UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÁRMACO E MEDICAMENTOS
ÁREA INSUMOS FARMACÊUTICOS

Aplicação de modelagem molecular e de formalismos do CAMD (*Computer-Aided Molecular Design*) na elucidação do mecanismo de ação de inibidores de metaloproteinases de matriz

Kely Medeiros Turra

Tese para obtenção do grau de DOUTOR

Orientador:
Profa. Dra. Silvia Berlanga de Moraes Barros
Co-orientador:
Profa. Dra. Kerly Fernanda Mesquita Pasqualoto

SÃO PAULO
2015

Kely Medeiros Turra

Aplicação de modelagem molecular e de formalismos do CAMD (Computer-Aided Molecular Design) na elucidação do mecanismo de ação de inibidores de metaloproteinases de matriz

Versão corrigida

Comissão Julgadora

da

Tese para obtenção do grau de Doutor

Profa. Dra. Silvia Berlanga de Moraes Barros

orientadora/presidente

Profa. Dra. Márcia Miguel Castro Ferreira

1° examinador

Prof. Dr. João Paulo dos Santos Fernandes

2° examinador

Prof. Dr. Salomão Dória Jorge

3° examinador

Profa. Dra. Kesley Moraes Godinho de Oliveira

4° examinador

São Paulo, 27 de março de 2015

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

Turra, Kely Medeiros
T958a Aplicação de modelagem molecular e de formalismos do CAMD (*Computer-Aided Molecular Design*) na elucidação do mecanismo de ação de inibidores de metaloproteinases de matriz / Kely Medeiros Turra. -- São Paulo, 2015.
166p.

Tese (doutorado) - Faculdade de Ciências Farmacêuticas da Universidade de São Paulo. Departamento de Farmácia.
Orientador: Barros, Silvia Berlanga de Moraes
Co-orientador: Pasqualoto, Kerly Fernanda Mesquita

1. Modelagem molecular : Química farmacêutica 2. Relações quantitativas entre estrutura química e atividade biológica : Química farmacêutica I. T. II. Barros, Silvia Berlanga de Moraes, orientador, II, Pasqualoto, Kerly Fernanda Mesquita, co-orientador.

615.19 CDD

AGRADECIMENTOS

Profa. Dra. Silvia Berlanga de Moraes Barros
Profa. Dra. Kerly Fernanda Mesquita Pasqualoto

Laboratório de Citopatologia e amigos:

Profa. Dra. Silvya Stucchi Maria-Engler
Andréa Fruet
Bárbara Vaz
Carolina Catarino
Débora Alves
Diogo Rivelli
Érika Matheus
Fanny Palace
Fernanda Faião
Fernanda Fillipin
Nayane de Souza
Natália Tamachiro
Manoela Thiago
Michelli Discacciati
Octávio Gomes
Paula Anastácia
Paula Penacchi
Patrícia Casari
Renato Massaro
Raquel Haga
Silvia Romano
Silvana Sandri
Tatiana Pedrosa
Thalita Zanoni

Colaboradores

Prof. Dr. Euzébio Guimarães Barbosa
Profa. Dra. Silvia Helena Pires Serrano
Dr. Rafael Martos Buoro

Projeto Financiado pela Fundação de Amparo às Pesquisas do Estado de São Paulo (FAPESP). Processo n° 2010/19142-1

RESUMO

TURRA, K.M. **Aplicação de modelagem molecular e de formalismos do CAMD (*Computer-Aided Molecular Design*) na elucidação do mecanismo de ação de inibidores de metaloproteinases de matriz.** 2015. 166f. (Tese de Doutorado). Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2015.

As metaloproteinases de matriz (MMP) são enzimas superexpressas em quase todos os tumores humanos, sendo que os subtipos MMP-2 e MMP-9 têm sido associados ao potencial metastático e prognóstico desfavorável em neoplasias malignas como, por exemplo, melanoma metastático e glioma. Compostos capazes de inibir a atividade destas enzimas podem representar potenciais agentes terapêuticos. O composto 4-nerolidilcatecol (4-NC), isolado de plantas do gênero *Pothomorphe*, apresentou resultados promissores para o tratamento do melanoma e glioma e foi capaz de atuar em várias etapas bioquímicas importantes envolvidas na progressão dessas patologias, inclusive inibindo MMP-2 e MMP-9. No entanto, o mecanismo de ação do 4-NC não está completamente elucidado.

O presente estudo envolveu a aplicação de métodos de modelagem molecular e de formalismos do planejamento de novas moléculas auxiliado por computador, CAMD (*Computer-Aided Molecular Design)* a fim de explorar a interação entre esta molécula e as enzimas MMP-2 e MMP-9, além de planejar novos inibidores para estes alvos.

Análise exploratória de dados, que compreende a análise de agrupamentos hierárquicos e de componentes principais, foi desenvolvida para um conjunto de hidroxamatos (N=64) descritos como inibidores de MMP-2 e MMP-9, a fim de identificar as propriedades moleculares que mais influenciavam o processo de discriminação dos compostos. As propriedades termodinâmicas, eletrônicas e estéricas foram importantes para descrever os compostos mais ativos no conjunto de dados da MMP-2. Para a MMP-9, o coeficiente de distribuição (ClogD) em pH 1,5 foi relevante no processo de discriminação do conjunto. A presença de substituintes volumosos na porção R_3 parece ser crucial para o conjunto de inibidores investigados. Esta região está envolvida em interações moleculares com a cavidade S1 de ambas as enzimas, mas há um limite de volume a ser considerado para estes substituintes.

O formalismo QSAR-4D independente do receptor (IR) foi aplicado ao mesmo conjunto de dados e permitiu estabelecer o mapeamento do farmacóforo, além de explorar diferentes alinhamentos para a obtenção da hipótese de conformação bioativa prevista pelo melhor modelo de QSAR. Os modelos QSAR apresentaram boa capacidade de previsão, auxiliaram na proposição de novos inibidores e estimaram a atividade do 4-NC. Com o melhor modelo QSAR para MMP-9 (N=64), a atividade prevista para o 4-NC foi classificada na faixa dos inibidores com atividade moderada. Entretanto, o melhor modelo QSAR obtido para MMP-2 (N=38) não foi capaz de prever, de forma adequada, a atividade de compostos com arcabouço químico diferente daqueles utilizados na construção dos modelos.

Estudos de ancoramento molecular foram desenvolvidos para investigar a orientação do 4-NC no sitio catalítico das duas enzimas e as interações que poderiam ser estabelecidas nestes complexos. Duas conformações favoráveis foram encontradas. Simulações computacionais de dinâmica molecular foram desenvolvidas com os complexos mais promissores selecionados nos estudos de ancoramento, a fim de obter informações mais detalhadas e de maior confiabilidade sobre suas interações intermoleculares. O 4-NC tende a se orientar no sítio de forma a acomodar sua cadeia lateral no bolso S1' adjacente ao sítio catalítico em ambas as enzimas.

Ensaios de zimografia também foram realizados com o objetivo de elucidar possíveis contribuições da cadeia lateral e do núcleo catecólico do 4-NC na atividade inibitória frente às enzimas em estudo. O núcleo catecólico parece ser o responsável por sua atividade, pois o composto 1,2-dimetoxibenzeno, que possui as hidroxilas bloqueadas por grupos metil, não foi capaz de exercer atividade inibitória significante frente à MMP-2 e MMP-9. Estudos de voltametria reforçaram a hipótese de que o 4-NC tem a capacidade de quelar os íons zinco presentes no tampão de incubação.

Palavras-chave: 4-nerolidilcatecol, metaloproteinase de matriz, QSAR-4D IR, ancoramento molecular, simulações computacionais de dinâmica molecular.

ABSTRACT

TURRA, K.M. **Molecular modeling methods and computer-aided molecular design (CAMD) formalisms for elucidating the mechanismof action of matrix metalloproteinases inhibitors.** 2015. 166f. (Tese de Doutorado). Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2015.

Matrix metalloproteinases (MMP) enzymes are overexpressed in almost all human tumors, and MMP-2 and MMP-9 subtypes have been associated with metastatic potential and poor prognosis in malignant tumors, such as metastatic melanoma and glioma. Compounds capable of inhibiting the activity of theses enzymes would be considered as potential therapeutic agents. The 4-nerolidylcatechol compound (4-NC), isolated from plants of genus *Pothomorphe*, has showed promising results in the treatment of melanoma and glioma, and was able to act in several important biochemical steps involved in the progression of these diseases, as well as inhibiting MMP-2 and MMP-9. However, the 4-NC mechanism of action is not completely understood. This study has involved the application of molecular modeling methods and formalisms of computer-aided molecular design (CAMD) in order to explore the interaction between 4-NC and MMP-2/MMP-9, and to design new inhibitors for these targets.

Exploratory data analysis, which comprises hierarchical cluster analysis and principal components analysis, was performed to a set of hydroxamates (N=64), previously reported as MMP-2 and MMP-9 inhibitors, in order to identify the molecular properties that is most critical for the discrimination process regarding the investigated compounds. The thermodynamic, electronic, and steric properties were quite important to describe the highly active compounds in the data set of MMP-2, whereas the apparent partition coefficient (ClogD) at pH 1.5 was the property more relevant for MMP-9 data set. The presence of bulky substituents on the R_3 moiety seems to be crucial for this set of inhibitors due to the molecular interaction with the S1 subsite of both enzymes. However, there is a limit regarding the substituents' volume in this region.

Receptor independent (RI) *4D-QSAR analysis* was applied to the same data set and it was possible to establish the pharmacophore mapping, besides to explore different alignments in order to generate the hypothesized bioactive conformation through the best QSAR model. The QSAR models have presented good predictability, assisted in proposing new inhibitors, and estimated the activity of 4-NC. Regarding the best QSAR model for MMP-9 (N=64), the 4-NC predicted activity was classified in the range of the moderate active inhibitors. The best QSAR model obtained for MMP-2 (N=38), however, was not able to properly predict the activity for compounds with different chemical scaffold from those used to build up the QSAR model. Molecular docking studies have been developed to investigate the 4-NC binding mode into the catalytic site of the two enzymes and the interactions that could be established in those complexes. The results have shown two favorable conformers regarding the MMP inhibition. Molecular dynamics computational simulation were combined to molecular docking studies in order to obtain more detailed and reliable information regarding the intermolecular interactions of each complex. The 4-NC molecule tends to accommodate the side chain in the S1´ pocket adjacent to the catalytic site in both enzymes.

Experimental zymography assays were also performed to elucidate the possible contribution of the side chain and the catechol core in the 4-NC inhibitory activity against the MMP-2 and MMP-9 enzymes. The catechol core seems to be responsible for its activity, since the 1,2 dimethoxybenzene compound, which has the hydroxyl blocked by a methyl group, was not able to exert any significant inhibition on enzymes. Voltametric assays confirmed the hypothesis that 4-NC chelates zinc ions present in the incubation buffer.

Keywords: 4-nerolydilcathecol, matrix metalloproteinases, IR QSAR-4D, molecular docking, molecular dynamics simulations.

LISTA DE ABREVIATURAS/SIGLAS

3D	tridimensional
4D	quadridimensional
4-NC	4-nerolidilcatecol
A	Ampère
Å	Angstrom
AGL	Algoritmo Genético Lamarckiano
Any	classe não-diferenciada de elemento de interação farmacofórica, qualquer tipo de átomo da molécula (análise QSAR-4D)
Ar	descrição de elemento de interação farmacofórica para átomos em sistema aromático da molécula (análise QSAR-4D)
BSA	*Bovine Serum Albumin* (*Albumina Bovina Sérica*)
CAMD	*Computer-Aided Molecular Design (planejamento de moléculas auxiliado por computador)*
CI	concentração inibitória
PC	*Principal Components* (componentes principais)
DM	dinâmica molecular
DMEM	*Dulbecco's modified Eagle's medium* (meio de cultura de Eagle modificado por Dulbecco)
$E_{1,4}$	Contribuição intramolecular de energia do tipo 1-4
E_{ang}	Contribuição intramolecular de energia de deformação angular
EDTA	*Ethylenediamine tetraacetic acid* (ácido etilenodiamino tetra-acético)
E_{el}	Contribuição intramolecular de energia eletrostática
Eest	Contribuição intramolecular de energia de deformação axial/estiramento
E_{lH}	Contribuição intramolecular de energia de ligação de hidrogênio
E_{HOMO}	Energia de orbital molecular de fronteira, orbital de mais alta energia ocupado por um elétron *(HOMO, highest occupied molecular orbital)*
E_{LUMO}	Energia de orbital molecular de fronteira, orbital de mais baixa energia não ocupado (*LUMO, lowest unnocupied molecular orbital*)
Eq.	Equação
EROS	espécies reativas de oxigênio
E_{solv}	Contribuição intramolecular de energia de solvatação
E_{tors}	Contribuição intramolecular de energia torsional
E_{tot}	Energia potencial total
E_{vdW}	Contribuição intramolecular de energia de van der Waals
Fs	Fentossegundo
GCOD	*Grid cell occupancy descriptors* (descritores de ocupação das células da grade virtual 3D)
GFA	*Genetic function approximation* (algoritmo de aproximação da função genética)
Grid	grade virtual 3D

LISTA DE ABREVIATURAS/SIGLAS

hba	*hydrogen bond acceptor* (descrição de elemento de interação farmacofórica para átomos aceitadores de ligação de hidrogênio; QSAR-4D análise)
HCA	*Hierachical cluster analysis* (análise de agrupamentos hierárquicos)
hbd	*hydrogen bond donor* (descrição de elemento de interação farmacofórica para átomos doadores de ligação de hidrogênio; análise QSAR-4D)
HF	Método *ab initio* Hartree-Fock
His	Histidina
INTERV+C	soma das contribuições intermoleculares de energia eletrostática e de Van der Waals
IPE	*Interaction pharmacophore elements* (elementos de interação farmacofórica da análise QSAR-4D)
IR	independente do receptor
K	Kelvin
Kcal	Quilocaloria
L	Ligante
Leu	Leucina
LNO	*Leave-N-out* (N = 2, 3, 4..., etc)
LOF	*Lack of fit* (fator de falta de ajuste de Friedman)
LOO	*Leave-one-out* (N = 1)
L-R	ligante-receptor
LSE	*Least Square Error* (erro de mínimos quadrados)
MARS	*Multivariate adaptative regression splines* (regressão multivariada adaptativa por *splines*)
MLR	*Multiple Linear Regression* (regressão linear múltipla)
MMP	metaloproteinase de matriz
NCE	New Chemical Entity (nova entidade química)
nM	nanomolar
np	descrição de elemento de interação farmacofórica para átomos de caráter não polar/apolar da molécula (análise QSAR-4D)
ns	nanosegundo
OMS	Organizaçao Mundial de Saúde
p-	descrição de elemento de interação farmacofórica para átomos de caráter polar, negativamente carregados, da molécula (análise QSAR-4D)
p+	descrição de IPE para átomos de caráter polar, positivamente carregados, da molécula (análise QSAR-4D)
PAC	perfil de amostragem conformacional
PCA	*Principal Component Analysis* (análise de componentes principais)
PDB	*Protein Data Bank* (banco de dados de proteínas)
PLS	*Partial Least Square* (regressão de quadrados mínimos parciais)
ps	picossegundo

LISTA DE ABREVIATURAS/SIGLAS

$q^2_{(LOO)}$	coeficiente de correlação de validação cruzada (*leave-one-out* (*N* = 1))
QSAR	*Quantitative structure-activity relationships* (estudos de relações quantitativas entre estrutura química e atividade biológica)
QSPR	*Quantitative structure-property relationships* (estudos de relações quantitativas entre estrutura química e propriedades moleculares)
R	receptor/biomacromolécula alvo
R	coeficiente de correlação linear de Pearson
r^2	coeficiente de correlação de determinação múltipla
RMN	Ressonância magnética nuclear
RMSD	*Root Mean Square Deviation* (desvio quadrático médio de posições atômicas)
SDS	*sodium dodecyl sulfate* (dodecil sulfato de sódio)
T	Temperatura
Tyr	Tirosina
u.m.a	unidade de massa atômica
V	Volt

SUMÁRIO

1. INTRODUÇÃO 11

1.1. MELANOMA E GLIOMA 11
1.1.1. Metaloproteinases de matriz (MMP) 11
1.1.2. Tratamento 12
1.1.3. 4-Nerolidilcatecol (4-NC) 13
1.2. MODELAGEM MOLECULAR E FORMALISMOS DO CAMD 14

2. JUSTIFICATIVA E OBJETIVO 19

3. MATERIAL E MÉTODOS 20

3.1. CONJUNTO DE INIBIDORES DAS ENZIMAS MMP-2 E MMP-9 20
3.2. CONSTRUÇÃO E OTIMIZAÇÃO DOS MODELOS MOLECULARES TRIDIMENSIONAIS 24
3.3. SIMULAÇÕES DE DINÂMICA MOLECULAR 25
3.4. ESTUDOS DE ANÁLISE EXPLORATÓRIA 26
3.4.1. Cálculo das variáveis independentes (descritores) 26
3.4.2. Seleção de variáveis independentes 29
3.4.3. HCA e PCA 29
3.5. ESTUDOS DE QSAR-4D IR 31
3.5.1. Construção e seleção do PAC para a análise QSAR-4D IR 32
3.5.2. Natureza dos descritores da análise QSAR-4D 33
3.5.3. Definição dos alinhamentos investigados 33
3.5.4. Tipo de ocupação e cálculo dos descritores de ocupação na grade virtual 3D 34
3.5.5. Análise de regressão por mínimos quadrados parciais (PLS) 35
3.5.6. Otimização dos modelos de QSAR e avaliação dos modelos gerados 36
3.5.7. Validação externa 38
3.5.8. Conformação bioativa prevista para os compostos do conjunto de treinamento 38
3.5.9. Validação cruzada leave-N-out 39
3.5.10. Aleatorização do vetor y (y-randomization) 39
3.5.11. Interpretação do melhor modelo QSAR 40
3.5.11.1. Processamento da proteína para o ancoramento das conformações biotivas/ descritores 41
3.5.11.2. Obtenção da conformação da estrutura proteica para o ancoramento da conformação bioativa/descritores de cada ligante 41
3.5.11.3. Dinâmica de relaxamento dos complexos MMP-2 e MMP-9/conformação bioativa/descritores de cada ligante 43
3.5.12. Planejamento, construção e otimização de novos inibidores das enzimas MMP 44
3.6. CONSTRUÇÃO E OTIMIZAÇÃO DA ESTRUTURA TRIDIMENSIONAL DO 4-NC 44
3.7. PREVISÃO DA ATIVIDADE DO 4-NC A PARTIR DOS MODELOS QSAR-4D IR 44
3.8. DETERMINAÇÃO DO ESTADO DE PROTONAÇÃO DO 4-NC 44
3.9. ESTUDOS DE ANCORAMENTO MOLECULAR (DOCKING) E REDOCKING 45
3.9.1. Processamento do modelo molecular 3D das enzimas MMP 45
3.9.2. Redocking 45
3.9.2.1. Estudos de redocking do complexo MMP-2/SC-74020 45
3.9.2.2. Estudos de redocking do complexo MMP-9/7-MR 46
3.9.3. Estudos de ancoramento molecular do 4-NC/MMP-2 e MMP-9 47
3.10. SIMULAÇÕES DE DINÂMICA MOLECULAR DOS COMPLEXOS PROVENIENTES DO PROCESSO DE ANCORAMENTO 48
3.11. PREPARO DO EXTRATO HIDROALCOÓLICO DE POTHOMORPHE UMBELLATA 49
3.12. EXTRAÇÃO DO 4-NEROLIDILCATECOL 49
3.13. ENSAIOS DE ZIMOGRAFIA 50
3.13.1. Eletroforese das proteínas no gel de poliacrilamida e gelatina 50
3.13.2. Análise das imagens 51
3.14. ENSAIOS DE VOLTAMETRIA 51

4. RESULTADOS E DISCUSSÃO 53
4.1. SIMULAÇÕES DE DINÂMICA MOLECULAR DOS INIBIDORES DE MMP 53
4.2. ANÁLISE EXPLORATÓRIA DE DADOS (HCA E PCA) 64
4.2.1. Cálculo de propriedades moleculares e seleção de variáveis 64
4.2.2. HCA e PCA para a tabela de dados X_I (inibidores de MMP-2) 69
4.2.3. HCA e PCA para os dados da tabela X_{II} (inibidores de MMP-9) 78
4.3. ESTUDOS DE QSAR-4D IR PARA OS INIBIDORES DE MMP-2 E MMP-9 83
4.3.1. Aplicação do formalismo QSAR-4D IR ao conjunto de hidroxamatos com atividade inibitória frente à enzima MMP-2 83
4.3.1.1. Validação externa 95
4.3.1.2. Conformação bioativa prevista e interpretação dos descritores de ocupação do modelo QSAR 97
4.3.1.3. Processamento da proteína para os estudos de ancoramento molecular 100
4.3.1.4. Previsão da atividade do 4-NC a partir do melhor modelo de QSAR obtido para o conjunto de inibidores de MMP-2 107
4.3.2. Aplicação do formalismo QSAR-4D IR ao conjunto de hidroxamatos com atividade inibitória frente à enzima MMP-9 108
4.3.2.1. Validação externa 120
4.3.2.2. Conformação bioativa prevista e interpretação dos descritores de ocupação do modelo QSAR 122
4.3.2.3. PROCESSAMENTO DA PROTEÍNA PARA OS ESTUDOS DE ANCORAMENTO DAS CONFORMAÇÕES BIOATIVAS/DESCRITORES, GCOD 125
4.3.2.4. Previsão da atividade do 4-NC a partir do melhor modelo de QSAR obtido para o conjunto de inibidores de MMP-9 133
4.4. DETERMINAÇÃO DO ESTADO DE PROTONAÇÃO DO 4-NC 135
4.5. ESTUDOS DE REDOCKING 136
4.5.1. Estudos de redocking do complexo inibidor/MMP-2 136
4.5.2. Estudos de redocking do complexo inibidor/MMP-9 137
4.6. ESTUDOS DE ANCORAMENTO 138
4.6.1. Estudos de ancoramento do 4-NC no sítio catalítico da enzima MMP-2 138
4.6.2. Estudos de ancoramento do 4-NC no sítio catalítico da enzima MMP-9 140
4.7. SIMULAÇÕES DE DINÂMICA MOLECULAR DOS COMPLEXOS COM O 4-NC 142
4.8. EXTRAÇÃO DO 4-NEROLIDILCATECOL 145
4.9. ENSAIOS DE ZIMOGRAFIA 148
4.10. ENSAIOS DE VOLTAMETRIA 149
5. CONCLUSÃO 152
6. REFERÊNCIAS BIBLIOGRÁFICAS 153

1. INTRODUÇÃO

O câncer representa um conjunto de mais de cem doenças que têm em comum o crescimento desordenado das células, sendo importante causa de morte no mundo (INCA, 2014). O melanoma e o glioma são neoplasias de baixa incidência e correspondem a apenas 3% e 10%, respectivamente, do total dos tumores malignos registrados no país, entretanto se destacam pelos seus altos índices de letalidade (INCA, 2014; GREINERT, 2009, INCA, 2014).

1.1. Melanoma e glioma

As neoplasias de pele podem ser classificadas em três tipos: carcinoma basocelular, carcinoma espinocelular e melanoma (INCA, 2014; GREINERT, 2009).

O melanoma é um tipo de câncer que se origina nos melanócitos, células produtoras do pigmento que determina a cor da pele. É considerado o mais agressivo dos tumores de pele devido à alta letalidade e ao elevado potencial para o desenvolvimento de metástases (DE LA FUENTE GARCIA, OCAMPO-CANDIANI, 2010). Vários fatores estão relacionados com o desenvolvimento desta doença como, por exemplo, a exposição a agentes químicos, alteração genética, doenças imunossupressoras e, principalmente, a exposição excessiva aos raios ultravioletas do sol (INCA, 2014).

Gliomas são tumores cerebrais primários provenientes da proliferação descontrolada de células nervosas denominadas gliais. Esta neoplasia pode ser classificada de acordo com as células das quais se originam: astrócitos, oligodendrócitos e células ependimárias. De acordo com a classificação da Organização Mundial de Saúde (OMS), existem quatro graus de malignidade atribuídos aos gliomas, que se baseiam na presença de células atípicas, mitoses, proliferação endotelial e necrose (a LOUIS *et al*., 2007, b LOUIS *et al*., 2007).

Os processos de invasão e metástase, fatores agravantes para a progressão do glioma e melanoma, são constituídos por várias etapas como, por exemplo, a degradação da matriz extracelular e destruição da membrana basal. Por conta disso, as metaloproteinases de matriz (MMP) são enzimas consideradas alvos importantes nestes processos (PEREIRA *et al*., 2006).

1.1.1. Metaloproteinases de matriz (MMP)

MMP são enzimas zinco-dependentes que clivam componentes proteicos da matriz extracelular. Estas enzimas estão envolvidas em uma variedade de processos fisiológicos que incluem processos de cicatrização, crescimento ósseo, processos inflamatórios, dentre outros. Mediante condições fisiológicas normais, os níveis de

expressão das MMP são baixos, ou praticamente indetectáveis. No entanto, tornam-se substancialmente elevados em neoplasias malignas como o melanoma. Estudos recentes afirmam que a presença de MMP em neoplasias é sinal indicativo de alto grau de agressividade e malignidade tumoral, uma vez que estas enzimas são responsáveis pela degradação dos principais componentes da matriz extracelular e facilitam, assim, a invasão tumoral e a metástase (WERMA, HANSCH, 2007; PEREIRA *et al*., 2006).

Até o presente momento, 23 membros da família das MMP foram identificados em humanos. A classificação das MMP varia de acordo com critérios estruturais, funcionais e de especificidade aos substratos. Dentre os vários tipos, a MMP-2 e a MMP-9 têm sido associadas ao potencial metastático e prognóstico desfavorável em neoplasias malignas. Estes subtipos também foram identificados como fatores majoritários em gliomas de alto grau de invasão (LEVICAR *et al*., 2003, WANG et *al*., 2003; CHOE *et al*., 2002) e a MMP-2 tem sido associada ao processo de invasão e metástase no melanoma (HOFMANN *et al*, 1999, REDONDO *et al*., 2005). Então, a pesquisa para inibir a função destas enzimas constituiria alternativa promissora na terapia antitumoral (YADAV *et al*., 2014; GIALELI *et al*. 2011, PEREIRA *et al*., 2006).

1.1.2. Tratamento

A escolha para o tratamento do melanoma varia de acordo com suas fases de estadiamento. Quando a doença se apresenta em estágio inicial, o tratamento é feito por excisão cirúrgica. Entretanto, quimioterapia e radioterapia são recomendadas quando os pacientes apresentam a forma metastática da doença. Os quimioterápicos mais utilizados no tratamento do melanoma são o anticorpo monoclonal anti-CTLA4 (*cytotoxic T-lymphocyte antigen 4*), ipilimumab, o fármaco vermurafenibe (inibe a proteína BRAF com a mutação V600E, alteração genética presente em 60% dos casos de melanoma) e o agente alquilante temozolomida (National Cancer Institute, 2014).

O tratamento dos gliomas representa um grande desafio na clínica médica uma vez que estes tumores cerebrais são altamente agressivos e resistentes à radioterapia e à quimioterapia. Fármacos como temozolomida, bevacizumab (agente angiogênico), cisplatina e carmustina (agentes alquilantes), moduladores do sistema imune e metabolismo celular, são geralmente utilizados na terapia, no entanto, dependendo do tipo e da localização do tumor, a remoção cirúrgica pode ser recomendada (ODIA *et al*., 2014; BADKE *et al*., 2014; NATSUME, *et al*., 2005; NAJMAN, GADELHA, 2002).

Infelizmente, tanto o melanoma metastático quanto o glioma são amplamente refratários às terapias existentes que, na grande maioria dos casos, apresentam eficácia

muito pequena, comprometendo o prognóstico de sobrevida do paciente (ODIA *et al*., 2014; KASPER *et al*., 2007, NAJAMAN, GADELHA, 2002).

No Brasil, de acordo com dados do INCA (2010), a sobrevida média de pacientes com melanoma mestastático, em estádio IV, é em torno de oito meses e menos de 10% dos pacientes sobrevivem cinco anos desde o diagnóstico (BALCH *et al*., 2009). Já para o glioblastoma multiforme (tumor primário do sistema nervoso central mais frequente em adultos) é, geralmente, menor que um ano desde seu diagnóstico.

Desta forma, o desenvolvimento de novos fármacos para combater estas neoplasias se faz necessário.

Frente à carência de alternativas terapêuticas para o tratamento destes tumores, diferentes classes de compostos químicos estão sendo testados como potenciais agentes antitumorais e merecem destaque àqueles capazes de inibir a atividade das MMP (GIALELI *et al*. 2011; TIAN, LEE, 2010; XI *et al*., 2009; 2004;; NELSON, MELENDEZ, 2004, NAJMAN, GADELHA, 2002).

Brohem *et al*. (2009) e Massaro *et al*. (2009) apresentaram resultados promissores na utilização do composto 4-nerolidilcatecol (4-NC) como potencial quimioterápico para o tratamento do melanoma e glioma.

1.1.3. 4-Nerolidilcatecol (4-NC)

O 4-NC (Figura 1), isolado e identificado por Kijjoa e colaboradores em 1980 (KIJJOA *et al*., 1980), é o principal metabólito secundário presente nas raízes e folhas de plantas do gênero *Pothomorphe*, como na *Pothomorphe umbellata*, planta nativa da mata atlântica.

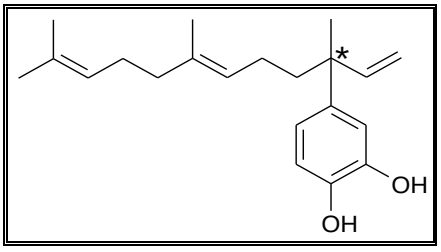


Figura 1. Estrutura química do 4-NC. (*: centro quiral).

Estudos realizados por Barros e colaboradores (1996) demonstraram o alto potencial antioxidante do 4-NC, que foi 10 vezes maior que a do α-tocoferol, um dos antioxidantes mais potentes descritos em literatura.

Ropke e colaboradores (2006) demonstraram que o 4-NC foi capaz de inibir a atividade gelatinolítica de enzimas, como a MMP-2 e MMP-9, em homogeneizados de pele de camundongos. Ainda, em estudos desenvolvidos com linhagens de melanoma

(SK-Mel 28 e SK-Mel 103), o 4-NC diminuiu a atividade de MMP e induziu morte celular por apoptose (BROHEM, 2009). Em modelos de peles artificiais, o 4-NC conseguiu conter a invasão do melanoma para a estrutura dérmica da pele reconstituída (BROHEM, 2010).

Em duas linhagens de células de glioblastoma multiforme (A172 e T98G), um tipo de glioma que apresenta alto grau de malignidade, o 4-NC foi capaz de inibir a invasão tumoral. Este composto, além de apresentar efeito antiproliferativo para ambas as linhagens, induziu as células à apoptose e, também, à autofagia. Estes dados sugerem que o 4-NC apresenta alto potencial de inibição de invasão e proliferação celular em gliomas (MASSARO *et al.*, 2009).

Outras atividades farmacológicas do 4-NC também foram evidenciadas, entre as quais se destacam: citotoxicidade sobre a linhagem de carcinoma humano bucal KB (MONGELII *et al*., 1996), atividade antimalárica (PINTO *et. al.*, 2006), inibição de mitotoxinas em serpentes (NUÑEZ *et al*., 2005), atividade fotoprotetora (ROPKE *et al*., 2003) e efeitos protetores no caso de genotoxicidade induzida por ciclofosfamida (VALADARES *et al.*, 2007).

O 4-NC tem se mostrado um agente muito promissor para ser utilizado no tratamento de glioma e de melanoma, já que é capaz de atuar em várias etapas bioquímicas no combate à estas neoplasias. Entretanto, seu mecanismo de ação em nível molecular ainda não está elucidado.

1.2. Modelagem molecular e formalismos do CAMD (*computer-aided molecular design*)

Modelagem molecular é a investigação das estruturas e das propriedades moleculares pelo uso da química computacional e técnicas de visualização gráfica, visando fornecer uma representação tridimensional, sob um dado conjunto de circunstâncias (SANT'ANNA, 2002).

Na década de 1980, o desenvolvimento de recursos computacionais e avanços em biologia estrutural viabilizaram o crescimento desta ferramenta de estudo *in silico*. Atualmente, a modelagem molecular é amplamente utilizada no planejamento racional de fármacos e auxilia na compreensão de interações entre fármaco, ou ligante, e os respectivos alvos biológicos (canais iônicos, proteínas transportadoras, enzimas, DNA) permitindo a compreensão dos mecanismos de ação em nível molecular (SILVA, 2003).

Em geral, a associação de duas abordagens é recomendada para a compreensão das interações tridimensionais entre ligante e biomacromolécula alvo: (a) ancoramento

molecular ou *docking* do ligante (L), ou micromolécula, no sítio de interação da biomacromolécula alvo (receptor, enzima, DNA, etc) e (b) simulações computacionais de dinâmica molecular (DM) do complexo L-R (ALONSO, BLIZNYUK, GREADY, 2006; MORGON, COUTINHO, 2007). Os estudos de ancoramento molecular consistem em prever possíveis orientações e confôrmeros da micromolécula no sítio de interação da biomacromolécula alvo, além de quantificar a afinidade do ligante em relação ao alvo (energia livre de interação de Gibbs, ΔG). No processo de planejamento e desenvolvimento de novas moléculas, os valores calculados de ΔG são comparados a dados experimentais de afinidade a fim de validar os dados obtidos e de auxiliar na construção e otimização de quimiotecas de compostos mais seletivos, com maior eficácia farmacodinâmica (WERMUTH, 2008; BULTINCK, WINTER, LANGERBAEKER, TOLLENAERE, 2004).

