chi</i>	Geométrico/ Descritores baseados em Matriz 2D	– 0,237

Maiores detalhes em *List of molecular descriptors calculated by Dragon* (TALETE srl, 1997) e Propriedades químico-quânticas empregadas em estudos das relações estrutura-atividade (ARROIO *et al.*, 2010).

^a GATEWAY: Geometry, Topology And Atom-Weights Assembly.

^b MoRSE: Molecule Representation of Structure based on Electron diffraction.

Em relação ao modelo B, a Figura 33 apresenta os coeficientes PLS dos descritores que compõem o modelo. Através da análise desta figura nota-se que a atividade é sumariamente relacionada com interações do tipo N1-N1 (grupos doadores de ligação Hidrogênio nos compostos) e interações do tipo DRY-O (Hidrofóbicas e grupos aceptores de ligação Hidrogênio nos compostos) uma vez que seus coeficientes PLS são os que possuem maiores valores (picos em azul e alaranjado).

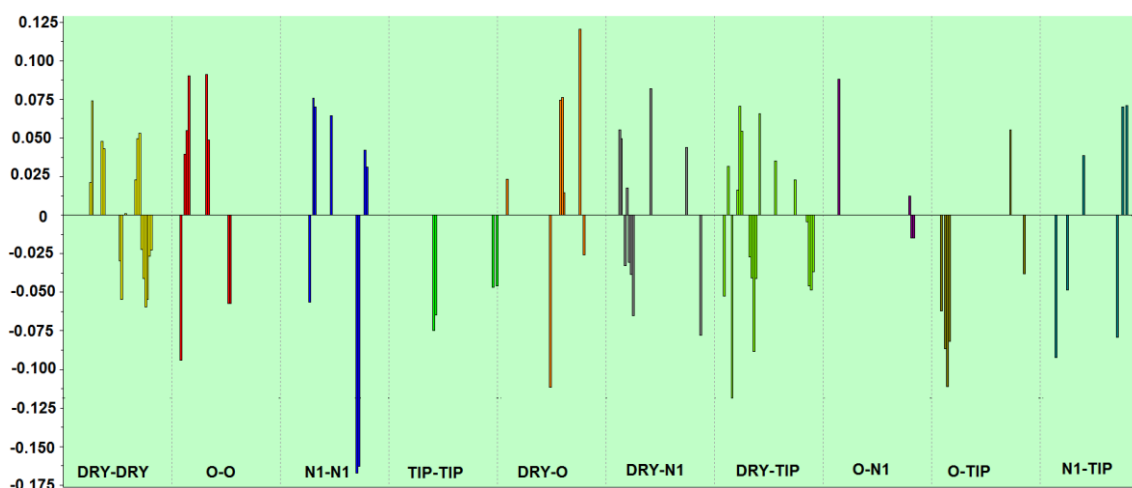


Figura 33 Coeficientes PLS mostrando os descritores diretamente (valores positivos) e inversamente (valores negativos) correlacionados com a atividade anti HIV-1 Integrase no modelo B.

A Tabela 7 apresenta os 10 descritores mais relacionados com a atividade biológica em ordem de importância e um detalhamento dos intervalos de distâncias das respectivas interações

Tabela 7 Descritores GRIND e suas correspondentes distâncias apresentados segundo a ordem de importância no modelo B. Tais variáveis foram identificadas como as mais correlacionadas com atividade biológica em estudo (segundo o coeficiente PLS).

Variável	Distância	Correlograma	Coeficiente PLS
148	15,2 – 16,6 Å	N1-N1	-0,167
261	16,4 – 16,8 Å	DRY-O	0,120
338	3,2 – 3,6 Å	DRY-TIP	-0,119
246	10,4 – 10,8 Å	DRY-O	-0,112
447	2,8 – 3,2 Å	O-TIP	-0,111
59	1,6 – 2,0 Å	O-O	-0,0942
502	2,8 – 3,2 Å	N1-TIP	-0,0927
72	6,8 – 7,2 Å	O-O	0,0909
349	7,6 – 8,0 Å	DRY-TIP	-0,0886
392	2,8 – 3,2 Å	O-N1	0,0881

Os descritores 148 e 261 são os mais relacionados com a atividade biológica, de acordo com seus coeficientes. Estes descritores são das classes de interações N1-N1 e DRY-O. Conforme anteriormente mencionado, estes tipos de interações são de extrema relevância para a atividade biológica segundo o modelo B. Considerando o exposto acima, uma análise mais aprofundada destes descritores foi realizada.

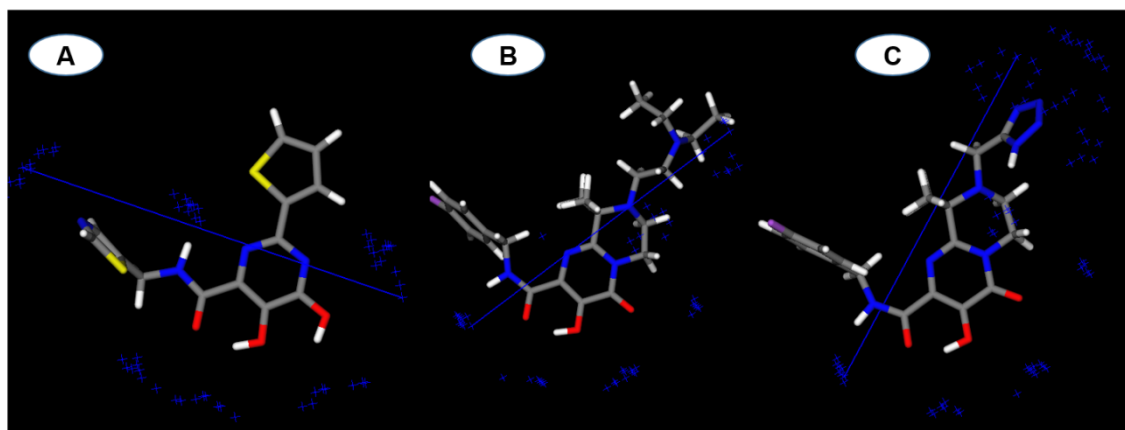


Figura 34 Detalhamento do descritor 148, intervalos de distância representados pelas linhas azuis. A, B e C representam a sobreposição das interações do tipo N1-N1 sobre as moléculas **174**, **45** e **76** com atividades baixa, intermediária e a mais alta respectivamente.

A interação N1-N1 dentro do intervalo 15,2 a 16,6 Å representada pelo descritor 148 se mostra direcionada para a posição horizontal na molécula menos ativa do conjunto de dados (**174**) (Figura 34). Ao ocorrer o deslocamento

desta interação no sentido vertical a atividade sofre aumento, conforme pode ser visto na Figura 34 para as moléculas com atividade intermediária e alta (**45** e **76**). Ao observar as moléculas com atividades mais altas nota-se que estas apresentam maior número de átomos de nitrogênio ao longo do esqueleto da molécula, dentro do intervalo representado pelo descritor 148 e posição deslocada, explicando o acréscimo em sua capacidade de inibir a enzima HIV-1 integrase. Isto se deve ao fato de que o coeficiente deste descritor é negativo. Desta maneira, quanto mais deslocado o descritor estiver, menor o seu valor e consequentemente maior a atividade biológica.

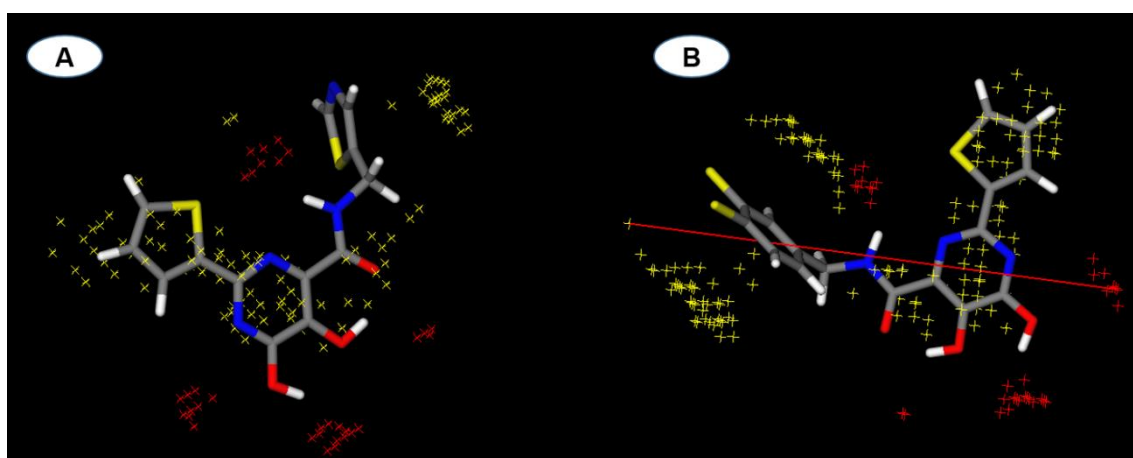


Figura 35 Detalhamento do descritor 261, intervalo de distância representado pela linha vermelha. A e B representam a sobreposição das interações do tipo DRY-O sobre as moléculas **174** e **190**, com atividades baixa e alta respectivamente.

Através da análise da Figura 35, percebe-se que o descritor 261 localiza a interação do grupo aceptor de ligação Hidrogênio próximo aos átomos de oxigênio, os quais quelam o metal Mg^{2+} , em uma distância dentro do intervalo 16,4 a 16,8 Å de um ponto de interação hidrofóbica (anel benzênico substituído com cloro). Com isto, a atividade do composto **190** é maior que o **174**, o qual não possui estas interações dentro deste intervalo. Nota-se também, que os átomos de oxigênio e o grupo hidrofóbico citados fazem parte do farmacóforo da classe de compostos aqui estudado, reforçando a importância deste tipo de interação para a atividade inibitória sobre a HIV-1 Integrase.

6. CONCLUSÃO

Neste estudo, foi possível obter três modelos QSAR multivariados empregando as metodologias de QSAR 3D e 2D, para um conjunto de 199 pirimidinas, pirimidonas e piridopirazinas carboxamidas as quais têm a capacidade de inibir *in vitro* a reação ST catalisada pela HIV 1-IN. Os testes de validação indicaram que tais modelos são significativos e não apresentam correlação ao acaso, porém somente o modelo A (2D) e B (3D) possuem boa capacidade de predição e robustez. Além disso, os modelos propostos validados satisfatoriamente foram utilizados para previsão da atividade inibitória sobre a HIV1-IN de um grande conjunto de dados, e os valores das previsões obtidos mostraram um resultado satisfatório. Finalmente, a interpretação mecanística reforça a confiabilidade dos modelos. Para o modelo A apesar da maioria dos descritores selecionados serem de origem topológica e geométrica, o descritor E_T (energia total da molécula) é consistente com achados bibliográficos prévios. Enquanto que para o modelo B, as interações entre N1-N1 e DRY-O se mostram altamente influentes sobre a atividade biológica. O modelo B apresentou um RMSD ligeiramente menor que A, e também um menor número de compostos com erro menor que uma unidade logarítmica, o que indica que ele apresenta um potencial de uso superior ao modelo A. Porém, a impossibilidade de exportação de dados pelo programa Pentacle impossibilitou a realização de testes de validação externa e de determinação de domínio de aplicabilidade, o que permitiria uma comparação mais adequada com o modelo A. Desta maneira, os modelos obtidos A e B podem ser úteis para a concepção de novos inibidores de HIV1-IN, com potencial para o desenvolvimento de compostos protótipos de novas drogas antirretrovirais.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ADAMSON, C. S.; FREED, E. O. Novel approaches to inhibiting HIV-1 replication. **Antiviral Research**, Frederick, v. 85, n. 1, p. 41-119, jan./set. 2010.

AFANI, S. A.; GALLARDO, O. A. M. Resistencia a la terapia antiretroviral en la infección por virus de inmunodeficiencia humana. **Revista chilena de infectología**, Santiago, v. 28, n. 5, p. 461-469, out. 2011.

ARROIO, A.; HONÓRIO, K. M.; SILVA, A. B. F. D. Propriedades químico-quânticas empregadas em estudos das relações estrutura-atividade. **Química Nova**, São Paulo, v. 33, n. 3, p. 694-699, fev. 2010.

BARONI, M.; COSTANTINO, G.; CRUCIANI, G.; RIGANELLI, D.; VALIGI, R.; CLEMENT, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. **Quantitative Structure-Activity Relationships**, Weinheim, v. 12, n. 1, p. 9-20, set. 1993.

BARRE-SINOUSSE, F.; CHERMANN, J. C.; REY, F.; NUGEYRE, M.T.; CHAMARET, S.; GRUEST, J.; DAUGUET, C.; AXLER-BLIN, C.; VÉZINET-BRUN, F.; ROUZIOUX, C.; ROZENBAUM, W.; MONTAGNIER, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). **Science**, Washington. DC, v. 220, n. 4599, p. 868-71, maio 1983.

BARRECA, M. L., DE LUCA, LAURA., FERRO, S., RAO, A., ; MONFORTE, A. M., CHIMIRRI, A. Computational and synthetic approaches for the discovery of HIV-1 integrase inhibitors. **ARKIVOC**, Gainesville, v. 7, p. 224-244, jan. 2006.

BARREIRO, E. J.; RODRIGUES, C. R. Modelagem Molecular: Uma Ferramenta para o Planejamento Racional de Fármacos em Química Medicinal. **Química Nova**, São Paulo, v. 20, p. 300-310, set. 1997.

BESALÚ, E.; VERA, L. Internal test set (ITS) method: a new cross-validation technique to assess the predictive capability of QSAR models. application to a benchmark set of steroids. **Journal of the Chilean Chemical Society**, Concepción, v. 53, p. 1576-1580, maio 2008.

BHATT, H.; PATEL, P.; PANNECOUQUE, C. Discovery of HIV-1 Integrase Inhibitors: Pharmacophore Mapping, Virtual Screening, Molecular Docking, Synthesis, and Biological Evaluation. **Chemical Biology & Drug Design**, Malden, v. 83, n. 2, p. 154-166, ago. 2014.

BLANCO, J. L.; MARTINEZ-PICADO, J. HIV integrase inhibitors in ART-experienced patients. **Current Opinion in HIV and AIDS**, Sydney, v. 7, n. 5, p. 415-21, set. 2012.

BRASIL, Ministério da saúde. **Recomendações para Terapia Anti-retroviral em Adultos Infectados pelo HIV**, Brasília, 2008.

_____. **MANUAL TÉCNICO PARA O DIAGNÓSTICO DA INFECÇÃO PELO HIV.** 2013a. Disponível em: <http://www.sbac.org.br/consulta_processual/manual_tecnico_hiv.pdf>. Acesso em 18 out, 2015.

_____. **Boletim Epidemiológico - Aids e DST-2013**, Brasília 2013b. Disponível em: <http://www.aids.gov.br/sites/default/files/anexos/publicacao/2013/55559/_p_boletim_2013_internet_pdf_p__51315.pdf>. Acesso em 18 out, 2015.

_____. **AIDS/Medicamentos/Quais são os antirretrovirais?** , 2014a. Disponível em: <http://www.aids.gov.br/pagina/quais-sao-os-antirretrovirais>. Acesso em 15 jun, 2015.

_____. **Relatório contabiliza casos de HIV/AIDS na América Latina.** 2014b. Disponível em: <http://www.brasil.gov.br/saude/2014/07/relatorio-contabiliza-casos-de-hiv-aids-na-america-latina>>. Acesso em 15 jun, 2015.

_____. **Antirretroviral dolutegravir é incorporado ao SUS.** 2015. Disponível em: <<http://www.aids.gov.br/noticia/2015/antirretroviral-dolutegravir-e-incorporado-ao-sus>>. Acesso em 07 dec, 2015.

BRODER, S. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. **Antiviral Research**, Frederick, v. 85, n. 1, p. 1-18, oct. 2010.

MONTANARI, C. A.;. **Química medicinal: Métodos e fundamentos em planejamento de fármacos.** 1. ed. São Paulo: Editora da Universidade de São Paulo, 2011. ISBN 978-85-314-1266-0.

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

CARCELLI, M.; ROGOLINO, D., SECHI, M., RISPOLI, G.; FISICARO, E.; COMPARI, C.; GRANDI, N.; CORONA, A.; TRAMONTANO, E.; PANNECOUQUE, C.; NAESENS, L. Antiretroviral activity of metal-chelating HIV-1 integrase inhibitors. **European Journal of Medicinal Chemistry**, Paris, v. 83, n. 0, p. 594-600, jun. 2014.

CHIRICO, N.; GRAMATICA, P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. **Journal of Chemical Information and Modeling**, Washington. DC, v. 51, n. 9, p. 2320-2335, jul. 2011.

CIHLAR, T.; RAY, A. S. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. **Antiviral Research**, Frederick, v. 85, n. 1, p. 39-58, set. 2010.

COTELLE, P. Patented HIV-1 integrase inhibitors (1998-2005). **Recent Patents on Anti-Infective Drug Discovery**, Emirate of Sharjah, v. 1, n. 1, p. 1-15, set. 2006.

CRUCIANI, G.; CRIVORI, P.; CARRUPT, P. A.; TESTA, B. Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. **Journal of Molecular Structure: THEOCHEM**, Amsterdam, v. 503, n. 1–2, p. 17-30, out. 2000.

CRUCIANI, G.; PASTOR, M.; GUBA, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. **European Journal of Pharmaceutical Sciences**, Amsterdam, v. 11, s. 2, p. S29-S39, out. 2000.

DAMALE, M. G.; HARKE, S. N.; KALAM KHAN, F.A.; SHINDE, D. B.; SANGSHETTI, J. N. Recent advances in multidimensional QSAR (4D-6D): a critical review. **Mini-Reviews in Medicinal Chemistry**, Beijing, v. 14, n. 1, p. 35-55, jan 2014.

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**, Washington, DC, v. 51, n. 5, p. 1136-1144, dec. 2008.

DE ALENCAR, T. M. D.; NEMES, M. I. B.; VELLOSO, M. A. From "acute AIDS" to "chronic AIDS": Body perception and surgical interventions in people living with HIV and AIDS. **Ciencia e Saude Coletiva**, Rio de Janeiro, v. 13, n. 6, p. 1841-1849, 2008.

DE MELO, E. B. Modeling physical and toxicity endpoints of alkyl (1-phenylsulfonyl) cycloalkane-carboxylates using the Ordered Predictors Selection (OPS) for variable selection and descriptors derived with SMILES. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 118, n. 0, p. 79-87, 2012.

DE 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**, Amsterdam, v. 44, n. 9, p. 3577-3583, 2009

DHUGURU, J.; GHEEWALA, C.; KUMAR, N. S. S.; WILSON, J. N. Highly Chromic, Proton-Responsive Phenyl Pyrimidones. **Organic Letters**, Washington, DC, v. 13, n. 16, p. 4188-4191, jul. 2011.

DI FRANCESCO, M. E. PACE, P.; FIORE, F.; NAIMO, F.; BONELLI, F.; ROWLEY, M.; SUMMA, V. Development of 2-t butyl-N-methyl pyrimidones as potent inhibitors of HIV integrase. **Bioorganic & Medicinal Chemistry Letters**, Amsterdam, v. 18, n. 8, p. 2709-13, mar. 2008.