O processo de reconhecimento molecular de uma micromolécula pela biomacromolécula alvo é dinâmico e envolve interações intermoleculares entre o ligante, a biomacromolécula e o meio solvente. Desta forma, o complexo L-R pode apresentar conformações variáveis e fatores entrópicos (ΔS) e entálpicos (ΔH) influenciam, significativamente, neste processo. Cabe mencionar que uma das limitações da metodologia de ancoramento molecular consiste no modo de considerar a flexibilidade da biomacromolécula alvo durante as análises, já que os graus de liberdade são atribuídos somente às cadeias laterais dos resíduos de aminoácido do sítio de interação. Sendo assim, a aproximação das variações entrópicas conformacionais das moléculas interagentes não é fidedigna (MORGON, COUTINHO, 2007; ALONSO, BLIZNYUK, GREADY, 2006). Para transpor esta limitação, a associação da abordagem de ancoramento molecular a simulações computacionais de DM, por exemplo, deveria ser considerada a fim de se obter modelos mais significativos do processo de interação L-R (MORGON, COUTINHO, 2007). As simulações computacionais de DM permitem incorporar flexibilidade ou movimento a todas as partículas do sistema investigado, refinando as interações intermoleculares e reforçando a complementaridade L-R (NAMBA, SILVA, SILVA, 2008). O comportamento dinâmico de proteínas e complexos, em diferentes escalas de tempo e de temperatura, pode ser explorado em presença de moléculas de solvente explícito possibilitando melhor aproximação na geração dos parâmetros termodinâmicos de entalpia e entropia (ALONSO, BLIZNYUK, GREADY, 2006).

Outra abordagem amplamente utilizada no planejamento racional de novos fármacos são os estudos quantitativos das relações entre estrutura química e atividade

biológica (QSAR, *Quantitative Structure-Activity Relationships*) ou entre estrutura química e propriedades moleculares (QSPR, *Quantitative Structure-Property Relationships*). Os formalismos QSAR têm como principal objetivo a construção de modelos matemáticos que correlacionem propriedades moleculares, que são diretamente dependentes da estrutura química, à atividade biológica de um conjunto de compostos que, em geral, diferem entre si pela presença de grupos substituintes diferentes em posições definidas em um arcabouço químico comum. As relações matemáticas encontradas devem ser capazes de prever propriedades físico-químicas importantes (QSPR) ou a atividade biológica de novos compostos potencialmente mais ativos (*hits*) (FERREIRA *et al*., 2002). Vários métodos foram desenvolvidos, explorando *n*-dimensões (D), como QSAR-2D, 3D e 4D, com intuito de auxiliar o planejamento e desenvolvimento de novas moléculas (TAVARES, 2004). No presente estudo, a abordagem QSAR-4D será aplicada para mapear o farmacóforo de um conjunto de inibidores de MMP-2 e MMP-9 e auxiliar na previsão de atividade inibitória do composto 4-NC nestas enzimas.

Em 1997, Hopfinger e colaboradores propuseram o formalismo QSAR-4D (*four-dimensional quantitative structure-activity relationships*, relações quantitativas quadridimensionais entre estrutura química e atividade biológica), que considera o farmacóforo, a liberdade conformacional e de alinhamento no desenvolvimento de modelos de QSAR-3D para os dados de estrutura química e atividade biológica do conjunto de treinamento, avaliando a média deste conjunto, a quarta "dimensão" (HOPFINGER *et al*., 1997). Esta abordagem pode ser dependente (DR) ou independente (IR) do receptor, se a biomacromolécula alvo for ou não considerada no processo de construção dos modelos. Quando o mecanismo de ação dos compostos a serem investigados não estiver completamente elucidado, é recomendado iniciar os estudos com a abordagem IR, mesmo que a estrutura 3D da biomacromolécula alvo esteja disponível em bancos de dados.

Os dados requeridos para se desenvolver a análise QSAR-4D IR incluem um conjunto de compostos (conjunto de treinamento) e os dados de atividade biológica (variável dependente), obtidos em um mesmo protocolo de ensaio biológico. Além disso, um conjunto de avaliação, constituído de compostos que não participaram na construção do modelo, deve ser utilizado para validar o modelo QSAR gerado com o conjunto de treinamento e para verificar a capacidade de previsão externa do modelo obtido. As etapas da análise QSAR-4D IR estão listadas na Tabela 1.

Tabela 1. Sequência de etapas da análise QSAR-4D IR (HOPFINGER, 1999 e 1997)

Etapa	Descrição
1	Seleção e construção dos modelos moleculares 3D dos compostos que serão analisados;
2	Gerar o perfil de amostragem conformacional (PAC) de cada composto (simulações de DM);
3	Seleção das classes de elementos de interação farmacofórica, IPE (*interaction pharmacophoric elements*), para a grade virtual3D; Seleção dos alinhamentos a serem testados;
4	Formação da grade virtual de referência e registro do perfil de ocupação das células da grade
5	para cada IPE, de acordo com a medida de ocupação escolhida e considerando o alinhamento selecionado;
6	Análise de regressão por quadrados mínimos parciais (PLS, *partial least squares*): redução de dados, considerando o conjunto total de descritores de ocupação, GCOD (*grid cell occupancy descriptors*) (variáveis independentes);
7	Determinar o tamanho do conjunto inicial de descritores, GCOD, considerando os melhores classificados na análise de regressão por PLS, bem como quaisquer outros descritores calculados pelo usuário (por ex.: ClogP), para desenvolver a otimização dos modelos com a análise de algoritmo genético (GFA, *genetic function approximation* – ROGERS, HOPFINGER, 1994);
8	Retornar à etapa 4 e repetir as etapas 4-7, a menos que todos os alinhamentos selecionados já tenham sido incluídos na análise;
9	Identificação dos melhores modelos QSAR-3D (medidas estatísticas) de acordo com os alinhamentos testados;
10	Hipótese da conformação bioativa prevista de cada composto do conjunto de treinamento, de acordo com o melhor modelo e alinhamento escolhidos.

A validação e interpretação dos descritores selecionados na análise de QSAR-4D IR (melhor modelo) pode ser desenvolvida por meio de ancoramento molecular das conformações bioativas previstas pelo programa 4D-QSAR 3.0 (HOPFINGER, 1999) e dos respectivos descritores (GCODs, *grid cell occupancy descriptors*) para cada um dos ligantes investigados, no sítio de interação do alvo em estudo, quando a estrutura 3D do alvo estiver disponível em banco de dados. Simulações curtas de DM (DM de relaxamento) são desenvolvidas nesta etapa para promover a "acomodação" da conformação bioativa dos ligantes + GCOD no sítio de interação da biomacromolécula alvo (ANDRADE *et al*., 2009; PASQUALOTO *et al*., 2004). Desta forma, pode-se verificar se os ligantes do conjunto de treinamento se ajustam de forma satisfatória ao modelo/alinhamento selecionado. Em caso afirmativo, além de mapear o farmacóforo, os dados gerados na análise (descritores) são utilizados para identificar os tipos de interações que se estabelecem entre os ligantes e os resíduos de aminoácido do sítio ativo (complementaridade), além de verificar quais modificações moleculares seriam

interessantes para gerar pontos adicionais de interação no sítio da biomacromolécula alvo e, então, permitir a proposição de novos *hits* ou compostos potencialmente mais ativos, inclusive de novas entidades químicas (NCEs, *new chemical entities*).

2. JUSTIFICATIVA E OBJETIVO

Considerando a elevada taxa de mortalidade e a carência de alternativas terapêuticas efetivas para o melanoma metastático e o glioma, faz-se necessário a descoberta e o desenvolvimento de novos agentes antitumorais.

Métodos de modelagem molecular, de química computacional e de quimiometria contribuem de forma relevante para o planejamento de novos fármacos, pois auxiliam na obtenção de modelos moleculares tridimensionais que permitem a compreensão das interações L-R, auxiliando na elucidação do mecanismo de ação, além de direcionarem o processo de desenvolvimento de compostos mais promissores.

O 4-NC é um composto promissor para ser utilizado no tratamento do glioma e do melanoma e as enzimas MMP-2 e MMP-9 são cruciais na progressão destas neoplasias.

Desta forma, este estudo teve como objetivo a aplicação de metodologias de modelagem molecular e de formalismos do planejamento de novas moléculas auxiliado por computador (CAMD, *computer-aided molecular design*) para identificar as interações intermoleculares entre os complexos 4-NC e as metaloproteinases (MMP-2 e MMP-9) e propor, assim, seu possível mecanismo de ação em nível molecular.

Aplicou-se a abordagem de ancoramento molecular ou *docking* do 4-NC no sítio de interação das biomacromoléculas MMP-2 e MMP-9, associada a simulações de dinâmica molecular (DM), a fim de se obter informações relacionadas ao provável mecanismo catalítico dos sistemas moleculares investigados.

Em paralelo, aplicou-se o formalismo QSAR-4D IR (HOPFINGER *et al.*, 1997) a um conjunto de inibidores de MMP-2 e MMP-9. Esta abordagem auxilia na identificação dos grupos funcionais importantes para o processo de reconhecimento molecular, identifica os prováveis tipos de interações estabelecidas entre os inibidores e os resíduos de aminoácidos presentes no sítio de interação das duas MMP e auxilia no planejamento de novos inibidores para estas enzimas. A abordagem QSAR-4D IR foi utilizada pois, apesar das estruturas 3D das MMP estarem disponíveis, o modo de interação dos ligantes investigados não está totalmente elucidado.

Além disso, estudos experimentais como ensaios de zimografia foram desenvolvidos a fim de avaliar a contribuição da cadeia lateral e do anel catecólico da estrutura química do 4-NC no processo de inibição destas enzimas. Ensaios de voltametria foram também realizados para avaliar a capacidade quelante do 4-NC frente aos íons de zinco, metal fundamental para a atividade das MMP.

3. MATERIAL E MÉTODOS

Neste estudo foram utilizados programas computacionais de licença colaborativa como o MOLSIM 3.2 (DOHERTY, 2002), o 4D-QSAR 3.0 (HOPFINGER, 1999) e a versão beta do Fedora-4D QSAR, desenvolvidos pelo The Chem21 Group, Inc., sob a coordenação do Prof. Dr. Anton J. Hopfinger (Universidade do Novo México, EUA. Estas licenças colaborativas foram concedidas à Profa. Dra. Kerly Fernanda Mesquita Pasqualoto.

3.1. Conjunto de inibidores das enzimas MMP-2 e MMP-9

Um conjunto de 64 β-*N*-biaril éter-sulfonamida hidroxamatos (Figura 2, Tabela 2), relatados como inibidores de MMP-2 e MMP-9, foi selecionado dos trabalhos previamente publicados por Yang e colaboradores (a YANG, 2008; b YANG, 2008). A atividade biológica dos compostos (variável dependente) foi avaliada sob mesmo protocolo de ensaio e baseia-se na concentração capaz de inibir em 50% a atividade das enzimas MMP-2 e MMP-9 (CI_{50}, valores expressos em nM). Os valores de atividade/concentração inibitória foram linearizados (pCI_{50} ou $-logCI_{50}$ ou $\log 1/CI_{50}$), uma vez que os valores experimentais apresentavam ampla faixa de variação (NANTASENAMAT *et al*., 2009; KUBINYI, 1993).

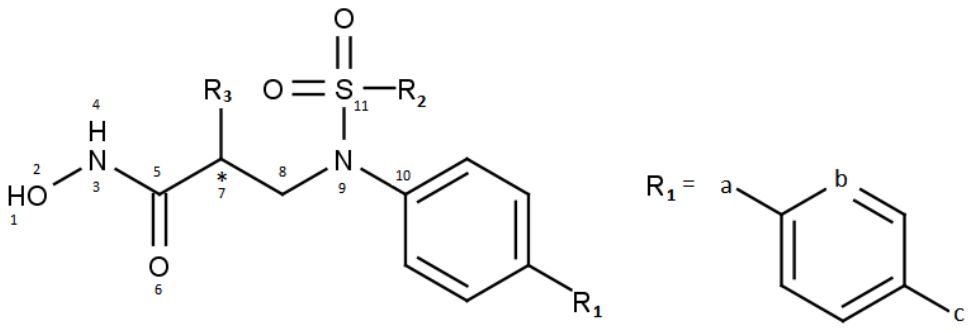


Figura 2. Características estruturais dos inibidores β-*N*-biaril éter-sulfonamida hidroxamatos. O carbono na posição 7(C_7*) é um centro quiral e pode apresentar configuração S/R.

Tabela 2. Valores de atividade biológica e substituintes **R_{1a}, R_{1b}, R_{1c}, R_2** e **R_3** do conjunto de inibidores β-*N*-biaril éter sulfonamida hidroxamatos

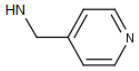
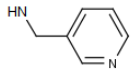
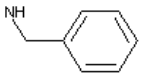
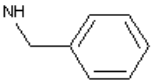
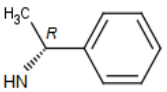
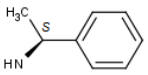
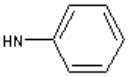
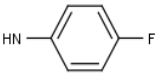
Compostos	C_7* (*R*/*S*)	pCI_{50} (MMP-2)	pCI_{50} (MMP-9)	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
3a	--	6,83	6,42	O	CH	H	Me	--
3b	--	7,14	7,51	O	CH	Me	Me	--
$5a_1$	*S*	7,21	8,11	O	CH	Me	Me	OH
$5a_2$	*R*	7,17	7,92	O	CH	Me	Me	OH
5b	*S*	6,85	7,55	O	CH	Me	Et	OH
5c	*S*	6,29	6,91	O	CH	Me	*n*-Pr	OH
5d	*S*	5,91	6,54	O	CH	Me	i-Pr	OH
5e	*S*	5,62	6,46	O	CH	Me	Ph	OH
5f	*S*	6,23	7,03	O	CH	Me	Bn	OH
5g	*S*	6,40	7,16	O	CH	Me	((piridina-3-il)CH_2	OH
5h	*S*	6,46	7,27	O	CH	Me	((piridina-4-il)CH_2	OH
5j	*S*	6,61	7,41	O	CH	H	Me	OH
5k	*S*	7,45	8,33	O	CH	Cl	Me	OH
5l	*S*	6,86	7,85	O	CH	CF_3	Me	OH
5m	*S*	5,60	6,33	O	N	CF_3	Me	OH
5n	*S*	6,53	7,27	S	CH	Me	Me	OH
5o	*S*	6,55	6,98	S	CH	F	Me	OH
$7a_1$	*S*	7,44	8,42	O	CH	Me	Me	O—CH_3
$7a_2$	*R*	6,99	7,89	O	CH	Me	Me	O—CH_3
8	--	7,21	8,18	O	CH	Me	Me	--
8a	*S*	8,04	8,62	O	CH	CF_3	Me	$NHCH(CH_3)_2$
$8b_1$	*S*	8,40	9,28	O	CH	Me	Me	$NHCH(CH_3)_2$
$8b_2$	*R*	6,75	7,51	O	CH	Me	Me	$NHCH_2(CH_3)_2$
8c	*S*	7,96	8,89	O	CH	Me	Me	$CH_2CH(CH_3)_2$
8d	*S*	8,00	8,77	O	CH	Me	Me	NH(ciclopentil)
8e	*S*	8,02	8,74	O	CH	Me	Me	
8f	*S*	7,77	8,55	O	CH	Me	Me	
8g	*S*	7,08	7,64	O	CH	Me	Me	
8h	*S*	6,69	8,15	O	CH	Me	Me	
8ir	*S*	7,62	8,82	O	CH	Me	Me	
8is	*S*	7,92	9,27	O	CH	Me	Me	
8j	*S*	7,89	9,28	O	CH	Me	Me	
8k	*S*	7,19	8,89	O	CH	Me	Me	

Tabela 2 (cont.). Valores de atividade biológica e substituintes **R_{1a}, R_{1b}, R_{1c}, R_2** e **R_3** do conjunto de inibidores β-*N*-biaril éter sulfonamida hidroxamatos

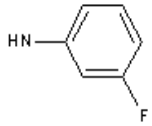
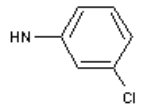
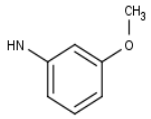
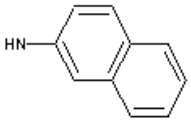
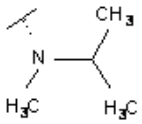
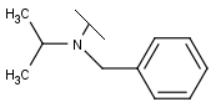
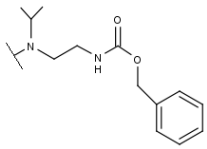
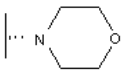
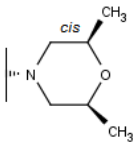
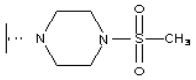
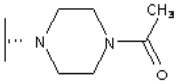
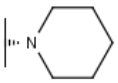
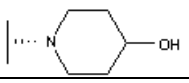
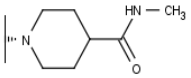
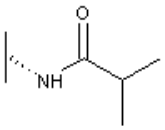
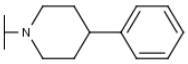
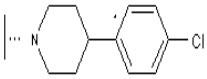
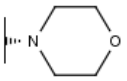
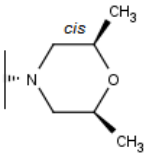
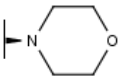
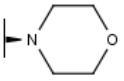
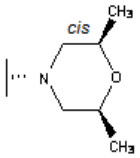
Compostos	*C7* (R/S)*	pCI_{50} (MMP-2)	pCI_{50} (MMP-9)	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
8l	*S*	7,27	9,05	O	CH	Me	Me	
8m	*S*	6,48	7,70	O	CH	Me	Me	
8n	*S*	7,57	9,21	O	CH	Me	Me	
8o	*S*	5,85	7,06	O	CH	Me	Me	
8p	*S*	7,62	8,05	O	CH	CF_3	Me	
8q	*S*	6,55	7,17	O	CH	CF_3	Me	
8r	*S*	6,58	7,54	O	CH	CF_3	Me	
8s	*S*	8,03	8,54	O	CH	CF_3	Me	
8t	*S*	8,25	8,96	O	CH	CF_3	Me	
8u	*S*	8,00	8,49	O	CH	CF_3	Me	
8v	*S*	8,06	8,32	O	CH	CF_3	Me	
8x	*S*	8,12	8,64	O	CH	CF_3	Me	
8y	*S*	8,24	8,72	O	CH	CF_3	Me	

Tabela 2 (cont.). Valores de atividade biológica e substituintes **R_{1a}**, **R_{1b}**, **R_{1c}**, **R_2** e **R_3** do conjunto de inibidores β-*N*-biaril éter sulfonamida hidroxamatos

Compostos	*C7* (R/S)*	pCI_{50} (MMP-2)	pCI_{50} (MMP-9)	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
8z	*S*	8,26	8,60	O	CH	CF_3	Me	
9a^{++}	*S*	5,90	6,03	O	CH	Cl	Me	OH
9b^{++}	*S*	6,54	6,62	O	CH	Cl	*i*-Pr	OH
11a	*S*	6,17	6,92	O	CH	Me	Me	NH_2
12a	*S*	7,15	8,27	O	CH	Me	Me	
13a	*S*	7,33	7,36	C≡C	CH	Me	Me	OH
13b	*S*	7,25	7,33	C≡C	CH	Cl	Me	OH
13c	*S*	7,49	7,04	C≡C	CH	OMe	Me	OH
14a	*S*	5,63	5,32	O	CH	F	Me	OH
14b	*S*	6,05	6,01	O—CH_2	CH	Cl	Me	OH
14c	*S*	5,53	5,83	O—CH_2	CH	CF_3	Me	OH
8aa	*S*	7,33	7,77	O	CH	CF_3	Me	
8bb	*S*	6,92	6,59	O	CH	CF_3	Me	
8cc	*S*	8,26	8,70	O	CH	Cl	Me	
8dd	*S*	8,60	9,34	O	CH	Cl	Me	
8ee$_1$	*S*	8,28	8,72	O	CH	Me	Me	
8ee$_2$	*R*	5,88	6,53	O	CH	Me	Me	
8ff	*S*	8,55	9,17	O	CH	Me	Me	

OBS: Os inibidores **3a** e **3b** não apresentam o carbono quiral na posição 7 (C_7). $^{++}$, os compostos **9a** e **9b** apresentam um grupo carbonila ligado ao nitrogênio n°9 e ao grupo R_2, no lugar do grupo sulfonil.

As faixas de dispersão para os valores de atividade biológica variam entre 5,53 a 8,60 e 5,32 a 9,34 para MMP-2 e MMP-9, respectivamente, conforme demonstrado nas Figuras 3 e 4.

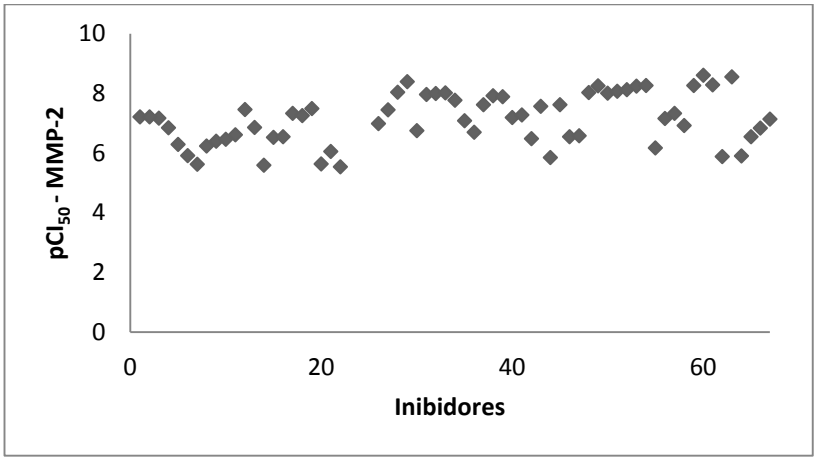


Figura 3. Dispersão dos valores de atividade biológica linearizados para os 64 inibidores hidroxamatos frente à enzima MMP-2.

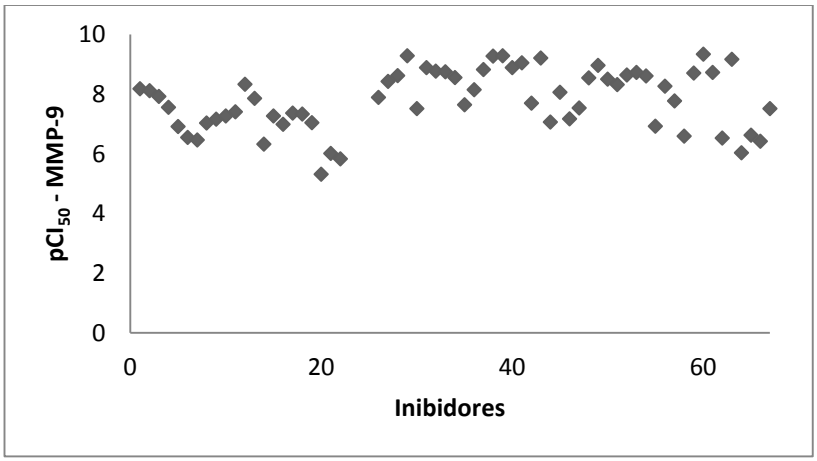


Figura 4. Dispersão dos valores de atividade biológica linearizados para os 64 inibidores hidroxamatos frente à enzima MMP-9.

Neste estudo, o gradiente de atividade biológica considerado para MMP-2 foi: valores de pCI_{50} entre 5,53 a 7,56 para compostos menos ativos e com atividade moderada; valores de pCI_{50} entre 7,57 a 8,60 para compostos muito ativos. Para MMP-9, consideraram-se compostos menos ativos e com moderada atividade aqueles com valores de pCI_{50} entre 5,30 a 8,40 e compostos muito ativos aqueles que apresentavam valores de pCI_{50} entre 8,41 a 9,34. Estes gradientes foram considerados de acordo com o estudo de Yang e colaboradores (a YANG *et al*., 2008; b YANG *et al*., 2008).

3.2. Construção e otimização dos modelos moleculares tridimensionais

Os modelos moleculares 3D dos 64 inibidores, em sua forma neutra, foram construídos utilizando o programa HyperChem 8.0 (HyperChem, 2007). A estrutura 3D de um inibidor hidroxamato (SC-74020) no sítio de interação da enzima MMP-2, cujas coordenadas cartesianas do complexo foram depositadas no banco de dados *Protein Data Bank* (**PDB**; BERMAN, 2000) (**PDB ID** 1HOV; RMN, FENG *et al*., 2002), foi

considerada como geometria de referência para a construção dos inibidores do conjunto em estudo. Adicionalmente, a configuração do carbono α (C_7**,** Figura 2) e de outros substituintes foi considerada e está indicada na Tabela 2.

A otimização de geometria de cada estrutura foi realizada empregando campo de força empírico MM+, adaptado do campo de força MM2 (ALLINGER, 1977), sem restrições. As cargas atômicas parciais dos modelos moleculares foram calculadas utilizando o método quântico semi-empírico AM1 (DEWAR, 1985), implementado no programa HyperChem 8.0 (HyperChem, 2007). Este método tem aplicação em sistemas que contenham os elementos H, Na, K, Rb, Zn, Hg, B, Al, C, Si, Ge, Sn, N, P, O, S, F, Cl e Br (SANT'ANNA, 2009) e alguns deles estão presentes nos compostos do conjunto em estudo.

Os modelos moleculares otimizados foram utilizados como dado de entrada para o processo de minimização de energia, com os métodos de declive máximo (*Steepest descent*) e de gradiente conjugado (*conjugate gradient*) com 500 e 20.000 iterações, respectivamente, utilizando o programa MOLSIM 3.2 (DOHERTY, 2002). A combinação destes métodos evita que o sistema entre em *loop* e permite que o mínimo local seja alcançado mais rapidamente (LEACH, 2001). Utilizou-se critério de convergência de 0,01 kcal/mol e constante dielétrica de 3,5, que melhor simula as condições do ambiente de membranas biológicas (ADMET/MI-QSAR, The Chem21 Group, Inc., 2001) e de bolsos ou cavidades no interior de enzimas (constante dielétrica 2-5). A temperatura foi correspondente à do ensaio biológico, 298 K (a YANG, 2008; b YANG, 2008). Os modelos moleculares com energia minimizada foram empregados como estruturas de entrada às simulações de DM e, consequente, construção do perfil de amostragem conformacional, PAC.

3.3. Simulações de Dinâmica Molecular

Para construção do PAC de cada um dos inibidores, empregou-se protocolo de simulação de DM de 1 ns (1.000.000 passos, 1 fs cada passo) a 298 K. Aos átomos da porção hidroxâmica (1, 2, 3, 4, 5, 6), carbono α (C_7*), nitrogênio (9) e enxofre (11) da porção sulfonamida foram conferidas massas fictícias elevadas, de 5000 u.m.a., a fim de restringir as coordenadas cartesianas e manter a integridade estrutural dos compostos em relação ao inibidor cristalografado no sítio da enzima. Arquivos de trajetória foram salvos a cada 20 passos de simulação, resultando em 50.000 conformações para cada inibidor.

Os arquivos de trajetória gerados nas simulações de DM foram analisados e as regiões que representaram a equilibração do sistema foram selecionadas para os estudos

de QSAR-4D IR. Para a análise exploratória de dados, o confôrmero de energia mais baixa de cada composto foi selecionado, pelo usuário, na região de equilibração para ser utilizado no cálculo dos descritores. O modelo de camada de hidratação proposto por Hopfinger (FORSYTHE, HOPFINGER, 1973) (solvente implícito) foi empregado para estimar as contribuições intramoleculares de energia de solvatação e de ligação de hidrogênio dos confôrmeros selecionados do PAC [MOLSIM 3.2 (DOHERTY, 1997)]. Nesta etapa, propriedades termodinâmicas foram geradas para cada composto.

3.4. Estudos de análise exploratória

3.4.1. Cálculo das variáveis independentes (descritores)

Os confôrmeros selecionados das simulações de DM, após minimização de energia, foram utilizados para o cálculo de diferentes propriedades moleculares ou descritores (variáveis independentes) utilizando programas apropriados. Descritores ou propriedades moleculares (termodinâmicas, eletrônicas, hidrofóbicas, topológicas, estéricas e de caráter misto) foram calculados com diferentes programas e métodos, listados na Tabela 3.

Tabela 3. Descritores, programas computacionais e métodos utilizados para calcular as propriedades moleculares do conjunto de inibidores β-*N*-biaril éter sulfonamida hidroxamatos para a análise exploratória de dados

Descritor	Descrição	Natureza	Métodos	Software
E_{est}	Contribuição de energia intramolecular de deformação axial ou estiramento	Descritores termodinâmicos	Simulações de DM (metodologia descrita em material e métodos)	*Molsim* 3.2
E_{ang}	Contribuição de energia intramolecular de deformação angular			
E_{tors}	Contribuição de energia intramolecular torsional			
$E_{1,4}$	Contribuição de energia intramolecular de interação do tipo 1-4 ou Lennard-Jones			
E_{vdW}	Contribuição de energia intramolecular de van der Waals			
E_{el}	Contribuição de energia intramolecular eletrostática			
E_{solv}	Contribuição de energia intramolecular de solvatação			
E_{LH}	Contribuição de energia intramolecular de ligação de hidrogênio			
E_T	Energia potencial total (soma de todas as contribuições energéticas)			
$ChelpG_1$	Carga de potencial eletrostático do átomo H_1	Descritores eletrônicos	*Ab-initio* HF/6-31G (d,p) (FOCK, 1930 ChelpG (BRENEMAN, 1990)	*Gaussian* 03W para Windows
$ChelpG_2$	Carga de potencial eletrostático do átomo O_2			
$ChelpG_3$	Carga de potencial eletrostático do átomo N_3			
$ChelpG_4$	Carga de potencial eletrostático do átomo H_4			
$ChelpG_5$	Carga de potencial eletrostático do átomo C_5			
$ChelpG_6$	Carga de potencial eletrostático do átomo O_6			
$ChelpG_7$	Carga de potencial eletrostático do átomo C_7			
$ChelpG_8$	Carga de potencial eletrostático do átomo C_8			
$ChelpG_9$	Carga de potencial eletrostático do átomo N_9			
$ChelpG_{10}$	Carga de potencial eletrostático do átomo C_{10}			
$ChelpG_{11}$	Carga de potencial eletrostático do átomoS_{11}			
$ChelpG_{12}$	Carga de potencial eletrostático do átomo O_{12}			
$ChelpG_{13}$	Carga de potencial eletrostático do átomoO_{13}			
$ChelpG_{14}$	Carga de potencial eletrostático do primeiro átomo ligado S_{11} (R_2)			
E_{lumo}	Energia do orbital molecular mais baixo desocupado (LUMO, thelowest unoccupied molecular orbital)			
E_{homo}	Energia do orbital molecular mais alto ocupado (HOMO, the highest occupied molecular orbital)			
μ_x	Componente X do momento de dipolo			
μ_y	Componente Y do momento de dipolo			
μ_z	Componente Z do momentode dipolo			
μ_{total}	Momento dipolar			
Platt	Índice topológico de Platt	Descritores topológicos	PLATT, 1947; PLATT, 1952	Marvin 5.8.0
Randic	Índice topológico de Randic		RANDIC, 1993	
Balaban	Índice topológico de Balaban		BALABAN,1982	
Harary	Índice topológico de Harary		PLAVSIC,1993	
Hyper Wiener	Índice topológico de Hyper-Wiener		RANDIC, 1993; LUKOVITS, 1994	

Tabela 3 (cont.). Descritores, programas computacionais e métodos utilizados para calcular as propriedades moleculares do conjunto de inibidores β-*N*-biaril éter sulfonamida hidroxamatos para a análise exploratória de dados

Descritor	Descrição	Natureza	Métodos	Software
Szeged	Índice topológico de Szeged	Descritores topológicos	KHADIKAR, 1995	Marvin 5.8.0
Wiener	Índice topológico de Wiener		WIENER,1947	
Wiener polarity	Índice topológico depolaridade deWiener		WIENER,1947; LUCIC*et al*., 2001	
ClogD_1,5	Coeficiente de distribuição calculado em pH 1,5	Descritores hidrofóbicos	KLOPMAN,1994	
ClogD_5,0	Coeficiente de distribuição calculado em pH 5,0			
ClogD_6,5	Coeficiente de distribuição calculado em pH 6,5			
ClogD_7,4	Coeficiente de distribuição calculado em pH 7,4			
ClogP	Coeficiente de partição *n*-octanol/água calculado		GHOSE,1989	Hyperchem 8.0
MSA	Área da superfície molecular	Descritores estéricos	FERRARA, 2002	Marvin 5.8.0
PSA	Área da superfície polar		ERTL, 2000	
V_{vdW}	Volume de van der Waals		HASEL,1988; STILL, 1990	
V	Volume molecular (estérico/intrínseco)		BODOR, 1989	Hyperchem 8.0
RM	Refratividade Molar	Estérico/ hidrofóbico	GHOSE, 1987	
α	Polarizabilidade	Estérico /eletrônico	MILLER,1990	
a_1	Ângulo entre os átomos O_6-C_5-C_7	Estrutural		

Os descritores termodinâmicos, gerados nas simulações de DM, correspondem às contribuições de energia intramolecular de deformação axial ou estiramento (Eest); de deformação angular (Eang); torsional (Etors); de interação do tipo 1-4 ou Lennard-Jones (E1,4); de van der Waals (EvdW); eletrostática (Eel); de ligação de hidrogênio (ELH); de solvatação (Esolv), que somadas correspondem à energia potencial total (ET). Estas contribuições de energia são referentes aos confôrmeros de energia mais baixa, que foram selecionados para os compostos dos respectivos PAC.

Descritores eletrônicos como, cargas parciais de potencial eletrostático (ESP ou CHELPG) (BRENEMAN, WIBERG, 1990), energias de orbitais moleculares de fronteira (E_{HOMO}, *Highest Occupied Molecular Orbital*, orbital molecular ocupado de maior energia; E_{LUMO}, *Lowest Unoccupied Molecular Orbital*, orbital molecular desocupado de menor energia), momento de dipolo total e nos eixos x, y, z, foram calculados com método *ab initio* Hartree-Fock (HF) (FOCK, 1930) utilizando o conjunto de bases 6-31G e funções de polarização (d, p) (BRENEMAN, WIBERG,1990).

Após o cálculo dos descritores, duas tabelas de dados foram geradas contendo as variáveis independentes calculadas (descritores) e os dados experimentais de atividade

em MMP (variável dependente). X_I: tabela de dados contendo os valores de atividade inibitória em MMP-2 e X_{II}: tabela de dados contendo os valores de atividade inibitória em MMP-9. As duas tabelas iniciais X_I e X_{II} apresentavam 64 linhas, correspondentes ao número de compostos ou amostras do conjunto investigado. O número de colunas equivale ao número total de descritores calculados (variáveis independentes, n = 49) mais o dado biológico (variável dependente, pCI_{50}).

3.4.2. Seleção de variáveis independentes

A seleção das variáveis independentes empregadas na análise exploratória de dados foi realizada por meio de dois critérios ou filtros: (1) coeficiente de correlação linear de Pearson (*R*) com o dado experimental; (2) inspeção visual da dispersão/distribuição do dado biológico (valores de pCI_{50}) para cada descritor calculado (FERREIRA, KIRALJ, 2009). Neste estudo, foi considerado como valor de corte $R > |0,25|$ e somente as variáveis/descritores que apresentaram boa distribuição e com tendência linear foram selecionadas para compor as tabelas finais X_I e X_{II}.

3.4.3. HCA e PCA

O método de HCA (*Hierarchical Cluster Analysis*, análise de agrupamentos hierárquicos) (JOHNSON, 1967) e de PCA (*Principal Components Analysis*, análise de componentes principais) (ABDI, WILLIAMS, 2010) foram aplicados às tabelas de dados finais, X_I e X_{II}, empregando-se o programa *Pirouette* 3.11 (Infometrix Inc, 1990-2003). Utilizou-se o autoescalamento, como método de pré-processamento das variáveis calculadas, devido às diferentes ordens de grandeza das variáveis calculadas (FERREIRA, KIRALJ, 2009; MOITA, MOITA, 1998).

HCA é um método multivariado para calcular e comparar distâncias entre os pares de amostras ou variáveis e aglomerá-las em agrupamentos que apresentam padrões e atributos similares. Os resultados são expressos como dendrogramas, que são mapas em formato de árvore construídos com base na proximidade existente entre as amostras no espaço. Para sua construção, calcula-se a distância entre todas amostras do conjunto, em pares, e então define-se uma matriz de similaridade cujos elementos são chamados índices de similaridade (MOITA, MOITA, 1998; JOHNSON, 1967). Estes índices variam entre zero a um, e quanto menor é a distância entre os pares de amostras ou variáveis, maior é o índice de similaridade.

Índice de similaridade $S_{ab} = 1{,}0 - d_{ab}/d_{max}$ (**eq.1**)

O cálculo é realizado depois que todas as amostras foram agrupadas, sendo:

d_{ab} = distância calculada entre duas amostras quaisquer (“a” e “b”)

d_{max} = maior distância calculada entre as amostras.

O método completo (*complete linkage)* e distância Euclideana (RENCHER, CHRISTENSEN, 2012) foram empregados neste estudo.

PCA é um método de compressão de dados baseado na correlação entre variáveis ou descritores. O objetivo é reduzir a dimensionalidade do conjunto de dados agrupando as variáveis correlacionadas e substituindo os descritores originais por novas variáveis denominadas fatores ou componentes principais. Assim, esta análise é capaz de transformar dados complexos e apresentá-los em uma nova perspectiva de forma que as informações mais importantes se tornem evidentes (FERREIRA, MONTANARI, GAUDIO, 2002; MOITA, MOITA, 1998). As componentes principais (PC), ou fatores, são variáveis não correlacionadas, construídas por combinações lineares simples a partir das variáveis originais e contêm a maior parte da variabilidade do conjunto de dados em um espaço dimensional muito menor. As PC são ortogonais e definidas em ordem decrescente da quantidade de variância que são capazes de explicar (FERREIRA, MONTANARI, GAUDIO, 2002).

PC1, ou fator 1, por exemplo, é definida na direção do máximo de variância de todo o conjunto de dados, enquanto a PC2, ou fator 2, é a direção que descreve o máximo da variância no subespaço perpendicular a PC1. As componentes seguintes são agrupadas perpendicularmente às primeiras e descrevem o máximo da variância restante. Estes eixos são ordenados pela sua importância, ou seja, pela ordem de variância. Uma vez que a redundância é removida, somente as primeiras PC são necessárias para descrever a maioria da informação contida no conjunto de dados original (FERREIRA, MONTANARI, GAUDIO, 2002; MOITA, MOITA, 1998, JOHNSON, 1967).

A matriz de dados $\mathbf{X}_{(I \times J)}$, que compreende I moléculas e J descritores, é decomposta em duas matrizes: **T** (matriz de escores) e **L** (matriz de pesos (*loadings*), sendo **X** = **TL**. A matriz **T** representa as posições das amostras no novo sistema de coordenadas, onde as PC são os eixos. A matriz **L** (de pesos), por sua vez, indica a importância ou contribuição das variáveis em cada PC (ABDI, WILLIAMS, 2010; FERREIRA, MONTANARI, GAUDIO, 2002). Neste estudo, sete PC foram consideradas

para processar os cálculos. A investigação de possíveis amostras atípicas (*outliers*) no conjunto de dados, implementada no software Pirouette 3.11 (Infometrix, Inc., 1990-2003), foi realizada com o emprego da distância de Mahalanobis (MAHALANOBIS, 1930).

O fluxograma referente às etapas de cálculo desenvolvidas está apresentado na figura 5.

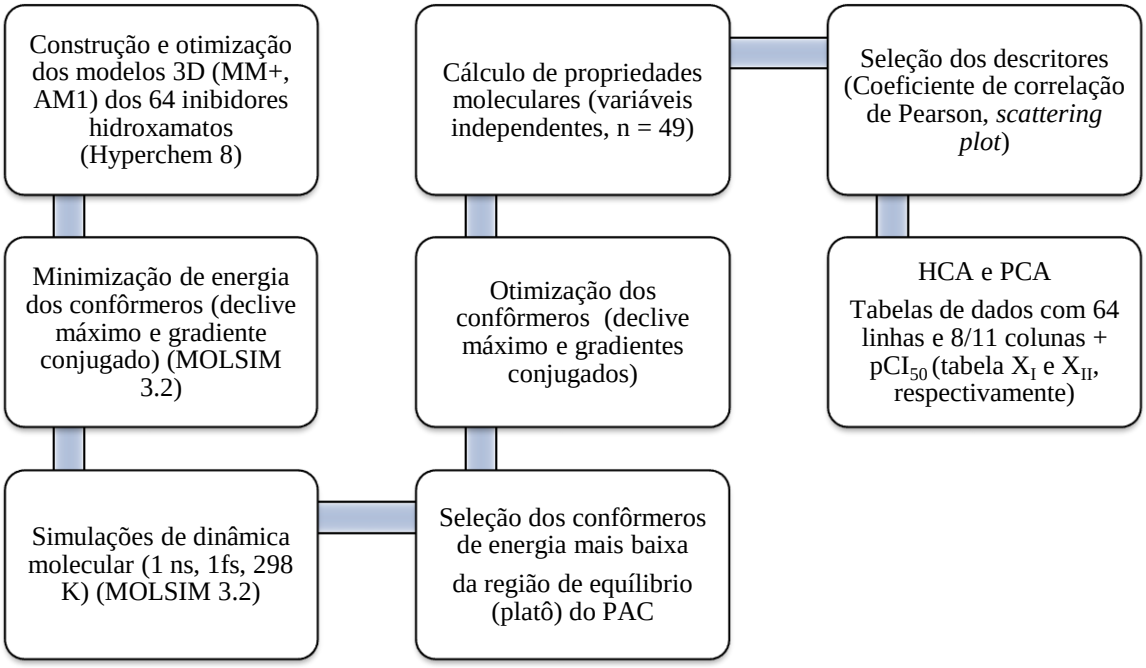


Figura 5. Fluxograma dos procedimentos para obtenção do PAC e etapas que precedem a análise exploratória de dados.

3.5. Estudos de QSAR-4D IR

Para o formalismo de QSAR-4D IR, selecionou-se um conjunto de 38 inibidores de MMP-2, dentre os 64 compostos do conjunto inicial (tabela 2) (a YANG, 2008; b YANG, 2008) utilizado na análise exploratória de dados, HCA e PCA, descrita no item **3.4**. Reduziu-se o conjunto de 64 para 38 compostos, pois se observou a presença de vários descritores de ocupação intercorrelacionados (GCOD), em análise preliminar, quando se utilizou o conjunto total de inibidores para gerar os modelos QSAR. Este conjunto de dados foi selecionado com o auxílio do método de HCA, de forma que as moléculas escolhidas representassem os inibidores contidos em cada um dos agrupamentos formados, além de apresentarem diversidade estrutural e variação no gradiente de atividade biológica (moléculas altamente ativas, com atividade moderada e menos ativas) (a YANG *et al*., 2008; b YANG *et al*., 2008).

Para a obtenção dos modelos com o conjunto de dados com atividade inibitória frente a MMP-9, utilizou-se os 64 compostos apresentados na tabela 2.

Primeiramente, modelos de QSAR foram construídos com os conjuntos totais de dados, N = 38 (MMP-2) e N = 64 (MMP-9), com o intuito de obter modelos que fossem estatisticamente mais representativos. Posteriormente, realizou-se a separação do conjunto em conjunto de treinamento, utilizado na construção dos modelos de QSAR, (N =30 para MMP-2 e N = 47 para a MMP-9), e de avaliação para a validação externa (N =10 para MMP-2 e N = 17 para MMP-9). Nesta etapa, o método de HCA foi aplicado para auxiliar a separação dos conjuntos e os critérios para dividir o conjunto de dados foram: (i) a diversidade estrutural/química e (ii) o gradiente de atividade biológica. É recomendado que o conjunto de avaliação contenha de 20 % a 30% do total do número de compostos (KIRALJ, FERREIRA, 2009; GOLBRAIKH, TROPSHA, 2002). A validação externa é essencial para a utilização de um modelo de QSAR no processo de desenvolvimento de novas moléculas ou de novos análogos, já que indica a capacidade de previsão do modelo considerando compostos que não participaram do processo de geração do modelo QSAR (KIRALJ, FERREIRA, 2009).

3.5.1. Construção e seleção do PAC para a análise QSAR-4D IR

A análise de QSAR-4D considera a liberdade conformacional (PAC) e de alinhamento de cada composto do conjunto investigado em função do tempo (quarta dimensão) no processo de construção dos modelos QSAR (HOPFINGER, 1999; HOPFINGER *et al*., 1997). Então, um conjunto de confôrmeros para cada ligante do conjunto é considerado ao invés de uma única conformação. Esta é uma das vantagens do método em relação ao QSAR-3D, onde apenas um único confôrmero de cada composto é considerado no desenvolvimento da análise.

A descrição da construção dos modelos moleculares 3D dos ligantes do conjunto investigado e do protocolo das simulações de DM, que geraram o PAC está nos itens **3.2** e **3.3**, respectivamente. Os arquivos de trajetória gerados nas simulações de DM foram avaliados e apenas porções de 100 ps da amostragem (5.000 conformações de cada ligante), que representavam o equilíbrio termodinâmico do sistema molecular, foram selecionadas e consideradas pelo programa 4D-QSAR 3.0 (HOPFINGER, 1999) na construção dos modelos matemáticos de previsão.

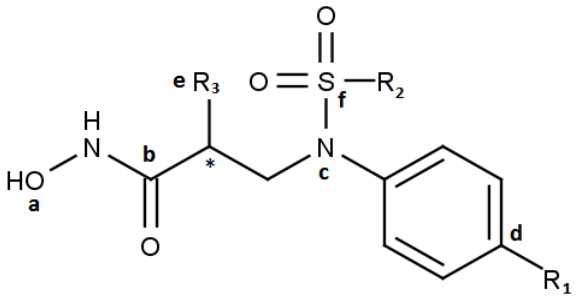
3.5.2. Natureza dos descritores da análise QSAR-4D (*interaction phamacophore elements*, IPE, elementos de interação farmacofórica)

O QSAR-4D utiliza sete classes de elementos de interação farmacofórica (IPE) para particionar os compostos investigados conforme a constituição química, a saber: qualquer tipo de átomo (*any*, classe não-diferenciada de IPE), átomos do tipo não-polar ou apolar (*np*), átomo de caráter polar positivamente carregado (p+), átomo de caráter polar negativamente carregado (p-), átomo aceitador de ligação de hidrogênio (*hba*), átomo doador de ligação de hidrogênio (*hbd*) e átomo em sistema aromático (*ar*). Estes IPE correspondem aos tipos de átomos que podem ocupar as células da grade virtual (*grid*), considerando um determinado alinhamento e identificam os tipos de interações que podem ocorrer entre os inibidores e o alvo, mapeando o farmacóforo (HOPFINGER, 1999; HOPFINGER *et al*., 1997). Todos os sete tipos de IPE foram utilizados neste estudo para explorar o conjunto de inibidores de MMP.

3.5.3. Definição dos alinhamentos investigados

O formalismo QSAR-4D utiliza alinhamentos de 3 átomos ordenados para comparar o PAC dos ligantes do conjunto de treinamento (HOPFINGER, 1999; HOPFINGER *et al*., 1997). A etapa do alinhamento é importante para determinar a orientação espacial dos compostos em relação ao sítio de ligação hipotético. A escolha dos alinhamentos a serem testados é de grande importância, uma vez que a resposta biológica está associada diretamente às interações entre os inibidores e o sítio de ação (HOPFINGER, 1999; HOPFINGER *et al*., 1997). Neste estudo, três alinhamentos foram selecionados de forma que contemplassem a estrutura comum aos inibidores, em toda a extensão, e o possível grupo farmacofórico, a porção hidroxâmica (a YANG *et al*., 2008). A porção sulfonamida também é considerada importante, pois está envolvida em interações com a cadeia polipetídica da enzima por intermédio de ligações de hidrogênio (b YANG *et al*., 2008). Os 3 átomos ordenados de cada alinhamento a ser investigado estão descritos na Tabela 4.

Tabela 4. Alinhamentos utilizados na construção dos modelos de QSAR-4D



Alinhamento	1° átomo	2° átomo	3° átomo
1	a	c	d
2	b	e	c
3	b	f	d

3.5.4. Tipo de ocupação e cálculo dos descritores de ocupação na grade virtual 3D

Todas as conformações provenientes dos PAC selecionados, de cada um dos inibidores, foram alinhadas de acordo com os átomos definidos para cada um dos alinhamentos em uma grade tridimensional virtual gerada pelo programa (HOPFINGER, 1999; HOPFINGER *et al*., 1997). Neste estudo, o tamanho selecionado para cada célula cúbica da grade virtual foi de 1 Å de aresta (resolução da grade virtual) e a grade virtual englobou todos os inibidores do conjunto de dados investigado. Considerou-se a ocupação absoluta das células da grade virtual para cada IPE utilizado (HOPFINGER *et al*., 1997) no cálculo dos descritores, denominados descritores de ocupação da célula da grade virtual, GCOD (*grid cell occupancy descriptors*). Ainda, nenhuma molécula de referência foi utilizada para o tipo de ocupação (HOPFINGER *et al*., 1997). As medidas de ocupação são normalizadas dividindo-se os valores amostrados pelo número de confôrmeros registrados na simulação de DM (HOPFINGER *et al*., 1999 e 1997). A ocupação absoluta, A_o, de uma célula da grade virtual (*i, j, k*) para uma determinada molécula é calculada como a soma de todos IPE (átomos) da molécula que estão localizados na célula da grade virtual considerando todas as conformações registradas no PAC (simulação de DM). "Estar" (O_t = m) ou "não estar" (O_t = 0) na célula da grade corresponde ao centro geométrico do átomo teste localizado em qualquer lugar dentro da célula da grade virtual. A ocupação absoluta é definida pela seguinte equação (**eq. 2**) (HOPFINGER *et al*., 1999 e 1997):

$$A_0\,(c, i, j, k, N) = \sum_{t=0}^{\tau} O_t\,(c, i, j, k) \qquad \textbf{(eq.2)}$$

τ = tempo decorrido na dinâmica molecular; *c* = molécula ou composto *c*; i, j, k célula da grade = x, y, z grade virtual; N = número de confôrmeros registrados de cada composto (PAC); t = 0 corresponde à primeira conformação do PAC; O = ocupação; O_t (*c*, i, j, k) = 0, se todos os átomos IPE de *c* não estiverem na célula da grade no tempo t; O_t (*c*, i, j, k) = m se m átomos IPE de *c* estiverem na célula (i, j, k) no tempo t.

3.5.5. Análise de regressão por mínimos quadrados parciais (PLS)

A análise QSAR-4D é multivariada e gera um grande número de descritores devido à quantidade de células da grade virtual (resolução da grade) e dos átomos IPE considerados. Por este motivo, a análise de regressão por mínimos quadrados parciais (*Partial Least Squares*, PLS) é uma das etapas deste formalismo. A compressão de dados considera a influência (peso) dos descritores calculados, GCOD, para cada um dos alinhamentos testados, em relação ao dado biológico (variável dependente, pCI_{50}). Este método de regressão agrupa os descritores que apresentam informações semelhantes em uma única variável latente e utiliza as variáveis que não estão intercorrelacionadas para a construção do modelo QSAR. Desta forma, além de evitar a colinearidade, este método funciona como ferramenta prévia de seleção de variáveis (FERREIRA, 2002). Neste estudo, testaram-seconjuntos de 200 a 700 descritores melhor classificados para gerar o conjunto inicial de descritores utilizados na etapa de otimização dos modelos QSAR.

O pré-processamento utilizado foi centrar os dados na média, já que os descritores apresentavam a mesma ordem de grandeza. Para reduzir o número de GCOD avaliados na análise de PLS, dois critérios (filtros) podem ser utilizados: a variância ou a ocupação máxima. Durante o processamento deste estudo, utilizou-se o critério de seleção de variância, onde descritores (GCOD/IPE) que apresentaram variância menor que 2,0 foram eliminados do conjunto de dados. Assim, descritores com maiores pesos individuais para as medidas de atividade biológica foram selecionados (HOPFINGER, 1999).

Cabe ressaltar que com o resultado obtido na análise de regressão PLS já seria possível gerar modelos de QSAR-3D, entretanto o objetivo do formalismo QSAR-4D é que apenas um pequeno número de descritores (menos de 15 dos milhares utilizados) seja significante no modelo de 3D-QSAR. Consequentemente, a análise de QSAR-4D fornece uma equação de 3D-QSAR mais simples em relação ao número de descritores/termos da

equação. Estes modelos compactos de 3D-QSAR são mais fáceis de interpretar e de serem utilizados para a previsão de novas estruturas químicas (HOPFINGER *et al*., 1999).

3.5.6. Otimização dos modelos de QSAR com algoritmo de aproximação da função genética (*GFA, genetic function approximation*) e regressão linear múltipla (MLR, multiple linear regression) e avaliação dos modelos QSAR gerados

Os descritores (GCOD/IPE) com maiores pesos individuais obtidos na análise de regressão PLS, para o alinhamento escolhido, foram aleatoriamente selecionados com a associação dos métodos de algoritmo genético (seleção de variáveis) e regressão linear múltipla (construção dos modelos de previsão) (ROGERS, HOPFINGER, 1994) para gerar o conjunto de dados dos modelos finais de QSAR-4D. O algoritmo de aproximação da função genética (GFA, *genetic function approximation*) (ROGERS, HOPFINGER, 1994; ROGERS, 1991) combina o algoritmo genético de Holland (AG -Holland's genetic algorithm) (HOLLAND, 1975) e algoritmo de Friedman (MARS, *multivariate adaptive regression splines*) (FRIEDMAN, 1991). A análise de regressão linear multidimensional ou linear múltipla (MLR, *multidimensional linear regression*) (PEARSON, LEE, 1908) é utilizada para reduzir os dados e construir os modelos finais de QSAR.

O algoritmo GFA utiliza uma população de múltiplos bons modelos e avalia somente os modelos finais, que agregam as melhores funções. Modelos melhores são gerados à medida que se desenvolvem as recombinações genéticas (*crossovers*), que têm como finalidade reorganizar os termos que conferem melhor perfil aos modelos. Os modelos iniciais são construídos de acordo com as especificações do usuário sobre o tipo de função a ser empregada. Estes são gerados por meio da seleção aleatória de um determinado número de características do conjunto de dados investigado (descritores melhores classificados previamente na análise PLS) e de funções matemáticas disponíveis no programa 4D-QSAR 3.0. A partir daí os modelos genéticos são gerados com sequências aleatórias daquelas funções. O algoritmo GFA pode gerar modelos que utilizam não somente funções lineares, mas também funções quadráticas (polinômios de segunda ordem), *splines* e outras funções não-lineares (ROGERS, HOPFINGER, 1994). Na população de um modelo, define-se também a porcentagem de mutação em cada ciclo de recombinação genética, para assegurar a diversidade no processo de evolução.

Neste estudo, consideraram-se os dez melhores modelos gerados com o programa WOLF 5.5 (ROGERS, 1994), incluso no pacote 4D-QSAR 3.0 (HOPFINGER, 1999). Os melhores modelos são classificados pelo fator de falta de ajuste de Friedman (LOF, *lack-of-fit*), que é a medida de mínimos quadrados penalizada. O fator de ajuste ou *smoothing*

factor (d), que faz parte da definição da medida estatística LOF, é o único parâmetro determinado pelo usuário (**eq. 3**; ROGERS, HOPFINGER, 1994) e altera o equilíbrio entre o número de variáveis independentes nos modelos e a redução no valor do erro de mínimos quadrados (LSE, *least squares error*). O valor considerado *default* para o fator de ajuste no programa 4D-QSAR 3.0 é 1,0 (ROGERS, 1994).

$$LOF = \frac{LSE}{\left(1 - \frac{c + dp}{M}\right)^2} \quad \textbf{(eq. 3)}$$

c = número de funções no modelo

d = fator de ajuste

p = número total de características contidas em todas as funções de base

M = número de amostras pertencentes ao conjunto de treinamento

A análise MLR, utilizada para reduzir os dados e construir modelos QSAR, apresenta a desvantagem de ser extremamente sensível à presença de colinearidade entre os descritores selecionados, gerando coeficientes de regressão instáveis e sem significado. Por este motivo, nas análises de QSAR o número de descritores presentes deve ser menor que o número de amostras (inibidores/compostos) em estudo. Em geral, a tabela de dados para gerar modelos de QSAR deve apresentar um descritor para cada cinco ou seis amostras (FERREIRA, 2002; MARTENS, NAE, 1989).

A etapa de otimização dos modelos QSAR, GFA-MLR, foi realizada utilizando o conjunto inicial de descritores previamente selecionado na etapa anterior (análise PLS). Empregou-se, neste estudo, somente a função linear e variaram-se os valores do fator de ajuste de 2,0 a 3,0 para determinar o número ótimo de descritores nos modelos. O número de ciclos de operações genéticas (*crossovers*) testado foi de 50.000 a 100.000 e a probabilidade de mutação de 10% foi considerada em cada ciclo de recombinação genética.

Avaliaram-se os dez melhores modelos de acordo com os seguintes parâmetros estatísticos: coeficiente de correlação de determinação múltipla (r^2), coeficiente de correlação de validação cruzada *leave-one-out* (LOO) (q^2), LSE e LOF. A matriz de correlação-cruzada entre pares de descritores também foi calculada para verificar se os descritores conferiam informação independente ao modelo. Pares de descritores que apresentassem valores de coeficiente de correlação linear de Pearson (*R*) maiores que 0,5 (método GFA-MLR) foram desconsiderados no processo de análise dos modelos QSAR.

Da mesma forma, calculou-se a matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para determinar se os dez modelos selecionados eram equivalentes ou distintos. Pares de modelos altamente correlacionados (R≈1) foram considerados equivalentes (modelos semelhantes), enquanto que os pares de modelos pouco correlacionados (R < 0,5) foram considerados diferentes entre si.

Neste estudo, ligantes ou compostos que apresentaram valor de resíduo [diferença entre valores de atividade experimental (pCI_{50_obs}) e calculada (pCI_{50_calc})] maior que dois desvios em relação ao desvio-padrão da média dos resíduos do modelo (limite de confiança de 95%) foram considerados como amostras atípicas (*outliers*).

3.5.7. Validação externa

A validação externa foi desenvolvida para verificar a capacidade de previsão externa do melhor modelo de QSAR construído a partir do conjunto de treinamento. Como já mencionado, a seleção dos compostos para compor o conjunto de avaliação foi realizada com auxílio do método HCA. Os critérios considerados para a seleção dos inibidores do conjunto de avaliação foram: representatividade de diversidade estrutural e de gradiente de atividade biológica do conjunto original. Recomenda-se a seleção de 20 a 30% de compostos correspondentes ao conjunto total de dados (KIRALJ, FERREIRA, 2009; GOLBRAIKH, TROPSHA, 2002). Os compostos do conjunto de avaliação foram submetidos às etapas 1 a 6 (Tabela 1) da análise QSAR-4D IR, a fim de gerar os valores de descritores de ocupação nas células da grade virtual, GCOD, para cada um dos inibidores.

Os valores calculados de pCI_{50} (pCI_{50_calc}) para cada composto do conjunto de avaliação foram obtidos a partir da substituição dos seus respectivos valores de GCOD na equação matemática obtida para o melhor modelo/alinhamento de QSAR com o conjunto de treinamento.

Consideraram-se compostos satisfatoriamente previstos pelo melhor modelo de QSAR aqueles que apresentaram valores de resíduo [pCI_{50_obs}- pCI_{50_calc}] menores ou iguais a um desvio em relação ao desvio-padrão da média dos resíduos do modelo (limite de confiança de 99%).

3.5.8. Conformação bioativa prevista para os compostos do conjunto de treinamento

A última etapa da análise QSAR-4D é gerar a conformação bioativa hipotética para cada um dos compostos do conjunto de treinamento, considerando o melhor modelo QSAR, do melhor alinhamento.

Neste estudo, selecionaram-se confôrmeros para cada ligante que apresentaram variação de energia (ΔE) de até 5,0 kcal/mol em relação à conformação de energia mais baixa de todo o PAC para gerar os descritores de ocupação. Os valores de GCOD de cada um dos confôrmeros foram substituídos na melhor equação/melhor alinhamento e o confôrmero cujos descritores de ocupação proporcionam o maior/melhor valor de atividade (pCI_{50}) foi considerado, pelo programa 4D-QSAR 3.0, como a conformação bioativa prevista (HOPFINGER, 1999).

3.5.9. Validação cruzada *leave-N-out*

A validação cruzada *leave-N-out* (LNO), também denominada *leave-many-out*, (BAUMANN, STIEFL, 2004; CLARK, FOX, 2004; BAUMANN, 2003), é um procedimento recomendado para testar a robustez do modelo QSAR gerado. Antes deste procedimento, tabela de dados contendo as amostras com os valores de descritores (variáveis independentes) e de atividade (variável dependente) deve ser aleatoriazada e recomenda-se que *N* (número de amostras a serem retiradas e utilizadas para a previsão a partir dos modelos gerados com as amostras restantes) represente uma fração significativa do conjunto de dados. Para conjuntos pequenos, 20 a 30% de amostras devem ser retiradas para avaliação dos valores de coeficiente de validação cruzada (q^2_{LNO}) gerados neste procedimento (GRAMATICA, 2007). O *N* crítico corresponde ao valor máximo de *N* no qual o valor de q^2_{LNO} ainda permanece estável, ou seja, semelhante ao valor de q^2_{LOO} (gerado na construção dos modelos QSAR). Para um bom modelo, q^2_{LNO} deve apresentar valores próximos ao q^2 obtido para o procedimento LOO. Neste estudo, a validação LNO foi realizada em triplicata. No caso de réplicas, o valor de 2 desvios padrão não deve ser maior que 0,1 para $N = 2, 3$, etc, incluindo o valor crítico de *N* (KIRALJ, FERREIRA, 2009).

3.5.10. Aleatorização do vetor *y* (*y-randomization*)

O objetivo do procedimento de aleatorização do vetor ***y*** (***y-randomization)*** (KIRALJ, FERREIRA, 2009; RÜCKER *et al*., 2007; BAUMANN, STIEFL, 2004; CLARK, FOX, 2004; ERIKSSON *et al*., 2003; WOLD, ERIKSSON, 1995) é detectar e quantificar correlação entre as variáveis dependentes e os descritores dos modelos gerados, neste caso, os descritores de ocupação, GCOD. O teste consiste em construir vários modelos considerando fixos os valores dos descritores (bloco **X**), mas permutando somente os valores de atividade biológica (vetor **y**). Os modelos obtidos, sob mesmas condições dos modelos originais, devem apresentar baixa qualidade e sem real significado, o que é refletido nas medidas estatísticas, q^2_{yrand} e r^2_{yrand}. Neste estudo,

considerou-se a abordagem de Eriksson and Wold (1995) para avaliar se o modelo real foi ou não obtido ao acaso. Consideraram-se os valores de q^2_{yrand} e r^2_{yrand} e suas relações $r^2_{yrand} > q^2_{yrand}$:

$q^2_{yrand} < 0,2$ e $r^2_{yrand} < 0,2 \rightarrow$ nenhuma correlação ao acaso;
qualquer q^2_{yrand} e $0,2 < r^2_{yrand} < 0,3 \rightarrow$ correlação ao acaso não significativa;
qualquer q^2_{yrand} e $0,3 < r^2_{yrand} < 0,4 \rightarrow$ correlação ao acaso tolerável;
qualquer q^2_{yrand} e $r^2_{yrand} > 0,4 \rightarrow$ correlação ao acaso importante/significativa.

A abordagem de Eriksson and Wold (1995) recomenda um número de 10 aleatorizações/corridas para qualquer conjunto de dados. Este teste seria suficiente uma vez que falsos bons modelos (pelo menos 10%) aparecem em uma ou mais corrida aleatorizadas de cálculo. Vários autores sugerem que as corridas sejam empregadas em centenas ou milhares de vezes, independente do tamanho do conjunto de dados estudado. Por outro lado, outros contestam que o número de corridas deve variar de acordo com o tamanho do conjunto. De acordo com Kiralj e Ferreira (2009), 10 e 1.000 corridas com os valores aleatorizados de **y** podem gerar a mesma informação qualitativa e espera-se que, mesmo com um pequeno número de corridas, modelos pobres apresentem um desempenho indesejado, enquanto que bons modelos estejam livres de quaisquer correlações ao acaso. Estes autores demonstraram, também, que 10 a 25 corridas são suficientes para o processo de validação dos modelos de QSAR. Neste estudo, realizaram-se 10, 30 e 50 corridas de valores aleatorizados de **y** para todos os conjuntos de dados, empregando-se o software QSAR modeling (FERREIRA, MARTINS, 2013).

3.5.11. Interpretação do melhor modelo QSAR

A conformação bioativa prevista e os respectivos descritores, de cada composto, gerados para o melhor modelo/alinhamento no programa 4D-QSAR 3.0 (HOPFINGER, 1999) foram ancorados manualmente às estruturas 3D das biomacromoléculas MMP-2 e MMP-9, considerando os 3 átomos ordenados do melhor alinhamento e o inibidor original de cada complexo como referência (programa HyperChem 8; Hypercube, Inc., 2007). O ancoramento molecular da conformação bioativa prevista com os descritores no alvo permite validar os tipos de interações previstas pelo melhor modelo/alinhamento gerado na análise QSAR-4D IR, considerando a complementaridade com os resíduos de aminoácidos do sítio ativo das enzimas alvo (ANDRADE *et al.*, 2010; ANDRADE *et al.*, 2009; PASQUALOTO *et al.*, 2004). Esta avaliação é importante para determinar a

natureza dos descritores de classe não-diferenciada (*any*) quando presentes no modelo final de QSAR-4D IR. Os modelos moleculares 3D das enzimas MMP foram previamente processados conforme descrito a seguir.