DI SANTO, R. Inhibiting the HIV integration process: Past, present, and the future. **Journal of Medicinal Chemistry**, Amsterdam, v. 57, n. 3, p. 539-566, fev. 2014.

DOMS, R. W.; MOORE, J. P. HIV-1 membrane fusion: targets of opportunity. **Journal of Cell Biology**, New York, v. 151, n. 2, p. F9-14, out. 2000.

DONGHI, M.; KINZEL, O. D.; SUMMA, V. 3-Hydroxy-4-oxo-4H-pyrido[1,2-a]pyrimidine-2-carboxylates--a new class of HIV-1 integrase inhibitors. **Bioorganic & Medicinal Chemistry Letters**, Amsterdam, v. 19, n. 7, p. 1930-4, fev. 2009.

DRAGON SOFTWARE VERSION 6.0. TALETE srl, Milano, Italy, 1997. Disponível em: http://www.taletе.mi.it/products/dragon_description.htm.

DU TOIT, K.; ELGORASHI, E. E.; MALAN, S. F.; DREWES, S. E.; VAN STADEN, J.; CROUCH, N.R.; MULHOLLAND, D.A. Anti-inflammatory activity and QSAR studies of compounds isolated from Hyacinthaceae species and Tachiadenus longiflorus Griseb. (Gentianaceae). **Bioorganic & Medicinal Chemistry**, Amsterdam, v. 13, n. 7, p. 2561-2568, abr. 2005.

DURÁN, Á.; MARTÍNEZ, G. C.; PASTOR, M. Development and Validation of AMANDA, a New Algorithm for Selecting Highly Relevant Regions in Molecular Interaction Fields. **Journal of Chemical Information and Modeling**, Washington. DC, v. 48, n. 9, p. 1813-1823, set. 2008.

ERIKETI, Z.-L.; NICOLAOU, C.; NICOLAIDES, A. KOSTRIKIS, LG. HIV-1 Integrase: From Biology to Chemotherapeutics. **Current HIV Research**, Emirate of Sharjah, v. 5, n. 4, p. 365-388, jul. 2007.

ERIKSSON, L.; JAWORSKA, J.; WORTH, A. P.; CRONIN, M. T. D.; MCDOWELL, R. M.; GRAMATICA, P. METHODS for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. **Environmental Health Perspectives**, New York, v. 111, n. 10, p. 1361-1375, ago. 2003.

ESTE, J. A.; CIHLAR, T. Current status and challenges of antiretroviral research and therapy. **Antiviral Research**, Frederick, v. 85, n. 1, p. 25-33, jan. 2010.

FANALES-BELASIO, E.; RAIMONDO, M.; SULIGOI, B.; BUTTÒ, S. HIV virology and pathogenetic mechanisms of infection: a brief overview. **Annali dell'Istituto Superiore di Sanità**, Roma, v. 46, p. 5-14, 2010.

FARIA, N. R. RAMBAUT, A.; SUCHARD, M. A.; BAELE, G.; BEDFORD T.; WARD, M. J.; TATEM, A. J.; SOUSA, J. D.; ARINAMINPATHY, N.; PÉPIN, J.; POSADA, D.; PEETERS, M.; PYBUS, O. G.; LEMEY, P. The early spread and epidemic ignition of HIV-1 in human populations. **Science**, Washington. DC, v. 346, n. 6205, p. 56-61, out. 2014.

FERRARA, M. FIORE, F.; SUMMA, V.; GARDELLI, C. Development of 2-pyrrolidinyl-N-methyl pyrimidones as potent and orally bioavailable HIV integrase inhibitors. **Bioorganic & Medicinal Chemistry Letters**, Amsterdam, v. 20, n. 17, p. 5031-4, set. 2010.

FERREIRA, M. M. C.; MONTANARI, C. A.; GAUDIO, A. C. Seleção de variáveis em QSAR. **Química Nova**, São Paulo, v. 25, p. 439-448, fev. 2002.

FERREIRA, M. S. A síndrome da imunodeficiência adquirida e as doenças endêmicas no Brasil. **Revista da Sociedade Brasileira de Medicina Tropical**, Brasília, v. 29, p. 531-535, jul. 1996.

FORATTINI, O. P. AIDS e sua origem. **Revista de Saúde Pública**, São Paulo, v. 27, p. 153-156, jun. 1993.

FORESMAN, J. B.; FRISCH, A. **Basis Set Effects. Exploring chemistry with electronic structure methods**. 1. ed. Pittsburg: Gaussian Inc, cap. 5, p.97-110, 2002. ISBN 0-9636769-3-8.

FRÖHLICH, A. C. et al. Associação de carcinoma broncogênico com síndrome de Pancoast e síndrome da imunodeficiência adquirida. **Jornal de Pneumologia**, São Paulo, v. 26, p. 269-272, set. 2000.

GALLO, R. C. The Early Years of HIV/AIDS. **Science**, Washington. DC, v. 298, n. 5599, p. 1728-1730, November 29, nov. 2002.

GALLO, R. C.; MONTAGNIER, L. Prospects for the Future. **Science**, Washington. DC, v. 298, n. 5599, p. 1730-1731, November 29, nov. 2002.

GARDELLI, C.; GARDELLI, C.; NIZI, E.; MURAGLIA, E.; CRESCENZI, B.; FERRARA, M.; ORVIETO, F.; PACE, P.; PESCATORE, G.; POMA, M.; FERREIRA, M. R.; SCARPELLI, R.; HOMNICK, C. F.; IKEMOTO, N.; ALFIERI, A.; VERDIRAME, M.; BONELLI, F.; PAZ, O. G.; TALIANI, M.; MONTEAGUDO, E.; PESCI, S.; LAUFER, R.; FELOCK, P.; STILLMOCK, K. A.; HAZUDA, D.; ROWLEY, M.; SUMMA, V. Discovery and synthesis of HIV integrase inhibitors: development of potent and orally bioavailable N-methyl pyrimidones. **Journal of Medicinal Chemistry**, Washington, DC, v. 50, n. 20, p. 4953-75, set. 2007.

GARVEY, E. P. et al. The naphthyridinone GSK364735 is a novel, potent human immunodeficiency virus type 1 integrase inhibitor and antiretroviral. **Antimicrobial Agents and Chemotherapy**, Washington, DC, v. 52, n. 5, p. 1899, maio. 2008.

GAUDIO, A. C.; ZANDONADE, E. Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. **Química Nova**, São Paulo, v. 24, p. 658-671, out. 2001.

GAUSSIAN SOFTWARE VERSION 09, Gaussian Inc, Wallingford, USA, 2009. Disponível em: http://www.gaussian.com/g_prod/g09.htm.

GAUSS VIEW SOFTWARE VERSION 5.0, Semichem Inc, Shawnee Mission, USA, 2009. Disponível em: http://www.gaussian.com/g_prod/gv5.htm.

GIR, E.; VAICHULONIS, C. G.; OLIVEIRA, M. D. D. Adesão à terapêutica anti-retroviral por indivíduos com HIV/AIDS assistidos em uma instituição do interior paulista. **Revista Latino-Americana de Enfermagem**, Ribeirão Preto, v. 13, p. 634-641, out. 2005.

GOLBRAIKH, A.; SHEN, M.;XIAO, Z.; XIAO, Y. D., LEE, K. H.; TROPSHA, A. Rational selection of training and test sets for the development of validated QSAR models. **Journal of Computer-Aided Molecular Design**, Berlim, v. 17, n. 2-4, p. 241-253, fev. 2003.

GOLBRAIKH, A.; TROPSHA, A. Beware of q2! **Journal of Molecular Graphics and Modelling**, Amsterdam, v. 20, n. 4, p. 269-76, jan. 2002.

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 Sciences of the United States of America**, Washington, DC, v. 96, n. 23, p. 13040-13043, 1999.

GOODFORD, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. **Journal of Medicinal Chemistry**, Amsterdam, v. 28, n. 7, p. 849-857, jul. 1985.

GREENE, W. C.; PETERLIN, B. M.; STREMLAU, M. H. Molecular biology of HIV: Implications for new therapies. In: **Sande's HIV/AIDS Medicine: Medical Management of AIDS 2013**. 2. ed. Vancouver: Elsevier, 2012. cap. 3, p.25-43.

HARE, S. VOS, A. M.; CLAYTON, R. F.; THURING, J. W.; CUMMINGS, M. D.; CHEREPANOV, P. Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, DC, v. 107, n. 46, p. 20057-62, nov. 2010.

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**, Washington, DC, v. 71, n. 9, p. 7005-7011, 1997.

HEALY, E. F. SANDERS, J.; KING, P. J.; ROBINSON, W. E. JR. A docking study of l-chicoric acid with HIV-1 integrase. **Journal of Molecular Graphics and Modelling**, Amsterdam, v. 27, n. 5, p. 584-589, jan. 2009.

HUGHES, A. K.; ADMIRAAL, K. R. A Systematic Review of HIV/AIDS Knowledge Measures. **Research on Social Work Practice**, v. 22, n. 3, p. 313-322, May 1, 2012 2012.

HYPERCHEM VERSION 7. HYPERCUBE INC. USA, 2002. Disponível em: <http://www.hyper.com/>.

AIDSinfo. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. Disponível em: <http://aidsinfo.nih.gov/guidelines>. Acesso em 15 jun, 2014.

JENS, S.; JOHANN, G.; CHRISTOF, H. S. **3D Structure Generation and Conformational Searching**. 1. ed. Computational Medicinal Chemistry for Drug Discovery: CRC Press, 2003. ISBN 978-0-8247-4774-9.

KIRALJ, R.; FERREIRA, M. M. C. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. **Journal of the Brazilian Chemical Society**, Campinas, v. 20, p. 770-787, 2009.

KUBINYI, H. **QSAR: Hansch Analysis and Related Approaches**. 1. ed: Wiley-VCH Verlag GmbH, 2008. p.1-13. ISBN 9783527616824.

LIGNANI JÚNIOR, L.; GRECO, D. B.; CARNEIRO, M. Avaliação da aderência aos anti-retrovirais em pacientes com infecção pelo HIV/Aids. **Revista de Saúde Pública**, São Paulo, v. 35, p. 495-501, ago. 2001.

LIMA, L. M. Química Medicinal Moderna: desafios e contribuição brasileira. **Química Nova**, São Paulo, v. 30, p. 1456-1468, 2007.

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 & Medicinal Chemistry Letters**, Amsterdam, v. 16, n. 14, p. 3817-3823, jul. 2006.

MAERTENS, G. N.; HARE, S.; CHEREPANOV, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. **Nature**, Londres, v. 468, n. 7321, p. 326-9, nov. 2010.

MAGALHAES UDE, O. SOUZA, A. M.; ALBUQUERQUE, M. G.; BRITO, M. A.; BELLO, M. L.; CABRAL, L. M.; RODRIGUES, C. R. Hologram quantitative structure-activity relationship and comparative molecular field analysis studies within a series of tricyclic phthalimide HIV-1 integrase inhibitors. **Journal of Drug Design, Development and Therapy**, Londres, v. 7, p. 953-61, set. 2013.

MAKHIJA, M. T.; KULKARNI, V. M. QSAR of HIV-1 integrase inhibitors by genetic function approximation method. **Bioorganic & Medicinal Chemistry**, Amsterdam, v. 10, n. 5, p. 1483-1497, maio. 2002.

MARTINS, J. P. A. **Apostila de Química computacional aplicada a QSAR**. São Paulo: Centro Nacional de Processamento de Alto Desempenho, São Paulo, 2010.

MARTINS, J. P. A.; FERREIRA, M. M. C. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. **Química Nova**, São Paulo, v. 36, p. 554-560, mar. 2013.

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**, Amsterdam, v. 47, n. 22, p. 5583-5586, out. 2004.

MEHELLOU, Y.; DE CLERCQ, E. Twenty-six years of anti-HIV drug discovery: where do we stand and where do we go? **Journal of Medicinal Chemistry**, Amsterdam, v. 53, n. 2, p. 521-38, jan. 2010.

MELO, E. B. D.; BRUNI, A. T.; FERREIRA, M. M. C. Inibidores da HIV-integrase: potencial abordagem farmacológica para tratamento da AIDS. **Química Nova**, São Paulo, v. 29, p. 555-562, jun. 2006.

METIFIOT, M.; VANDEGRAAFF, N.; MADDALI, K.; NAUMOVA, A.; ZHANG, X.; RHODES, D.; MARCHAND, C.; POMMIER, Y. Elvitegravir overcomes resistance to raltegravir induced by integrase mutation Y143. **AIDS**, Amsterdam, v. 25, n. 9, p. 1175-8, jun. 2011.

MURAGLIA, E.; KINZEL, O.; GARDELLI, C.; CRESCENZI, B.; DONGHI, M.; FERRARA, M.; NIZI, E.; ORVIETO, F.; PESCATORE, G.; LAUFER, R.; GONZALEZ-PAZ, O.; DI MARCO, A.; FIORE, F.; MONTEAGUDO, E.; FONSI, M.; FELOCK, P. J.; ROWLEY, M.; SUMMA, V. Design and synthesis of bicyclic pyrimidinones as potent and orally bioavailable HIV-1 integrase inhibitors. **Journal of Medicinal Chemistry**, Amsterdam, v. 51, n. 4, p. 861-74, fev. 2008.

NIZI, E. ORSALE, M. V.; CRESCENZI, B.; PESCATORE, G.; MURAGLIA, E.; ALFIERI, A.; GARDELLI, C.; SPIESER, S. A.; SUMMA, V. Dihydroxy-pyrimidine and N-methylpyrimidone HIV-integrase inhibitors: improving cell based activity by the quaternarization of a chiral center. **Bioorganic & Medicinal Chemistry Letters**, Amsterdam, v. 19, n. 16, p. 4617-21, aug. 2009.

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/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono%282007%292>>. Acesso em 15 jun, 2015.

OPEN BABEL VERSION 2.3.1. The Open Babel Package, 2011. Disponível em: <http://openbabel.org> .

PACE, P.; DI FRANCESCO, M. E.; GARDELLI, C.; HARPER, S.; MURAGLIA, E.; NIZI, E.; ORVIETO, F.; PETROCCHI, A.; POMA, M.; ROWLEY, M.; SCARPELLI, R.; LAUFER, R.; GONZALEZ, PAZ, O.; MONTEAGUDO, E.; BONELLI, F.; HAZUDA, D.; STILLMOCK, K. A.; SUMMA, V. Dihydroxypyrimidine-4-carboxamides as novel potent and selective HIV integrase inhibitors. **Journal of Medicinal Chemistry**, Amsterdam, v. 50, n. 9, p. 2225-39, maio. 2007.

PANNECOUQUE, C.; PLUYMERS, W.; VAN-MAELE, B.; TETZ, V.; CHEREPANOV, P.; DE CLERCQ, E.; WITVROUW, M.; DEBYSER, Z. New class of HIV integrase inhibitors that block viral replication in cell culture. **Current Biology**, Cambridge, v. 12, n. 14, p. 1169-77, jul. 2002.

PASTOR, M.; CRUCIANI, G.; MCLAY, I.; PICKETT, S.; CLEMENTI, S. GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-

dimensional molecular descriptors. **Journal of Medicinal Chemistry**, Amsterdam, v. 43, n. 17, p. 3233-43, aug. 2000.

PATEL, R. V.; KEUM, Y. S.; PARK, S. W. Sketching the historical development of pyrimidones as the inhibitors of the HIV integrase. **European Journal of Medicinal Chemistry**, Amsterdam, v. 97, n. 5, p. 649-663, jun. 2014.

PAU, A. K.; GEORGE, J. M. Antiretroviral Therapy: Current Drugs. **Infectious Disease Clinics of North America**, Washington, DC, v. 28, n. 3, p. 371-402, set. 2014.

PEÇANHA, E. P.; ANTUNES, O. A. C.; TANURI, A. Estratégias farmacológicas para a terapia anti-AIDS. **Química Nova**, São Paulo, v. 25, p. 1108-1116, dez. 2002.

PENTACLE SOFTWARE VERSION 1.0.7. Molecular Discovery, United Kingdom, 2015. Disponível em: <<http://www.moldiscovery.com/software/pentacle/>>.

PETROCCHI, A.; JONES, P.; ROWLEY, M.; FIORE, F.; SUMMA, V. N-(4-Fluorobenzyl)-3-hydroxy-9,9-dimethyl-4-oxo-6,7,8,9-tetrahydro-4H-pyrazino[1,2-a]pyrimidine-2-carboxamides a novel class of potent HIV-1 integrase inhibitors. **Bioorganic & Medicinal Chemistry Letters**, Amsterdam, v. 19, n. 15, p. 4245-4249, aug. 2009.

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**, Amsterdam, v. 17, n. 2, p. 350-353, jan. 2007.

PHILLIPS, 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**, Paris, v. 43, n. 5, p. 1095-1104, maio. 2008.

PORTELA, M. C.; LOTROWSKA, M. Assistência aos pacientes com HIV/Aids no Brasil. **Revista de Saúde Pública**, São Paulo, v. 40, p. 70-79, mar. 2006.

PRATIM ROY, P.; PAUL, S.; MITRA, I.; ROY, K. On Two Novel Parameters for Validation of Predictive QSAR Models. **Molecules**, Basel, v. 14, n. 5, p. 1660-1701, jan. 2009.

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**, Amsterdam, v. 38, n. 6, p. 890-897, mar.1995.

REQUEJO, H. I. Z. Worldwide molecular epidemiology of HIV. **Revista de Saúde Pública**, São Paulo, v. 40, p. 331-345, jan. 2006.

RERKS-NGARM, S. PITISUTTITHUM, P.; NITAYAPHAN, S.; KAEWKUNGWAL, J.; CHIU, J.; PARIS, R.; PREMSRI, N.; NAMWAT, C.; DE SOUZA, M.; ADAMS, E.; BENENSON, M.; GURUNATHAN, S.; TARTAGLIA, J.; MCNEIL, J. G.; FRANCIS, D. P.; STABLEIN, D.; BIRX, D. L.; CHUNSUTTIWAT, S.; KHAMBOONRUANG, C.; THONGCHAROEN, P.; ROBB, M. L.; MICHAEL, N. L.; KUNASOL, P.; KIM, J. H.; MOPH-TAVEG INVESTIGATORS. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. **The New England Journal of Medicine**, Waltham, v. 361, n. 23, p. 2209-20, dez. 2009.

ROGOLINO, D.; CARCELLI, M.; SECHI, M.; NEAMATI, N. Viral enzymes containing magnesium: Metal binding as a successful strategy in drug design. **Coordination Chemistry Reviews**, Amsterdam, v. 256, n. 23–24, p. 3063-3086, jul. 2012.

ROY, K. PRATIM, C.; MITRA, I.; OJHA, P. K.; KAR, S.; DAS, R. N. Some case studies on application of “rm2” metrics for judging quality of quantitative structure–activity relationship predictions: Emphasis on scaling of response data. **Journal of Computational Chemistry**, Hoboken, v. 34, n. 12, p. 1071-1082, maio. 2013.

ROY, K.; MITRA, I. On the use of the metric rm(2) as an effective tool for validation of QSAR models in computational drug design and predictive toxicology. **Mini-Reviews in Medicinal Chemistry**, Beijing v. 12, n. 6, p. 491-504, jun. 2012.

ROY, K.; MITRA, I.; KAR, S.; OJHA, P. K.; DAS, R. N.; KABIR, H. Comparative studies on some metrics for external validation of QSPR models. **Journal of Chemical Information and Modeling**, Washington. DC, v. 52, n. 2, p. 396-408, fev. 2012.

ROY, K.; NARAYAN DAS, R. A Review on Principles, Theory and Practices of 2D-QSAR. **Current Drug Metabolism**, Emirate of Sharjah, v. 15, n. 4, p. 346-379, maio. 2014.

ROY, P. P.; ROY, K. On Some Aspects of Variable Selection for Partial Least Squares Regression Models. **QSAR & Combinatorial Science**, Hoboken, v. 27, n. 3, p. 302-313, jul. 2008.

RUSTICI, V. C. F.; CARAMORI, G. F.; GALEMBECK, S. E. Efeitos de substituintes na ligação de hidrogênio do 3-hidroxipropenal. **Química Nova**, São Paulo, v. 29, p. 1187-1192, dez. 2006.

SAIZ-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**, Paris, v. 42, n. 1, p. 64-70, out. 2007.

SANT'ANNA, C. M. R. Glossário de termos usados no planejamento de fármacos (recomendações da IUPAC para 1997). **Química Nova**, São Paulo, v. 25, p. 505-512, maio. 2002.

SANT'ANNA, C. M. R. Métodos de modelagem molecular para estudo e planejamento de compostos bioativos: Uma introdução. **Revista. Virtual de Química**, Niterói, v. 1, n. 1, p. 49-57, fev. 2009.

SANTOS FILHO, O. A.; ALENCASTRO, R. B. D. Modelagem de proteínas por homologia. **Química Nova**, São Paulo, v. 26, p. 253-259, set. 2003.

SCHULZ, T. F.; BOSHOFF, C. H.; WEISS, R. A. HIV infection and neoplasia. **The Lancet**, v. 348, n. 9027, p. 587-591, set. 1996.

SERRAO, E.; XU, Z. L.; DEBNATH, B.; CHRIST, F.; DEBYSER, Z.; LONG, Y. Q.; NEAMATI, N. Discovery of a novel 5-carbonyl-1H-imidazole-4-carboxamide class of inhibitors of the HIV-1 integrase-LEDGF/p75 interaction. **Bioorganic & Medicinal Chemistry**, Amsterdam, v. 21, n. 19, p. 5963-72, out. 2013.

SMITH, P. L.; TANNER, H.; DALGLEISH, A. Developments in HIV-1 immunotherapy and therapeutic vaccination. **F1000Prime Reports**, Londres, v. 6, jun. 2014.

SOSA, N. Avances en VIH/SIDA y complicaciones de la terapia antirretroviral. **Acta Medica Colombiana**, Santa Fé de Bogotá, v. 32, p. 174-176, set. 2007.

STIERAND, K.; MAASS, P. C.; RAREY, M. Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. **Bioinformatics**, Oxford, v. 22, n. 14, p. 1710-6, abri. 2006.

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**, Amsterdam, v. 49, n. 23, p. 6646-9, nov. 2006.

TALETE srl. *List of molecular descriptors calculated by Dragon*. Disponível em: <http://www.talete.mi.it/products/dragon_molecular_descriptor_list.pdf>. Acesso em 07 out, 2015.

TAVARES, L. C. QSAR: a abordagem de Hansch. **Química Nova**, São Paulo, v. 27, p. 631-639, aug. 2004.

TODESCHINI, R.; CONSONNI, V. **Molecular Descriptors for Chemoinformatics**. 2. ed. Weinheim, Germany: Wiley-VCH, 2009. ISBN 978352731852-0.

TOSCO, P.; BALLE, T. Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. **Journal of Molecular Modeling**, Amsterdam, v. 17, n. 1, p. 201-8, jan. 2011.

TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. **Molecular Informatics**, Hoboken, v. 29, n. 6-7, p. 476-488, jul. 2010.

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**, Hoboken, v. 22, n. 1, p. 69-77, abril. 2003.

VERMA, J.; KHEDKAR, V. M.; COUTINHO, E. C. 3D-QSAR in drug design--a review. **Current Topics in Medicinal Chemistry**, Emirate of Sharjah, v. 10, n. 1, p. 95-115, nov. 2010.

VERMA, R. P.; HANSCH, C. An approach toward the problem of outliers in QSAR. **Bioorganic & Medicinal Chemistry**, Amsterdam, v. 13, n. 15, p. 4597-4621, jan. 2005.

VOET, A. R. D. MAEYER, M.; CHRIST, F.; DEBYSER, Z. Targeting Integration beyond Strand Transfer: Development of Second-Generation HIV Integrase Inhibitors. In: 1. ed. **Antiviral Drug Strategies**: Wiley-VCH Verlag GmbH & Co. KGaA, 2011. p.51-71. ISBN 9783527635955.

WAI, J. S.; KIM, B.; FISHER, T. E.; ZHUANG, L.; EMBREY, M. W.; WILLIAMS, P. D.; STAAS, D. D.; CULBERSON, C.; LYLE, T. A.; VACCA, J. P.; HAZUDA, D. J.; FELOCK, P. J.; SCHLEIF, W. A.; GABRYELSKI, L. J.; JIN, L.; CHEN, I. W.; ELLIS, J. D.; MALLAI, R.; YOUNG, S. D. Dihydropyridopyrazine-1,6-dione HIV-1 integrase inhibitors. **Bioorganic & Medicinal Chemistry Letters**, Amsterdam, v. 17, n. 20, p. 5595-9, aug. 2007.

WIELENS, J.; CROSBY, I. T.; CHALMERS, D. K. A three-dimensional model of the human immunodeficiency virus type 1 integration complex. **Journal of Computer-Aided Molecular Design**, Berlim, v. 19, n. 5, p. 301-17, maio. 2005.

WORLD HEALTH ORGANIZATION. **O uso de medicamentos antirretrovirais para tratamento e prevenção da infecção pelo HIV. Resumo das principais características e recomendações**, 2013. Disponível em: <http://apps.who.int/iris/bitstream/10665/85322/14/WHO_HIV_2013.7_por.pdf>. Acesso em 20 jun, 2014.

_____. HIV/AIDS. 2014a. Disponível em: <<http://www.who.int/hiv/en/>>. Acesso em 25 jun, 2014.

_____. HIV/AIDS-Global Health Observatory (GHO) 2014b. Disponível em: <http://www.who.int/gho/hiv/en/>. Acesso em 25 jun, 2014.

WOLD, S.; ERIKSSON, L.; CLEMENTI, S. Statistical Validation of QSAR Results. In: 1.ed. **Chemometric Methods in Molecular Design**: Wiley-VCH Verlag GmbH, 2008. p.309-338. ISBN 9783527615452.

8. Anexos

8.1 Anexo 1

Tabela 1 Detalhamento dos valores preditos pelo modelo B e observados para a variável independente para o conjunto de predição ($N=145$).

Composto	y real	y pred	Resíduo
1	7,699	7,349	0,350
2	8,000	7,273	0,727
3	7,824	7,153	0,671
4	8,000	7,124	0,876
5	8,000	7,147	0,853
6	7,824	7,170	0,654
7	8,046	7,115	0,931
8	8,000	7,294	0,706
9	7,678	7,195	0,483
11	7,569	7,293	0,276
13	7,698	7,365	0,333
15	7,208	7,130	0,078
16	8,000	7,178	0,822
18	7,000	7,204	-0,204
20	6,699	7,101	-0,402
22	7,046	7,361	-0,315
23	7,699	7,201	0,498
24	6,887	7,351	-0,464
25	7,523	7,202	0,321
27	7,699	7,152	0,547
28	7,699	7,209	0,490
29	7,523	7,171	0,352
30	7,699	7,210	0,489
31	7,699	7,420	0,279
32	7,523	7,290	0,233
33	7,494	7,339	0,155
34	8,301	7,307	0,994
37	8,222	7,251	0,971

38	7,699	7,324	0,375
39	8,301	7,247	1,054
41	8,000	7,210	0,790
42	7,824	7,212	0,612
43	8,097	7,286	0,811
44	7,208	7,486	-0,278
47	8,301	7,142	1,159
48	7,678	7,170	0,508
49	8,097	7,426	0,671
50	8,155	7,325	0,830
51	8,155	7,378	0,777
52	7,678	7,274	0,404
53	8,097	7,031	1,066
54	8,301	7,344	0,957
55	8,301	7,329	0,972
56	8,046	7,382	0,664
57	7,921	7,362	0,559
58	8,222	7,513	0,709
59	8,301	7,239	1,062
60	8,155	7,264	0,891
61	8,046	7,383	0,663
62	8,222	7,374	0,848
63	8,155	7,204	0,951
64	8,523	7,229	1,294
68	7,678	7,304	0,374
69	8,301	7,322	0,979
70	8,097	7,296	0,801
73	8,222	7,401	0,821
74	8,699	7,266	1,433
75	8,222	7,310	0,912
77	8,301	7,310	0,991
78	8,222	7,277	0,945
79	8,000	7,053	0,947

80	8,000	7,012	0,988
81	8,000	7,109	0,891
82	7,699	6,903	0,796
83	7,699	6,979	0,720
85	8,000	7,253	0,747
87	7,699	7,198	0,501
88	7,699	7,045	0,654
89	7,699	7,053	0,646
90	7,699	7,234	0,465
91	7,699	6,990	0,709
92	7,155	7,017	0,138
93	7,155	7,106	0,049
94	7,046	7,152	-0,106
96	6,745	7,074	-0,329
99	8,097	7,260	0,837
100	8,301	7,195	1,106
101	7,721	7,393	0,328
102	8,155	7,187	0,968
104	8,000	7,103	0,897
105	8,155	7,083	1,072
106	8,154	7,301	0,853
107	8,154	7,420	0,734
108	7,721	7,308	0,413
109	7,921	7,266	0,655
112	7,585	7,242	0,343
113	7,444	7,221	0,223
115	7,678	7,238	0,440
116	7,721	7,209	0,512
117	7,602	7,095	0,507
118	7,678	7,242	0,436
119	7,678	7,095	0,583
120	7,656	7,192	0,464
122	7,444	7,226	0,218

123	7,155	7,187	-0,032
124	7,745	7,309	0,436
125	8,046	7,081	0,965
126	6,770	7,127	-0,357
128	7,620	7,314	0,306
129	6,733	7,289	-0,556
131	7,301	7,119	0,182
132	7,678	7,270	0,408
134	7,745	7,223	0,522
136	7,000	7,003	-0,003
137	7,699	7,167	0,532
138	7,699	6,972	0,727
139	7,301	7,040	0,261
141	6,824	7,083	-0,259
142	6,824	7,231	-0,407
145	7,222	6,894	0,328
146	7,155	7,018	0,137
148	6,699	7,022	-0,323
150	7,301	7,230	0,071
151	7,222	7,235	-0,013
152	7,398	7,320	0,078
153	8,000	7,208	0,792
154	7,284	7,105	0,179
156	7,097	7,180	-0,083
158	7,222	7,028	0,194
159	7,097	7,108	-0,011
160	6,699	6,952	-0,253
161	7,301	7,094	0,207
162	7,071	6,953	0,118
163	7,770	7,044	0,726
166	7,398	7,192	0,206
168	6,678	7,052	-0,374
169	7,699	7,035	0,664

170	6,215	7,043	-0,828
173	6,699	7,041	-0,342
175	8,000	7,212	0,788
176	7,301	7,071	0,230
177	7,699	7,155	0,544
178	7,046	7,067	-0,021
179	7,301	6,899	0,402
182	7,398	7,086	0,312
183	7,155	7,088	0,067
184	6,886	7,081	-0,195
185	7,301	7,286	0,015
186	7,699	7,207	0,492
189	8,000	7,053	0,947
192	7,000	6,864	0,136
193	7,398	6,892	0,506
194	6,638	6,870	-0,232
198	7,398	6,986	0,412
199	7,699	7,092	0,607

8.2 Anexo 2

Artigo revista: Chemometrics and Intelligent Laboratory Systems

Relação Estrutura Atividade Quantitativa de pirimidonas, pirimidinas e piridopirazinas carboxamidas inibidoras da HIV 1 Integrase usando Seleção por preditores ordenados (OPS) e Regressão por Quadrados Parciais Mínimos (PLS)

Luana Janaína de Campos^a, Eduardo Borges de Melo^{a*}

^aLaboratório de Química Medicinal e Ambiental Teórica (LQMAT), Departamento de Farmácia, Universidade Estadual do Oeste do Paraná (UNIOESTE), Rua Universitária 2069, Cascavel, Paraná, 85819-110, Brasil.

*Autor correspondente. Fone: (55 45) 3220-3256; e-mail: eduardo.b.de.melo@gmail.com

Neste estudo foram modelados 199 compostos derivados de pirimidonas, pirimidinas e piridopirazinas carboxamidas com atividade inibitória sobre a HIV 1 Integrase. Posteriormente, um estudo QSAR multivariado foi realizado com 54 moléculas empregando Seleção por Preditores Ordenados (OPS) e Quadrados Mínimos Parciais (PLS) para a seleção de variáveis e construção dos modelos, respectivamente. As classes de descritores utilizados foram constitucionais, topológicos, eletrotológicos e geométricos ou moleculares. O modelo real selecionado mostrou-se robusto e livre de correlação ao acaso, além disto demonstrou qualidade estatística interna e externa satisfatória ($R^2 = 0,756$; $RMSEC = 0,371$; $F_{(2,641)} = 27,111$; $Q^2_{LOO} = 0,632$; $RMSECV = 0,456$; $Average\ r_m^2_{LOO-scaled} = 0,510$; $\Delta r_m^2_{LOO-scaled} = 0,199$; $R^2_{pred} = 0,735$; $Average\ r_m^2_{pred-scaled} = 0,645$; $\Delta r_m^2_{pred-scaled} = 0,053$; $K = 1,049$; $K' = 0,950$; $|R^2_0 - R'^2_0| = 0,126$). Uma vez validado estatisticamente, o modelo de treinamento foi empregado para predição das atividades de um segundo conjunto de dados (conjunto de predição, $N = 145$). O RMSD entre os valores observados e preditos foi de 0,698. Apesar de ser um valor fora dos padrões, apenas 15 (10,34%) das amostras exibiram valores de resíduos superior a 1 unidade logarítmica, resultado considerado aceitável. Os resultados para os domínios de aplicabilidade de Willians e Euclidiano em relação as amostras do conjunto de predição revelam que as predições não ocorreram por extrapolação e que o modelo é representativo do espaço químico dos compostos em estudo. Além disto, o descritor selecionado E_T (energia total da molécula) é consistente com achados bibliográficos prévios.

Palavras Chaves: QSAR; Integrase; Carboxamidas; AIDS; OPS; PLS.

1. Introdução

Uma das abordagens mais utilizadas no campo do planejamento de fármacos auxiliado por computador são os estudos de relação estrutura atividade quantitativa (QSAR). Esses estudos se baseiam no fato de que as propriedades de moléculas orgânicas dependem de sua estrutura química, e podem ser quantitativamente descritas por modelos matemáticos multiparamétricos [1, 2]. Esta abordagem encontra ampla aplicação para a avaliação do impacto em potencial de produtos químicos na saúde humana, em sistemas ecológicos e em processos tecnológicos [3].

Considerando a necessidade do desenvolvimento de novos fármacos antirretrovirais, uma classe de compostos detentores de atividade anti-HIV (*Human Immunodeficiency Virus*) amplamente estudada são os derivados de dicetoácidos (DCAs). Estes compostos são conhecidos por inibirem a enzima HIV-integrase (HIV-IN), pois podem quelar o cofator Mg^{2+} [4, 5]. Atualmente, os fármacos Raltegravir [6, 7], Elvitegravir [8] e Dolutegravir [9] (Fig. 1), capazes de inibir a reação de transferência de fita (*strand transfer*, ST) catalizada por esta enzima, já são utilizados na terapêutica.

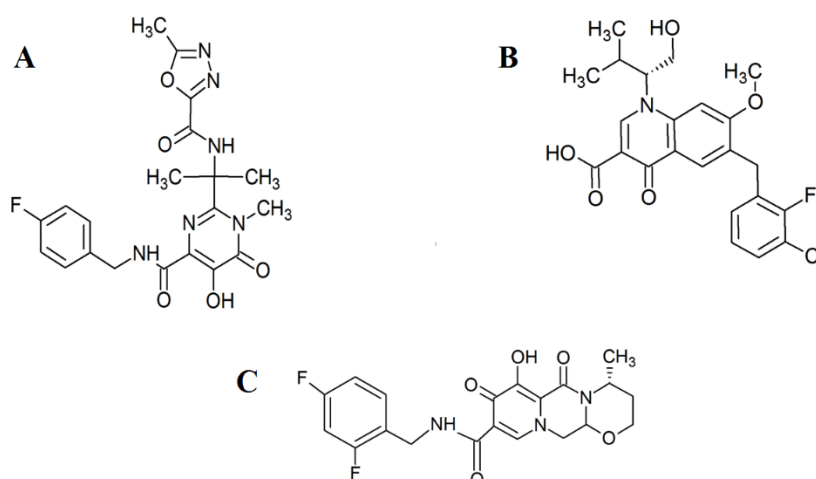


Fig. 1. Estruturas químicas dos inibidores da HIV-IN em uso clínico (A: raltegravir; B: elvitegravir; C: dolutegravir).

Os fármacos mostrados na figura 1 podem ser enquadrados na classe dos inibidores de transferência de fita da integrase (*integrase strand transfer inhibitors*, INSTI). Pirimidinas, pirimidonas e piridopirazinas são estruturas heterocíclicas contendo os núcleos básicos representados na Fig. 2, os quais na presença do grupamento carboxamida apresentam o farmacóforo dos DCAs. A partir destes núcleos bases, uma grande quantidade de derivados já encontra-se descrita na literatura [10, 11]. Um dos fatos que torna esta classe terapêutica atrativa para o desenvolvimento de novos

antirretrovirais é que as células de mamíferos não apresentam enzimas homólogas a integrase, o que aumenta a especificidade destes fármacos contra o vírus, apresentando baixo potencial de toxicidade [12].

As metodologias de QSAR baseiam-se na hipótese de que, utilizando modelos matemáticos é possível prever a atividade biológica de novos análogos, auxiliando na seleção daqueles de maior interesse em sua preparação [2, 13]. Assim, o objetivo deste trabalho foi obter um modelo QSAR estatisticamente validado, com alta capacidade de predição de novos inibidores da HIV-IN derivados de pirimidinas, pirimidonas e piridopirazinas carboxamidas, potencialmente útil como ferramenta de apoio ao planejamento de novos derivados. Desta maneira, foi selecionado da literatura um conjunto de 199 compostos descritos como inibidores da reação de transferência de fita [14-24]. O estudo foi realizado utilizando a metodologia de seleção de variáveis Ordered Predictors Selection (OPS) [25], sendo o modelo construído utilizando quadrados mínimos parciais (PLS).