3.5.11.1. Processamento da proteína para o ancoramento das conformações biotivas/ descritores, GCOD

Os modelos moleculares 3D do sítio catalítico das enzimas MMP-2 (ressonância magnética nuclear, RMN) e MMP-9 (difração de raios-X), depositados no banco de dados **PDB**, com os códigos 1HOV (FENG *et al*. 2002) e 2OW1 (resolução 2,20 Å; TOCHOWICZ *et al*., 2007), respectivamente, foram utilizados para auxiliar na interpretação e na validação do melhor modelo/alinhamento obtido com a abordagem QSAR-4D IR. Utilizou-se apenas o sítio catalítico destas enzimas, a fim de reduzir o tempo computacional nas simulações de dinâmica molecular de relaxamento. Neste caso, a utilização apenas do sítio catalítico é plausível, uma vez que somente este domínio é necessário para hidrolisar substratos sintéticos e proteínas não colágenas (NAGASE, VISSE, MURPHY, 2006). As moléculas de solvente e os pares de elétrons desemparelhados (*lone pair*) foram retirados da estrutura manualmente e as estruturas foram corrigidas com a biblioteca de rotâmeros Drunback (DUNBRACK, 2002) no programa UCSF Chimera 1.7 (PETTERSEN *et al*., 2004). Somente os inibidores complexados, íons de zinco e de cálcio foram mantidos no sítio ativo para a construção dos modelos. Para a estrutura da enzima MMP-9, o aminoácido glutamina (Q) na posição 402 foi mutado para ácido glutâmico (Glu), utilizando o programa *Discovery Studio Visualization 3.5* (Accelrys Software Inc., 2005). O resíduo Glu402 faz parte da porção catalítica no sítio ativo da MMP-9. Após a adição de hidrogênios aos resíduos, cargas parciais AMBER foram atribuídas a todos os átomos (WEINER *et al*., 1984) (software Hyperchem 8.0; Hypecube, 2007). Grupos metila (*block groups*) foram adicionados ao primeiro e último resíduos da cadeia polipeptídica, com cargas parciais iguais à zero.

3.5.11.2. Obtenção da conformação da estrutura proteica para o ancoramento da conformação bioativa de cada ligante/descritores, GCOD

O programa MOLSIM 3.2 (DOHERTY, 2002) foi empregado para desenvolver as simulações computacionais de DM dos modelos moleculares 3D dos complexos provenientes do PBD. As estruturas processadas, como descritas no item anterior, foram minimizadas com os métodos de declive máximo e de gradientes conjugados na temperatura de 298 K. O critério de convergência estabelecido foi de 0,5 kcal/mol. Massas fictícias de 5.000 u.m.a. foram atribuídas aos átomos da cadeia polipeptídica

principal a fim de preservar a integridade estrutural dos sistemas moleculares durante as simulações. As massas fictícias equivalem a restrições cartesianas e são utilizadas principalmente quando os sistemas moleculares apresentam alto peso molecular (TOKARSKI, HOPFINGER, 1997).

O protocolo empregado para as simulações de DM consistiu de um esquema de aquecimento gradativo *(warming-up*), seguido de simulação mais longa na temperatura do ensaio biológico. Empregaram-se simulações curtas (50 ps; cada passo 1 fs) em 50, 100, 200 e 300 K. Este tipo de esquema é aplicado para gerar o estado de menor energia do complexo de forma mais rápida, permitindo que o sistema supere barreiras energéticas (TOKARSKI, HOPFINGER, 1997). A conformação de energia mais baixa do complexo à temperatura de 50K foi selecionada na região de equilíbrio do PAC, minimizada, e então utilizada como dado de entrada para a próxima simulação de 50 ps em T mais elevada (100 K). O mesmo procedimento foi repetido até à temperatura de 300 K. O confôrmero de energia mais baixa do complexo selecionado em 300 K, após minimização de energia, foi utilizado em simulação de 2ns a 298 K, temperatura de determinação da proteína (Figura 6). O arquivo trajetória foi registrado a cada 20 passos de simulação. Utilizou-se constante dielétrica de 3,5 (SANTOS-FILHO, MISHRA, HOPFINGER, 2001; TOKARSKI, HOPFINGER, 1997). Os PAC foram analisados e os confôrmeros de energia mais baixa foram selecionados para análise e visualização, após a minimização de energia.

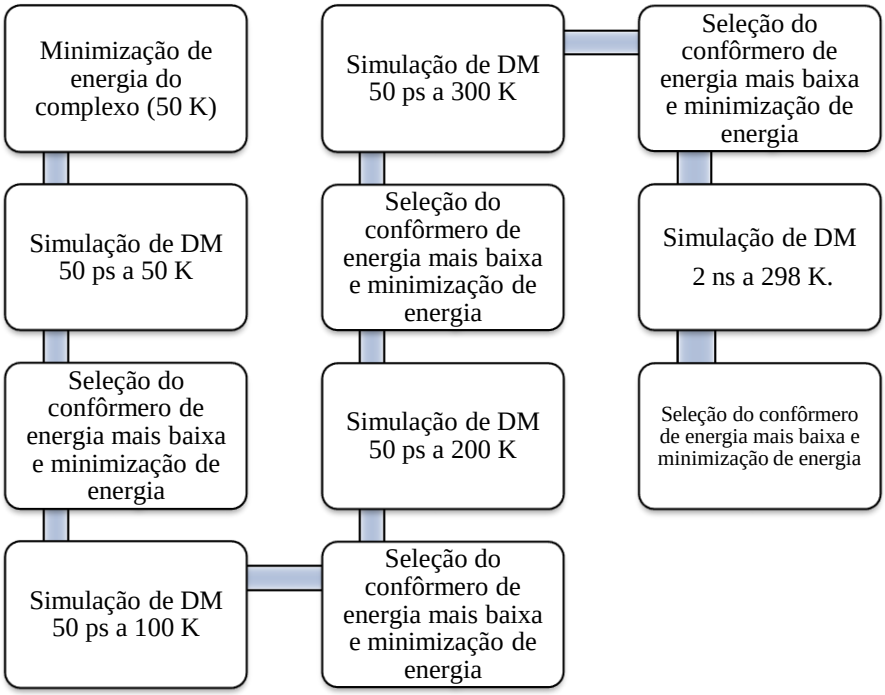


Figura 6. Fluxograma das simulações de dinâmica molecular para a preparação dos complexos provenientes do **PDB**.

3.5.11.3. Dinâmica de relaxamento dos complexos MMP-2 e MMP-9/conformação bioativa/descritores de cada ligante

A conformação bioativa de cada composto do conjunto de treinamento e seus respectivos descritores de ocupação, GCOD, foram ancorados manualmente no sítio catalítico do modelo molecular 3D da MMP-2 e da MMP-9, provenientes das simulações de DM descritas no item anterior. As coordenadas dos inibidores hidroxamatos 7-MR (**PDB ID** 2OW1, TOCHOWICZ *et al*., 2007) e SC-74020 (**PDB ID**1HOV, FENG *et al*. 2002) foram utilizadas como referência no processo de ancoramento considerando o melhor modelo/alinhamento para o conjunto de treinamento. Os complexos tiveram, então, a energia minimizada (MOLSIM 3.2; DOHERTY, 2002) empregando os métodos de declive máximo e de gradientes conjugados a 298 K e foram submetidos a simulações de DM de relaxamento (simulações curtas, de 50 ps, a 298 K) para melhor acomodar as conformações bioativas nos sítios de interação das MMP. Os confôrmeros de energia mais baixa dos complexos foram selecionados do PAC, minimizados, para posterior análise e visualização.

3.5.12. Planejamento, construção e otimização de novos inibidores das enzimas MMP considerando os modelos finais de QSAR-4D IR

Novos inibidores para as enzimas MMP-2 e MMP-9 foram planejados/previstos a partir dos melhores modelos obtidos nos estudos de QSAR-4D IR. O procedimento realizado para a construção e previsão de atividade biológica destas estruturas foi o mesmo descrito nos itens **3.2**, **3.3** e **3.5.7** em Material e Métodos.

3.6. Construção e otimização da estrutura tridimensional do 4-NC

O modelo molecular 3D do 4-NC foi construído, na forma protonada, com o módulo Build do software Hyperchem8.0 (Hypecube, 2007). Empregaram-se, como referência, as coordenadas cartesianas do ácido dodecanóico e do catecol, depositadas no banco de dados **PDB** (BERMAN *et al*., 2000), com os códigos de entrada 1EK4 (OLSEN *et al*., 2001; resolução 1,85 Å) e 1XEP (GRAVES, BRENK, SHOICHET, 2005; resolução 1,55 Å), respectivamente. A isomeria geométrica da molécula foi definida como E (PUBCHEM, 2014). Além disso, a molécula possui um centro quiral. O estudo deve ser conduzido com os dois enantiômeros, já que não há informações disponíveis sobre qual seria o enântiomero predominante na natureza. No entanto, inicialmente, os resultados apresentados foram obtidos apenas para o enantiômero *S*. A geometria foi otimizada em campo de força empírico MM+ (ALLINGER, 1977) e as cargas atômicas parciais foram calculadas com o método semi-empírico AM1 (DEWAR *et al*., 1985), implementado no Hyperchem 8.0 (Hypecube, 2007).

3.7. Previsão da atividade do 4-NC a partir dos modelos QSAR-4D IR estatisticamente validados

O valor de atividade biológica foi calculado para a estrutura do 4-NC a partir dos modelos QSAR-4D IR gerados com o conjunto de treinamento com atividade inibitória frente às enzimas MMP-2 e MMP-9. O procedimento foi realizado utilizando-se as mesmas etapas descritas no item **3.5.7**.

3.8. Determinação do estado de protonação do 4-NC

Para dar início aos estudos de ancoramento, foram realizadas análises da micromolécula, 4-NC, no programa Marvin Beans 5.8.0 (ChemAxonLtd., 2009), com a finalidade de compreender o comportamento em diferentes valores de pH, tanto no alvo (pH 7,4) como em outros compartimentos orgânicos, como o intestino delgado (pH 8,0), por exemplo. A estrutura foi desenhada no módulo Marvin Sketch e visualizada no módulo Marvin View (pacote Marvin Beans 5.8.0; ChemAxonLtd., 2012). A porcentagem de distribuição das microespécies do 4-NC passíveis de serem formadas em

diferentes valores de pH foram obtidas com a opção *pka plugin*, nas seguintes condições de análise: limite inferior de pH = 0; limite superior de pH = 14,0; intervalo de incremento no pH de 0,2; temperatura de 298 K.

3.9. Estudos de ancoramento molecular (*docking*) e *redocking*

3.9.1. Processamento do modelo molecular 3D das enzimas MMP

O processamento das enzimas MMP utilizadas nos estudos de ancoramento está descrito no item 3.5.11.1 em Material e Métodos.

3.9.2. *Redocking*

Estudo de *redocking* é o processo no qual um composto previamente ligado à biomacromolécula alvo é retirado da estrutura do complexo e, novamente, ancorado no mesmo alvo. É utilizado, frequentemente, para verificar se os parâmetros ou opções utilizados/selecionados no processo de ancoramento molecular estão adequados e são suficientes para reproduzir os dados do complexo original, bem como as interações obtidas experimentalmente (MORRIS *et al*., 2009). Um problema importante relacionado aos estudos de ancoramento envolvendo metaloproteinases é a determinação dos parâmetros para os íons de zinco. No software Autodocktools4 (MORRIS *et al*., 2009) os átomos são parametrizados em campo de força AMBER94 (WEINER *et al*., 1984). No campo de força AMBER94 não há parâmetro adequado disponível para os átomos de zinco. A parametrização do metal é necessáriapara se determinar os valores de raio (r) e a profundidade do poço de potencial (ε), a fim de se gerar a caixa virtual e os demais arquivos com os parâmetros para o ancoramento molecular (HU, SHELVER, 2003). Neste estudo, utilizaram-se os valores de r e de ε propostos pelos autores Hu e Shelver (2003), que avaliaram a otimização destes valores para 14 complexos envolvendo MMP.

3.9.2.1. Estudos de *redocking* do complexo MMP-2/SC-74020

Com o objetivo de verificar se os parâmetros utilizados no processo de ancoramento do 4-NC no sítio catalítico da enzima MMP-2 seriam capazes de reproduzir os dados de um complexo obtido experimentalmente, o inibidor hidroxamato (SC-74020) foi extraído do sítio catalítico (**PDB ID** 1HOV; FENG *et al*., 2002), na conformação experimental, e otimizado em método de campo de força empírico (MM+; Allinger, 1977). Empregou-se também o software AutodockTools4 (MORRIS *et al*., 2009) para a atribuição de cargas atômicas Gasteiger e hidrogênios polares ao modelo da biomacromolécula alvo e do ligante. O ligante foi ancorado novamente no sítio catalítico correspondente, utilizando o programa Autodock 4 (HUEY *et al*., 2007). Para desenvolver o procedimento de ancoramento molecular, consideraram-se todos os

resíduos de aminoácidos que compõem o sítio catalítico da MMP-2. O tamanho da grade/caixa virtual 3D, centrada no átomo de zinco catalítico do modelo, foi de 90, 90, 90 (x, y, z) em 6,92 x 22,92 x 19,58 pontos da grade (x, y, z), e o espaço considerado entre os pontos da grade foi 0,38 Å. O tamanho da grade considerou os aminoácidos que compõem o sítio catalítico e propiciou espaço suficiente para o movimento rotacional e translacional do ligante. Empregou-se o algoritmo genético Lamarckiano nas simulações de ancoramento e o algoritmo Solis Wets foi utilizado para a otimização local subsequente. A população foi composta por 150 indivíduos, o número máximo de avaliações de energia foi definido em 2.500.000, o número máximo de gerações em 270.000 e o número de corridas em 50. Taxa de mutação máxima em 0,02 foi estabelecida, elitismo de 1, e a taxa de crossover foi de 0,8. Os parâmetros atribuídos aos íons de zinco foram r = 0,87 Å, ε = 0,35 kcal/mol, carga dos íons de zinco = +0,95 (HU, SHELVER, 2003). As conformações obtidas foram avaliadas em termos de desvio quadrático médio (RMSD, *root mean squares deviation*) de posições atômicas em relação à conformação do complexo determinada experimentalmente. Os valores de RMSD foram considerados de acordo com os seguintes critérios (DADRASS *et al*., 2004): excelente (valores < 1Å); bom (entre 1,0 – 3,0 Å); regular (entre 3,1- 5,0 Å) e ruim (valores > 5,1 Å).

3.9.2.2. Estudos de *redocking* do complexo MMP-9/7-MR

Para avaliar os parâmetros utilizados no processo de ancoramento do 4-NC no sítio catalítico da enzima MMP-9, extraiu-se o inibidor RO-206-0222 (um derivado pirimidínico) do sítio ativo do cristal (2OVX; TOCHOWICZ *et al*., 2007), disponível no banco de dados **PDB** (BERMAN *et al*., 2000). Empregou-se também o software AutodockTools4 (MORRIS *et al*., 2009) para a atribuição de cargas atômicas Gasteiger e hidrogênios polares ao modelo da biomacromolécula alvo e do ligante. O inibidor foi otimizado em método de campo de força empírico MM+ (ALLINGER, 1977) e foi ancorado novamente no sítio catalítico correspondente utilizando o software Autodock 4 (HUEY *et al*., 2007). Para o ancoramento molecular do ligante à MMP-9, o tamanho da grade ou caixa virtual (*grid*) foi 90, 90, 90 (x, y, z) em 28,01 x 9,00 x 46,79 pontos da grade. O espaço considerado entre os pontos da grade foi 0,38 Å. Os parâmetros empregados no processo de *redocking* e os critérios considerados na avaliação dos resultados foram os mesmos descritos para a MMP-2 no item **3.9.2.1**.

3.9.3. Estudos de ancoramento molecular do 4-NC no sítio catalítico das MMP-2 e MMP-9

Os estudos de ancoramento molecular foram empregados com o intuito de encontrar o confôrmero do 4-NC com orientação mais adequada ao sítio catalítico das enzimas MMP. Estas simulações foram realizadas com o programa Autodock4 (HUEY e*t al*., 2007) e os confôrmeros obtidos foram utilizados como estrutura de entrada (*input*) nas simulações de DM.

Empregou-se o software AutodockTools 4 (MORRIS *et al*., 2009) para a atribuição de cargas atômicas Gasteiger e hidrogênios polares ao modelo da biomacromolécula alvo e do ligante. Mantiveram-se as ligações rotáveis (10) da estrutura do 4-NC sem quaisquer restrições durante o processo de ancoramento. Ainda, consideraram-se todos os resíduos de aminoácido que compõem o sítio catalítico das enzimas. Na estrutura da MMP-2, o tamanho da grade ou caixa virtual (*grid*), centrada no átomo de zinco catalítico do modelo molecular do complexo, corresponde a 90, 90, 90 (x, y, z) em 6,92 x 22,92 x 19,58 pontos da grade (x, y, z). O espaço considerado entre os pontos da grade foi de 0,38 Å. O tamanho da caixa utilizada foi cerca de duas vezes maior que o tamanho do ligante. Para o ancoramento do ligante à enzima MMP-9, o tamanho da grade foi de 90, 90, 90 (x, y, z) em 28,01 x 9,00 x 46,79 (x, y, z) pontos da grade.

O algorítimo genético Lamarckiano foi escolhido na busca dos melhores confôrmeros (DADRAS *et al*., 2004). A população foi composta por 150 indivíduos, o número máximo de avaliações de energia definido em 2.500.000, o número máximo de gerações em 270.000 e o número de corridas em 50. A taxa de mutação máxima considerada foi 0,02, com elitismo de 1 e a taxa de crossover de 0,8. Os parâmetros atribuídos aos íons de zinco foram r = 0,87 Å, ε = 0,35 kcal/mol, carga dos íons de zinco = +0,95 (HU, SHELVER, 2003).

A estrutura do 4-NC, previamente processada no programa Hyperchem 8.0 (HyperChem, 2007) (item **3.6**), foi selecionada para dar início ao processo. O ancoramento molecular foi realizado trinta vezes, em condições idênticas para cada um dos complexos. Cada corrida gerou 50 confôrmeros, classificados em ordem crescente de energia livre de interação de Gibbs (ΔG). Para a avaliação dos confôrmeros consideraram-se a frequência de orientação das porções da cadeia lateral e do núcleo catecólico do 4-NC no sítio catalítico e a função de ranqueamento (*scoring*) de energia.

3.10. Simulações de dinâmica molecular dos complexos provenientes do processo de ancoramento

O programa MOLSIM 3.2 (DOHERTY, 2002) foi empregado para o processamento das simulações de DM das estruturas tridimensionais dos complexos MMP-2/4-NC e MMP-9/4-NC, previamente obtidos com o software Autodock 4 (HUEY e*t al*., 2007). As estruturas foram minimizadas empregando-se os métodos de declive máximo e de gradientes conjugados. O critério de convergência estabelecido foi de 0,5 kcal/mol. Atribuíram-se massas fictícias de 5.000 u.m.a. aos átomos da cadeia polipeptídica principal a fim de preservar a integridade estrutural dos sistemas moleculares durante as simulações computacionais (TOKARSKI, HOPFINGER, 1997).

O protocolo empregado para as simulações de DM consistiu de um esquema de aquecimento lento *(warming-up*) seguido de simulação mais longa na temperatura do ensaio biológico. Empregaram-se simulações curtas (50 ps; cada passo 1 fs) em temperaturas de 50, 100, 150, 200, 250 e 300 K. Este tipo de esquema é aplicado para gerar o estado de menor energia do complexo de forma mais rápida, permitindo que o sistema supere barreiras energéticas (TOKARSKI, HOPFINGER, 1997).

A conformação de energia mais baixa do complexo à temperatura de 50K foi selecionada, minimizada, e o complexo de energia mais baixa resultante foi o dado de entrada para a próxima simulação de 50 ps a 100 K. O mesmo procedimento foi repetido até à temperatura de 300 K. O confôrmero de energia mais baixa do complexo selecionado em 300 K, após minimização de energia, foi utilizado em simulação de 5ns a 310 K, temperatura do ensaio biológico (Figura 7). O arquivo trajetória foi registrado a cada 20 passos da simulação. Utilizou-se constante dielétrica de 3,5 (SANTOS-FILHO, MISHRA, HOPFINGER, 2001; TOKARSKI, HOPFINGER, 1997). Os PAC foram analisados e os confôrmeros de energia mais baixa foram selecionados, minimizados, e considerados para análise e visualização. Preliminarmente, considerou-se a energia potencial total do sistema molecular para a avaliação dos complexos e orientação do 4-NC no sítio catalítico das MMP.

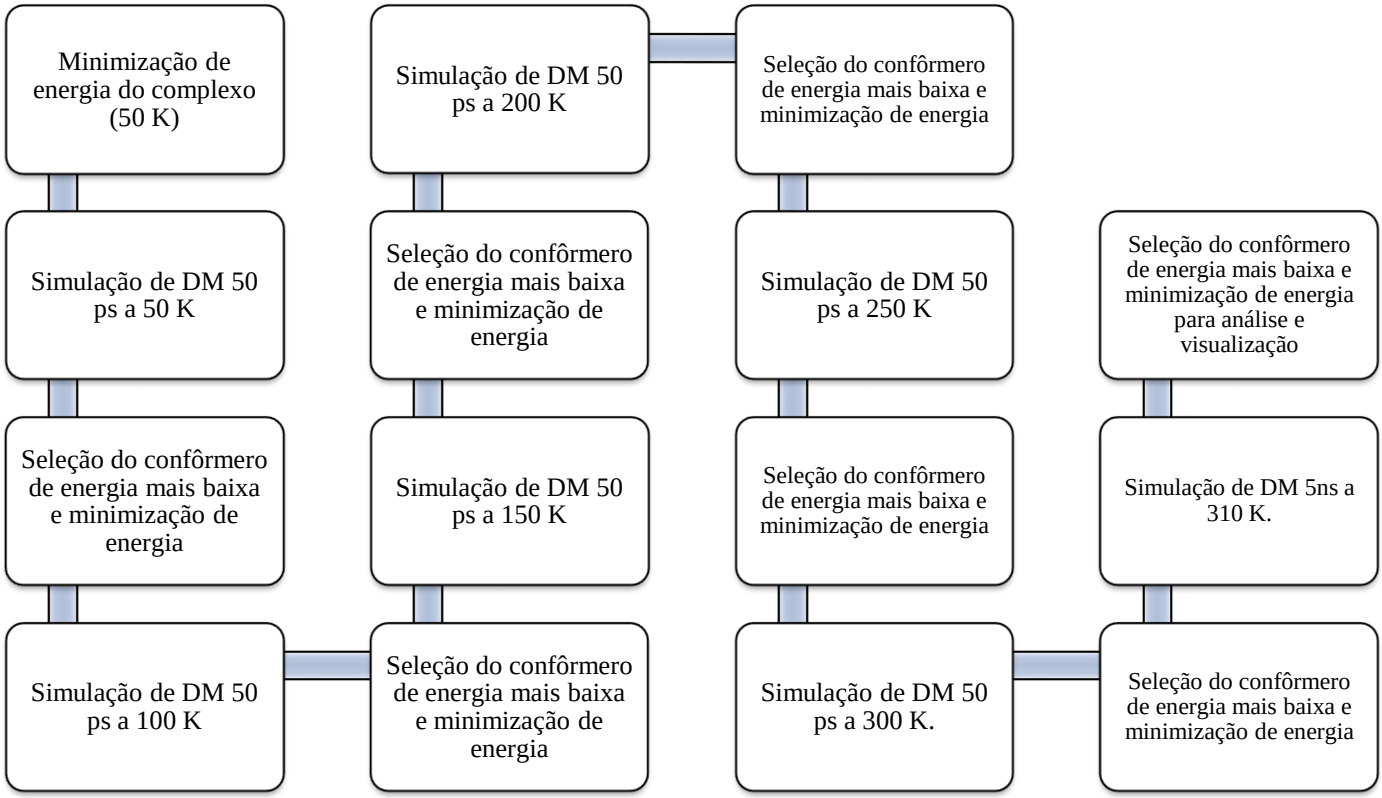


Figura 7. Fluxograma do procedimento utilizado para as simulações de DM dos complexos MMP-2/4-NC e MMP-9/4-NC.

3.11. Preparo do extrato hidroalcoólico de *Pothomorphe umbellata*

As raízes secas de *Pothomorphe umbellata* (p.u.), doadas pelo grupo Anidro do Brasil S/A, foram trituradas em moinhos de martelos e facas, sendo em seguida tamizadas em 40 mesh. O material obtido foi percolado (etanol:água 50%) até o esgotamento da droga, e o extrato foi concentrado e liofilizado (ROPKE, 2003).

3.12. Extração do 4-Nerolidilcatecol

O extrato liofilizado de p.u. foi solubilizado em etanol:água 50% (10 g/L) e esta solução foi extraída com volume igual de hexano por 10 vezes. A fração hexânica foi rotaevaporada e o material resultante foi purificado por cromatografia líquida de alta eficiência (CLAE). A fase móvel consistia de metanol: água (9:1), fluxo em 5 mL/min e detecção em 282 nm. A separação foi realizada à temperatura ambiente em uma coluna semi-preparativa Phenomenex Luna C18 10 µm (10x250mm) (Torrance, CA, EUA) (ROPKE *et al*., 2003). O sistema constituiu-se de: bomba TSP modelo 3500 (Fremont, CA, EUA), loop 500 µL, injetor Rheodyne 7161 (Rohnert Park, CA, USA) e detector LabAlliance 525 (San Jose, CA, EUA). Os solventes utilizados eram de padrão analítico provenientes da Merck (Darmstadt, Alemanha).

3.13. Ensaios de zimografia

Ensaios de zimografia foram também desenvolvidos com intuito de auxiliar na compreensão da possível contribuição da cadeia lateral e do núcleo catecólico da estrutura química do 4-NC em sua atividade inibitória frente às enzimas em estudo. Para estes ensaios foram utilizados os compostos 4-NC, nerolidol, catecol e 1,2-dimetoxibenzeno (Figura 8) e linhagem celular de fibrosarcoma (HT1080) que secreta MMP no meio de cultura (SHEU *et al*., 2003; KATO, YAMASHITA, ISHIKAWA, 2002; OH *et al*., 2001).

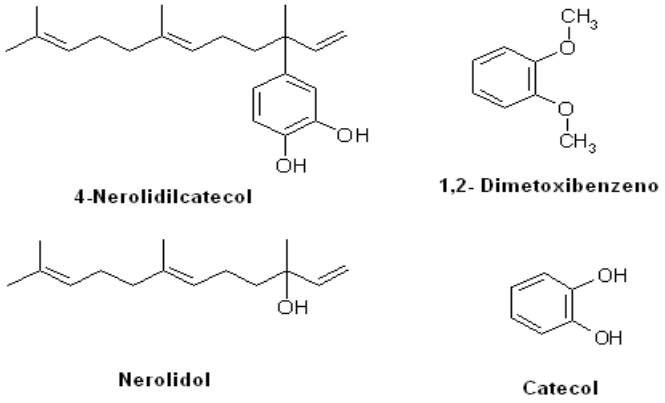


Figura 8. Estrutura química dos compostos avaliados nos ensaios de zimografia.

As células HT1080 utilizadas nestes ensaios foram adquiridas na American Type Culture Collection (ATCC, Manassas, VA, EUA), e apresentavam entre 3 a 7 passagens. Células na ordem de $1x10^6$ foram cultivadas em placas p100 contendo meio de cultura Eagle modificado por Dulbeccos (DMEM- *Dulbecco's modified Eagle's medium*) suplementado com 0,1% de albumina sérica bovina (BSA- *bovine sérum albumin*). As células foram cultivadas durante 48 horas em incubadora a 37°C, atmosfera contendo 5% de CO2 e umidade saturada até que atingissem 70% de confluência. Após esse período, o sobrenadante foi retirado e centrifugado (3000 rpm por 5 minutos) para remoção de células mortas. A concentração de proteína foi determinada pelo método colorimétrico de Lowry (LOWRY, 1951). As amostras de sobrenadantes foram congeladas em alíquotas em freezer -20°C até a realização dos ensaios.

3.13.1. Eletroforese das proteínas no gel de poliacrilamida e gelatina

Para realizar a separação das metaloproteinases provenientes do meio de cultura, prepararam-se géis de poliacrilamida (10%), contendo dodecil sulfato de sódio (SDS, Sodiumdodecyl sulfate) (10%), copolimerizado com gelatina (0,5%).

Em cada poço foi aplicado 20 μg de proteína com tampão de amostra 5X (Tris-HCl 62,5 mM pH 6,8, SDS 2%, glicerol 10%, azul de bromofenol 0,1%, DTT 50mM) e

para cada gel foi utilizado marcador de peso molecular variando de 250-10 kDa (Precision Plus Standard Dual Core -Bio-Rad, Hercules, CA, EUA). As amostras de proteína foram submetidas à eletroforese, em sistema Mini-Protean- III (BioRad, Hercules, CA, EUA), por 30 minutos a 50V, (temperatura ambiente) para compactar as proteínas, e em seguida a 100V por 80 minutos para a separação das mesmas. Ao final da corrida, o gel foi cortado em tiras e, para remover o SDS, foi lavado 2X com 2,5 % Triton X-100, por 15 minutos cada lavagem. As tiras foram incubadas, separadamente, em banho-maria em recipiente fechado à 37°C por 18h com tampão de incubação (Tris-HCl 0,05M, 5mM CaCl2, 5 μM $ZnCl_2$, pH 8,0) juntamente com os compostos 4-NC, nerolidol, catecol e 1,2-dimetoxibenzeno nas concentrações 1,59 mM. Como controle negativo da atividade das metaloproteinases, algumas tiras foram incubadas em tampão de incubação contendo ácido etilenodiamino tetra-acético (EDTA, Ethylenediamine Tetraacetic Acid, um inibidor de MMP) a 0,5 mM e outras, para controle positivo, foram incubadas somente com tampão. Após incubação, os géis foram então corados com 0,5% de Coomassie Brilliant Blue R-250 por 15 min, sob agitação, e descorados com solução aquosa contendo 10% de metanol e 10 % de ácido acético.

A atividade total das MMP-2 e 9 (formas pró e ativa) são reveladas no gel de zimografia em regiões (bandas) sem o corante devido a degradação da gelatina (INOMATA *et al*., 2003).

3.13.2. Análise das imagens

As imagens foram analisadas no software ImageJ 1.44p (Wayne Rasband, National Institute of Health, USA). A atividade das metaloproteinases foi calculada pela diferença de pixels entre o gel corado (considerado como fundo) e a banda que apresenta a cor branca (local de digestão pelas proteinases). Os valores maiores são atribuídos às bandas de maior digestão, portanto maior atividade das enzimas.

3.14. Ensaios de voltametria

O comportamento redox de uma solução tampão contendo cloreto de zinco e 4-NC foi investigado por voltametria cíclica e voltametria de pulso diferencial para avaliar a possível formação de complexo entre íons zinco (Zn^{2+}) e o 4-NC. Utilizou-se, para aquisição e tratamento dos dados, um potenciostato PGStat101 da EcoChemie (Autolab) e um microcomputador operando com software NOVA versão 1.10.4.

Empregou-se também um sistema convencional de três eletrodos composto por um eletrodo de carbono vítreo (diâmetro de 3 mm) como eletrodo trabalho, um eletrodo

de Ag/AgCl, $KCl_{(sat)}$ usado como eletrodo de referência e um fio de platina usado como eletrodo auxiliar ou contra-eletrodo em cela contendo 20 mL de solução. A superfície do eletrodo de carbono vítreo foi polida em feltro metalográfico com suspensão de diamante de granulometria 1 e 0,1 µm, respectivamente, nessa ordem. Após este procedimento, o eletrodo foi lavado com água Mili-Q por 30 segundos. As condições experimentais para a para a voltametria de pulso diferencial foram: amplitude do pulso: 50 mV, largura de pulso de 50 ms, degrau de potencial de 2 mV e intervalo de pulso de 400 ms, o que resultaem uma velocidade de varredura final de 5mV s^{-1}. Os voltamogramas cíclicos foram obtidos no intervalo de $-0,4 \leq E_{aplicado} \leq +1,0$ V, usando velocidade de varredura de 100 mVs^{-1} com degrau de potencial de 2 mV e foram apresentados sem qualquer tratamento matemático ou correções de linha de base.

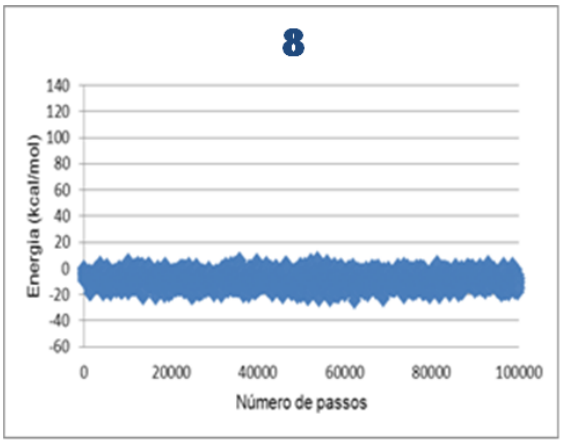
Os experimentos foram realizados no Departamento de Química Fundamental do Instituto de Química da USP, em colaboração com o Profa. Dra. Silvia Helena Pires Serrano e o Dr. Rafael Martos Buoro. Três soluções foram caracterizadas eletroquimicamente para comparação: (1) solução tampão contendo Tris-HCl, $ZnCl_2$ e $CaCl_2$, (2) solução tampão contendo Tris-HCl, $ZnCl_2$, $CaCl_2$ e 4-NC e (3) solução contendo Tris-HCl, $CaCl_2$ e 4-NC. Todas soluções apresentavam pH 8,0 e as concentrações de seus constituintes foram as mesmas utilizadas nos ensaios de zimografia (item **3.13**).