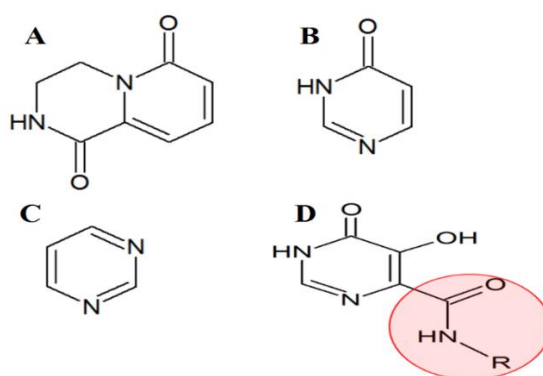


Fig. 2. Núcleo básico das piridopirazinas em A; pirimidonas em B; pirimidinas em C; inserção do grupamento carboxamidas à pirimidona.

2. Materiais e métodos

2.1 Conjunto de dados

Foi selecionado da literatura um conjunto de dados formado por 199 compostos descritos como INSTR em relação a HIV-1 IN, sendo este conjunto formado por: (i) 134 compostos derivados de pirimidonas sintetizadas e testadas por Ferrara *et al.* [14], Muraglia *et al.* [15], Gardelli *et al.* [16], Di Francesco *et al.* [17], Nizi *et al.* [18], Petrocchi *et al.* [19] e Donghi *et al.* [20]; (ii) 56 derivados de pirimidinas Pace *et al.* [21], Summa *et al.* [22] e Petrocchi *et al.* [23]; e (iii) 9 derivados de piridopirazinas Wai *et al.* [24]. Segundo os autores, todos os compostos foram ensaiados pelo mesmo protocolo [26]. Derivados descritos como misturas racêmicas foram excluídos do estudo. A estrutura base

do conjunto é apresentada na Fig. 3. No material suplementar (Tabelas 1 a 9) são disponibilizados os compostos utilizados no estudo e suas respectivas atividades biológicas.

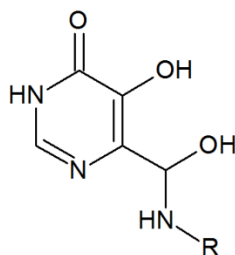


Fig. 3. Estrutura básica. Representação 2D.

A inibição enzimática foi medida quanto à concentração necessária (em nanomolar, nM, ou micromolar, μ M, dependendo da referência) para reduzir em 50% a reação de ST (IC_{50}). Os valores observados foram convertidos em seus correspondentes $-\log IC_{50}$ (ou pIC_{50}), e desta maneira as atividades ficaram distribuídas dentro do intervalo de 3.067 unidades logarítmicas (pIC_{50} 5.602 a 8.699).

O conjunto de dados foi dividido em dois subconjuntos. O primeiro foi destinado a modelagem dos dados, com 54 compostos (material suplementar, ver destaques nas Tabelas 1 a 9) selecionados de modo a representar adequadamente a variabilidade estrutural e a faixa de atividade biológica do conjunto de dados. O segundo, com os demais compostos, foi utilizado como um conjunto de predição, visando avaliar o resultado hipotético do uso do modelo obtido na anterior proposição destes derivados, levando em conta que este é o objetivo principal de estudos QSAR aplicados a planejamento de novos fármacos.

2.2 Modelagem molecular

As estruturas dos 199 compostos do conjunto de dados foram construídas a partir de uma estrutura cristalográfica similar (código DOTRUZ) obtida no *Cambridge Structural Database* (CSD) [27]. Todas as moléculas foram inicialmente otimizadas em mecânica molecular (MM+), com otimizações alternadas com ciclos de dinâmica molecular (1 ps, 300 K). O procedimento foi repetido até que a energia obtida não variasse de maneira significativa, indicando a obtenção de uma possível estrutura de energia mínima. Em seguida, os compostos foram otimizados por mecânica quântica, na sequência dos níveis de teoria *Austin Model 1* (AM1), *Hartree-Fock* (HF/6-31Gd,p) e, finalmente, pela Teoria do Funcional de Densidade (B3LYP/6-311G++d,p). As funções

difusas (++) foram utilizadas porque a adição de orbitais s e p muito difusos melhora a descrição de pares de elétrons de alta energia [28], o que as tornam interessantes quando é considerado o mecanismo de ação dos compostos aqui estudados.

2.3 Descritores moleculares

Utilizando as estruturas otimizadas, foram obtidos os seguintes descritores eletrônicos: cargas parciais de Mulliken, NBO (*Natural Bond Orders*) e *CHELPG* (*Charges from Electrostatic Potential Grid-based*), energia dos orbitais moleculares de valência ($E_{\text{HOMO}-1}$, E_{HOMO} , E_{LUMO} , $E_{\text{LUMO}+1}$), momento de dipolo total (D) e nos eixos x, y e z (D_x , D_y e D_z), e energia total (E_T). Além disso, utilizando as equações descritas por Todeschini e Consoni [29] e os valores obtidos para os descritores de energia dos orbitais moleculares, foram derivados mais 10 descritores: GAP (diferença energética entre E_{HOMO} e E_{LUMO}), índice de energia de ativação (*activation energy index*, AEI), fração de energia HOMO/LUMO (*energy fraction*, $f(\text{H/L})$), (*hardness*, n), potencial de ionização (*ionization potential*, IP), afinidade eletrônica (*electronic affinity*, EA), (*softness*, S), índice de eletrofilicidade no estado fundamental (*electrophilicity index in the ground state*, ω_{GS}), índice de eletrofilicidade (*eletrofilicity*, ω) eletronegatividade molecular (*molecular electronegativity*, χ). Deste modo, foram utilizados no estudo 52 descritores de caráter eletrônico.

As geometrias de menor energia também foram utilizadas no programa Dragon 6 para a obtenção de descritores das seguintes classes: descritores constitucionais (*constitutional descriptors*), descritores de contagem de grupos funcionais (*functional groups counts*), descritores de carga (*charge descriptors*), propriedades moleculares (*molecular properties*), contagem de circuitos e caminhos (*walk and path counts*), índices de informação (*information indices*), índices de adjacência de ponta (*edge adjacency indices*), índices de cargas topológicas (*topological charge indices*), descritores topológicos (*topological descriptors*), índices de conectividade (*connectivity indices*), descritores geométricos (*geometrical descriptors*), auto correlações 2D (*2D autocorrelations*), auto correlações 3D (*3D autocorrelations*), pares de átomos 2D e 3D (*2D e 3D atom pairs*), autovalores de Burden (*Burden eigenvalues*), descritores 3D de Morse (*3D-MoRSE descriptors*), autovalores baseados em índices (*eigenvalues-based indices*), descritores GETAWAY (*GETAWAY descriptors*), descritores RDF (*RDF descriptors*), descritores WHIM (*WHIM descriptors*), perfis moleculares de Randic

(*Randic molecular profiles*), índices ETA (*ETA indices*), impressões digitais CATS2D (*CATS2D finger prints*). De modo geral, estas classes podem ser descritas como descritores constitucionais, topológicos, eletrotológicos e geométricos ou moleculares.

A matriz de descritores obtida passou por um processo de redução de variáveis, devido a grande quantidade obtida. Assim, ainda no Dragon 6, o total foi reduzido através da eliminação de um membro dentre pares com alta correlação entre si (maior do que 0,9), de descritores invariantes ou quase-invariantes, e aqueles que apresentavam desvio padrão menor que 0,001. Finalmente, foi realizada uma redução manual, com intuito de excluir descritores com baixa variação não eliminados pelo programa. Em seguida, utilizando o programa QSAR Modeling [30], foram excluídas aquelas variáveis que apresentassem valores absolutos dos coeficientes de correlação de Pearson com a atividade biológica ($|r|$) abaixo de 0,3, visando eliminar descritores que não apresentassem informação relevante para a construção dos modelos. Este processo resultou em uma matriz de descritores formada por 319 descritores.

2.4 Estudo QSAR

Durante o desenvolvimento de modelos QSAR, os descritores moleculares constituem a ferramenta básica que transforma a informação química em informação numérica adequada para a aplicação de procedimentos computacionais [29]. Os descritores podem ser obtidos por procedimentos experimentais [31] ou teóricos. Estes últimos, mais utilizados, são obtidos através de diferentes teorias, como química quântica, teoria de grafos, dentre outras [29]. Devido ao grande número de descritores, faz-se necessário utilizar metodologias que automatizem a seleção daqueles, dentre os disponíveis, que mais contribuem para a atividade biológica. Estes descritores serão aqueles utilizados para construir os modelos mais adequados para descrever atividade biológica em investigação [2, 32]. Neste estudo foi utilizado o método OPS [25], um algoritmo iterativo para seleção de variáveis que utiliza regressão por PLS [33] para a construção de modelos. Este método seleciona as variáveis através de um rearranjo da matriz de descritores de acordo com sua relevância. Deste modo, a variável mais importante é disposta na primeira coluna da matriz, a segunda mais relevante na coluna subsequente, e assim por diante [25, 30]. Esta reorganização é realizada segundo um vetor informativo (vetor de regressão do PLS, vetor de correlação, e o produto entre eles) [25].

Os descritores selecionados são utilizados para a construção de modelos também utilizando a regressão por PLS. Esta abordagem é interessante para problemas quimiométricos, uma vez que projeta o grande ambiente possível de informação contida nos descritores para um espaço pequeno, na forma de variáveis latentes (LVs), mutuamente ortogonais entre si [2, 34, 35]. A construção dos modelos, também foi necessário pré-processar os dados, pois conjuntos de dados de QSAR geralmente consistem em variáveis que possuem diferentes unidades e/ou grandes faixas de variação. O método básico para estudos QSAR, e que foi utilizado neste estudo, é o autoescalamamento [30]. Esta abordagem consiste em subtrair de cada elemento de uma coluna da matriz de dados o valor médio da respectiva coluna, e dividir o resultado pelo desvio padrão da mesma. Assim, a influência de uma variável dominante é minimizada.

Os modelos obtidos através do uso da abordagem OPS foram classificados segundo critérios estatísticos. Inicialmente, os modelos foram classificados pela raiz quadrada do erro da validação cruzada (*Root Mean Square Error of Cross-Validation, RMSECV*), com o objetivo de obter um conjunto reduzido de descritores que poderiam levar a modelos com baixo erro. Nos ciclos subsequentes, o coeficiente de determinação da validação cruzada *leave-one-out* (Q^2_{LOO}) foi utilizado para auxiliar na seleção do subconjunto de descritores que levassem a melhor capacidade de predição. Este processo foi repetido de maneira iterativa até que o modelo com maior valor de Q^2_{LOO} e menor número de descritores quanto possível fosse encontrado.

2.5 Validação dos modelos

O uso de um modelo de predição é condicionado a sua prévia aprovação em uma série de testes estatísticos. Isto visa garantir que as variáveis dependentes de interesse sejam previstas de maneira mais confiável possível. O método mais empregado de validação de modelos QSAR consiste em duas etapas: (i) testes usados para avaliar a capacidade de previsão dos modelos em relação aos compostos que auxiliaram em sua construção, ou validação interna; (ii) e testes de avaliação da previsão para compostos que não foram empregados na obtenção, ou validação externa [36, 37].

Em relação à validação interna, o grau de ajuste do modelo foi avaliado através do coeficiente de determinação múltipla (R^2) e do erro associado medido pelo correspondente raiz quadrada do erro da calibração (*RMSEC*). A significância da regressão avaliada utilizando o teste de Fischer (F , 95% de confiança, $\alpha=0,05$). O grau de previsibilidade do modelo avaliado pela validação cruzada LOO, onde é obtido o Q^2_{LOO} ,

e seu associado, expresso pelo *RMSECV*, que deve apresentar o menor valor possível [2, 36]. Também foram utilizados os testes *RmSquare* aplicados a validação cruzada (*Average r_m^2 LOO-scaled* e *Δr_m^2 LOO-scaled*) [38, 39].

Nesta etapa, também deve ser avaliada a robustez dos modelos, através da validação cruzada *leave-N-out* (LNO), com uso do Q^2_{LNO} (coeficiente de determinação entre as atividades biológicas observadas e previstas durante o processo de validação cruzada *leave-N-out*). Este teste foi aplicado para $N=1-14$ (26% do conjunto de dados) e repetido seis vezes para cada “ N ”. Também foi avaliado a presença de correlação ao acaso nas informações explicadas e previstas, através do teste de randomização do y (y -randomização) [36], onde apenas este vetor é randomizado. Foi utilizada a abordagem sugerida por Eriksson e colaboradores, onde o $|r|$ entre o vetor y original e os vetores randomizados é utilizado para avaliar a presença de correlações espúrias [40].

Para a validação externa, um conjunto-teste de 14 compostos (35, 45, 66, 72, 86, 97, 98, 121, 147, 157, 164, 171, 195 e 196), totalizando 26 % do conjunto de treinamento, os quais são representativos da faixa de variação da atividade biológica e da variabilidade estrutural das amostras [30]. Como a atividade biológica das amostras do *test set* é conhecida, pode-se fazer uma comparação entre o valor previsto pelo modelo e o valor real, utilizando parâmetros estatísticos [41]. Assim, a validação externa é muito mais confiável para assegurar a capacidade preditiva dos modelos QSAR, pois as amostras não foram utilizadas na construção dos modelos [37, 42]. A qualidade da predição foi avaliada através do coeficiente de determinação da validação externa (R^2_{pred}), da raiz quadrada do erro quadrático médio da predição externa (*RMSEP*), das estatísticas de Gobraikh e Tropsha (k , k' e $|R^2_o - R'^2_o|$) [41] e *RmSquare* aplicados a validação externa (*Average r_m^2 -scaled* e *Δr_m^2 -scaled*) [43, 44].

Os parâmetros utilizados foram selecionados para avaliar de maneira bastante rigorosa a capacidade preditiva dos modelos QSAR, determinando a proximidade entre a atividade observada experimentalmente e a prevista. O uso destes critérios visa garantir uma mínima confiabilidade, qualidade e eficácia dos modelos de regressão para fins práticos [36, 38, 45]. As equações utilizadas para obter os parâmetros estão disponíveis na Tabela 1.

Tabela 1

Parâmetros estatísticos empregados para a validação interna/externa.

PARÂMETRO	SIGNIFICADO	EQUAÇÃO
R^2	Coeficiente de determinação ^a	$R^2 = 1 - \frac{\sum_i (y_{obsi} - y_{ci})^2}{\sum_i (y_{obsi} - \bar{y}_{ci})^2}$
$RMSEC$	Raiz quadrada do erro da calibração ^a	$RMSEC = \sqrt{\frac{\sum_i (y_{obsi} - y_{ci})^2}{n}}$
F	Teste F^b	$F = \frac{\frac{\sum_i (y_{obsi} - y_{ci})^2}{k}}{\frac{\sum_i (y_{obsi} - \bar{y}_{obs})^2}{n - p - 1}}$
Q^2_{LOO}	Coeficiente de determinação da validação cruzada ^{c,d}	$Q^2_{LOO} = 1 - \frac{\sum_i (y_{obsi} - y_{vi})^2}{\sum_i (y_{obsi} - \bar{y}_{obs})^2}$
$RMSECV$	Raiz quadrada do erro da validação cruzada ^c	$RMSECV = \sqrt{\frac{\sum_i (y_{obsi} - y_{vi})^2}{n}}$
R^2_{pred}	Coeficiente de determinação da validação externa ^e	$R^2_{pred} = 1 - \frac{\sum_i (y_{obsi} - y_{evi})^2}{\sum_i (y_{obsi} - \bar{y}_{obs})^2}$
$RMSEP$	Raiz quadrada do erro quadrático médio da predição externa	$RMSEP = \sqrt{\frac{\sum_i (y_{obsi} - y_{evi})^2}{n}}$
k e k'	Inclinações das linhas das regressões lineares de predição ^e	$k = \frac{\sum_i (y_{obsi} - y_{evi})}{\sum_i y_{evi}}$ $k' = \frac{\sum_i (y_{obsi} - y_{evi})}{\sum_i y_{obsi}}$

$ R^2_0 - R'^2_0 $	Diferença absoluta entre os coeficientes de determinação múltipla ^e	$R_0^2 = \left(1 - \frac{\sum(y_{obsi} - ky_{obsi})^2}{\sum(y_{obsi} - \bar{y}_{evi})^2}\right)$
		$R'^2_0 = \left(1 - \frac{\sum(y_{obs} - ky_{obs})^2}{\sum(y_{obs} - \bar{y}_{obsi})^2}\right)$
Average r_m^2 -scaled	Média de r^2 modificado escalado ^e	$Média\ r_m^2 = \frac{(r_m^2 + r'^2_m)}{2}$
Δr_m^2 -scaled	Variação de r^2 modificado escalado ^e	$\Delta r_m^2 = r_m^2 - r'^2_m $
Y_i scaled	Valor da atividade biológica observado ou predito escalado ^f	$Y_i\ scaled = \frac{y - y_{min}(obs\ ou\ evi)}{y_{max}(obs\ ou\ evi) - y_{min}(obs\ ou\ evi)}$

^a qualidade do ajuste; ^b significância; ^c validação cruzada, se $n = 1$, e teste de robustez *leave-N-out*, se $n > 1$; ^d qualidade da predição interna; ^e validação externa; ^f pré-tratamento de dados; y : atividade biológica; $\bar{Y}$: atividade biológica média; obs: valores experimentais; ci: atividade estimada no modelo de regressão para o conjunto de treinamento completo; vi: valor obtido na validação cruzada LOO; 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; evi: valor obtido na validação externa; r_m^2 : coeficiente de correlação modificado obtido a partir da regressão com dados observados e preditos; r'^2_m : coeficiente de correlação modificado obtido da regressão com a troca de eixos.

2.6 Teste dos modelos

Após o modelo representado pela equação 1 ser validado interna e externamente, seu respectivo modelo de treinamento ($n=54$) foi utilizado para a previsão da atividade de 145 compostos análogos do conjunto de dados estudado, obtidos das mesmas fontes bibliográficas. O uso do modelo de treinamento (equação 2) para a predição da atividade biológica deste segundo conjunto de predição se deve ao fato de que este modelo contém maior quantidade de informação

química a respeito da classe de compostos aqui estudados em relação ao seu respectivo modelo real (equação 1). Este teste teve como objetivo avaliar o potencial do modelo como ferramenta de apoio a síntese de novos derivados. Para isto, estes compostos foram avaliados quanto a adequação no domínio de aplicabilidade do modelo. Foram utilizados tanto os testes de domínio de aplicabilidade euclidiano quanto o de Willians [46, 47].

2.7 Softwares

A construção do conjunto de dados e as otimizações aos níveis mecânica molecular e semi-empírico foram realizadas no programa HyperChem 7 [48]. As otimizações em HF e DFT foram realizadas no programa Gaussian 09 [49]. Os descritores eletrônicos foram obtidos no programa Gauss View 5 [50], enquanto para os demais foi utilizado o programa Dragon 6 [51]. O processo de redução de variáveis por alta correlação entre si e baixo desvio padrão também foi realizada através do programa Dragon 6. O programa OpenBabel 2.3.1 [52] foi usado para realizar todas as conversões de formatos químicos, quando necessário.

As etapas de redução de variáveis por baixa correlação com a atividade biológica, seleção de variáveis (OPS), construção dos modelos (PLS) e validação interna foram realizadas no programa QSAR Modeling [30] (<http://lqta.igq.unicamp.br>). Os modelos foram refinados com o auxílio do programa Pirouette 4 [53]. A validação externa foi realizada com o *Xternal Validation* 1.0, enquanto o domínio de aplicabilidade euclidiano foi obtido utilizando o programa *Euclidean Applicabilty Domain 1.0* (ambos disponíveis em <http://dtclab.webs.com/software-tools>). Finalmente, o domínio de aplicabilidade de Willians e alguns parâmetros da validação interna e externa foram calculados através de uma tabela Excel “*in house*”.

3. Resultados e discussão

O melhor modelo obtido é apresentado na equação 1. A análise dos resíduos de Student pelos respectivos *leverages* mostrou que o modelo não apresenta nenhum *outlier* (i.e., amostras anômalas) [30, 32, 54]. Desta maneira, o modelo real obtido foi formado por 7 descritores, que deram origem a quatro VLs (informação acumulada: 77,438%; VL1: 34,142%; VL2: 6,744%; VL3: 12,657%; VL4: 23,894). Os valores dos descritores dos 54 compostos do conjunto de dados, além dos valores observados e previstos (validação cruzada e externa) são apresentados na Tabela 2. Os resultados dos testes de validação interna e externa são disponibilizados na Tabela 3.

$$pIC_{50} = 7,503 + 1,740 (VE1_B(p)) - 30,263 (R5m+) - 1,111 (Eig09_EA(bo)) + 0,827 (H2m) - 0,140 (Mor28s) - 0,001 (E_T) - 26,036 (SpPosA_X)$$

(1)

A qualidade do ajuste dos dados, ou variância explicada pelo modelo, equivale a 75,6%, sendo o mínimo recomendado é 60% ($R^2 > 0,6$). O *RMSEC* representa a variabilidade na atividade não explicada pelo modelo e, conseqüentemente, deve ser o menor possível, também apresenta valor aceitável [36, 45]. A razão entre a variabilidade explicada pelo modelo e a variabilidade que permanece sem explicação é expressa pelo resultado do teste *F* (95% de confiança, $\alpha=0,05$). Considerando que o valor obtido é maior que seu valor crítico ($F_c=2,641$, para $p=4$, $n-p-1=35$), e que quanto maior o valor de *F*, mais significativo é o modelo, o modelo 1 pode então ser classificado como significativo [30, 55].

A previsibilidade interna foi testada pela validação cruzada LOO, que avalia a quantidade de informação ou variabilidade que um modelo pode prever. Se esta apresentar elevado grau de previsibilidade, seu Q^2_{LOO} será o mais próximo possível de 1, com mínimo recomendado de 0,5 (i.e., 50%), enquanto o *RMSECV* deve apresentar o menor valor possível. Os resultados mostram que o modelo é capaz de prever 63,2% de informação, com um *RMSECV* associado de 0,456. Para os parâmetros *average $r_m^2_{LOO-scaled}$* e *$\Delta r_m^2_{LOO-scaled}$* , modelos com boa previsibilidade devem apresentar valores superiores a 0,5 para o primeiro e menores que 0,2 para o segundo, resultados que foram aqui obtidos [56]. Finalmente, a diferença entre os valores de R^2 e Q^2_{LOO} é de apenas 0,124, o que indica que a probabilidade de estar ocorrendo sobreajuste dos dados é muito pequena [30, 55, 57]. Assim, considerando o exposto, o modelo 1 possui boa capacidade de predição, é significativo, e possui e baixa tendência de sofrer sobreajuste.

Tabela 2

Valores dos descritores selecionados para as amostras do conjunto de dados e resultados dos processos de predição (validação cruzada e externa).

Compounds	VE1_B(p)	R5m+	Eig09_EA(bo)	H2m	Mor28s	E_T	SpPosA_X	pIC₅₀ real	pIC₅₀ pred	Residue
10	3,941	0,034	1,948	2,215	1,907	-1915,737	0,292	7,620	7,402	0,218
12	3,924	0,028	1,948	1,809	1,532	-1846,779	0,292	7,494	7,174	0,320
14	3,979	0,026	1,948	2,292	-0,927	-1990,986	0,300	7,721	8,164	-0,443
17	3,785	0,028	1,500	1,629	-0,336	-1282,780	0,292	6,721	6,927	-0,206
19	3,791	0,025	1,500	1,635	-0,776	-1358,006	0,299	6,745	7,010	-0,265
21	3,786	0,030	1,266	1,626	1,137	-1604,213	0,295	7,097	7,209	-0,112
26	3,799	0,027	1,734	1,807	-0,027	-1451,533	0,289	7,215	7,082	0,133
35*	4,153	0,045	1,500	1,870	-1,143	-1713,448	0,274	8,155	7,987	0,168
36	3,888	0,030	1,388	1,560	-0,151	-1204,151	0,288	7,796	6,994	0,802
40	4,376	0,026	1,875	1,676	-3,267	-1564,887	0,288	8,222	8,578	-0,356
45*	4,025	0,024	1,663	1,642	-1,836	-1495,413	0,298	7,119	7,284	-0,165
46	4,030	0,023	1,687	1,782	-1,563	-1569,431	0,293	8,523	7,829	0,694
65	4,065	0,030	1,751	2,110	-0,601	-1811,369	0,291	8,046	8,165	-0,119
66*	4,075	0,029	1,751	1,879	-1,789	-1488,389	0,291	8,398	7,487	0,911
67	4,092	0,028	2,046	1,828	-2,545	-1523,892	0,289	7,602	7,698	-0,096
71	4,080	0,029	2,046	1,878	-1,611	-1527,685	0,289	7,208	7,593	-0,385

72*	4,096	0,028	2,046	1,927	-1,711	-1543,745	0,289	7,886	7,362	0,524
76	4,084	0,028	1,751	2,043	-2,083	-1500,575	0,291	8,699	7,852	0,847
84	3,840	0,050	1,713	1,833	-0,506	-1698,641	0,270	7,222	7,527	-0,305
86*	3,782	0,025	1,108	1,253	-0,422	-1204,603	0,287	7,698	6,924	0,774
95	3,957	0,039	1,935	2,324	5,017	-1695,212	0,275	7,046	7,369	-0,323
97*	3,977	0,040	1,927	1,551	-1,279	-1474,812	0,286	6,638	6,562	0,076
98*	3,998	0,039	1,946	1,568	0,526	-1569,484	0,285	6,310	6,490	-0,180
103	4,300	0,022	1,500	1,953	0,722	-1604,218	0,291	8,301	8,575	-0,274
110	4,184	0,020	1,751	1,702	1,699	-1204,152	0,297	7,456	7,016	0,440
111	4,184	0,022	1,751	1,787	0,499	-1886,785	0,297	7,921	8,127	-0,206
114	4,372	0,022	1,982	1,896	0,459	-1641,131	0,300	7,745	7,940	-0,195
121*	3,909	0,027	1,500	1,555	0,608	-1201,734	0,304	6,886	6,309	0,577
127	4,004	0,022	2,072	1,767	1,718	-1565,863	0,295	6,733	6,964	-0,231
130	4,202	0,025	2,073	2,065	0,750	-1680,405	0,300	7,638	7,558	0,080
133	4,224	0,037	2,071	2,345	0,452	-1922,338	0,301	7,678	7,823	-0,145
135	3,657	0,048	1,298	1,602	-0,119	-1445,307	0,300	6,301	6,237	0,064
140	3,920	0,028	1,866	1,706	-1,422	-1452,917	0,303	6,824	6,871	-0,047
143	3,631	0,027	1,242	1,575	0,377	-1204,149	0,301	5,910	6,586	-0,676
144	3,565	0,030	0,720	1,536	0,602	-1126,698	0,290	6,699	7,070	-0,371
147*	3,709	0,026	1,500	1,635	0,614	-1223,810	0,298	7,301	6,235	1,066
149	3,575	0,028	1,866	1,796	0,193	-1397,132	0,299	5,699	6,409	-0,710

155	3,787	0,026	1,500	1,468	0,248	-1282,771	0,294	7,056	6,675	0,381
157*	3,878	0,026	1,500	1,659	-0,543	-1358,000	0,296	7,585	6,897	0,688
164*	3,698	0,041	1,291	1,362	0,380	-1366,664	0,299	6,000	5,918	0,082
165	4,171	0,035	1,766	1,531	0,427	-1520,341	0,297	7,523	6,937	0,586
167	3,960	0,045	1,899	1,850	-0,997	-1598,985	0,297	6,215	6,915	-0,700
171*	3,600	0,045	1,293	1,655	0,441	-1422,026	0,304	5,780	5,783	-0,003
172	3,536	0,052	0,927	1,816	0,045	-1726,748	0,301	7,000	6,756	0,244
174	3,538	0,076	0,927	1,859	0,247	-1742,796	0,301	5,602	6,369	-0,767
180	3,616	0,044	1,293	1,719	-0,040	-1520,543	0,306	6,260	6,327	-0,067
181	3,626	0,042	1,293	1,600	-0,076	-1445,311	0,299	6,959	6,349	0,610
187	3,643	0,050	1,293	1,711	1,162	-1505,255	0,299	6,796	6,100	0,696
188	3,620	0,044	1,293	1,801	0,684	-1505,254	0,297	7,155	6,462	0,693
190	3,632	0,059	1,293	2,644	0,819	-2325,226	0,299	8,000	7,439	0,561
191	3,777	0,037	1,018	1,338	1,795	-1129,316	0,293	5,983	6,506	-0,523
195*	3,878	0,049	0,894	1,457	1,832	-1450,356	0,293	6,432	6,545	-0,113
196*	3,852	0,043	0,884	1,803	1,160	-1450,358	0,290	7,398	7,151	0,247
197	3,845	0,051	0,902	1,716	1,924	-1450,358	0,292	6,420	7,063	-0,643

* Compostos selecionados para a validação externa.

Tabela 3

Resultados dos parâmetros estatísticos para o modelo 1.

Parâmetro	Valor
R^2	0,756
F	27,111
Valor referência do teste F	2,641
$RMSEC$	0,371
$RMSECV$	0,456
Q^2_{LOO}	0,632
$R^2 - Q^2_{LOO}$	0,124
$Average\ r_m^2_{LOO-scaled}$	0,510
$\Delta r_m^2_{LOO-scaled}$	0,199
Informação acumulada	77,438%
<i>Outlier</i>	0
R^2_{pred}	0,735
$RMSEP$	0,422
k	1,049
k'	0,950
$ R^2_0 - R'^2_0 $	0,126
$Average\ r_m^2_{pred-scaled}$	0,645
$\Delta r_m^2_{pred-scaled}$	0,053

Na Figura 4 são apresentados os gráficos correspondentes aos resultados dos testes de validação cruzada LNO e randomização do y . A robustez dos modelos foi avaliada através do procedimento de *leave-N-out*, outro processo de validação cruzada, onde um maior número de amostras é retirado do conjunto de dados (geralmente, entre 20 a 30%). Como resultado, avalia-se a oscilação nos valores de cada Q^2_{LNO} e do Q^2_{LNO} médio em relação ao valor de Q^2_{LOO} , e do desvio padrão de cada replicata [36]. Neste estudo, foi

realizado um processo de *leave-14-out* (26% do conjunto de treinamento). Na figura 4a, pode-se observar que o modelo pode ser considerado robusto, pois mantém seu Q^2_{LNO} (0.610) em apenas 0.022 unidades menor que o Q^2_{LOO} , algumas variações são observadas nos resultados, entretanto o valor para a maior variação individual em relação ao Q^2_{LOO} foi de apenas 0,091 unidades ($Q^2_{L13O} = 0.541$), e o maior desvio padrão obtido para as hexaplicatas correspondentes a cada “N” foi de apenas 0,059 (Q^2_{L6O}). Assim, o resultado pode ser considerado aceitável.

Já o teste de randomização do y (Figura 4b) é realizado através de modelos paralelos, os quais mantêm as colunas de descritores originais enquanto os valores do vetor y são aleatoriamente alterados. Os modelos resultantes devem apresentar baixa qualidade estatística. Segundo Eriksson et al. (2003) [58], deve-se realizar uma regressão entre os valores absolutos das correlações de Pearson, $|r|$, entre o vetor y original e cada novo vetor obtido após uma aleatorização, com os valores de R^2 ou Q^2_{LOO} correspondentes. As intersecções das retas obtidas devem apresentar interceptos menores que 0,05 para Q^2_{LOO} e menores que 0,3 para R^2 para que o modelo possa ser considerado livre de correlações espúrias [30, 36]. O resultado obtido mostra que o modelo obtido apresenta baixa probabilidade de sofrer de correlação ao acaso.

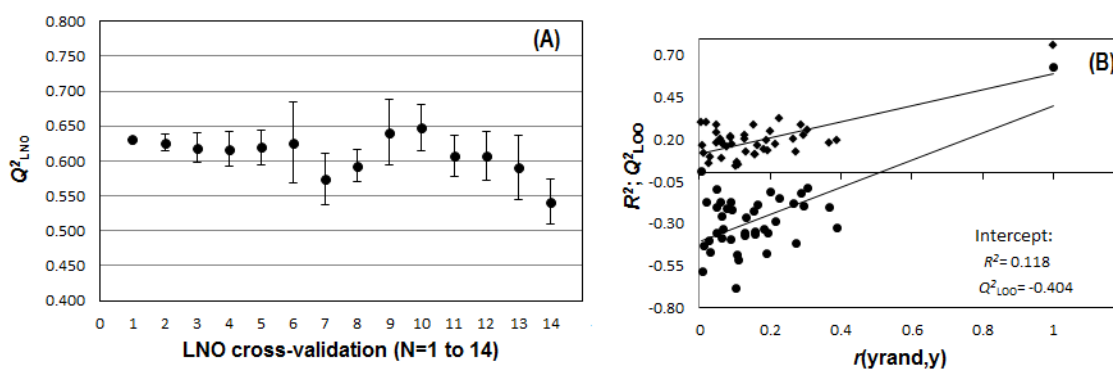


Figura 4. Resultados da validação cruzada LNO (A) e do teste de randomização do y (B).

Quanto a validação externa, os resultados apresentados na Tabela 3 mostram que o modelo real foi aprovado em todos os testes adotados. Alguns autores sugerem que o valor limite de R^2_{pred} deve ser maior que 0,6 [41], enquanto outros propõem valores maiores que 0,5 [44]. Apesar de o último valor ser o mais comumente encontrado na literatura, o valor obtido neste estudo foi de 0,735, mostrando uma ótima capacidade preditiva.

Independente dos valores adotados, a maioria dos autores atualmente concorda que o uso somente do R^2_{pred} não é suficiente para avaliar a capacidade preditiva externa de um modelo.

Golbraikh *et al.* (2003) [41] e Roy *et al.* (2012) [38, 44] propuseram alguns parâmetros para avaliação do poder de predição externa de modelos QSAR baseados nas inclinações de regressões centradas na origem (ou seja, o resultado de uma previsão ideal). Pode-se observar pelos resultados dos parâmetros k , k' (resultados esperados: dentro do intervalo de 0,85 a 1,15), $|R^2_0 - R'^2_0|$ (resultado esperado: $< 0,3$) [42], *average r_m^2 pred-scaled* (resultado esperado: $> 0,5$), e Δr_m^2 pred-scaled (resultado esperado: $< 0,2$) [38, 39] que o modelo obtido possui boa capacidade de predição [56].

Com a aprovação nos testes de validação interna e externa, o modelo de treinamento (equação 2) foi utilizado para a predição das atividades do segundo *test set* (conjunto de predição, $n=145$) (ver material suplementar tabela 1 a 10). O *RMSD* (*root mean square deviation*) entre os valores preditos e observados foi de 0,698. Apesar de estar acima dos valores obtidos para os parâmetros *RMSEC*, *RMSECV* e *RMSEP*, apenas 15 (10,34%) das amostras apresentaram resíduos acima de 1 unidade logarítmica (ver material suplementar, Tabela 10), o que pode ser considerado aceitável. Na Figura 5, são apresentados os domínios de aplicabilidade (DA) pelas abordagens de Willians (Fig. 5a) e pela distância euclidiana (Fig. 5b). No primeiro, apenas 6 amostras (4,14% do segundo *test set*) ficaram localizados fora do domínio, apesar da amostra **103** (material suplementar Tabela 5) estar muito próxima da linha limite superior de resíduos. Já no domínio de aplicabilidade Euclidiano, nenhuma das amostras apresentou uma distância média normalizada (*normalized mean distance*) maior que a linha limite. Assim, é possível propor que as predições das atividades realizadas para as amostras que compõem o segundo *test set* não são frutos de extrapolação, e que o modelo é representativo do espaço químico das moléculas em estudo.

$$\text{pIC}_{50} = 8,131 + 1,549 (\text{VE1_B(p)}) - 36,773 (\text{R5m+}) - 1,195 (\text{Eig09_EA(bo)}) + 1,251 (\text{H2m}) - 0,164 (\text{Mor28s}) - 0,001 (\text{E}_T) - 25,000 (\text{SpPosA_X}) \quad (2)$$

$n= 54$; outliers: 0; LVs= 4; informação acumulada: 79.755%; $R^2 = 0.759$; *RMSEC* = 0.372; $F(2.790)= 50.538$; $Q^2_{LOO} = 0.679$; *RMSECV*= 0.430;

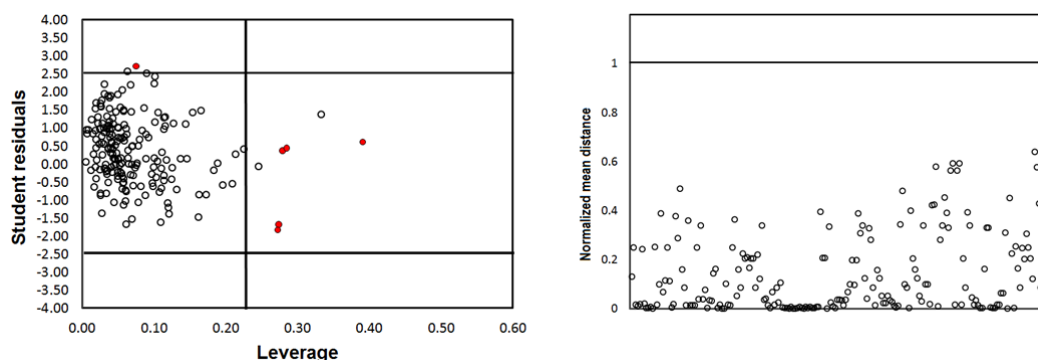


Figura 5. Resultados dos testes de domínio de aplicabilidade. Todas as amostras do estudo ($n = 199$) são apresentadas. A: resíduos de *Student* versus valores de *leverage* (DA de Willian). As amostras do segundo *test set* fora do espaço químico são representados como pontos vermelhos. A linha vertical é o valor limite (0,222). B: DA euclidiano. A linha horizontal é o valor limite de distância média normalizada.