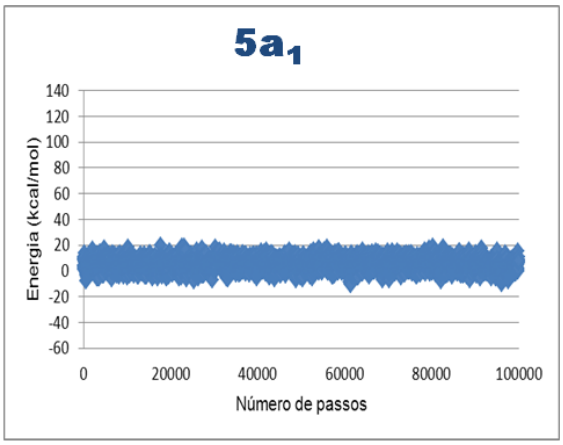
4. RESULTADOS E DISCUSSÃO

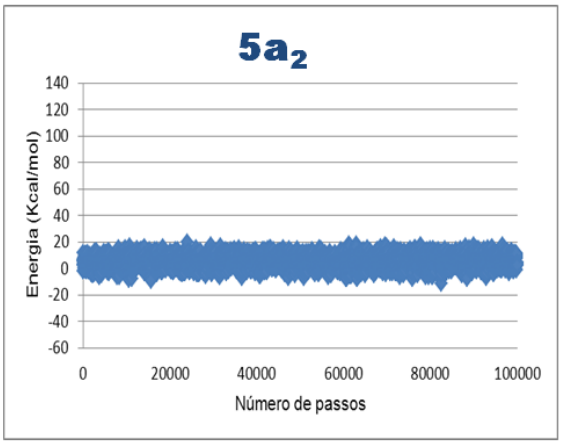
Os modelos moleculares 3D dos ligantes, previamente construídos e otimizados (HyperChem 8.0, Hypercube, Inc. 2007; MOLSIM 3.2, DOHERTY, 2002), foram utilizados como estruturas de partida para o desenvolvimento das simulações computacionais de DM (1 ns a 298 K), para gerar o PAC de cada um dos 64 inibidores de MMP.

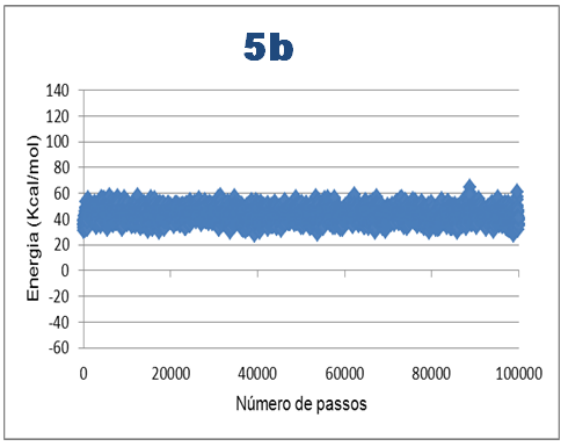
4.1. Simulações de Dinâmica Molecular dos inibidores de MMP

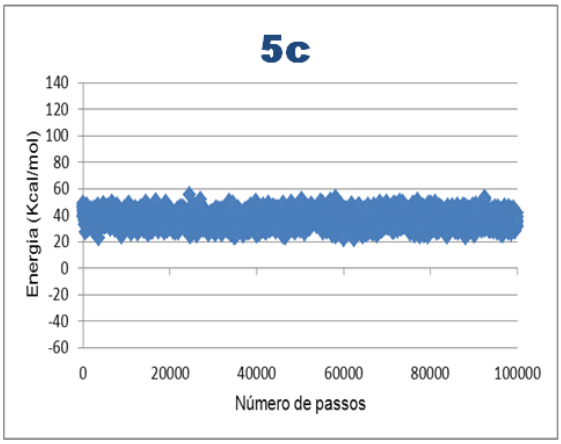
As regiões mais estáveis do perfil de 1 ns obtidas nas simulações de DM dos ligantes foram selecionadas para a aplicação do formalismo QSAR-4D IR e, também, para a seleção do confôrmero de energia mais baixa de cada inibidor, que após minimização de energia, foi utilizado nos cálculos das variáveis independentes para a análise exploratória de dados (HCA e PCA). A porção selecionada do PAC (100 ps de 1ns) para cada um dos 64 inibidores está apresentada na Figura 9. A energia, no eixo y, corresponde à soma dos parâmetros termodinâmicos referentes às contribuições de energia intramolecular dos ligantes (E_{est}, E_{ang}, E_{tors}, $E_{1,4}$, E_{vdW} e E_{el}), exceto energia intramolecular de ligação de hidrogênio (E_{Hb}) e de solvatação (E_{solv}). Estas duas contribuições são consideradas nos valores de energia potencial total (E_T) de cada ligante, que foram calculados para os confôrmeros de energia mais baixa selecionados do PAC, utilizados na análise exploratória.











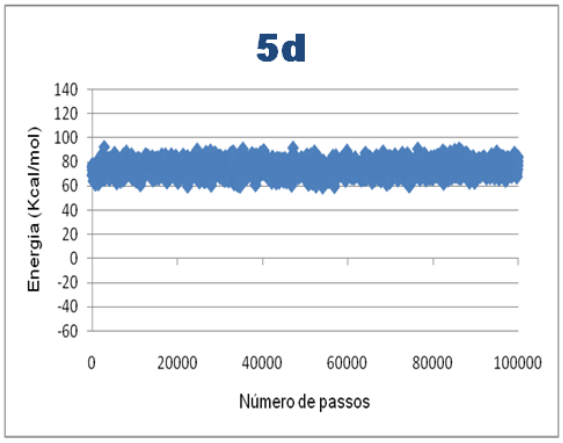
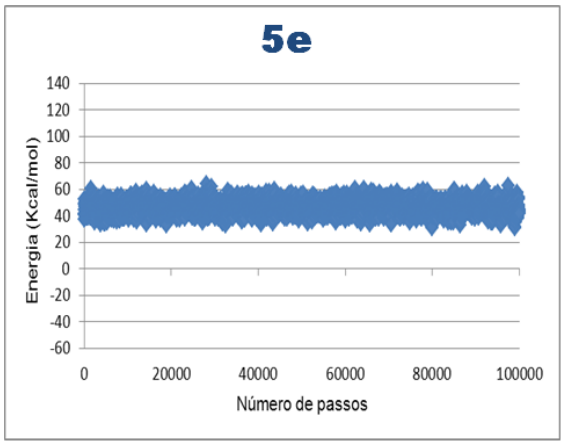
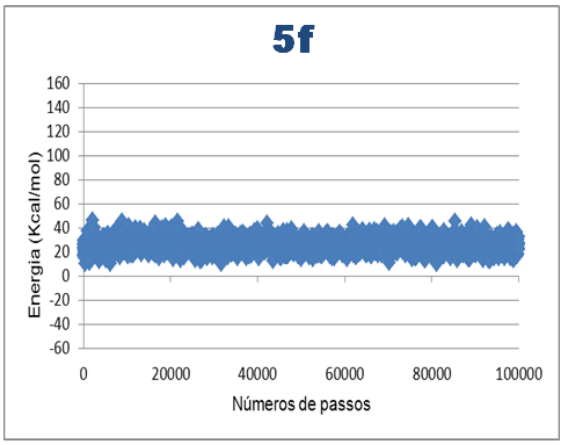
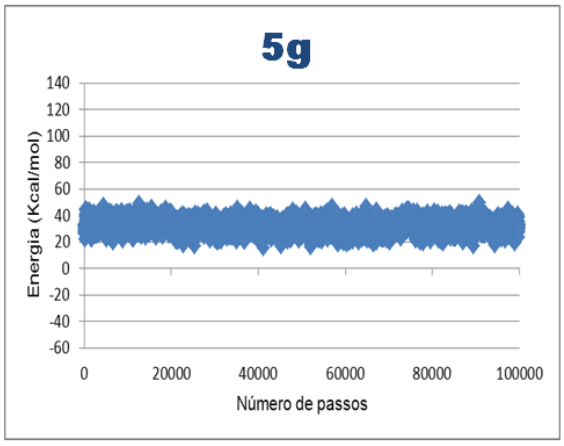
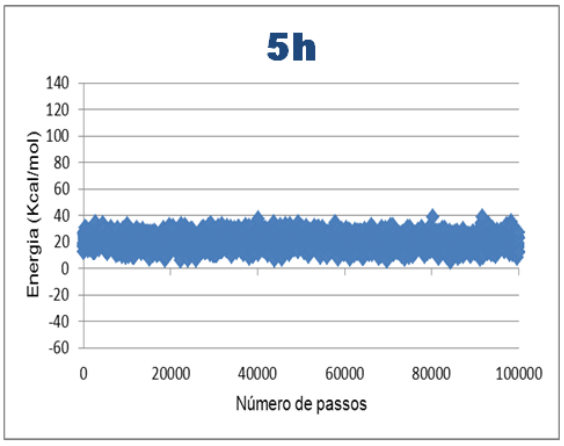


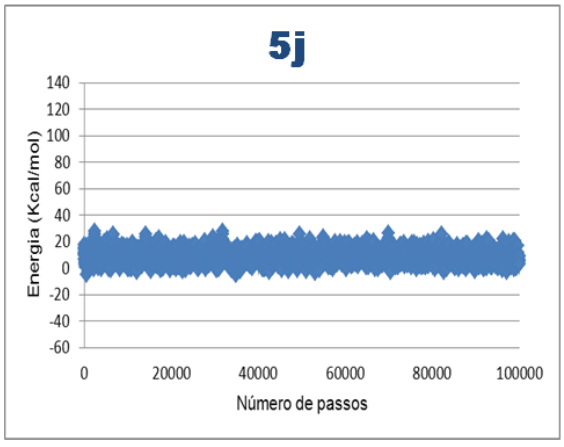
Figura 9. Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1 ns a 298K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











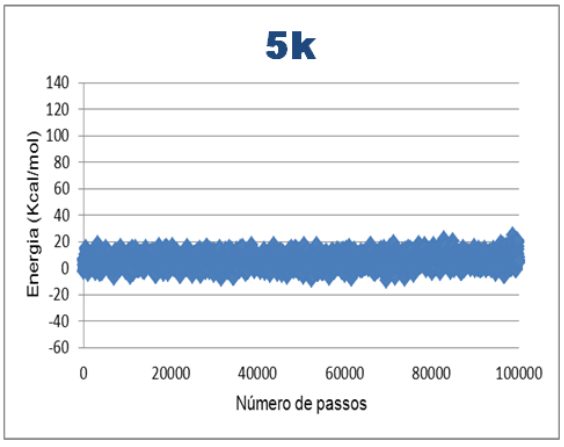
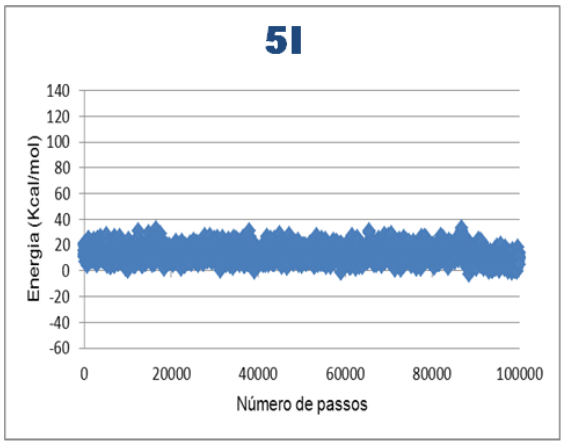
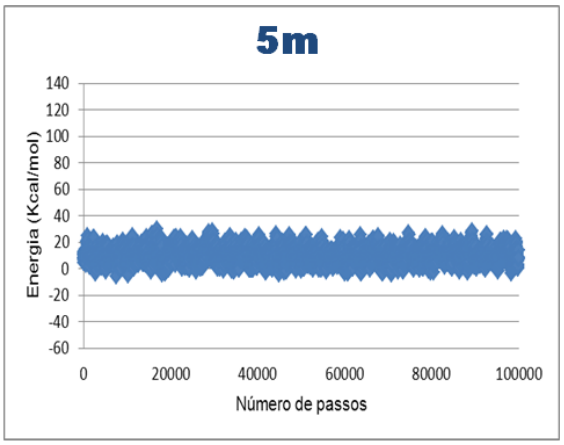
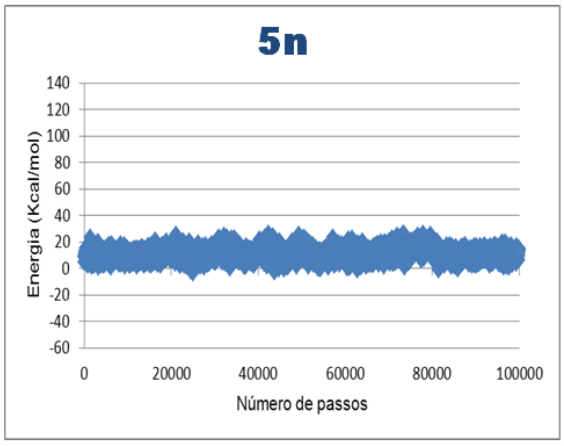
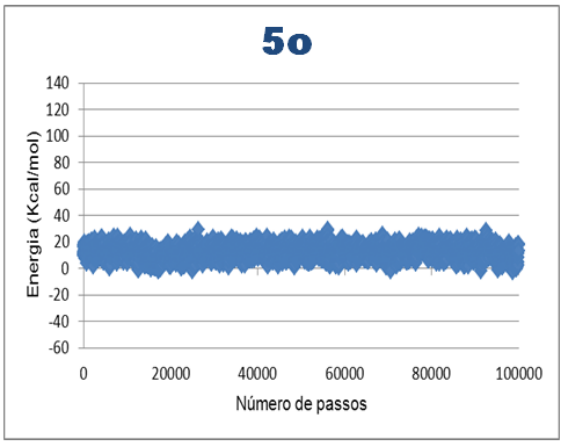


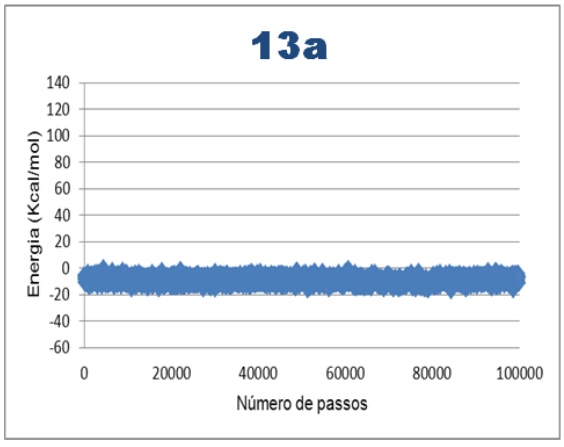
Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











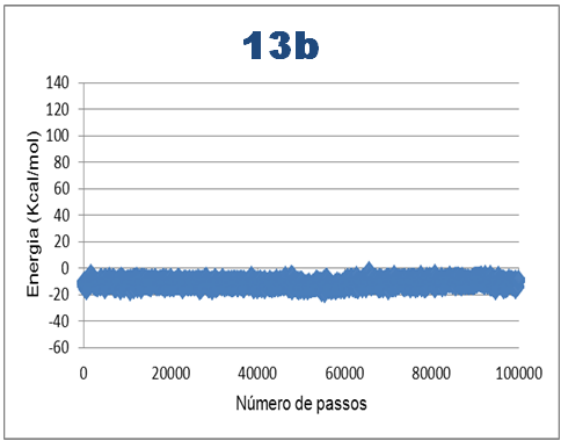
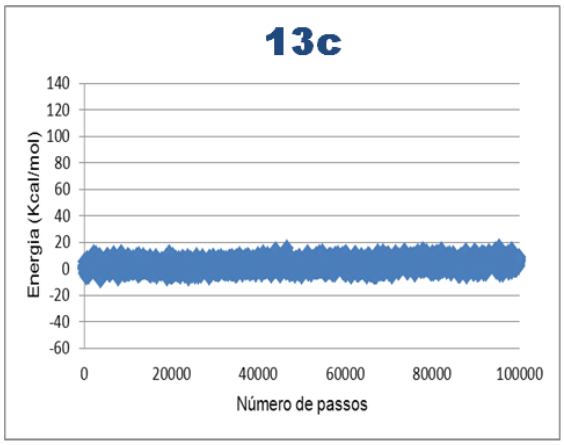
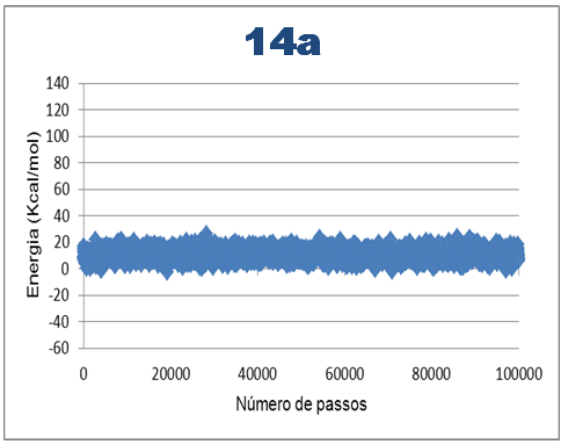
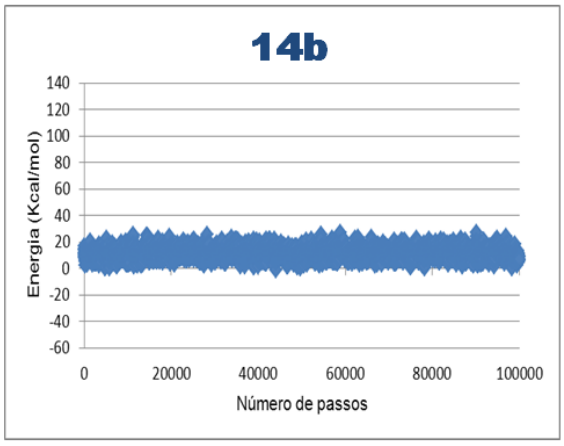
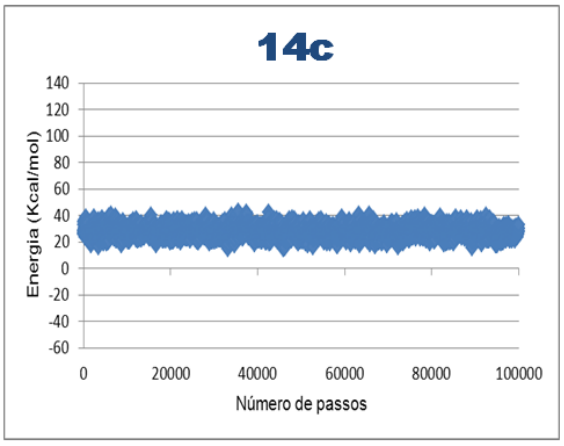


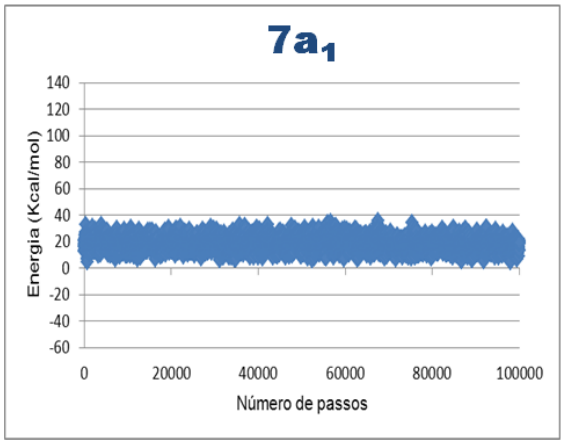
Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











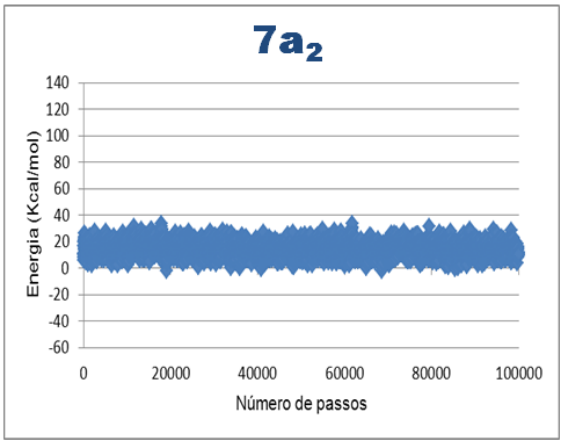
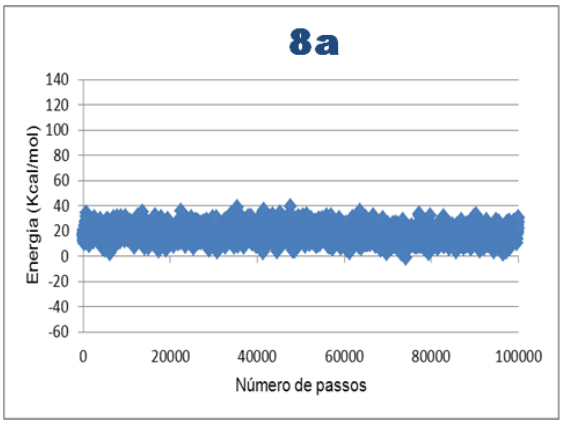
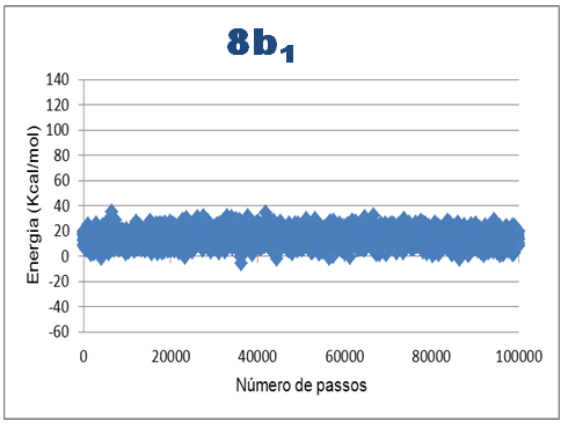
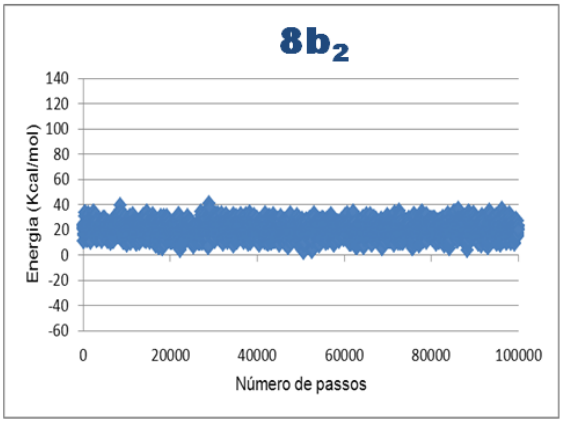
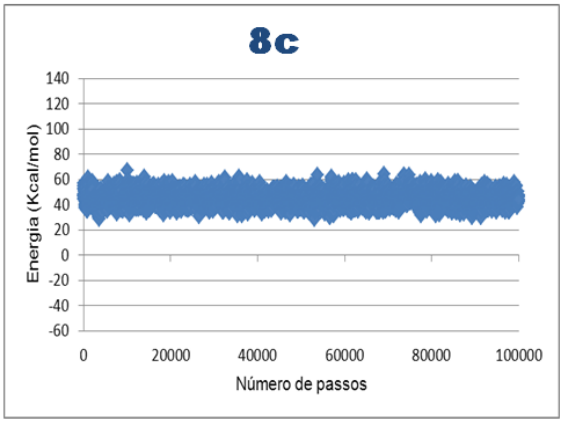


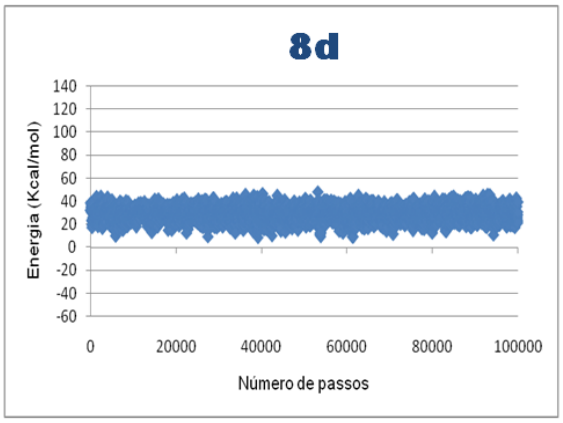
Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











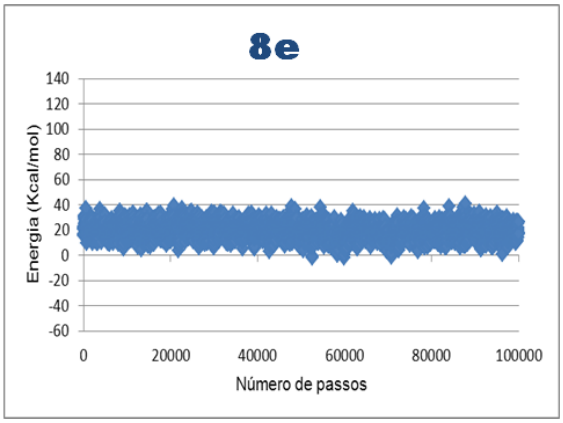
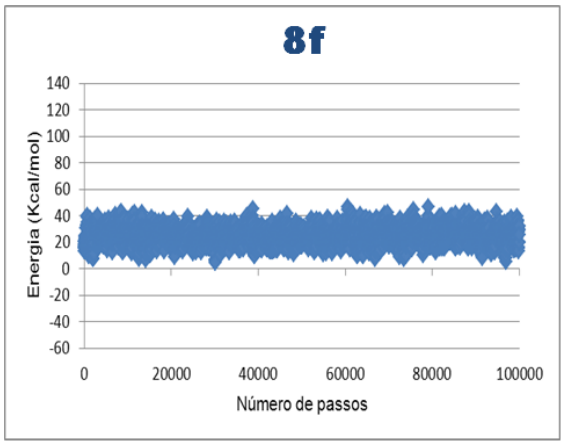
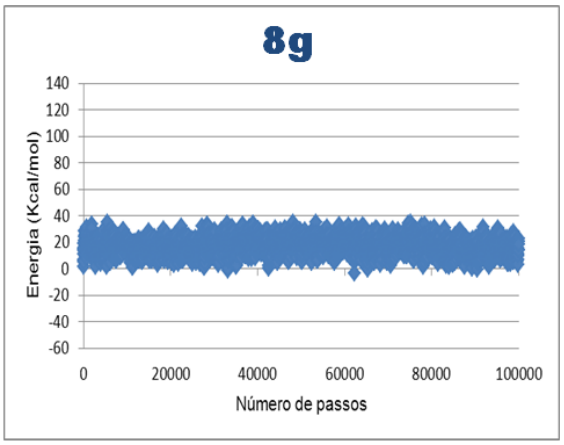
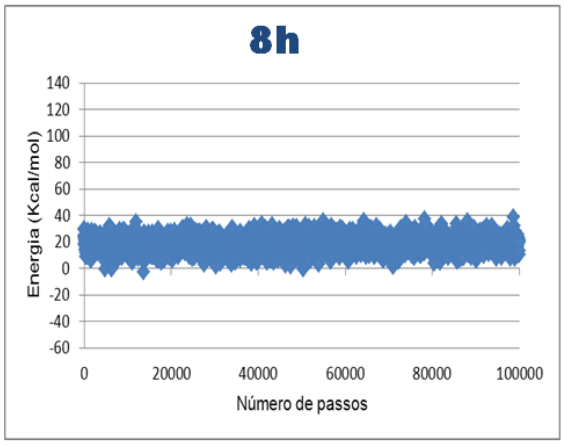
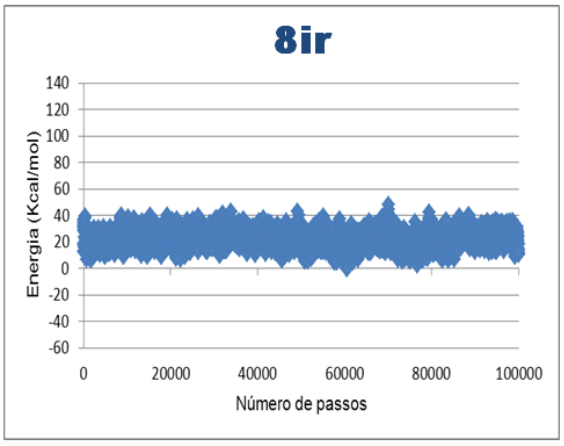


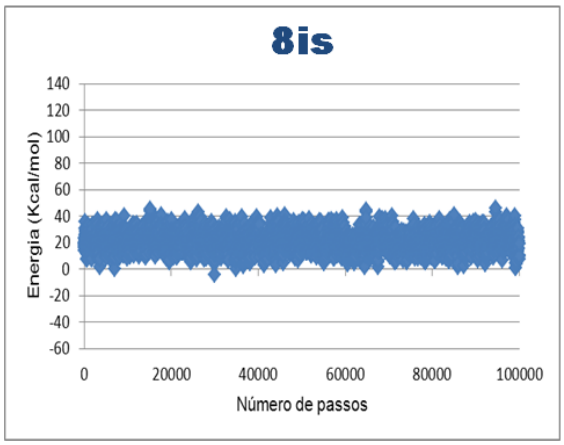
Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











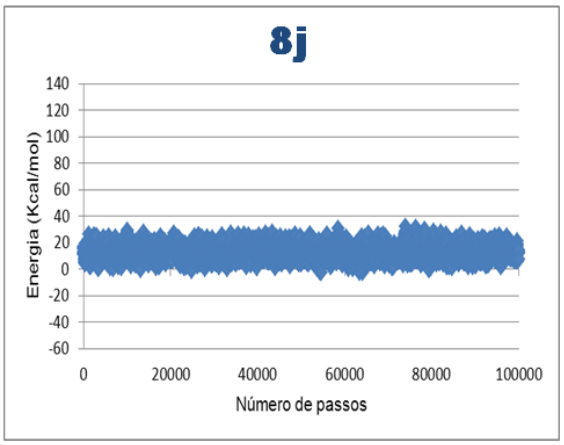
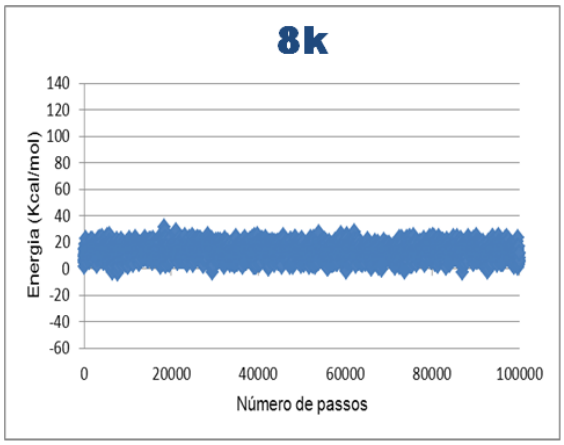
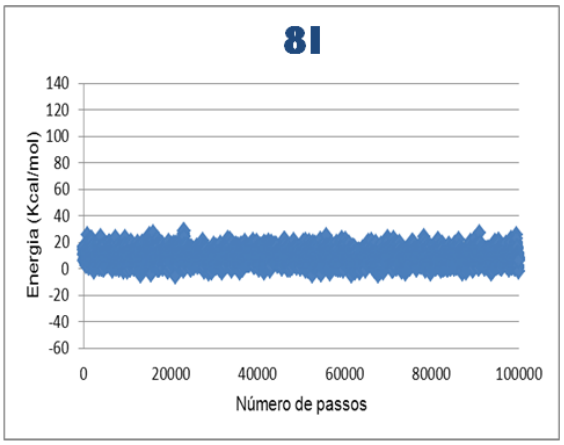
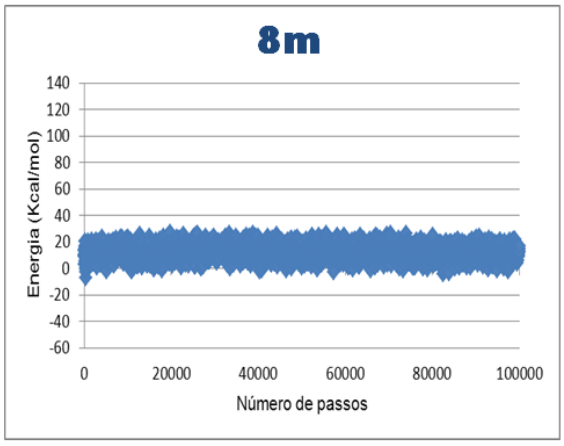
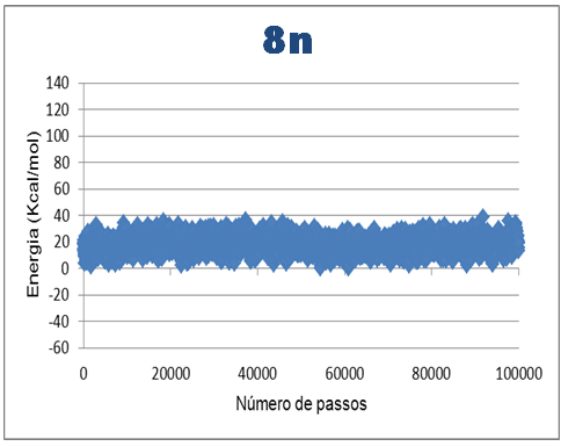