Embora o modelo obtido tenha-se mostrado adequado para fins de predição, quando possível deve-se realizar a interpretação mecanística do modelo em relação ao mecanismo de ação das moléculas em estudo, quando este for conhecido [59]. Pirimidonas, pirimidinas e piridopirazinas carboxamidas são compostos relacionados aos DCA's, que inibem a HIV-1 IN por quelação dos cofatores Mg^{2+} presentes no sítio de ligação através de ligações com a tríade catalítica formada pelos resíduos Asp64, Asp116, and Glu152. Também estão presentes outras interações, em especial interações hidrofóbicas da cadeia lateral aromática com uma alça desordenada formada pelos resíduos 140-145 [5, 60, 61]. Os descritores selecionados pertencem a categorias diversas (Tabela 4), sendo dois topológicos, quatro geométricos e um eletrônico. Os coeficientes PLS autoescalados mostram que todos os descritores são relevantes para o modelo. Apesar das duas primeiras categorias serem de difícil interpretação, o descritor molecular E_T (e outros descritores termodinâmicos) pode ser relacionado com a estabilidade molecular [62-64]. De modo interessante, este descritor foi anteriormente selecionado no modelo previamente publicado por de Melo e Ferreira (2009) [65] para um conjunto de 33 dihidroxipirimidinas carboxamidas, obtido pela mesma abordagem (OPS/PLS).

Tabela 4

Descritores selecionados no modelo 1 e seus respectivos coeficientes/coeficientes padronizados (abordagem de autoescalamento).

Símbolo	Descritor	Classe	Coeficiente autoescalado
VE1_B(p)	Coeficiente da soma do último autovetor da matriz Burden/ ponderado pela polarizabilidade	Topológico/ Descritores baseados na matrix 2D	0,552
R5m+	Autocorrelação R máxima entre átomos separados por 5 ligações químicas/ ponderada pela massa	Geométrico/ Descritores <i>GETAWAY</i> ^a	– 0,484
Eig09_EA(bo)	Autovalor n.9 da matriz de adjacência de ponta/ ponderado pela ordem de ligação	Topológico/ Índices de adjacência de ponta.	– 0,550
H2m	Autocorrelação H entre átomos separados por 2 ligações químicas/ ponderada pela massa	Geométrico/ Descritores <i>GETAWAY</i> ^a	0,288
Mor28s	Sinal 28/ ponderado pelo I-estado	Geométrico/ Descritores 3D-MoRSE ^b	– 0,265
E _T	Energia total da molécula	Eletrônico/ Descritor termodinâmico	– 0,397
SpPosA_X	Soma positiva espectral normalizada a partir da matriz <i>Chi</i>	Geométrico/ Descritores baseados em Matriz 2D	– 0,237

Maiores detalhes em *List of molecular descriptors calculated by Dragon* [66] e Propriedades químico-quânticas empregadas em estudos das relações estrutura-atividade[67].

^a *GATEWAY: Geometry, Topology And Atom-Weights Assembly.*

^b *MoRSE: Molecule Representation of Structure based on Electron diffraction.*

4. Conclusão

Neste estudo, foi possível obter um modelo QSAR multivariado utilizando um conjunto de 199 pirimidinas, pirimidonas e piridopirazinas carboxamidas as quais têm a capacidade de inibir *in vitro* a reação ST catalisada pela HIV-IN. O modelo obtido apresentou excelente ajuste, poder de predição interna e externa, o desempenho na validação cruzada LNO mostra que este é robusto e no teste de randomização do y é

perceptível que o modelo não apresenta correlação acaso. Além disso, a equação proposta foi utilizada para previsão da atividade inibitória sobre a HIV-IN de um grande conjunto de dados, e os valores obtidos mostraram um resultado aceitável, incluindo domínio de aplicabilidade, tanto por abordagem baseada nas predições (método de Willians) quanto nos descritores moleculares selecionados (método euclidiano). Finalmente, a interpretação mecanística reforça a confiabilidade do modelo, apesar da maioria dos descritores selecionados serem de origem topológica e geométrica. Assim, o modelo obtido pode ser útil para a concepção de novos inibidores de HIV 1-IN, com potencial para o desenvolvimento de compostos protótipos de novas fármacos antirretrovirais.

Agradecimentos

Fundação Araucária, Coordenação de aperfeiçoamento de pessoal de Nível Superior (CAPES).

Referências

- [1] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction, *Chem Rev*, 110 (2010) 5714-5789.
- [2] M.M.C. Ferreira, C.A. Montanari, A.C. Gaudio, Seleção de variáveis em QSAR, *Quím. Nova*, 25 (2002) 439-448.
- [3] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloeth, J. Gasteiger, A. Richard, A. Tropsha, QSAR Modeling: Where Have You Been? Where Are You Going To?, *J. Med. Chem*, 57 (2014) 4977-5010.
- [4] Y. Goldgur, R. Craigie, G.H. Cohen, T. Fujiwara, T. Yoshinaga, T. Fujishita, H. Sugimoto, T. Endo, H. Murai, D.R. Davies, Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design, *Proc. Natl. Acad. Sci. U.S.A.*, 96 (1999) 13040-13043.
- [5] M. Carcelli, D. Rogolino, M. Sechi, G. Rispoli, E. Fisicaro, C. Compari, N. Grandi, A. Corona, E. Tramontano, C. Pannecouque, L. Naesens, Antiretroviral activity of metal-chelating HIV-1 integrase inhibitors, *Eur. J. Med. Chem*, 83 (2014) 594-600.
- [6] C. Hicks, R.M. Gulick, Raltegravir: The First HIV Type 1 Integrase Inhibitor, *Clin. Infect. Dis*, 48 (2009) 931-939.
- [7] V. Summa, A. Petrocchi, F. Bonelli, B. Crescenzi, M. Donghi, M. Ferrara, F. Fiore, C. Gardelli, O. Gonzalez Paz, D.J. Hazuda, P. Jones, O. Kinzel, R. Laufer, E. Monteagudo, E. Muraglia, E. Nizi, F. Orvieto, P. Pace, G. Pescatore, R. Scarpelli, K. Stillmock, M.V. Witmer, M. Rowley, Discovery of Raltegravir, a Potent, Selective Orally Bioavailable HIV-Integrase Inhibitor for the Treatment of HIV-AIDS Infection, *J. Med. Chem*, 51 (2008) 5843-5855.
- [8] M. Sato, T. Motomura, H. Aramaki, T. Matsuda, M. Yamashita, Y. Ito, H. Kawakami, Y. Matsuzaki, W. Watanabe, K. Yamataka, S. Ikeda, E. Kodama, M. Matsuoka, H.

- Shinkai, Novel HIV-1 Integrase Inhibitors Derived from Quinolone Antibiotics, *J. Med. Chem.*, 49 (2006) 1506-1508.
- [9] D. Dow, J. Bartlett, Dolutegravir, the Second-Generation of Integrase Strand Transfer Inhibitors (INSTIs) for the Treatment of HIV, *Infect. Dis. Ther.*, 3 (2014) 83-102.
- [10] R.V. Patel, Y.S. Keum, S.W. Park, Sketching the historical development of pyrimidones as the inhibitors of the HIV integrase, *Eur. J. Med. Chem.*, (2014).
- [11] G.N. Maertens, S. Hare, P. Cherepanov, The mechanism of retroviral integration from X-ray structures of its key intermediates, *Nature*, 468 (2010) 326-329.
- [12] J.L. Blanco, J. Martinez-Picado, HIV integrase inhibitors in ART-experienced patients, *Curr Opin HIV AIDS*, 7 (2012) 415-421.
- [13] L.C. Tavares, QSAR: a abordagem de Hansch, *Quím. Nova*, 27 (2004) 631-639.
- [14] M. Ferrara, F. Fiore, V. Summa, C. Gardelli, Development of 2-pyrrolidinyl-N-methyl pyrimidones as potent and orally bioavailable HIV integrase inhibitors, *Bioorg. Med. Chem. Lett*, 20 (2010) 5031-5034.
- [15] E. Muraglia, O. Kinzel, C. Gardelli, B. Crescenzi, M. Donghi, M. Ferrara, E. Nizi, F. Orvieto, G. Pescatore, R. Laufer, O. Gonzalez-Paz, A. Di Marco, F. Fiore, E. Monteagudo, M. Fonsi, P.J. Felock, M. Rowley, V. Summa, Design and synthesis of bicyclic pyrimidinones as potent and orally bioavailable HIV-1 integrase inhibitors, *J. Med. Chem.*, 51 (2008) 861-874.
- [16] C. Gardelli, E. Nizi, E. Muraglia, B. Crescenzi, M. Ferrara, F. Orvieto, P. Pace, G. Pescatore, M. Poma, R. Ferreira Mdel, R. Scarpelli, C.F. Homnick, N. Ikemoto, A. Alfieri, M. Verdirame, F. Bonelli, O.G. Paz, M. Taliani, E. Monteagudo, S. Pesci, R. Laufer, P. Felock, K.A. Stillmock, D. Hazuda, M. Rowley, V. Summa, Discovery and synthesis of HIV integrase inhibitors: development of potent and orally bioavailable N-methyl pyrimidones, *J. Med. Chem.*, 50 (2007) 4953-4975.
- [17] M.E. Di Francesco, P. Pace, F. Fiore, F. Naimo, F. Bonelli, M. Rowley, V. Summa, Development of 2-t butyl-N-methyl pyrimidones as potent inhibitors of HIV integrase, *Bioorg. Med. Chem. Lett*, 18 (2008) 2709-2713.
- [18] E. Nizi, M.V. Orsale, B. Crescenzi, G. Pescatore, E. Muraglia, A. Alfieri, C. Gardelli, S.A. Spieser, V. Summa, Dihydroxy-pyrimidine and N-methylpyrimidine HIV-integrase inhibitors: improving cell based activity by the quaternarization of a chiral center, *Bioorg. Med. Chem. Lett*, 19 (2009) 4617-4621.
- [19] A. Petrocchi, P. Jones, M. Rowley, F. Fiore, V. Summa, N-(4-Fluorobenzyl)-3-hydroxy-9,9-dimethyl-4-oxo-6,7,8,9-tetrahydro-4H-pyrazino[1,2-a]pyrimidine-2-carboxamides a novel class of potent HIV-1 integrase inhibitors, *Bioorg. Med. Chem. Lett*, 19 (2009) 4245-4249.
- [20] M. Donghi, O.D. Kinzel, V. Summa, 3-Hydroxy-4-oxo-4H-pyrido[1,2-a]pyrimidine-2-carboxylates--a new class of HIV-1 integrase inhibitors, *Bioorg. Med. Chem. Lett*, 19 (2009) 1930-1934.
- [21] P. Pace, M.E. Di Francesco, C. Gardelli, S. Harper, E. Muraglia, E. Nizi, F. Orvieto, A. Petrocchi, M. Poma, M. Rowley, R. Scarpelli, R. Laufer, O. Gonzalez Paz, E. Monteagudo, F. Bonelli, D. Hazuda, K.A. Stillmock, V. Summa, Dihydroxypyrimidine-4-carboxamides as novel potent and selective HIV integrase inhibitors, *J. Med. Chem.*, 50 (2007) 2225-2239.
- [22] V. Summa, A. Petrocchi, V.G. Matassa, C. Gardelli, E. Muraglia, M. Rowley, O.G. Paz, R. Laufer, E. Monteagudo, P. Pace, 4,5-dihydroxypyrimidine carboxamides and N-alkyl-5-hydroxypyrimidinone carboxamides are potent, selective HIV integrase inhibitors with good pharmacokinetic profiles in preclinical species, *J. Med. Chem.*, 49 (2006) 6646-6649.

- [23] A. Petrocchi, U. Koch, V.G. Matassa, B. Pacini, K.A. Stillmock, V. Summa, From dihydroxypyrimidine carboxylic acids to carboxamide HIV-1 integrase inhibitors: SAR around the amide moiety, *Bioorg. Med. Chem. Lett*, 17 (2007) 350-353.
- [24] J.S. Wai, B. Kim, T.E. Fisher, L. Zhuang, M.W. Embrey, P.D. Williams, D.D. Staas, C. Culberson, T.A. Lyle, J.P. Vacca, D.J. Hazuda, P.J. Felock, W.A. Schleif, L.J. Gabryelski, L. Jin, I.W. Chen, J.D. Ellis, R. Mallai, S.D. Young, Dihydroxypyridopyrazine-1,6-dione HIV-1 integrase inhibitors, *Bioorg. Med. Chem. Lett*, 17 (2007) 5595-5599.
- [25] R.F. Teófilo, J.P.A. Martins, M.M.C. Ferreira, Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression, *J. Chemometr*, 23 (2009) 32-48.
- [26] D.J. Hazuda, P.J. Felock, J.C. Hastings, B. Pramanik, A.L. Wolfe, Differential divalent cation requirements uncouple the assembly and catalytic reactions of human immunodeficiency virus type 1 integrase, *J. Virol*, 71 (1997) 7005-7011.
- [27] Cambridge Structural Database Software Version 5.29, Cambridge Crystallographic Data Centre, England, 2007.
- [28] J.B. Foresman, A. Frisch, Basis Set Effects, *Exploring chemistry with electronic structure methods.*, Gaussian Inc, Pittsburg, 2002, pp. 97-110.
- [29] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, second ed., Wiley-VCH, Weinheim, Germany, 2009.
- [30] J.P.A. Martins, M.M.C. Ferreira, QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR, *Quím. Nova*, 36 (2013) 554-560.
- [31] C.K. Hattotuwegama, D.R. Flower, Empirical prediction of peptide octanol-water partition coefficients, *Bioinformation*, 1 (2006) 257-259.
- [32] J.P.A. Martins, *Apostila de química computacional aplicada a QSAR*, Centro Nacional de Processamento de Alto Desempenho em São Paulo, São Paulo, 2010.
- [33] P. Liu, W. Long, Current Mathematical Methods Used in QSAR/QSPR Studies, *Int. J. Mol. Sci*, 10 (2009) 1978-1998.
- [34] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst*, 58 (2001) 109-130.
- [35] L. Eriksson, P. Andersson, E. Johansson, M. Tysklind, Megavariate analysis of environmental QSAR data. Part I – A basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD), *Mol. Divers*, 10 (2006) 169-186.
- [36] R. Kiralj, M.M.C. Ferreira, Basic validation procedures for regression models in QSAR and QSPR studies: theory and application, *J. Braz. Chem. Soc*, 20 (2009) 770-787.
- [37] A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, *Mol. Inf*, 29 (2010) 476-488.
- [38] K. Roy, P. Chakraborty, I. Mitra, P.K. Ojha, S. Kar, R.N. Das, Some case studies on application of “rm2” metrics for judging quality of quantitative structure–activity relationship predictions: Emphasis on scaling of response data, *J. Comput. Chem*, 34 (2013) 1071-1082.
- [39] K. Roy, I. Mitra, On the use of the metric rm(2) as an effective tool for validation of QSAR models in computational drug design and predictive toxicology, *Mini. Rev. Med. Chem*, 12 (2012) 491-504.
- [40] L. Eriksson, J. Jaworska, A.P. Worth, M.T. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs, *Environ. Health. Perspect*, 111 (2003) 1361-1375.

- [41] A. Golbraikh, M. Shen, Z. Xiao, Y.-D. Xiao, K.-H. Lee, A. Tropsha, Rational selection of training and test sets for the development of validated QSAR models, *J. Comput. Aided Mol. Des.*, 17 (2003) 241-253.
- [42] A. Golbraikh, A. Tropsha, Beware of q^2 !, *J. Mol. Graph. Model.*, 20 (2002) 269-276.
- [43] P.P. Roy, K. Roy, On Some Aspects of Variable Selection for Partial Least Squares Regression Models, *QSAR. Comb. Sci.*, 27 (2008) 302-313.
- [44] P. Pratim Roy, S. Paul, I. Mitra, K. Roy, On Two Novel Parameters for Validation of Predictive QSAR Models, *Molecules*, 14 (2009) 1660-1701.
- [45] E.B. de Melo, Modeling physical and toxicity endpoints of alkyl (1-phenylsulfonyl) cycloalkane-carboxylates using the Ordered Predictors Selection (OPS) for variable selection and descriptors derived with SMILES, *Chemom. Intell. Lab. Syst.*, 118 (2012) 79-87.
- [46] A. Tropsha, P. Gramatica, V.K. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models, *QSAR. Comb. Sci.*, 22 (2003) 69-77.
- [47] S. Wold, L. Eriksson, S. Clementi, Statistical Validation of QSAR Results, *Chemometric Methods in Molecular Design*, Wiley-VCH Verlag GmbH 2008, pp. 309-338.
- [48] HyperChem Software Version. 7, Hypercube Inc, USA, 2002. <http://www.hyper.com/>
- [49] Gaussian software Version 09, Gaussian Inc, Wallingford, USA, 2009. <http://www.gaussian.com/gprod/g09.htm>
- [50] GaussView software Version 5.0, Semichem Inc, Shawnee Mission, USA, 2009. <http://www.gaussian.com/gprod/gv5.htm>
- [51] Dragon Software Version 6.0, TALETE srl, Milano, Italy, 1997. <http://www.talete.mi.it/products/dragondescription.htm>
- [52] The Open Babel Package Version 2.3.1, 2011. <http://openbabel.org>
- [53] Pirouette Software, Infometrix Inc, Woodinville, USA, 2011. <http://infometrix.com/pirouette/>
- [54] R.P. Verma, C. Hansch, An approach toward the problem of outliers in QSAR, *Bioorg. Med. Chem.*, 13 (2005) 4597-4621.
- [55] A.C. Gaudio, E. Zandonade, Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica, *Quím.Nova*, 24 (2001) 658-671.
- [56] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, H. Kabir, Comparative studies on some metrics for external validation of QSPR models, *J. Chem. Inf. Model.*, 52 (2012) 396-408.
- [57] E. Besalú, L. Vera, Internal test set (ITS) method: a new cross-validation technique to assess the predictive capability of QSAR models. application to a benchmark set of steroids, *J. Chil. Chem. Soc.*, 53 (2008) 1576-1580.
- [58] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs, *Environ. Health. Perspect.*, 111 (2003) 1361-1375.
- [59] Organization for Economic Co-Operation and Development (OECD), Guidance Document on the Validation of (Quantitative) Structure–Activity Relationship [(QSAR) Models], OECD, Paris, 2007.
- [60] M. Metifiot, N. Vandegraaff, K. Maddali, A. Naumova, X. Zhang, D. Rhodes, C. Marchand, Y. Pommier, Elvitegravir overcomes resistance to raltegravir induced by integrase mutation Y143, *AIDS*, 25 (2011) 1175-1178.