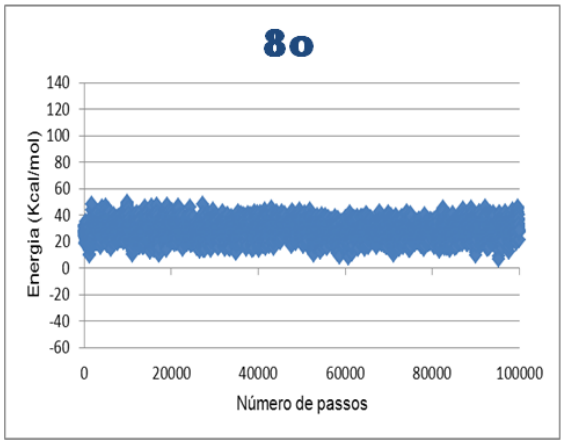
Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











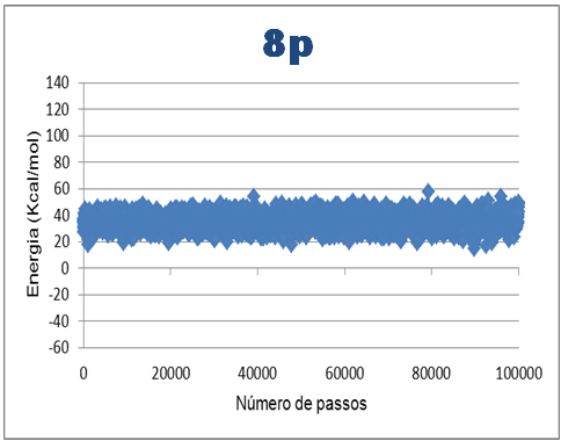
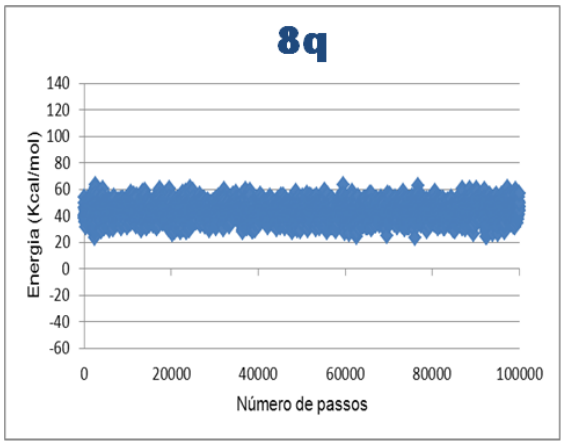
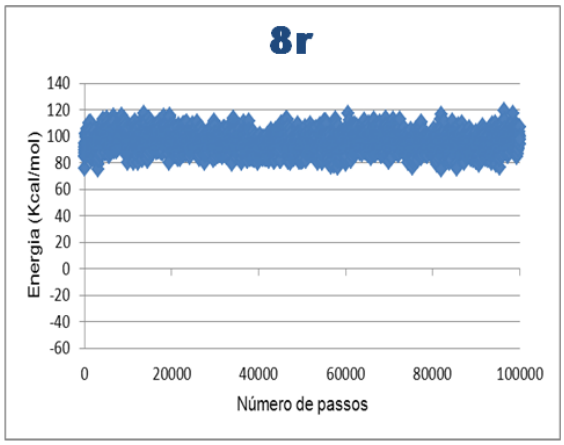
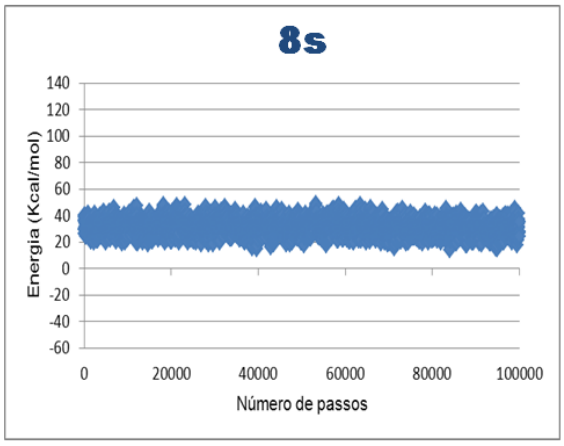
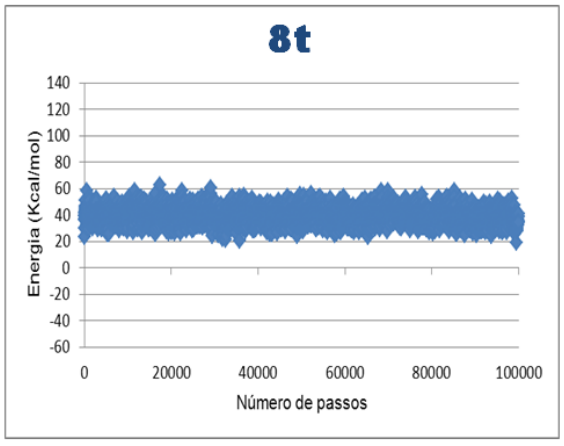


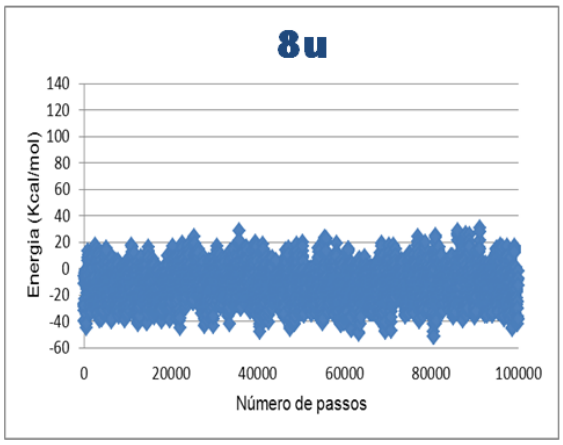
Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











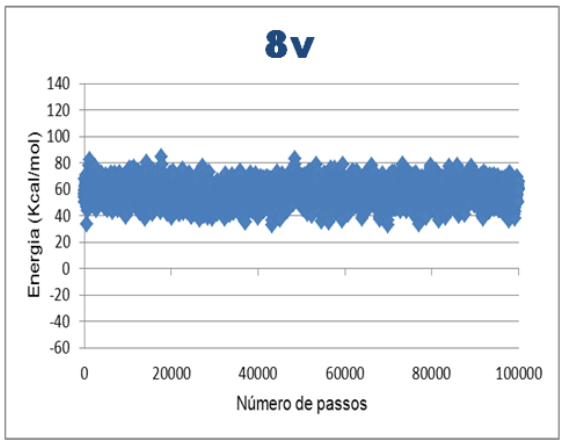
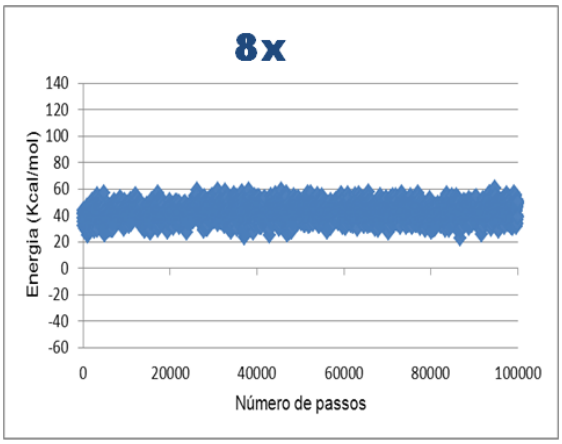
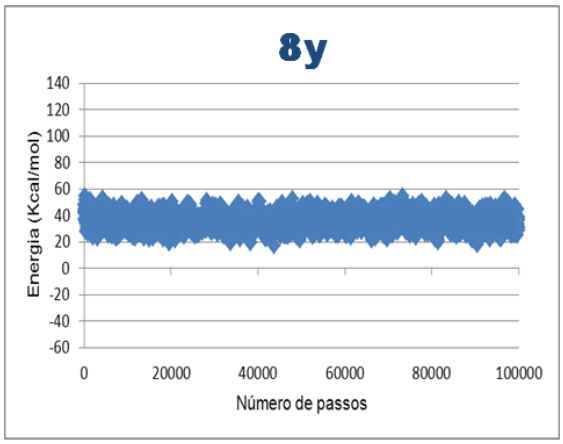
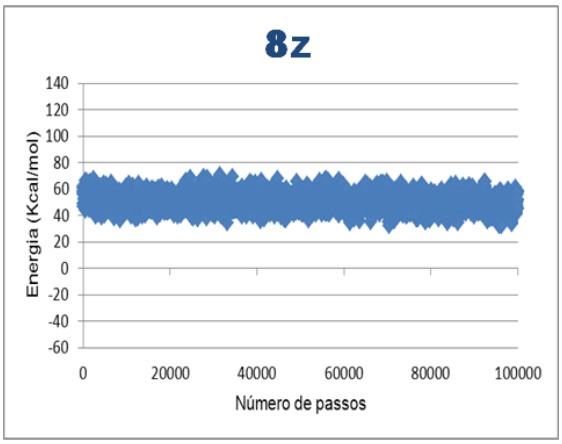
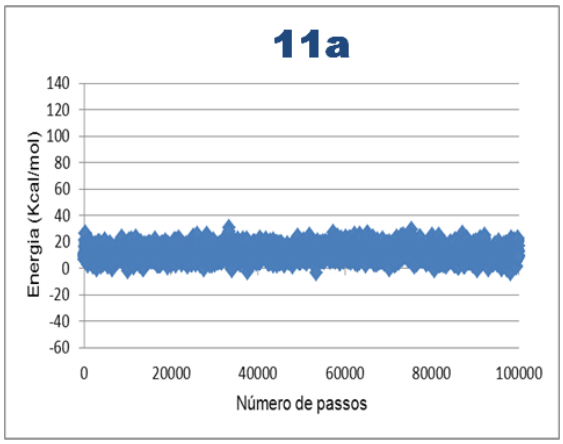


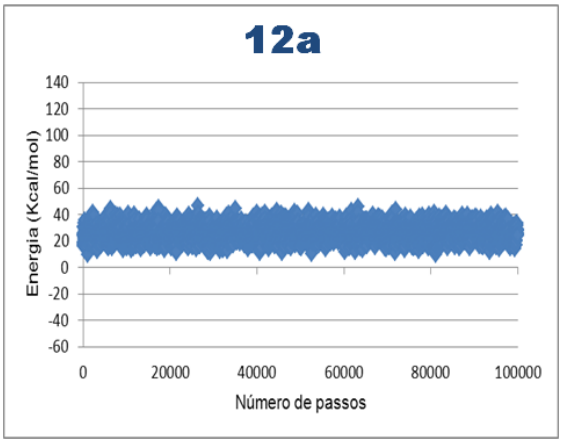
Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











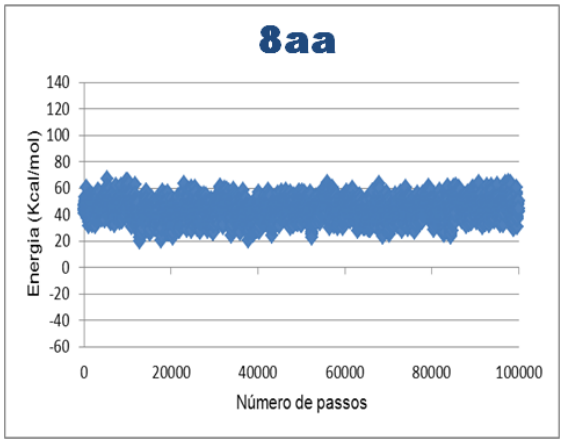
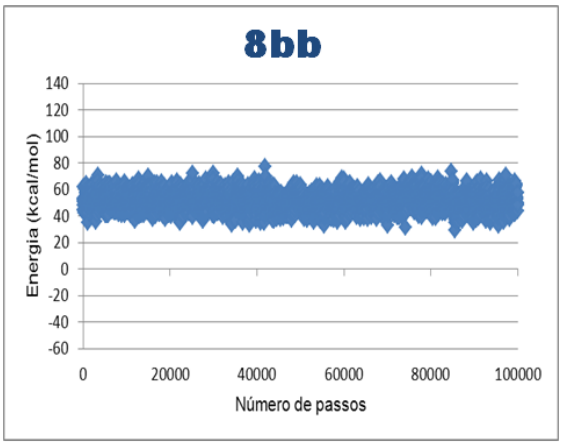
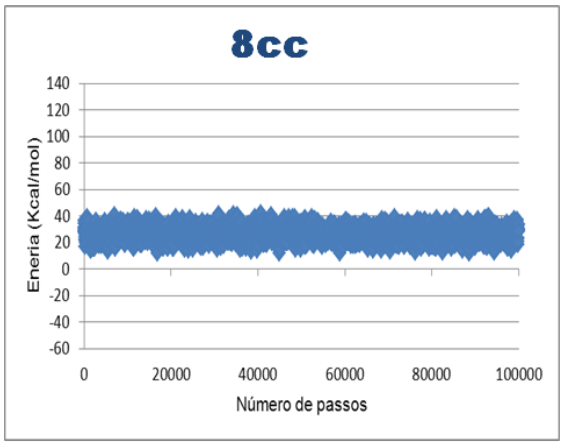
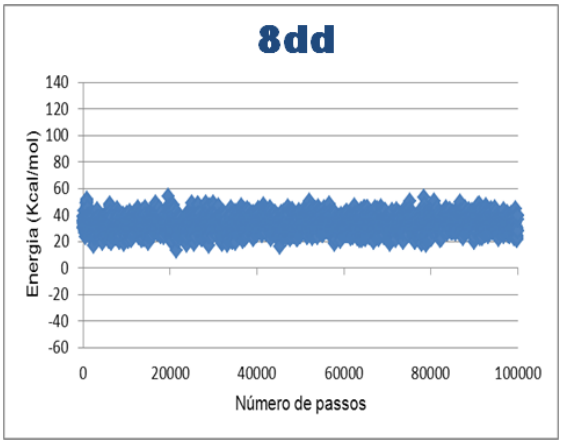
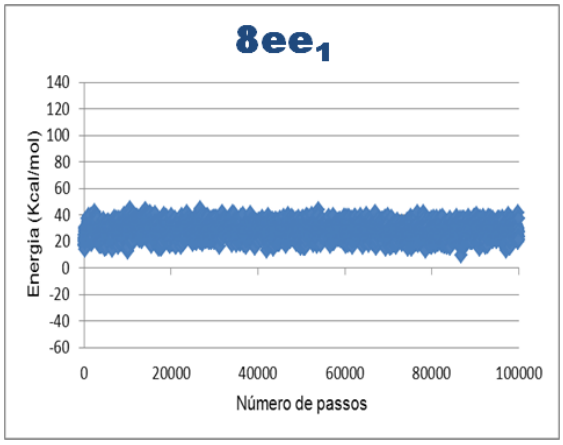


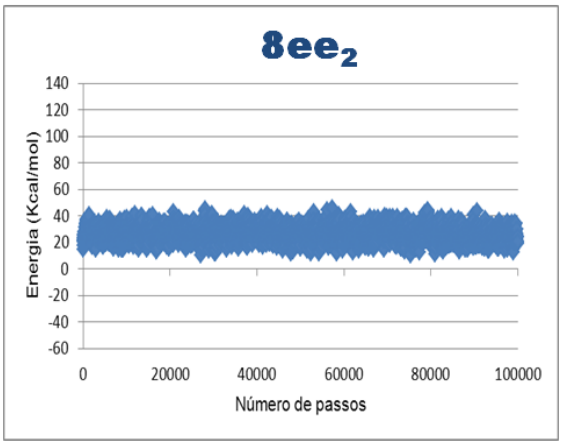
Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.











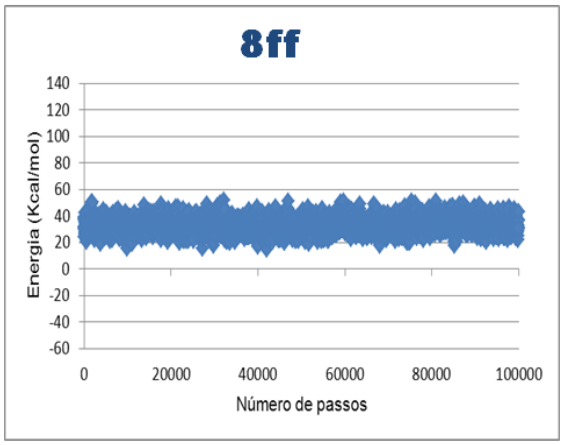


Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.

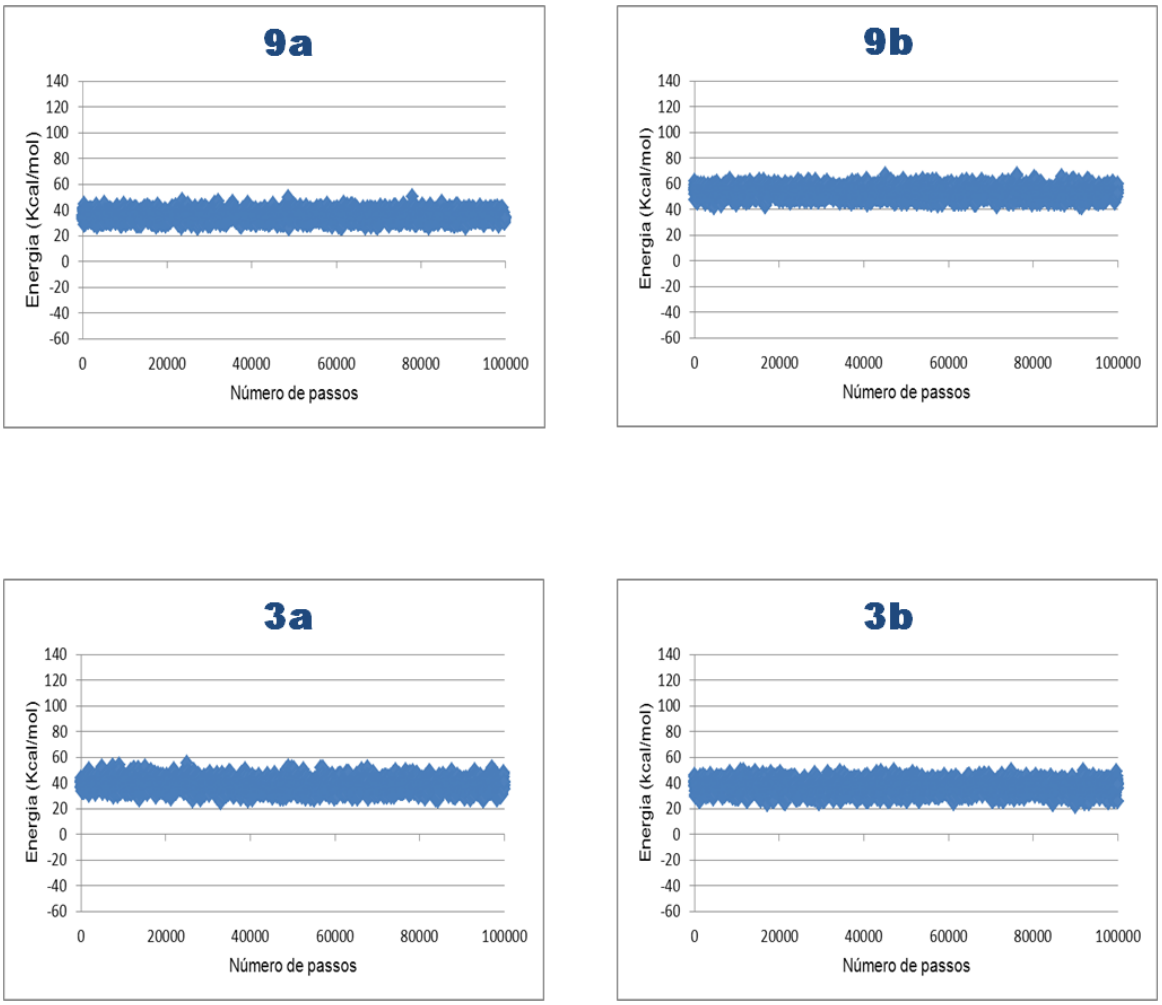


Figura 9 (cont.). Região de equilibração do PAC (100 ps) selecionada da simulação de DM (1ns a 298 K) de cada inibidor para ser utilizada na aplicação do formalismo QSAR-4D IR.

4.2. Análise exploratória de dados (HCA e PCA)

Como já mencionado, a partir da região do PAC apresentada na Figura 9, selecionou-se o confôrmero de energia mais baixa para cada um dos inibidores. Estes confôrmeros, após minimização de energia (MOLSIM 3.2; DOHERTY, 2002), foram utilizados para o cálculo de propriedades moleculares de natureza distinta (eletrônica, estrutural, lipofílica, estérica, topológica) com diferentes programas e métodos computacionais, listados na Tabela 3.

4.2.1. Cálculo de propriedades moleculares e seleção de variáveis

Calcularam-se 49 descritores ou propriedades moleculares (variáveis independentes), que estão apresentados na tabela 1A (Anexo A). A seleção dos descritores considerados na construção das tabelas utilizadas como dados de entrada para os estudos de HCA e PCA foi realizada com base em dois critérios: o coeficiente de

correlação linear de Pearson, em relação à variável dependente (valores de pCI_{50}), e a inspeção visual dos gráficos do dado biológico *versus* cada propriedade calculada. Neste estudo, descritores considerados adequados foram aqueles que apresentaram valores de *R*> |0,25| e distribuição/dispersão em relação ao dado biológico com tendência linear (gráficos de dispersão/distribuição em relação aos valores de pCI_{50}), pois o programa Pirouette 3.11 (Infometrix Inc, 1990-2003) considera somente função linear no tratamento dos dados. Os gráficos dos descritores calculados (eixo x) *versus* o dado biológico para MMP-2 e MMP-9 (eixo y), bem como os valores de *R* para cada descritor estão apresentados nas Figuras 10 e 11.

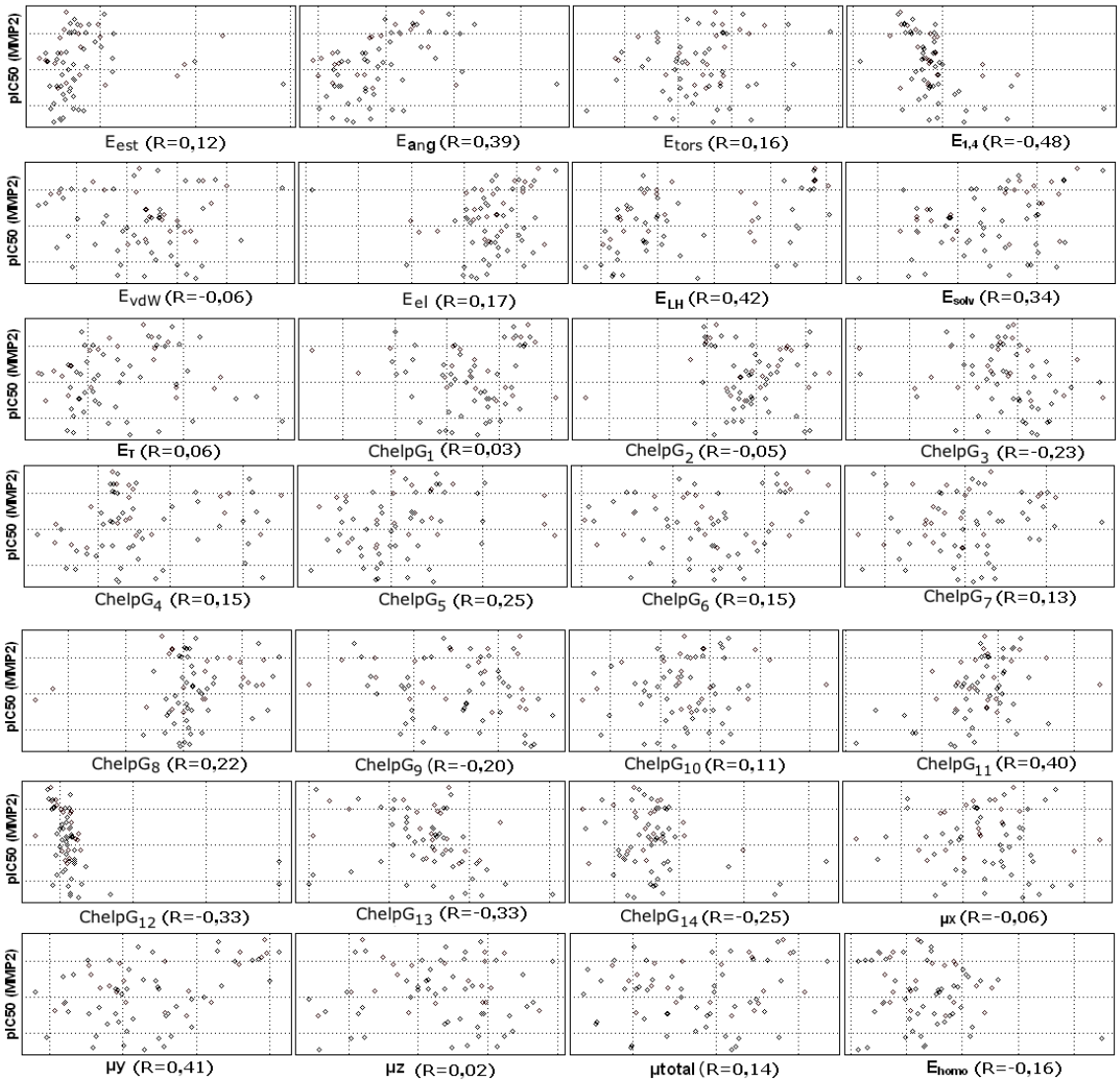


Figura 10. Gráficos de dispersão dos valores de cada descritor *versus* o dado biológico (valores de pCI_{50} para MMP-2), com os respectivos valores de *R*.

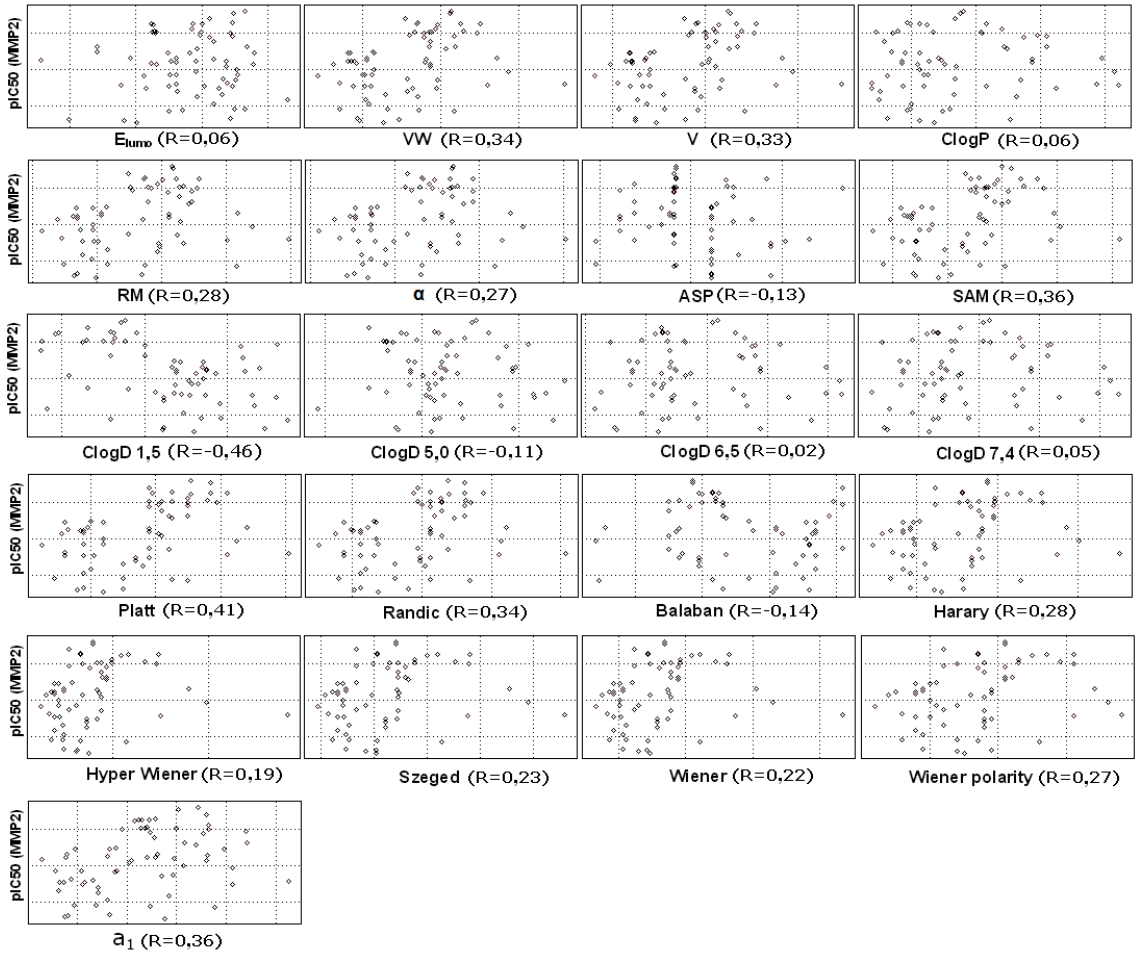


Figura 10 (cont.). Gráficos de dispersão dos valores de cada descritor *versus* o dado biológico (valores de pCI_{50} para MMP-2), com os respectivos valores de *R*.

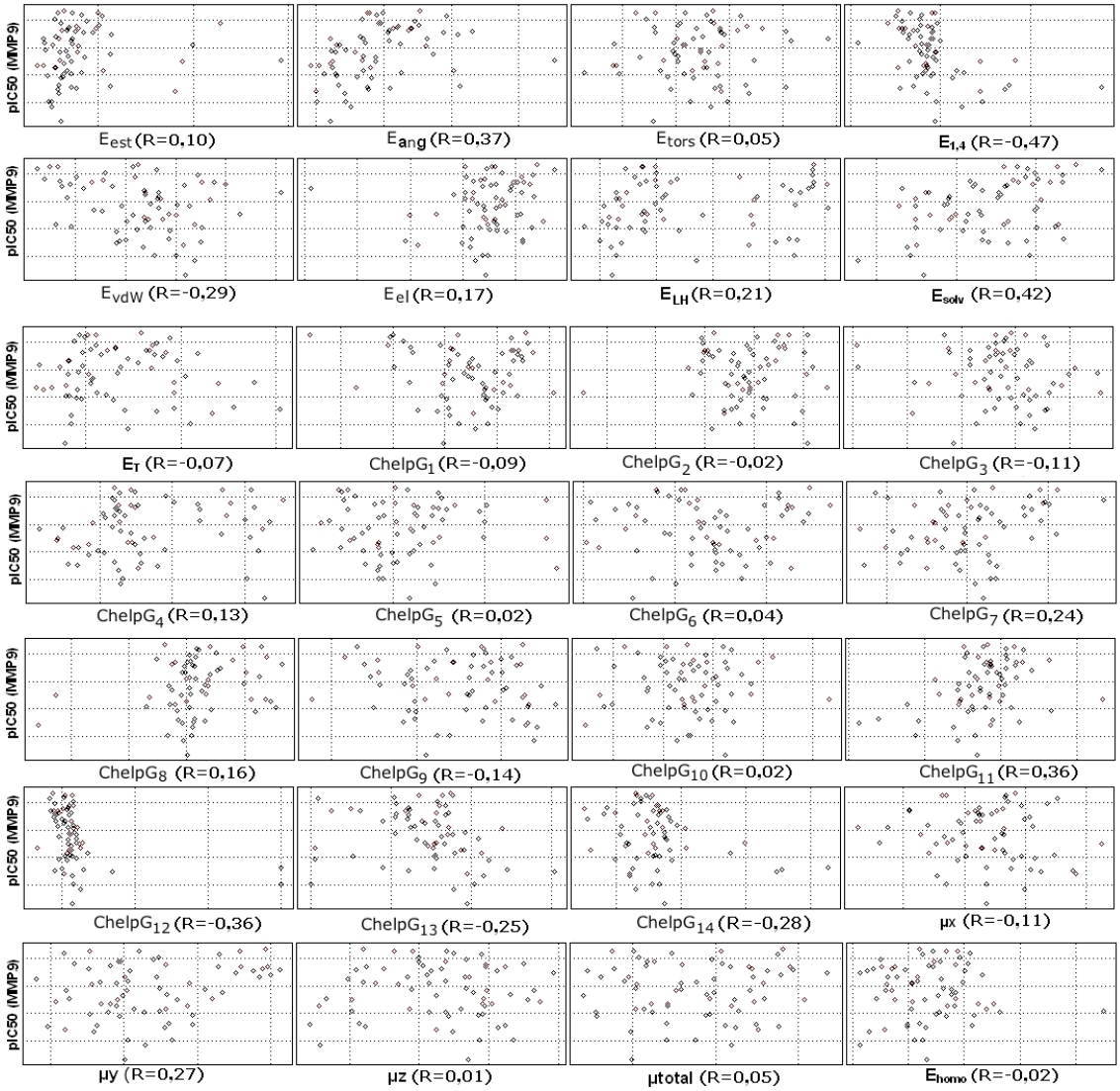


Figura 11. Gráficos de dispersão dos valores de cada descritor *versus* o dado biológico (valores de pCI_{50} para MMP-9), com os respectivos valores de *R*.