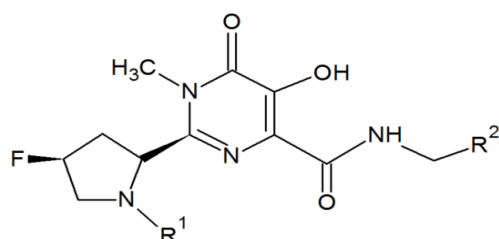
- [61] S. Hare, A.M. Vos, R.F. Clayton, J.W. Thuring, M.D. Cummings, P. Cherepanov, Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance, *Proc. Natl. Acad. Sci. USA*, 107 (2010) 20057-20062.
- [62] K. du Toit, E.E. Elgorashi, S.F. Malan, S.E. Drewes, J. van Staden, N.R. Crouch, D.A. Mulholland, Anti-inflammatory activity and QSAR studies of compounds isolated from Hyacinthaceae species and *Tachiadenus longiflorus* Griseb. (Gentianaceae), *Bioorg. Med. Chem.*, 13 (2005) 2561-2568.
- [63] B.B. Lohray, N. Gandhi, B.K. Srivastava, V.B. Lohray, 3D QSAR studies of N-4-arylacryloylpiperazin-1-yl-phenyl-oxazolidinones: A novel class of antibacterial agents, *Bioorg. Med. Chem. Lett.*, 16 (2006) 3817-3823.
- [64] O.A. Phillips, E.E. Udo, S.M. Samuel, Synthesis and structure-antibacterial activity of triazolyl oxazolidinones containing long chain acyl moiety, *Eur. J. Med. Chem.*, 43 (2008) 1095-1104.
- [65] E.B. de Melo, M.M.C. Ferreira, Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors, *Eur. J. Med. Chem.*, 44 (2009) 3577-3583.
- [66] Talete. Dragon molecular descriptors. In: List of molecular descriptors calculated by Dragon. http://www.talete.mi.it/products/dragon_molecular_descriptor_list.pdf, accessed in September 01 2015.
- [67] A. Arroio, K.M. Honório, A.B.F.d. Silva, Propriedades químico-quânticas empregadas em estudos das relações estrutura-atividade, *Quím. Nova*, 33 (2010) 694-699.

8.3 Anexo 3

Material suplementar Artigo

Tabela 1

Compostos descritos por Ferrara *et al.* (2010), pertencentes à classe das 2-pirrolidinil-*N*-metil pirimidonas carboxamidas.

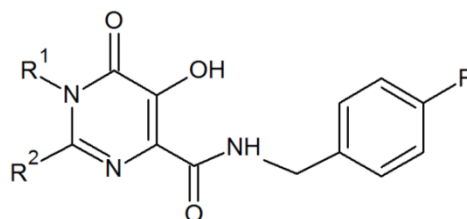


Composto	R ¹	R ²	pIC ₅₀	Composto	R ¹	R ²	pIC ₅₀
1	c-Pr		7.699	8	CO-pirazina		8.000
2	CH ₂ (3-isoxazol)		8.000	9	COCONMe ₂		7.678
3	COMe		7.824	10 CT	COMe		7.620
4	CO ₂ Me		8.000	11	COMe		7.569
5	CONMe ₂		8.000	12 CT	COMe		7.494
6	SO ₂ Me		7.824	13	CO ₂ Me		7.698
7	SO ₂ Me		8.046	14 CT	CO ₂ Me		7.721

CT: composto pertencente ao conjunto de treinamento.

Tabela 2

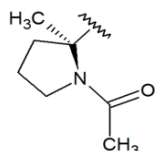
Compostos descritos por Gardelli *et al.* (2007), pertencentes à classe das *N*-metil pirimidonas carboxamidas e compostos descritos por Nizi *et al.* (2009), pertencentes à classe das dihidroxipirimidinas carboxamidas e *N*-metil pirimidonas carboxamidas.



Composto	R ¹	R ²	pIC ₅₀	Composto	R ¹	R ²	pIC ₅₀
15 ^a	Me		7.208	24 ^a	Me		6.887
16 ^a	Me		8.000	25 ^a	Me		7.523
17 ^a CT	Me		6.721	26 ^a CT	Me		7.215
18 ^a	Me		7.000	27 ^a	Me		7.699
19 ^a CT	Me		6.745	28 ^a	Me		7.699
20 ^a	Me		6.699	29 ^a	Me		7.523
*21S ^b CT	H		7.097	*30R ^b	Me		7.699
*22S ^b	H		7.046	*31S ^b	Me		7.699

***23S^b**

H

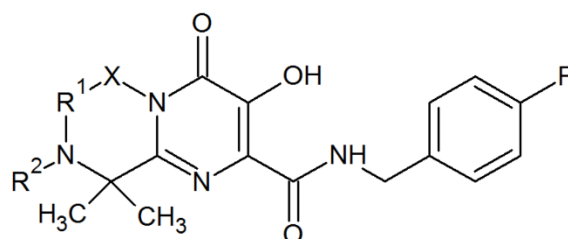


7.699

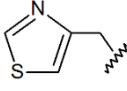
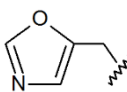
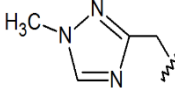
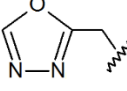
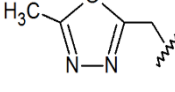
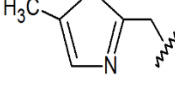
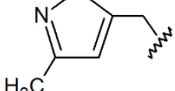
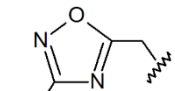
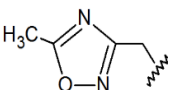
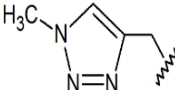
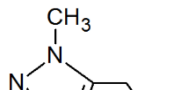
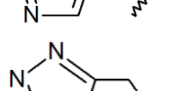
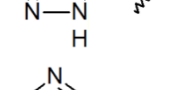
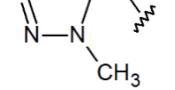
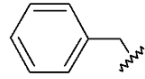
^a Gardelli *et al.* (2007) e ^b Nizi *et al.* (2009).

* R e S enantiômeros.

CT: composto pertencente ao conjunto de treinamento.

Tabela 3Compostos descritos por Petrochi *et al.* (2009), pertencentes à classe das pirazino pirimidonas carboxamidas.

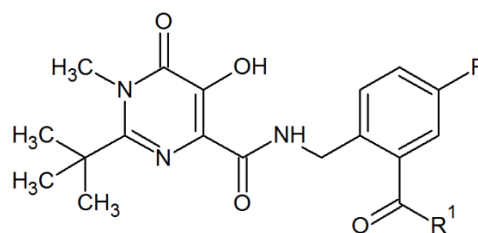
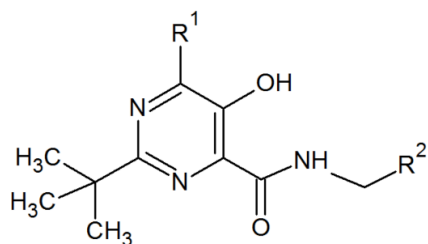
Composto	R ¹	R ²	X	pIC ₅₀	Composto	R ¹	R ²	X	pIC ₅₀
32	CH ₂	Me	0	7.523	56	CH ₂		0	8.046
33	CO	H	0	7.494	57	CH ₂		0	7.921
34	CO	H	1	8.301	58	CH ₂		0	8.222
35 CT	SO ₂	H	1	8.155	59	CH ₂		0	8.301
36 CT	CH ₂	H	1	7.796	60	CH ₂		0	8.155
37	CH ₂	Me	1	8.222	61	CH ₂		0	8.046
38	CH ₂	H	2	7.699	62	CH ₂		0	8.222
39	CH ₂	COMe	0	8.301	63	CH ₂		0	8.155
40 CT	CH ₂	COCON(Me) ₂	0	8.222	64	CH ₂		0	8.523

41	CH ₂	SO ₂ Me	0	8.000	65 CT	CH ₂		0	8.046
42	CH ₂	CONHCH ₂ CH ₃	0	7.824	66 CT	CH ₂		0	8.398
43	CH ₂	SO ₂ N(Me) ₂	0	8.097	67 CT	CH ₂		0	7.602
44	CH ₂	CH ₂ N(Me) ₂	0	7.208	68	CH ₂		0	7.678
45 CT	CH ₂	CH ₂ N(CHCH ₃) ₂	0	7.119	69	CH ₂		0	8.301
46 CT	CH ₂	CH ₂ -morfolina	0	8.523	70	CH ₂		0	8.097
47	CH ₂	CONH ₂	0	8.301	71 CT	CH ₂		0	7.208
48	CH ₂	CONHMe	0	7.678	72 CT	CH ₂		0	7.886
49	CH ₂	CON(Me) ₂	0	8.097	73	CH ₂		0	8.222
50	CH ₂	CO-morfolina	0	8.155	74	CH ₂		0	8.699
51	CH ₂	CO-pirrolidina	0	8.155	75	CH ₂		0	8.222
52	CH ₂	CO-(4-CH ₃ -piperazina)	0	7.678	76 CT	CH ₂		0	8.699
53	CH ₂	COOH	0	8.097	77	CH ₂		0	8.301
54	CH ₂	CN	0	8.301	78	CH ₂		0	8.222
55	CH ₂		0	8.301					

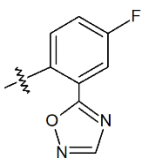
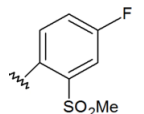
CT: composto pertencente ao conjunto de treinamento.

Tabela 4

Compostos descritos por Di Francesco *et al.* (2008), pertencentes à classe das *N*-metil pirimidonas carboxamidas.



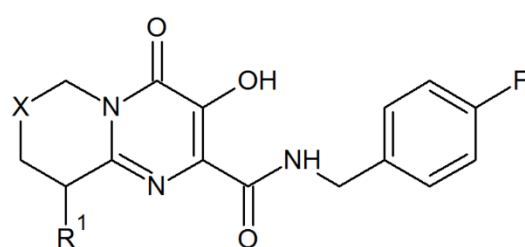
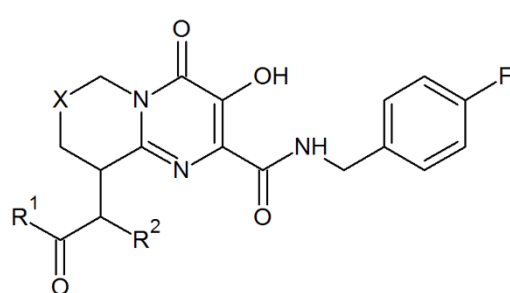
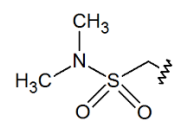
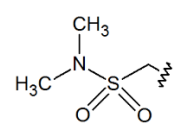
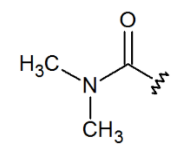
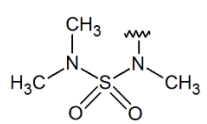
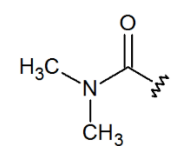
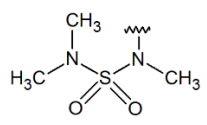
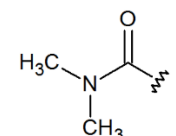
Composto	R ¹	R ²	pIC ₅₀	Composto	R ¹	pIC ₅₀
79	OH		8.000	91	NH ₂	7.699
80	OH		8.000	92	NHCH ₂ Me	7.155
81	OH		8.000	93	NH(CH ₂) ₂ Me	7.155
82	OH		7.699	94	NHCH(Me) ₂	7.046
83	OH		7.699	95 CT	NHCH ₂ CF ₃	7.046
84 CT	OH		7.222	96	N(Me) ₂	6.745
85	O		8.000	97 CT		6.638
86 CT	O		7.698	98 CT		6.310
87	O		7.699			
88	O		7.699			

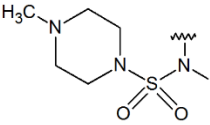
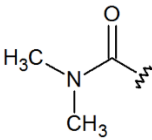
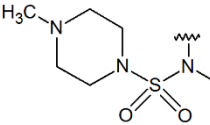
89	O		7.699
90	O		7.699

CT: composto pertencente ao conjunto de treinamento.

Tabela 5

Compostos descritos por Muraglia *et al.* (2008), pertencentes à classe das pirimidonas bicíclicas carboxamidas.

								
Composto	R ¹	X	pIC ₅₀	Composto	R ¹	R ²	X	pIC ₅₀
99	H	0	8.097	*106R		Me	1	8.154
100	H	1	8.301	*107S		Me	1	8.154
101	H	2	7.721	*108R		Me	1	7.721
*102R		1	8.155	*109S		Me	1	7.921
*103S CT		1	8.301	*110R CT		Me	2	7.456

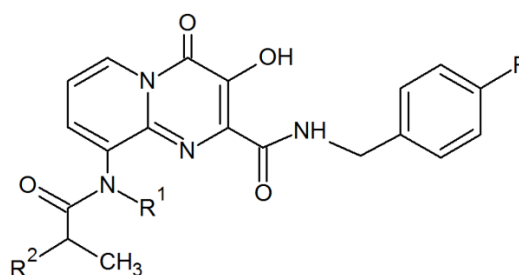
*104R		1	8.000	*111S CT		Me	2	7.921
*105S		1	8.155					

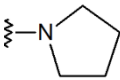
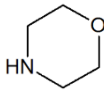
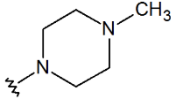
* R e S enantiômeros.

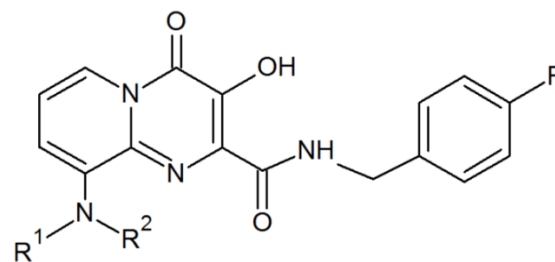
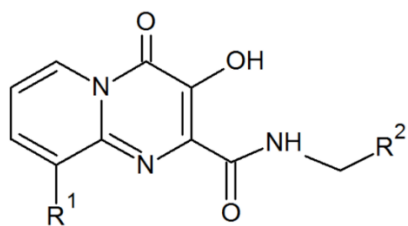
CT: composto pertencente ao conjunto de treinamento.

Tabela 6

Compostos descritos por Donghi *et al.* (2009), pertencentes à classe das pirimidinas carboxamidas.



Composto	R ¹	R ²	pIC ₅₀	Composto	R ¹	R ²	pIC ₅₀
112	Me	NMe ₂	7.585	116	Et	NEt ₂	7.721
113	Et	NMe ₂	7.444	117	Et		7.602
114 CT	<i>n</i> Pr	NMe ₂	7.745	118	Et		7.678
115	<i>i</i> Pr	NMe ₂	7.678	119	Et		7.678

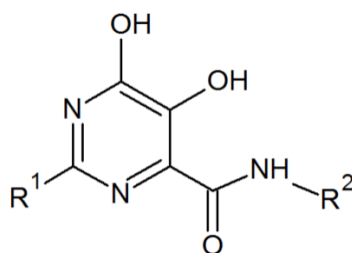


Composto	R ¹	R ²	pIC ₅₀	Composto	R ¹	R ²	pIC ₅₀
120	H		7.656	129		Et	6.733
121 CT	NHMe		6.886	130 CT		Et	7.638
122	NHEt		7.444	131		Et	7.301
123	NMe ₂		7.155	132		Et	7.678
124	NHCOPh		7.745	133 CT		Et	7.678
125	Me		8.046	134		Et	7.745
126			6.770				
127 CT			6.733				
128	CH ₂ NHCOPh		7.620				

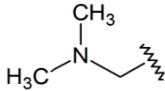
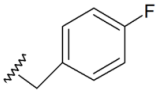
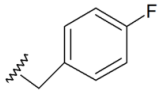
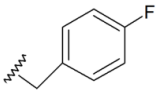
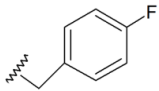
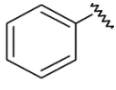
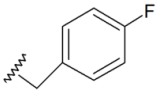
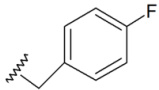
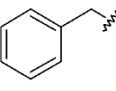
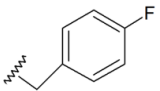
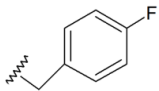
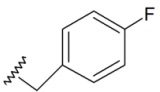
CT: composto pertencente ao conjunto de treinamento.

Tabela 7

Compostos descritos por Pace *et al.* (2007), e compostos descritos por Summa *et al.* (2006), pertencentes à classe das dihidroxi pirimidinas carboxamidas.



Composto	R ¹	R ²	pIC ₅₀	Composto	R ¹	R ²	pIC ₅₀
135^c CT		CH(CH ₃)Ph	6.301	149^c CT			5.699
136^c	<i>i</i> Pr		7.000	150^c			7.301
137^c	<i>p</i> -tolil		7.699	151^c			7.222
138^c	2-tiazolil		7.699	152^c			7.398
139^c	2-piridil		7.301	153^c			8.000
140^c CT			6.824	154^c			7.284
141^c			6.824	155^c CT			7.056
142^c			6.824	156^c			7.097
143^c CT			5.910	157^c CT			7.585

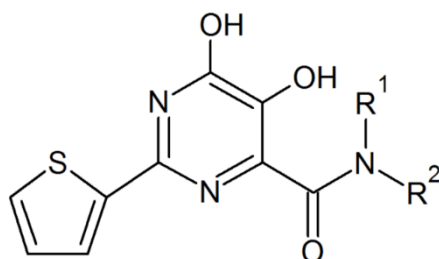
144^c CT			6.699	158^d	H		7.222
145^d	Me		7.222	159^d	CH(Me)NMe ₂		7.097
146^d			7.155	160^d	CH ₂ NMe ₂		6.699
147^d CT			7.301	161^d	C(Me) ₂ NMe ₂		7.301
148^d	CH(Ph)NMe ₂		6.699				

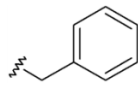
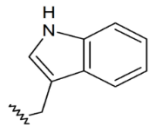
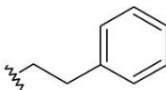
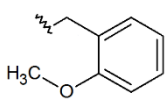
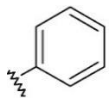
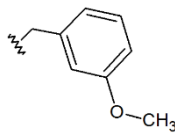
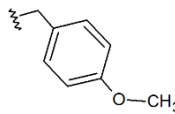
^c Pace *et al.* (2007) e ^d Summa *et al.* (2006).

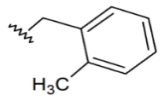
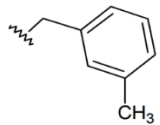
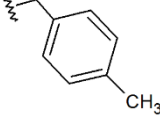
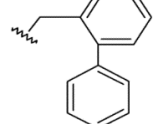
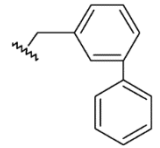
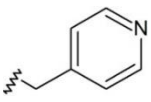
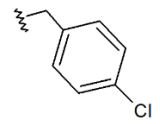
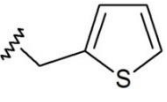
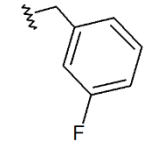
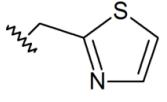
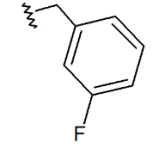
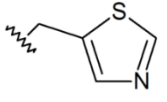
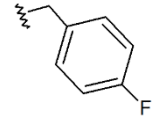
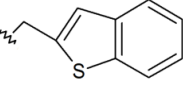
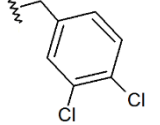
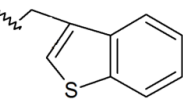
CT: composto pertencente ao conjunto de treinamento.

Tabela 8

Compostos descritos por Petrochi *et al.* (2007), pertencentes à classe das dihidroxipirimidinas carboxamidas.



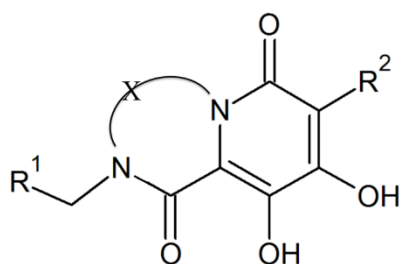
Compost o	R ¹	R ²	pIC ₅₀	Composto	R ¹	R ²	pIC ₅₀
162		H	7.071	177	H		7.699
163		H	7.770	178	H		7.046
164 CT		H	6.000	179	H		7.301
165 CT	1-Naftaleno	H	7.523	180 CT	H		6.260

166	(S)-CH(CH ₃)-2-naftil	H	7.398	181 CT	H		6.959
167 CT	(R)-CH(CH ₃)-2-naftil	H	6.215	182	H		7.398
168	1-Naftaleno	CH ₃	6.678	183	H		7.155
169	(1S)-1-Indane	H	7.699	184	H		6.886
170	(1R)-1-Indane	H	6.215	185	H		7.301
171 CT	H		5.780	186	H		7.699
172 CT	H		7.000	187 CT	H		6.796
173	H		6.699	188 CT	H		7.155
174 CT	H		5.602	189	H		8.000
175	H		8.000	190 CT	H		8.000
176	H		7.301				

CT: composto pertencente ao conjunto de treinamento.

Tabela 9

Compostos descritos por Wai *et al.* (2007), pertencentes à classe das dihidroxi-*piridopirazinas* carboxamidas.



Composto	R ¹	R ²	X	pIC ₅₀	Composto	R ¹	R ²	X	pIC ₅₀
191 CT		H	(CH ₂) ₃	5.983	196 CT		H	(CH ₂) ₂	7.398
192		H	(CH ₂) ₂	7.000	197 CT		H	(CH ₂) ₂	6.420
193		H	CH=CH	7.398	198		H	(CH ₂) ₂	7.398
194		H	(CH ₂) ₂	6.638	199		CN	(CH ₂) ₂	7.699
195 CT		H	(CH ₂) ₂	6.432					

CT: composto pertencente ao conjunto de treinamento.

Tabela 10

Detalhamento dos valores dos descritores presentes no modelo (equação 1) para os compostos do conjunto de predição ($N=145$), valores preditos pelo modelo e observados para a variável independente.

Composto	VE1_B(p)	R5m+	Eig09_EA(bo)	H2m	Mor28s	(E_T)	SpPosA_X	y real	y pred	Resíduo
1	3,810	0,027	1,557	1,727	0,380	-1420,128	0,294	7,699	7,347	0,352
2	3,851	0,024	1,900	2,141	1,665	-1587,626	0,294	8,000	7,586	0,414
3	3,899	0,029	1,500	1,814	2,143	-1456,117	0,291	7,824	7,411	0,413
4	3,937	0,028	1,500	1,912	1,717	-1531,368	0,300	8,000	7,549	0,451
5	3,978	0,026	1,670	1,893	1,614	-1550,804	0,293	8,000	7,670	0,330
6	4,018	0,030	1,552	2,057	1,207	-1891,381	0,283	7,824	8,589	-0,765
7	4,162	0,025	1,673	1,889	2,784	-1986,052	0,287	8,046	8,377	-0,331
8	4,208	0,025	1,991	2,127	1,829	-1679,976	0,298	8,000	7,942	0,058
9	4,101	0,026	1,991	2,020	-0,402	-1664,176	0,294	7,678	8,055	-0,377
11	3,946	0,027	1,948	1,762	1,646	-1495,446	0,292	7,569	7,052	0,517
13	3,962	0,028	1,948	1,872	0,626	-1931,037	0,300	7,698	7,581	0,117

15	3,778	0,029	1,158	1,594	-0,417	-1243,455	0,297	7,208	7,414	-0,206
16	3,785	0,026	1,500	1,598	0,561	-1282,779	0,292	8,000	7,135	0,865
18	3,787	0,026	1,500	1,704	0,290	-1318,695	0,292	7,000	7,351	-0,351
20	3,792	0,024	1,500	1,623	-1,374	-1397,336	0,297	6,699	7,558	-0,859
22	3,857	0,031	1,401	1,604	1,519	-1243,473	0,294	7,046	6,942	0,104
23	4,007	0,031	1,500	1,653	-1,213	-1356,865	0,291	7,699	7,754	-0,055
24	3,799	0,022	1,993	1,782	-0,085	-1589,115	0,299	6,887	7,182	-0,295
25	3,809	0,027	1,798	2,094	2,075	-1886,798	0,280	7,523	8,056	-0,533
27	3,788	0,027	1,500	1,699	1,915	-1342,726	0,292	7,699	7,067	0,632
28	3,788	0,027	1,500	1,774	1,325	-1342,723	0,292	7,699	7,258	0,441
29	3,808	0,029	1,500	1,961	0,818	-1442,004	0,284	7,523	7,832	-0,309
30	4,245	0,026	1,717	2,147	1,152	-1926,089	0,284	7,699	9,022	-1,323
31	4,082	0,038	1,687	2,105	1,233	-1886,776	0,286	7,699	8,209	-0,510
32	3,970	0,031	1,500	1,568	-1,967	-1243,449	0,290	7,523	7,626	-0,103
33	3,787	0,034	1,500	1,802	0,344	-1238,893	0,283	7,494	7,316	0,178

34	4,022	0,039	1,500	1,757	0,193	-1278,214	0,283	8,301	7,504	0,797
37	3,970	0,028	1,500	1,528	0,309	-1243,462	0,290	8,222	7,313	0,909
38	3,837	0,028	1,500	1,578	-0,086	-1243,463	0,290	7,699	7,234	0,465
39	4,129	0,028	1,500	1,601	-2,254	-1356,835	0,285	8,301	8,309	-0,008
41	4,269	0,029	1,500	1,942	-1,745	-1792,098	0,276	8,000	9,492	-1,492
42	4,175	0,031	1,500	1,720	-1,329	-1451,533	0,290	7,824	8,237	-0,413
43	4,417	0,028	1,593	1,977	-2,081	-1886,776	0,281	8,097	9,716	-1,619
44	4,021	0,026	1,645	1,583	-1,758	-1416,768	0,286	7,208	7,973	-0,765
47	4,069	0,029	1,599	1,756	-1,134	-1412,207	0,286	8,301	8,102	0,199
48	4,078	0,028	1,609	1,729	-1,958	-1451,525	0,294	7,678	8,081	-0,403
49	4,107	0,028	1,647	1,750	-2,372	-1490,832	0,288	8,097	8,364	-0,267
50	4,123	0,029	1,912	2,004	-2,981	-1643,501	0,294	8,155	8,456	-0,301
51	4,107	0,030	1,899	1,952	-1,066	-1568,272	0,291	8,155	8,030	0,125
52	4,128	0,029	2,061	1,941	-2,230	-1662,950	0,290	7,678	8,203	-0,525
53	4,074	0,028	1,599	1,712	-0,776	-1432,078	0,286	8,097	8,052	0,045

54	4,045	0,030	1,502	1,641	-1,844	-1335,706	0,294	8,301	7,840	0,461
55	4,112	0,028	2,007	1,726	-1,733	-1474,556	0,293	8,301	7,665	0,636
56	4,143	0,029	2,007	1,795	-1,785	-1490,596	0,293	8,046	7,788	0,258
57	4,063	0,034	2,051	1,687	-0,747	-1529,916	0,293	7,921	7,161	0,760
58	4,047	0,026	1,655	1,744	-1,470	-1510,245	0,293	8,222	8,074	0,148
59	4,069	0,028	1,751	1,793	-1,638	-1468,511	0,291	8,301	8,017	0,284
60	4,109	0,029	1,781	1,846	-2,059	-1507,833	0,290	8,155	8,206	-0,051
61	4,081	0,027	2,046	1,769	-1,705	-1507,836	0,289	8,046	7,790	0,256
62	4,070	0,028	1,751	1,949	-2,549	-1484,540	0,291	8,222	8,379	-0,157
63	4,058	0,028	1,751	1,785	-2,266	-1468,510	0,291	8,155	8,093	0,062
64	4,069	0,028	1,751	1,866	-1,378	-1488,389	0,291	8,523	8,086	0,437
68	4,090	0,029	1,751	2,002	-1,600	-1504,419	0,291	7,678	8,304	-0,626
69	4,097	0,030	2,046	1,971	-1,048	-1543,756	0,289	8,301	7,885	0,416
70	4,096	0,028	2,046	1,911	-1,496	-1527,724	0,289	8,097	7,940	0,157
73	4,088	0,028	2,046	1,979	-2,062	-1543,746	0,289	8,222	8,121	0,101

74	4,075	0,030	2,046	1,845	-2,130	-1523,864	0,289	8,699	7,851	0,848
75	4,109	0,025	1,781	1,920	-1,576	-1523,859	0,290	8,222	8,382	-0,160
77	4,126	0,029	1,781	2,028	-2,543	-1539,897	0,290	8,301	8,571	-0,270
78	4,092	0,029	2,046	1,867	-2,313	-1539,904	0,289	8,222	7,988	0,234
79	3,631	0,029	0,707	1,438	-1,530	-1110,677	0,283	8,000	7,930	0,070
80	3,662	0,026	1,131	1,287	-2,236	-1165,295	0,289	8,000	7,413	0,587
81	3,741	0,025	1,500	1,337	-3,002	-1143,018	0,284	8,000	7,422	0,578
82	3,765	0,043	1,571	1,551	-1,623	-1318,751	0,288	7,699	6,830	0,869
83	3,778	0,037	1,861	1,643	-1,352	-1410,973	0,282	7,699	7,037	0,662
85	3,773	0,029	0,726	1,463	0,186	-1149,986	0,283	8,000	7,916	0,084
87	3,907	0,025	1,500	1,349	-0,913	-1182,327	0,284	7,699	7,391	0,308
88	3,940	0,042	1,572	1,530	0,007	-1358,061	0,288	7,699	6,882	0,817
89	3,957	0,042	1,933	1,821	-0,771	-1450,289	0,282	7,699	7,211	0,488
90	4,047	0,048	1,715	1,827	0,274	-1737,940	0,270	7,699	7,814	-0,115
91	3,935	0,052	1,476	1,591	0,017	-1318,743	0,279	7,699	6,882	0,817

92	3,942	0,035	1,626	1,556	0,131	-1397,387	0,285	7,155	7,204	-0,049
93	3,942	0,035	1,677	1,505	-0,310	-1436,712	0,289	7,155	7,091	0,064
94	3,951	0,034	1,929	1,555	-0,399	-1436,714	0,278	7,046	7,193	-0,147
96	3,983	0,041	1,811	1,494	-0,135	-1397,370	0,281	6,745	6,892	-0,147
99	3,581	0,035	0,843	1,466	0,826	-1070,135	0,298	8,097	6,702	1,395
100	3,709	0,034	1,045	1,469	1,394	-1109,460	0,301	8,301	6,571	1,730
101	3,663	0,033	1,226	1,451	2,889	-1148,774	0,299	7,721	6,142	1,579
102	4,300	0,023	1,500	1,876	0,792	-1886,788	0,291	8,155	8,982	-0,827
104	4,361	0,023	1,960	1,837	1,837	-2058,901	0,293	8,000	8,429	-0,429
105	4,361	0,022	1,960	2,095	1,675	-2058,900	0,293	8,155	8,815	-0,660
106	4,166	0,026	1,780	2,000	-0,776	-2039,484	0,289	8,154	8,945	-0,791
107	4,166	0,023	1,780	1,874	-1,769	-2039,484	0,289	8,154	9,060	-0,906
108	4,236	0,024	1,697	1,832	1,016	-1564,900	0,298	7,721	8,022	-0,301
109	4,236	0,023	1,697	1,802	0,549	-1564,902	0,298	7,921	8,098	-0,177
112	4,341	0,025	1,977	1,925	0,446	-1562,480	0,298	7,585	8,021	-0,436

113	4,362	0,023	1,980	1,891	1,396	-1601,806	0,299	7,444	7,939	-0,495
115	4,430	0,023	2,000	1,924	-1,302	-1641,130	0,292	7,678	8,719	-1,041
116	4,374	0,022	2,048	1,814	0,483	-1680,459	0,305	7,721	7,896	-0,175
117	4,396	0,022	2,178	1,996	0,771	-1773,921	0,300	7,602	8,173	-0,571
118	4,361	0,022	2,106	1,982	0,809	-1679,247	0,302	7,678	8,037	-0,359
119	4,396	0,022	2,178	1,989	1,364	-1773,921	0,300	7,678	8,067	-0,389
120	3,782	0,036	1,324	1,591	1,051	-1107,037	0,301	7,656	6,483	1,173
122	3,925	0,026	1,500	1,600	0,507	-1241,061	0,301	7,444	7,097	0,347
123	3,981	0,029	1,500	1,682	1,380	-1241,031	0,295	7,155	7,183	-0,028
124	4,183	0,028	2,108	2,129	1,284	-1506,902	0,300	7,745	7,521	0,224
125	3,858	0,035	1,500	1,655	0,871	-1146,366	0,297	8,046	6,676	1,370
126	3,981	0,026	1,670	1,763	2,013	-1433,031	0,299	6,770	7,179	-0,409
128	4,037	0,030	2,148	1,970	0,628	-1546,220	0,302	7,620	7,072	0,548
129	4,189	0,025	1,944	1,822	1,514	-1518,603	0,294	6,733	7,577	-0,844
131	4,384	0,024	2,160	2,017	1,245	-1601,563	0,303	7,301	7,804	-0,503

132	4,417	0,025	2,160	2,030	0,848	-1601,568	0,303	7,678	7,899	-0,221
134	4,303	0,026	2,085	2,148	1,117	-1599,320	0,301	7,745	7,927	-0,182
136	3,545	0,029	0,703	1,424	-0,601	-1071,354	0,294	7,000	7,317	-0,317
137	3,888	0,028	1,500	1,632	0,477	-1223,818	0,296	7,699	7,119	0,580
138	3,637	0,045	1,293	2,015	0,742	-1521,298	0,298	7,699	7,035	0,664
139	3,865	0,037	1,500	1,745	-0,194	-1200,529	0,301	7,301	6,855	0,446
141	3,879	0,029	1,500	1,673	0,238	-1279,175	0,302	6,824	7,064	-0,240
142	3,928	0,026	1,863	1,689	0,340	-1357,814	0,292	6,824	7,148	-0,324
145	3,429	0,037	0,317	1,389	0,153	-992,706	0,295	7,222	7,033	0,189
146	3,796	0,027	1,500	1,322	-0,493	-1085,222	0,307	7,155	6,371	0,784
148	3,947	0,026	1,842	1,725	1,445	-1357,802	0,295	6,699	6,991	-0,292
150	3,759	0,027	1,212	1,473	-0,043	-1205,339	0,287	7,301	7,393	-0,092
151	3,777	0,026	1,426	1,484	-0,957	-1244,663	0,297	7,222	7,155	0,067
152	3,682	0,028	0,710	1,463	-0,198	-1166,030	0,296	7,398	7,585	-0,187
153	3,683	0,027	1,500	1,489	-0,375	-1204,102	0,292	8,000	6,879	1,121

154	3,847	0,027	1,500	1,494	0,153	-1243,421	0,284	7,284	7,292	-0,008
156	3,861	0,026	1,500	1,590	0,121	-1322,103	0,296	7,097	7,254	-0,157
158	3,394	0,037	0,098	1,325	0,265	-953,373	0,301	7,222	6,953	0,269
159	3,649	0,031	0,818	1,485	0,558	-1166,022	0,295	7,097	7,223	-0,126
160	3,565	0,030	0,720	1,555	0,138	-1126,696	0,290	6,699	7,489	-0,790
161	3,759	0,028	1,212	1,489	-0,828	-1205,339	0,287	7,301	7,505	-0,204
162	3,589	0,043	1,293	1,563	0,261	-1405,986	0,304	7,071	6,282	0,789
163	3,541	0,047	1,295	1,618	0,298	-1445,312	0,302	7,770	6,211	1,559
166	3,960	0,046	1,899	1,861	-1,045	-1598,989	0,297	7,398	6,978	0,420
168	4,540	0,052	1,879	1,766	-0,593	-1559,637	0,302	6,678	7,322	-0,644
169	3,646	0,042	1,408	1,726	-0,082	-1483,428	0,298	7,699	6,758	0,941
170	3,646	0,042	1,408	1,726	-0,082	-1483,428	0,298	6,215	6,758	-0,543
173	3,544	0,044	0,927	1,806	-0,034	-1742,799	0,301	6,699	7,378	-0,679
175	3,545	0,041	1,840	1,914	0,819	-1880,433	0,298	8,000	6,606	1,394
176	3,625	0,047	1,694	2,121	0,494	-1880,432	0,299	7,301	6,971	0,330

177	3,627	0,044	1,694	1,767	-0,200	-1537,590	0,299	7,699	6,413	1,286
178	3,645	0,039	1,293	1,599	-0,327	-1520,544	0,306	7,046	6,722	0,324
179	3,622	0,038	1,293	1,608	-0,144	-1520,543	0,305	7,301	6,730	0,571
182	3,610	0,038	1,293	1,586	0,314	-1445,314	0,297	7,398	6,733	0,665
183	3,606	0,043	1,293	1,608	-0,077	-1445,314	0,298	7,155	6,610	0,545
184	3,711	0,031	1,834	1,872	0,115	-1637,093	0,304	6,886	6,908	-0,022
185	3,661	0,031	1,809	1,805	0,036	-1637,098	0,303	7,301	6,814	0,487
186	3,605	0,066	1,293	2,281	0,000	-1865,608	0,298	7,699	7,012	0,687
189	3,615	0,045	1,293	1,931	0,727	-1505,254	0,298	8,000	6,882	1,118
192	3,867	0,042	0,902	1,423	1,299	-1090,004	0,292	7,000	6,856	0,144
193	3,907	0,042	0,926	1,583	1,489	-1088,784	0,292	7,398	7,057	0,341
194	3,812	0,043	0,883	1,125	1,478	-990,736	0,297	6,638	6,130	0,508
198	3,941	0,043	0,904	2,066	0,853	-1549,622	0,292	7,398	8,268	-0,870
199	3,977	0,038	1,367	1,516	1,934	-1182,266	0,298	7,699	6,572	1,127