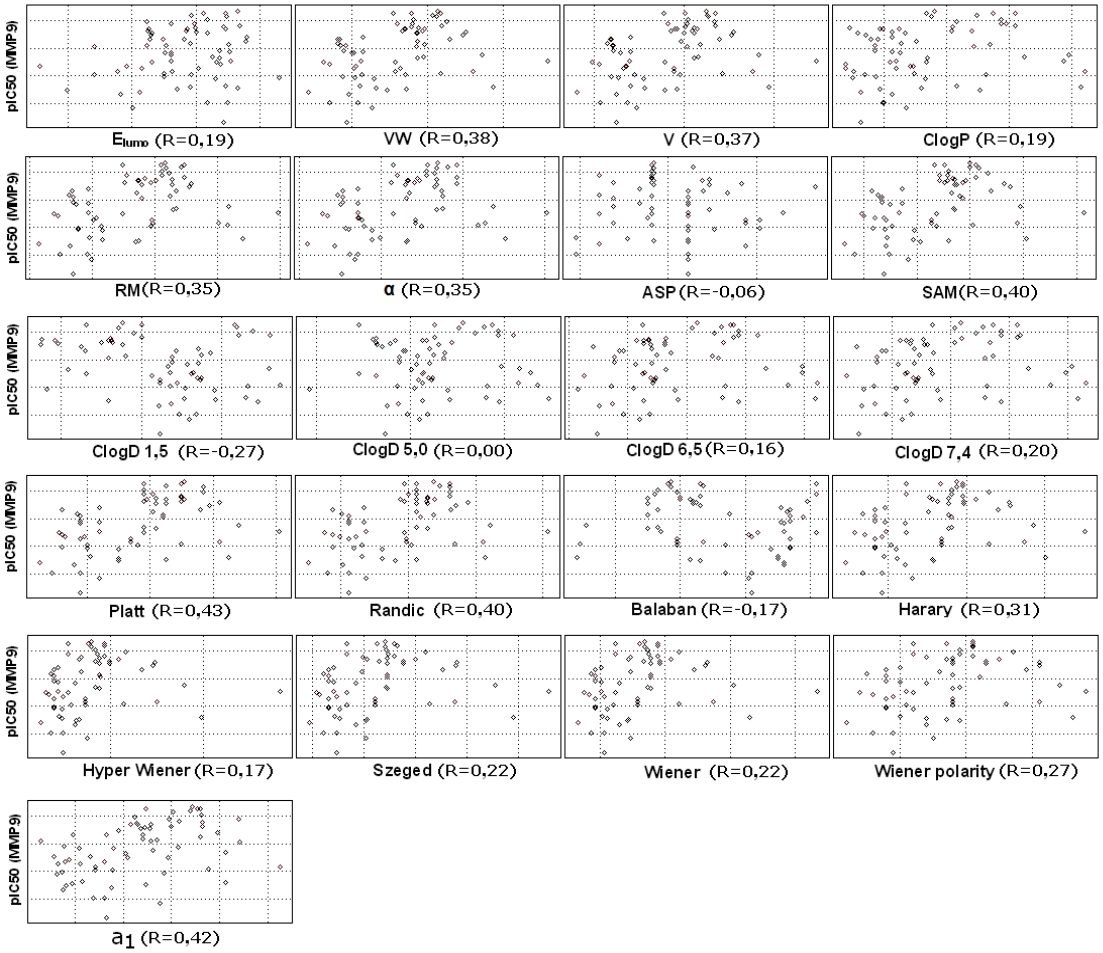


Figura 11 (cont.). Gráficos de dispersão dos valores de cada descritor *versus* o dado biológico (valores de pCI_{50} para MMP-9), com os respectivos valores de *R*.

Descritores que apresentaram boa dispersão/distribuição em relação ao dado biológico, embora não essencialmente com tendência linear (V, V_{vdw}, RM e α, por exemplo) foram mantidos no conjunto de variáveis selecionadas, uma vez que apresentaram papel importante na discriminação das amostras durante estudos preliminares.

As variáveis independentes selecionadas para compor as tabelas de dados finais (bloco **X**), $\mathbf{X}_I$ (MMP-2) e $\mathbf{X}_{II}$ (MMP-9)**,** utilizadas nos estudos de HCA e PCA, foram aquelas que apresentaram maior relevância quando correlacionadas à atividade biológica e, portanto, apresentariam maiores chances de gerar bons resultados na análise exploratória de dados. A tabela final $\mathbf{X}_I$ foi composta por 64 linhas (número de inibidores) e 9 colunas (8 descritores, que incluem E_{LH}, E_{solv}, μy, V, V_{vdw}, RM, α, a_1, mais os valores de pCI_{50} para MMP-2, **y**). A tabela final $\mathbf{X}_{II}$, apresentou o mesmo número de linhas (compostos ou amostras), mas 12 colunas (11 descritores, E_{vdw} , E_{solv} , $ChelpG_{13}$, μy , V, V_{vdw} , RM , α , SAM , CLog_D 1.5, a_1, maisos valores de pCI_{50} para MMP-9, **y**).

4.2.2. HCA e PCA para a tabela de dados X_I (inibidores de MMP-2)

Os resultados obtidos com os métodos HCA e PCA foram complementares. Para o método PCA, considerou-se até 7 fatores, ou componentes principais (PC), no cálculo para a descrição do conjunto de dados e os resultados estão demonstrados na Figura 12. Este método permite descrever o conjunto de dados a partir da combinação linear dos dados originais. De acordo com a tabela de seleção de fatores ou PC, os dois primeiros fatores descrevem 78,45 % da variância total dos dados originais (PC1 (ou fator1) com 60,42 % e a PC2 (ou fator 2) com 18,03 %). PC3 a PC7 descreveram menos de 10 % a praticamente zero, respectivamente (Figura 12, A). Então, as duas primeiras PC, que juntas representam uma porcentagem significativa da variância total dos dados originais, foram consideradas para representar e discutir os resultados obtidos com o conjunto de inibidores de MMP-2.

O gráfico de escores para as duas primeiras PC, PC1 (60,42 %) e PC2 (18,03 %), está apresentado na Figura 12, B. Do lado esquerdo do eixo de PC1 (valores negativos) estão a maioria dos inibidores com grupos substituintes de menor volume na posição R_3, classificados como moderadamente e menos ativos. A grande parte dos compostos mais ativos está do lado direito do eixo de PC1 (valores positivos) e corresponde aos compostos com grupos substituintes volumosos em R_3. As amostras ou compostos mais ativos estão destacados por um quadrado azul na porção superior direita do gráfico de escores e a seta vermelha indica o composto mais ativo (**8dd**).

Na tabela de pesos (*loadings*) para as duas primeiras PC, ou fatores (Figura 12, C), está a informação sobre quais descritores, ou propriedades moleculares, mais influenciaram no processo de separação/discriminação dos compostos (amostras). Em PC1, as propriedades moleculares de natureza estérica/intrínseca (volume molecular, V, e volume de van der Waals, V_{vdW}), estérica/hidrofóbica (caráter misto; refratividade molar, RM) e estérica/eletrônica (polarizabilidade, α) tiveram maior contribuição na descrição do conjunto de inibidores de MMP-2 (valores de pesos iguais a 0,4). Em PC2, descritores termodinâmicos (E_{Hb}) e eletrônicos (μ_y) apresentaram maiores valores de pesos (0,67 e 0,61, respectivamente).

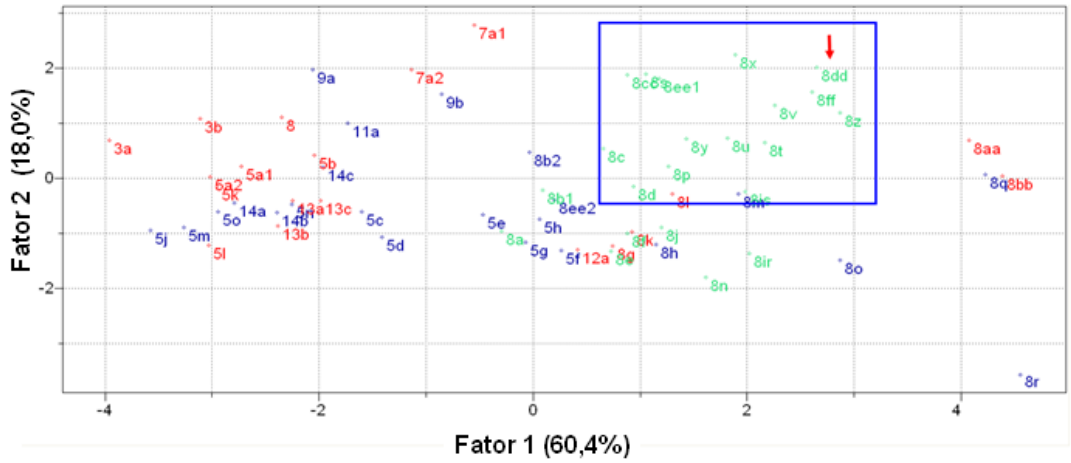
A **Seleção de fatores**

	Variância	%	Cumulativa
Fator 1	304,52	60,42	60,42
Fator 2	90,87	18,03	78,45
Fator 3	48,71	9,66	88,11
Fator 4	34,26	6,79	94,91
Fator 5	23,56	4,67	99,59
Fator 6	1,72	0,34	99,93
Fator 7	0,29	0,05	99,99

C **Tabela de pesos (*loadings*)**

Variável	Fator 1	Fator 2
E_{LH}	0,18	0,67
E_{solv}	0,31	0,14
μy	0,16	0,61
V_{VDW}	0,44	-0,14
V	0,44	-0,16
RM	0,43	-0,21
α	0,44	-0,20
a_1	0,29	0,17

B **Gráfico de escores**



D **Diagnóstico de *outliers*/Gráfico de Mahalanobis**

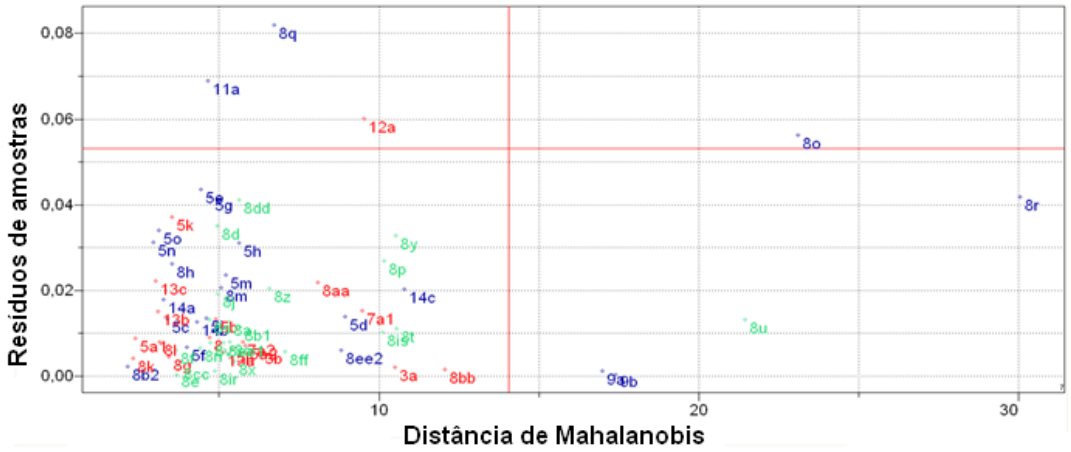


Figura 12. Resultados obtidos com o método PCA, com a seleção de variáveis para os inibidores de MMP-2. (A) Seleção de fatores ou PC: as duas primeiras PC descreveram 78,45 % da variância total dos dados originais; (B) gráfico de escores para as duas primeiras PC, P1 (60,42 %) e PC2 (18,03 %): do lado direito do eixo de PC1 estão os inibidores com substituintes mais volumosos em R_3, os inibidores representados na cor verde são os mais ativos (quadrado azul; o composto mais ativo, **8dd**, está indicado com a seta vermelha), na cor vermelha são moderadamente ativos e na cor azul são menos ativos; (C) tabela de pesos ou *loadings*: descritores de natureza estérica/intrínseca (V e V_{vdW}), estérica/hidrofóbica (RM) e estérica/eletrônica (α) influenciaram mais a descrição dos dados em PC1, enquanto que em PC2 descritores de natureza eletrônica (μ_y) e termodinâmica (E_{Hb}) apresentaram os maiores valores de pesos; (D) diagnóstico de *outliers* ou

gráfico da distância de Mahalanobis *versus* resíduos das amostras, utilizando limite de confiança de 95 % (linhas em vermelho), configurado internamente no programa *Pirouette* 3.11 (Infometrix Inc, 1990-2003).

O composto **8r** (menos ativo; canto inferior direito do gráfico de escores, Fig. 12 B) ficou mais distante de todos os outros inibidores de MMP-2. Então, o gráfico que correlaciona as distâncias de Mahalanobis e os resíduos de amostras foi avaliado para diagnosticar possíveis amostras atípicas ou *outliers* (Figura 12, D). Utilizou-se o limite de confiança de 95 %, configurado internamente no programa Pirouette 3.11 (Infometrix, Inc., 1990-2003), e as linhas em vermelho demonstram esta restrição. Assim, 5% das amostras consideradas normais poderiam ficar fora deste corte e, por esta razão, em um conjunto grande de amostras aquelas que excedem somente um dos limites, não devem ser consideradas atípicas. Para o conjunto de inibidores de MMP-2, os compostos menos ativos **8q**, **8r**, **9a**, **9b**, **11a**, o composto com atividade moderada **12a** e o composto ativo **8u** ficaram fora de apenas um dos dois limites do gráfico de Mahalanobis. Já o composto **8o** (menos ativo) ficou fora dos dois limites do gráfico. Este composto, mesmo com a presença de grupo volumoso em R_3, não apresentou atividade inibitória interessante frente à MMP-2 e os descritores selecionados provavelmente não foram suficientes para descrever esta característica/comportamento diferente em relação ao conjunto de inibidores investigados. O composto menos ativo **8r**, que ficou fora de apenas um dos limites do gráfico de Mahalanobis, mas bem distante dos demais, possui o grupo R_3 mais volumoso de todo o conjunto de dados e esta propriedade do grupo R_3 reflete no volume molecular (descritores V ou V_{vdW}), que dificultaria a interação com a cavidade S1 da enzima MMP-2 (figura 13).

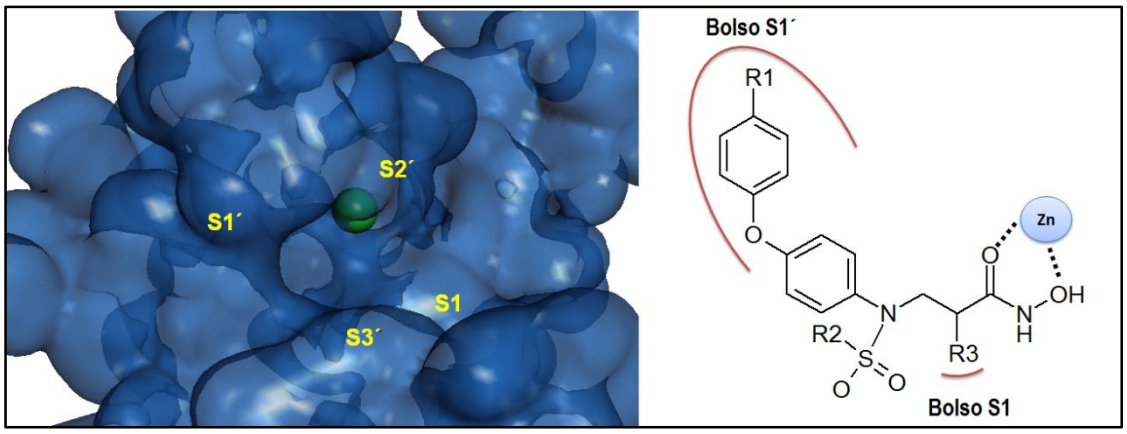


Figura 13. (A) Representação dos bolsos adjacentes ao sítio catalítico na MMP-2 (**PBD** 1HOV; FENG *et al*., 2002) (Figura adaptada de RUSH, POWERS, 2004). A superfície molecular da proteína está demonstrada em azul. (B) Representação bidimensional da estrutura geral do

inibidor hidroxamato interagindo com os bolsos adjacentes ao sítio catalítico (Figura adaptada de FENG *et al*., 2002).

O valor calculado para os descritores V e V_{vdW}, que apresentaram valores de pesos importantes em PC1, justificaria a posição do composto **8r** tão além de um dos limites no gráfico de Mahalanobis. De qualquer modo, os descritores selecionados na análise foram efetivos para discriminar a grande parte do conjunto, particularmente os inibidores mais ativos, o que é importante para se iniciar e direcionar o processo de planejamento de novas estruturas.

Para o método HCA, o dendrograma de amostras (compostos) obtido com o conjunto de 64 derivados β-*N*-biaril éter sulfonamida hidroxamatos indica a divisão dos compostos em 2 agrupamentos principais, A e B (Figura 14), de acordo com as propriedades moleculares (descritores calculados) selecionadas na tabela $\mathbf{X}_I$. Os inibidores do grupo A compartilham 35 % de similaridade, enquanto que os compostos do grupo B compartilham 30 % de similaridade. Neste último grupo estão os compostos mais ativos salientados no gráfico de escores da análise PCA (quadrado azul, Fig. 12 B).

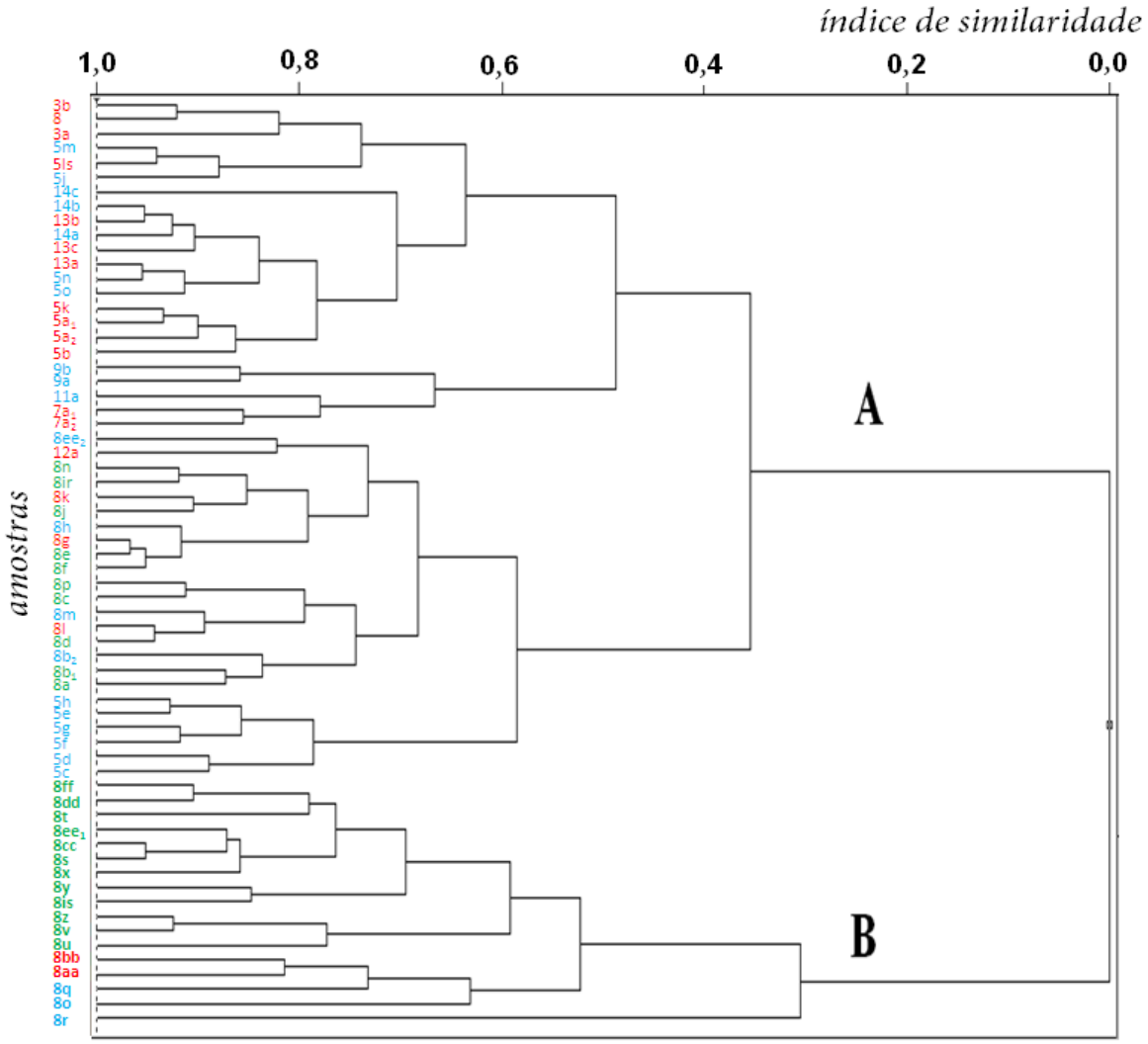


Figura 14. Dendrograma de amostras para o conjunto de inibidores de MMP-2 (tabela $\mathbf{X_I}$). Os inibidores foram divididos em dois agrupamentos principais, A e B. Os compostos mais ativos estão apresentados na cor verde, os de atividade moderada em vermelho e os menos ativos em azul. No agrupamento B estão os compostos mais ativos salientados no quadrado azul do gráfico de escores da análise PCA.

De acordo com a Figura 15, o agrupamento A possui 2 subgrupos, A' e A''. Os compostos do subgrupo A' compartilham 48 % de similaridade (ou possuem índice de similaridade de 0,48). Este subgrupo compreende exclusivamente compostos moderadamente ativos e menos ativos. O subgrupo A'' (índice de similaridade de 0,58) apresenta uma miscelânea de compostos, altamente ativos (**8n**, **8ir**, **8j**, **8e**, **8f**, **8p**, **8c**, **8d**, **8b$_1$**, **8a**), moderados (**12a**, **8k**, **8g**, **8l**) e menos ativos (**8h**, **8m**, **8b$_2$**, **5c-h**).

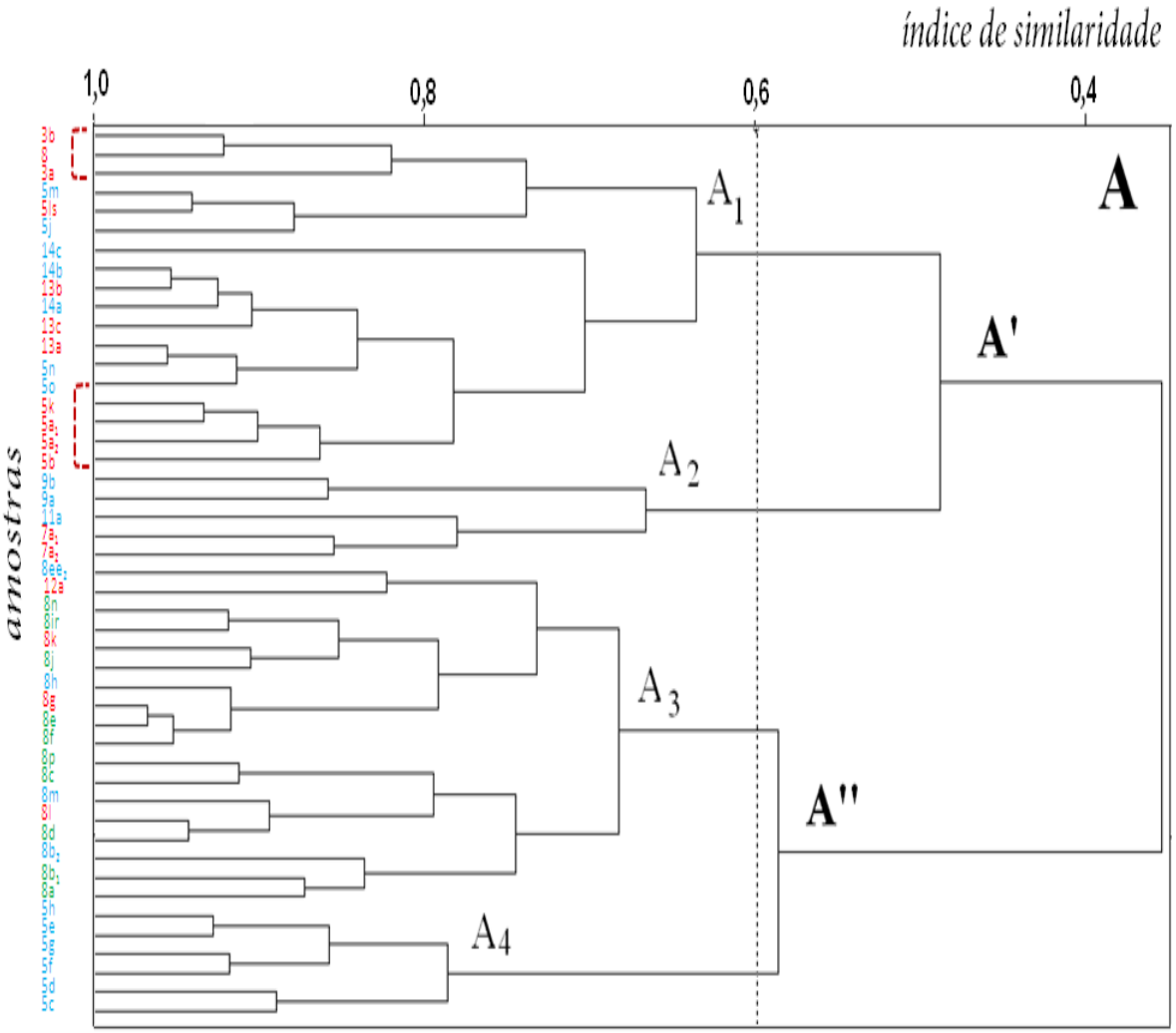


Figura 15. Visualização do agrupamento A do dendrograma de amostras para os inibidores de MMP-2. Dois subgrupos principais estão representados: A' (48 % similaridade) e A'' (58 % similaridade). Considerando o cursor de similaridade em 0,6 (linha cinza pontilhada), o agrupamento A pode ser dividido em quatro subgrupos: A_1, A_2, A_3, A_4. No subgrupo A_1 foram detectados dois grupos constituídos exclusivamente de compostos com atividade moderada (linha pontilhada vermelha). O subgrupo A_4 é composto por amostras com atividade inibitória mais baixa. No subgrupo A_3 estão os compostos mais ativos. Os compostos mais ativos estão apresentados na cor verde, os de atividade moderada em vermelho e os menos ativos em azul.

Considerando o cursor de similaridade em 0,6 (Figura 15), o agrupamento A pode ser dividido em 4 subgrupos: A_1, A_2, A_3 e A_4 com similaridade de 63 %, 66 %, 68 % e 78 %, respectivamente.

O subgrupo A_1 apresenta dois grupos constituídos somente de compostos com atividade moderada (linha tracejada vermelha). O primeiro grupo é formado pelos compostos **3b**, **8** e **3a**. O composto **8** não é substituído na posição R_3 (R_3 = H) e, por causa disso, o carbono na posição 7 não corresponde a um centro quiral. Os compostos **3a** e **3b** não possuem o carbono da posição 7 e, consequentemente, não apresentam o grupo substituinte R_3. Os compostos **8** e **3b** compartilham 92 % de similaridade, enquanto que

o composto **3a** apresenta 82% de similaridade em relação aos dois primeiros. Segundo Yang e colaboradores (2008), derivados que apresentam pequenos grupos α substituintes na porção R_3 possivelmente não interagem de modo eficiente com a cavidade ou bolso S1 da enzima. Logo, os compostos **8**, **3a** e **3b** provavelmente não estabeleceriam interações com a cavidade S1 da MMP-2.

Os compostos **5k**, **5a$_1$**, **5a** e **5b** estão no segundo grupo de inibidores moderadamente ativos de A_1. Os compostos **5a$_1$** e **5a$_2$** são enantiômeros e o composto **5b** difere destes dois na porção R_2, onde possui um grupo etil ao invés de metil. O composto **5k** difere de **5a$_1$** e de **5a$_2$** na porção R_{1c}, onde o grupo substituinte é cloro ao invés de metil. Os substituintes 4-Cl e 4-CH_3 conferem caráter mais lipofílico à molécula uma vez que o valor da constante π de Hansch destes substituintes é positivo (π_p = 0,71 e π_p = 0,52, respectivamente). Além disso, estes grupos substituintes não alteraram drasticamente o volume da molécula (V ou Vv_{dW}; Anexo A, Tabela 1A). Por outro lado, os substituintes apresentam caráter eletrônico distinto, 4-Cl é um grupo retirador de elétrons enquanto 4-CH_3 é um grupo doador de elétrons. A visualização do mapa de potencial eletrostático dos compostos **5a$_1$** e **5k** demonstra o caráter eletrônico dos substituintes 4-CH_3 e 4-Cl, respectivamente. A cor mais amarelada (mais negativa) no centro do anel aromático para o composto **5a$_1$** e mais esverdeada (neutra) para o composto **5k** são dados que corroboram o caráter eletrônico dos grupos substituintes na posição *para* do anel aromático (Figura 16). O composto **5k** poderia apresentar interação adicional do tipo dipolo-dipolo entre o átomo de cloro e resíduos de aminoácido do sítio de interação da MMP-2. Estes tipos de interação não seriam possíveis com o grupo metila, por exemplo. Os compostos **5a$_1$** e **5k** apresentam 93 % de similaridade entre si; o composto **5a$_2$**, por sua vez, apresenta 90% de similaridade em relação a ambos e o composto **5b** compartilha 86 % de similaridade com todo o grupo (Figura 15).

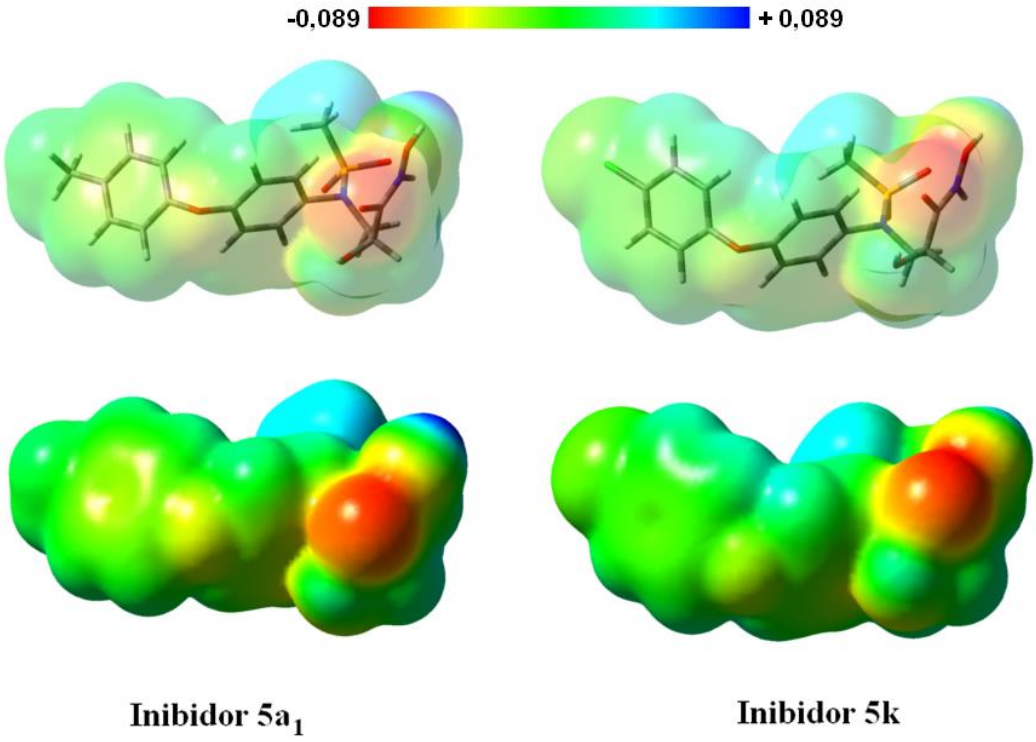


Figura 16. Mapas de potencial eletrostático (MPE) calculados na superfície molecular dos compostos **5a₁** e **5k** (R_1 = CH_3 e Cl, respectivamente, na posição *para* do anel aromático). Utilizou-se a faixa de cores de -0,089 (vermelho intenso; maior distribuição de densidade eletrîonica) a 0,089 (azul intenso; menor distribuição de densidade eletrônica). Os modelos moleculares estão apresentados em tubo, onde os átomos de carbono estão em cinza, oxigênio em vermelho, nitrogênio em azul, enxofre em amarelo, cloro em verde e os de hidrogênio em branco.

O subgrupo A_4 é composto somente por compostos menos ativos (**5h**, **5e**, **5g**, **5f**, **5d**, **5c**) (Figura 15). Estes compostos diferem entre si principalmente na posição R_2. As amostras **5h** e **5e** tanto quanto as amostras **5g** e **5f** possuem substituintes aromáticos na porção R_2 e compartilham 93 % e 92 % de similaridade, respectivamente. A similaridade é de 86 % entre os dois pares de compostos. Os compostos **5c** e **5d** possuem 89% de similaridade.

O agrupamento B, por sua vez, é dividido em dois subgrupos, B’ e B’’, como apresentado na Figura 17, que compartilham 52 % de similaridade entre si. O subgrupo B’ (índice de similaridade de 0,59) é composto por inibidores mais ativos, que apresentam grupos volumosos na porção R_3 e, provavelmente, possuem características moleculares que propiciam a interação com a cavidade S1 da enzima MMP-2. Por outro lado, o subgrupo B’’ (índice de similaridade de 0,63) compreende os inibidores menos ativos (**8o** e **8q**) e de atividade moderada (**8aa** e **8bb**). A porção R_3 quando demasiadamente volumosa, parece desfavorecer a atividade inibitória. A amostra **8o** possui um anel 2-aminonaftaleno na posição R_3 e ultrapassou os dois limites do gráfico de Mahalanobis na

análise PCA (Fig. 12 D). O composto **8q** (ultrapassou um dos limites no gráfico de Mahalanobis, mas ficou bem distante dos demais) possui um anel benzil(propan-2-il) amina na porção R_3. Já os compostos **8aa** e **8bb** apresentam um anel fenilpiperidina não substituído e substituído, respectivamente, na mesma região.

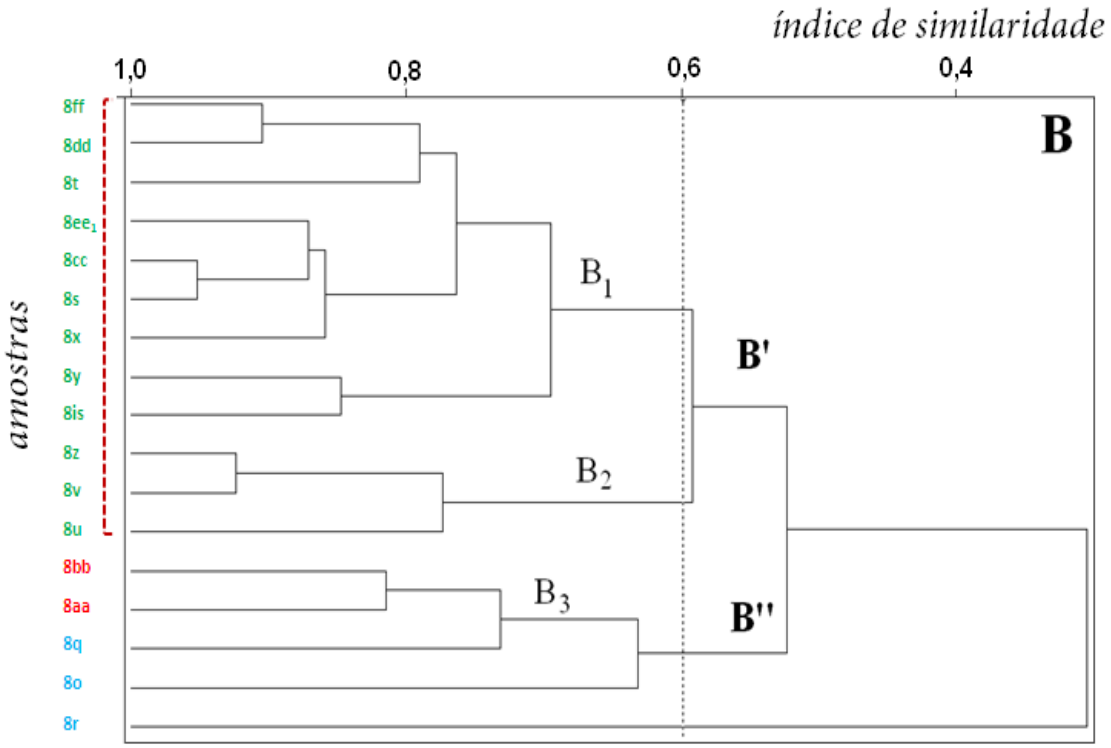


Figura 17. Visualização do agrupamento B do dendrograma de amostras para os inibidores de MMP-2. Dois subgrupos principais podem ser visualizados B' e B''. O subgrupo B' apresenta compostos mais ativos (linha tracejada vermelha). Considerando o cursor de similaridade em 0,6 (linha cinza pontilhada), o agrupamento B pode ser dividido em três subgrupos: B_1, B_2 e B_3. O inibidor mais ativo (composto **8dd**) está no subgrupo B_1. Os compostos mais ativos estão apresentados na cor verde, os de atividade moderada em vermelho e os menos ativos em azul.

Considerando o cursor de similaridade em 0,6 (linha cinza pontilhada; Fig.17), o agrupamento B pode ser dividido em 3 subgrupos: B_1 (índice de similaridade de 0,69), B_2 (índice de similaridade de 0,77) e B_3 (índice de similaridade de 0,63). O inibidor mais ativo (composto **8dd**) está no subgrupo B_1 e compartilha 90 % de similaridade com o composto **8ff**. O composto **8t** apresenta 79% de similaridade em relação a ambos. Os inibidores **8ff**, **8dd** e **8t** diferem entre si na porção R_{1c}, cujos substituintes são 4-CH_3, 4-Cl e 4-CF_3, respectivamente. O caráter estérico, eletrônico e hidrofóbico dos grupos substituintes 4-CH_3 e 4-Cl já foi previamente discutido. O grupo R_3 dos inibidores **8ff**, **8dd** e **8t** é um anel 2,6 cis-dimetilmorfolina.

As amostras **8ee$_1$**, **8cc**, **8s** e **8x** encontram-se agrupadas no subgrupo B_1. Os compostos **8cc** e **8s** compartilham índice de similaridade de 0,95 e o composto **8ee$_1$** possui 87% de similaridade com ambos. Estes compostos possuem grupo substituinte similar em R_3 (anel morfolino). Os inibidores **8ee$_1$**, **8cc** e **8s** diferem entre si na porção R_{1c} (substituintes 4-CH_3, 4-Cl e 4-CF_3, respectivamente). O composto **8x** compartilha 86 % de similaridade com os outros 3 compostos, possui o substituinte 4-CF_3 na porção R_{1c} e o anel piperidina na porção R_3. Os inibidores **8ff**, **8dd**, **8t**, **8ee$_1$**, **8cc**, **8s** e **8x** compartilham 76% de similaridade.

Os compostos **8y** e **8is** estão agrupados separadamente no subgrupo B_1 (Figura 17), mas possuem índice de similaridade de 0,85. Estes compostos diferem entre si em relação aos substituintes da porção R_{1c} (4-CF_3 e 4-CH_3, respectivamente) e da porção R_3 (anel piperidina-4-ol e 1-feniletano-1-amina).

O subgrupo B_2 é constituído das amostras **8z**, **8v** e **8u**, que diferem entre si somente na posição R_3. Os inibidores **8z** e **8v** apresentam índice de similaridade de 0,92, enquanto o composto **8u** compartilha 77% de similaridade com ambos. O anel substituinte em R_3 é piperidina-4-metilcarboxamida, 1-piperaziniletanona e 1-sulfonilpiperazina para os compostos **8z**, **8v** e **8u**, respectivamente. Os subgrupos B_2 e B_1 compartilham 59 % de similaridade.

O composto **8r** (mais distante no gráfico de escores na análise PCA, Fig. 12 B) não foi agrupado com nenhum dos subgrupos formados no dendrograma de amostras (Figura 17), provavelmente devido ao grupo substituinte R_3[benzil N-(2- propan-2-ilamino) etil carbamato], que contribuiu para gerar o maior valor calculado de V_{vdW} (659,190 $Å^2$). O volume excessivo em R_3 pode prejudicar a interação com a cavidade S1 da enzima MMP-2 (a YANG, 2008; b YANG, 2008).

4.2.3. HCA e PCA para os dados da tabela X_{II} (inibidores de MMP-9)

Novamente os resultados obtidos com os métodos HCA e PCA foram complementares. De acordo com a tabela de seleção de fatores ou PC, os dois primeiros fatores discriminaram 66,60 % da variância total dos dados originais (PC1 (ou fator 1) com 52,16 % e a PC2 (ou fator 2) com 14,43 %). PC3 a PC7 descreveram menos de 9 % e 4%, respectivamente (Figura, 18 A). Então, as duas primeiras PC, que juntas representam uma porcentagem significativa da variância total dos dados originais, foram consideradas para representar e discutir os resultados obtidos com o conjunto de inibidores de MMP-9.

O gráfico de escores para as duas primeiras PC, PC1 (52,16 %) e PC2 (14,43 %), está apresentado na Figura 18, B. Do lado esquerdo do eixo de PC1(valores negativos) estão a maioria dos inibidores com grupos substituintes de menor volume na posição R_3, classificados como moderadamente e menos ativos. A grande parte dos compostos mais ativos está do lado direito do eixo de PC1 (valores positivos) e corresponde aos compostos com grupos substituintes volumosos em R_3. As amostras ou compostos mais ativos estão destacados por um retângulo azul no quadrante inferior direito do gráfico de escores e a seta vermelha indica o composto mais ativo (**8dd**). Do lado direito do eixo da PC1 há, também, amostras de atividade moderada e menos ativas.

Na tabela de pesos (*loadings*) para as duas primeiras PC, ou fatores (Figura 18, C), está a informação sobre quais descritores, ou propriedades moleculares, mais influenciaram no processo de separação/discriminação dos compostos (amostras) mais ativos dos demais. Em PC1, as propriedades moleculares de natureza estérica/intrínseca (volume molecular, V, e volume de van der Waals, V_{vdW}), estérica/hidrofóbica (caráter misto; refratividade molar, RM), estérica/eletrônica (polarizabilidade, α) e geométrica (área superficial molecular, SAM) tiveram maior contribuição na descrição do conjunto de inibidores de MMP-9 (valores de pesos iguais a 0,4). Em PC2, descritores hidrofóbicos (ClogD_pH1,5) e eletrônicos (μ_y, $CHELPG_{13}$) apresentaram maiores valores (em módulo) de pesos (0,52, 0,51 e 0,50, respectivamente).

A

Seleção de fatores

	Variância	%	Cumulativa
Fator 1	361,52	52,16	52,16
Fator 2	100,05	14,43	66,60
Fator 3	62,13	8,96	75,57
Fator 4	58,41	8,42	84,00
Fator 5	42,12	6,07	90,07
Fator 6	39,83	5,74	95,82
Fator 7	25,94	3,74	99,57

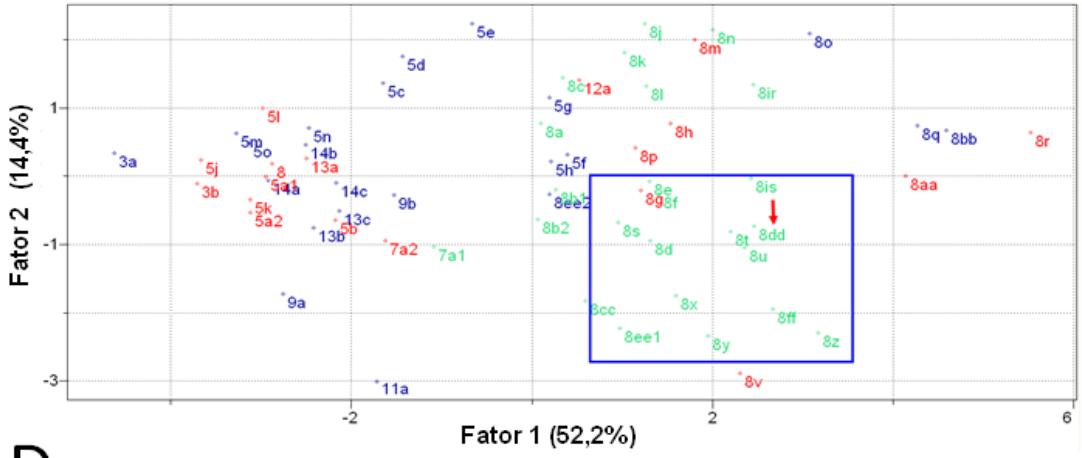
C

Tabela de pesos (*loadings*)

Variáveis	Fator1	Factor2
E_{VDW}	-0,07	-0,41
E_{SOLV}	0,26	0,02
$ChelpG_{13}$	-0,06	0,50
μy	0,13	-0,51
V_{VDW}	0,40	0,01
V	0,41	0,04
RM	0,40	0,10
α	0,40	0,09
SAM	0,40	0,00
ClogD_1,5	-0,09	0,52
a_1	0,24	0,03

B

Gráfico de escores



D

Diagnóstico de *outliers*/Gráfico de Mahalanobis

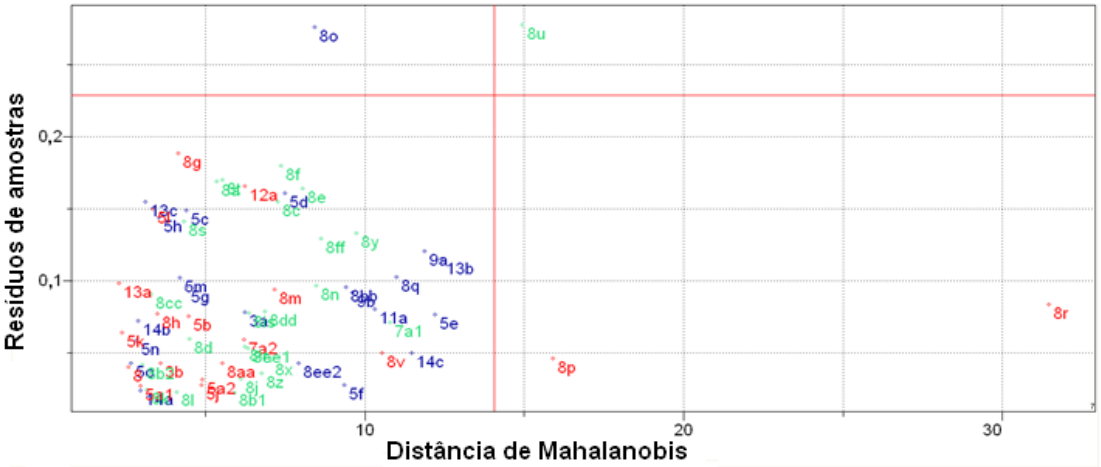


Figura 18. Resultados obtidos com o método PCA, com a seleção de variáveis para os inibidores de MMP-9. (A) Seleção de fatores ou PC: as duas primeiras PC descreveram 66,60 % da variância total dos dados originais; (B) gráfico de escores para as duas primeiras PC, P1 (52,16 %) e PC2 (14,43 %): do lado direito do eixo de PC1 estão os inibidores com substituintes mais volumosos em R_3, os inibidores representados na cor verde são os mais ativos (retângulo azul; o composto mais ativo, **8dd**, está indicado com a seta vermelha), na cor vermelha são moderadamente ativos e na cor azul são menos ativos; (C) tabela de pesos ou *loadings*: descritores de natureza estérica/intrínseca (V e V_{vdW}), estérica/hidrofóbica (RM), estérica/eletrônica (α) e geométrica (SAM) influenciaram mais a descrição dos dados em PC1, enquanto que em PC2 descritores de natureza eletrônica (μ_y, $CHELPG_{13}$) e hidrofóbicos (ClogD_pH1,5) apresentaram os maiores

valores de pesos; (D) diagnóstico de *outliers* ou gráfico da distância de Mahalanobis *versus* resíduos das amostras, utilizando limite de confiança de 95 % (linhas em vermelho), configurado internamente no programa *Pirouette* 3.11 (Infometrix Inc, 1990-2003).

O gráfico que correlaciona as distâncias de Mahalanobis e os resíduos de amostras foi avaliado para diagnosticar possíveis amostras atípicas ou *outliers* para a tabela de dados $\mathbf{X}_{II}$ (Figura 18, D). As condições foram as mesmas utilizadas para a tabela de dados $\mathbf{X}_{I}$ (limite de confiança de 95 %, configurado internamente no programa Pirouette 3.11 (Infometrix, Inc., 1990-2003)). Os compostos de atividade moderada **8p** e **8r** e o composto menos ativo **8o** ficaram de fora de um dos dois limites, enquanto o composto ativo **8u** ficou fora das duas linhas limítrofes do gráfico de Mahalanobis. O volume do grupo substituinte em R_3 parece ser responsável por estes resultados, considerando as variáveis de maior peso em PC1, por exemplo.

O dendrograma de amostras (HCA) obtido para a tabela de dados $\mathbf{X}_{II}$ está apresentado na Figura 19. O conjunto de inibidores de MMP-9 foi agrupado em 2 agrupamentos principais, A e B. O grupo A compartilha 24 % de similaridade e o grupo B apresenta 52 % de similaridade.

Considerando o cursor de similaridade em 0,60 (linha cinza pontilhada), o agrupamento A pode ser dividido em 9 subgrupos (A_1 a A_9) com valores de índices de similaridade de 0,67, 0,65, 0,70, 0,67, 0,71, 0,72, 0,68, 0,72 e 0,64, respectivamente.

O agrupamento B é formado por apenas três compostos (**8bb**, **8aa** e **8q**) e apresenta índice de similaridade de 0,75.

Os subgrupos A_4 a A_6 são os que apresentam maior número de compostos altamente ativos. O subgrupo A_4 é composto somente por inibidores ativos (**8ff**, **8y**, **8d**, **8dd**, **8s**, **8is**), incluindo o composto mais ativo do conjunto de dados (**8dd**). Os inibidores **8ff** e **8y** apresentam índice de similaridade de 0,79 e o composto **8d** compartilha 77% de similaridade com ambos. O composto **8dd** apresenta 82% de similaridade com o **8s** e o composto **8is** é 78% similar a ambos inibidores, considerando as propriedades moleculares selecionadas para a análise. Os inibidores **8v** e **8g**, moderadamente ativos para MMP-9, fazem parte do subgrupo A_5 (**8ee$_1$**, **8cc**, **8x, 8t**, **8z, 8v**) e A_6 (**8u, 8g, 8f**, **8e**), respectivamente. A diferença estrutural entre eles consiste na substituição das porções R_{1c} e R_3, conforme já discutido para o conjunto de dados com valores de inibição para MMP-2. Um grupo menor de inibidores altamente ativos (**8l**, **8n, 8k, 8j**,**8ir**) foi encontrado no subgrupo A_9 (linha vermelha tracejada, Figura 19). Nesta análise (cursor em 0,6), observou-se que o composto **8r** não se agrupa em A, nem em B. A substituição em R_3 pode ser a responsável por este comportamento. Somente dois subconjuntos constituídos

de inibidores moderadamente ativos foram detectados no subgrupoA_1 (linha vermelha tracejada na Figura 19). Um grupo é formado pelos compostos **8** e **3b** e o outro pelos compostos **5k**, **5a**$_1$, **5a**$_2$ e **5b**. O subgrupo A_8 é formado por compostos menos ativos, como **5h**, **5g**, **5e**, **5d**, **5c**. O agrupamento B é composto por amostras com atividade baixa e moderada.

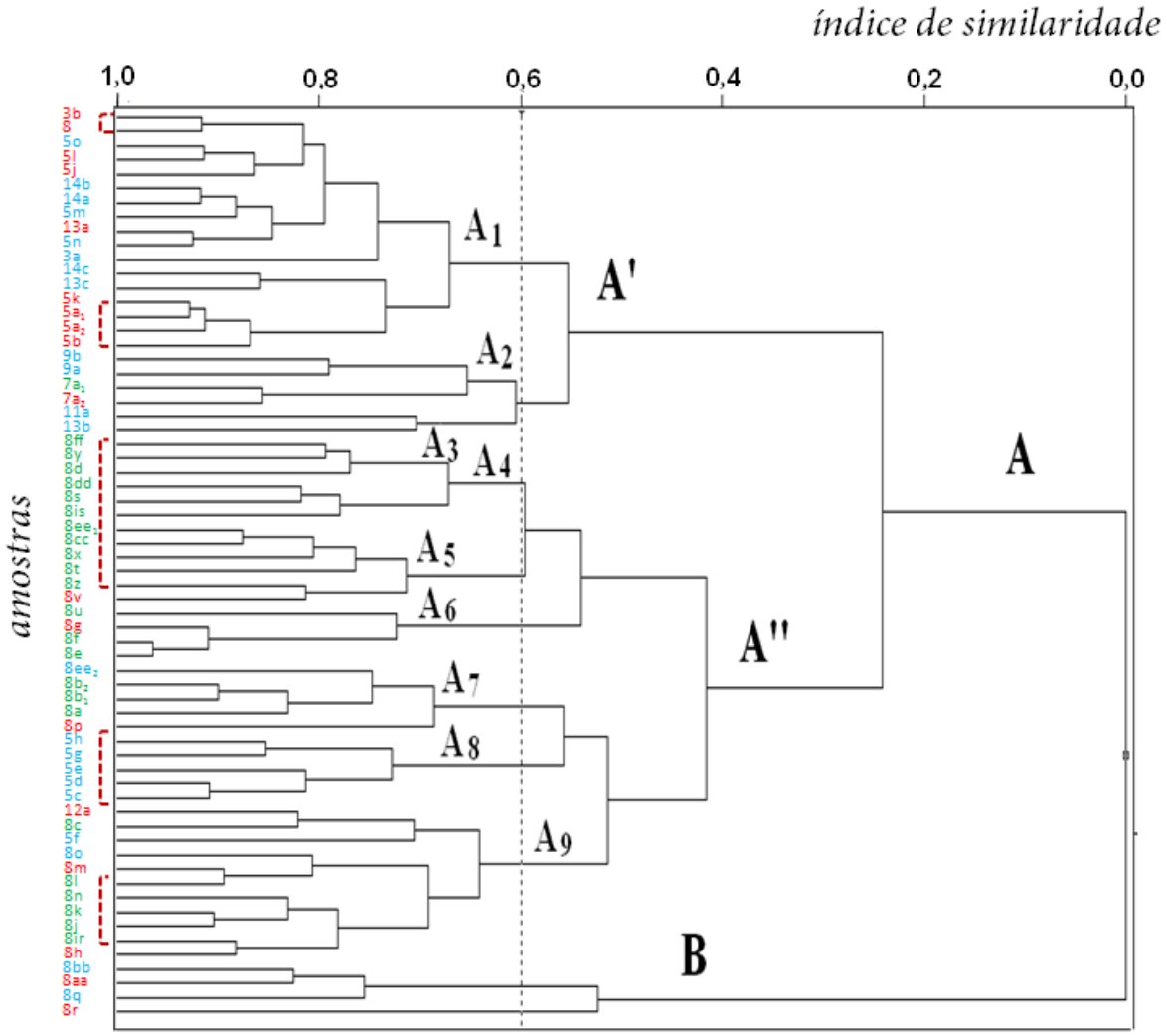


Figura 19. Dendrograma de amostras para os inibidores de MMP-9 (tabela $\mathbf{X_{II}}$). As amostras ou compostos foram separadas em dois agrupamntos principais, A e B. Considerando o cursor de similaridade em 0,6 (linha cinza pontilhada), o agrupamento A pode ser dividido em nove subgrupos, $A_{1\text{-}9}$. Dois subconjuntos constituídos de compostos moderadamente ativos foram detectados em A_1 (linha vermelha tracejada). Nos subgrupos A_4, A_5 e A_6 estão presentes, em grande parte, compostos altamente ativos (linha pontilhada vermelha). Estes compostos estão em destaque no gráfico de escores da análise PCA (retângulo azul). A_8 é formado por compostos menos ativos. Os compostos mais ativos estão apresentados na cor verde, os de atividade moderada em vermelho e os menos ativos em azul.

O estudo de análise exploratória foi apresentado no Simpósio Brasileiro de Química Teórica de 2011 e publicado no *International Journal of Quantum Chemistry*, em 2012 (**Anexo B**).

4.3. Estudos de QSAR-4D IR para os inibidores de MMP-2 e MMP-9

4.3.1. Aplicação do formalismo QSAR-4D IR ao conjunto de hidroxamatos com atividade inibitória frente à enzima MMP-2

A análise de QSAR-4D IR foi aplicada a um conjunto inicial de dados constituído por 64 moléculas com atividade inibitória (baixa, média e alta) frente à enzima MMP-2 (a YANG, 2008; b YANG, 2008), previamente utilizado na análise exploratória de dados. As medidas estatísticas (r^2, q^2, LOF, LSE) e o número de descritores correlacionados para os 10 melhores modelos, após a otimização dos modelos com GFA-MLR, para cada um dos três alinhamentos, estão apresentados na Tabela 5. As opções selecionadas para gerar os 10 melhores de cada alinhamento na etapa GFA-MLR foram as seguintes: 70.000 operações genéticas, 10% de mutação em cada ciclo de recombinação (*crossover*) e fator de ajuste (*smoothing factor*) de 2,0 a 3,0.

Tabela 5. Medidas estatísticas, número de descritores de ocupação (GCOD) e de GCOD correlacionados no modelo (GFA-MLR), para os 10 melhores modelos de QSAR (N = 64) de cada alinhamento (**1**,**2**,**3**), com fator de ajuste de 2,0 a 3,5

Alinhamento	**1**			**2**			**3**		
Fator de ajuste	**2,0**	**2,5**	**3,0**	**2,0**	**2,5**	**3,0**	**2,0**	**2,5**	**3,0**
n° de GCOD no modelo	14-15	12-13	6-8	14-15	12-14	6-8	15	9-11	8-10
r^2	0,91-0,92	0,88-0,90	0,74-0,78	0,90-0,91	0,88-0,91	0,75-0,78	0,92-0,93	0,83-0,85	0,78-0,82
q^2	0,87-0,88	0,82-0,84	0,69-0,73	0,86-0,86	0,83-0,85	0,71-0,72	0,88	0,77-0,79	0,71-0,73
LSE	0,06	0,07-0,08	0,15-0,18	0,06	0,06-0,08	0,15-0,17	0,05	0,10-0,12	0,12-0,15
LOF	0,19-0,22	0,28-0,32	0,34-0,39	0,19-0,21	0,27-0,29	0,33-0,39	0,18-0,19	0,29-0,34	0,32-0,44
N ° de GCOD correlacionados nos modelos GFA-MLR	5-7	3-5	1-2	3-5	2-3	3-5	5-7	2-5	1-3

Então, considerando-se os resultados gerados com o conjunto total de dados em estudo, optou-se por reduzir o conjunto de 64 para 38 inibidores. A seleção dos compostos foi realizada com base nos dados de HCA de amostras, previamente obtidos na análise exploratória de dados, e considerou-se como critério a diversidade estrutural dos inibidores e seus gradientes de atividade inibitória, conforme apresentado na figura 20. Os inibidores que compuseram o conjunto inicial de dados foram: **5a**, **5b**, **5c**, **5f**, **5g**, **5h**, **5j**, **5l**, **5o**, **13a**, **13b**, **14b**, **5k**, **7a**, **8a**, **8b$_1$**, **8b$_2$**, **8c**, **8d**, **8e**, **8f**, **8ir**, **8is**, **8k**, **8n**, **8q**, **8r**, **8t**, **8x**, **8z**, **11a**, **8aa**, **8cc**, **8dd**, **8ee$_2$**, **9a**, **9b** e **3b** (tabela 2).

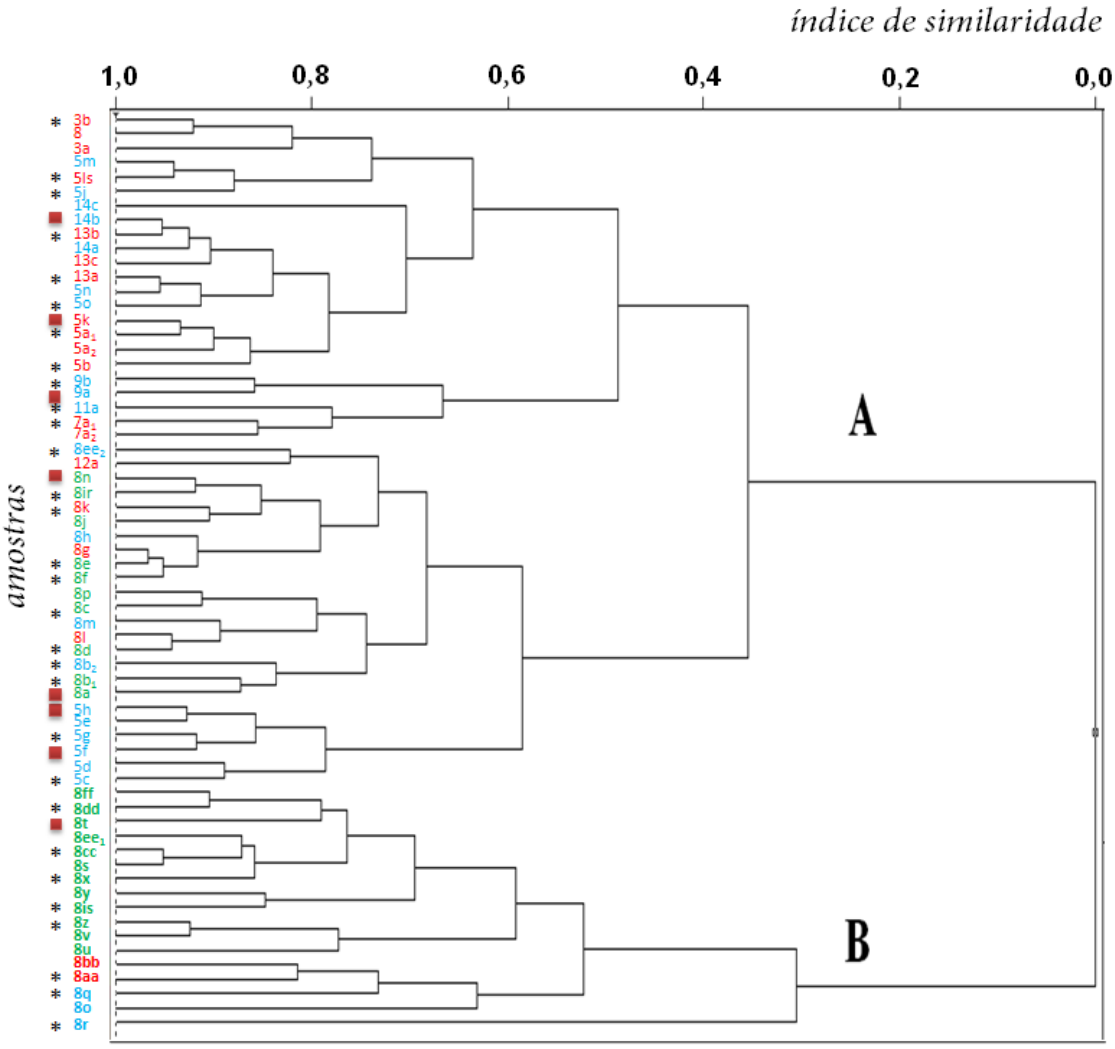


Figura 20. Dendrograma de amostras do conjunto de derivados hidroxamatos com atividade inibitória frente à enzima MMP-2 (N = 38). Os asteriscos (*) indicam os compostos que constituem o conjunto de treinamento (N = 30) e os quadrados preenchidos em vermelho (■) indicam os compostos que compõem o conjunto de avaliação (N = 8; 22 % do total de compostos). Compostos representados em verde correspondem aos mais ativos, em vermelho aos de atividade moderada e em azul aos menos ativos do conjunto de dados.

Antes de obter modelos apenas com os compostos do conjunto de treinamento, aplicou-se o formalismo do QSAR-4D IR ao conjunto total de dados (N = 38), afim de gerar modelos estatisticamente mais representativos e de otimizar as condições de cálculo para gerar os modelos matemáticos com o conjunto de treinamento (N = 30).

Os modelos de QSAR foram construídos para cada um dos 3 alinhamentos propostos, conforme apresentado na tabela 4. Estes alinhamentos exploraram a estrutura comum às moléculas, em toda sua extensão, o grupo farmacofórico representado pela porção hidroxamato, que é responsável por quelar o átomo de zinco das MMP, e a porção sulfonamida que permite a interação dos inibidores com a cadeia polipeptídica das enzimas por meio de ligações de hidrogênio (a YANG, 2008).

Avaliaram-se, então, os modelos de QSAR gerados em relação às medidas estatísticas, número de descritores (GCOD) e número de amostras atípicas (*outliers*). As medidas estatísticas para os 10 melhores modelos de cada alinhamento utilizando 70.000 operações genéticas, 10% de mutação em cada ciclo de recombinação (*crossover*) e fator de ajuste (*smoothing fator*) de 2,5, na otimização GFA-MLR, estão apresentados na Tabela 6.

Tabela 6. Medidas estatísticas, número de descritores de ocupação (GCOD) e de amostras atípicas (*outliers*) para os 10 melhores modelos de QSAR (N = 38) obtidos para cada alinhamento (**1**,**2**,**3**), com fator de ajuste de 2,5

Alinhamento	n° GCOD	r^2	q^2	LSE	LOF	amostras atípicas (*outliers*)
1	9	0,94-0,95	0,90-0,93	0,02-0,03	0,17-0,21	1
2	**6-7**	**0,88-0,90**	**0,83**	**0,06-0,07**	**0,19-0,21**	**0-1**
3	8-9	0,94-0,95	0,90-0,91	0,03-0,04	0,17-0,22	1-2

Todos os alinhamentos apresentaram modelos com valores altos de q^2 (0,83 a 0,93) e r^2 (0,88 a 0,95) e valores baixos de erro (LSE (0,02 a 0,07) e LOF (0,17 a 0,22). O alinhamento **2**, entretanto, apresentou modelos de QSAR sem quaisquer amostras atípicas. Além disso, o número de descritores obtido é apropriado ao tamanho do conjunto de dados (N = 38), já que os modelos QSAR finais são otimizados com a abordagem GFA-MLR (relação de 5 ou 6 amostras para cada descritor). Então, o alinhamento **2** foi considerado o melhor alinhamento. A tabela 7 apresenta as medidas estatísticas, o número de descritores de ocupação (GCOD) e de amostras atípicas para os 10 melhores modelos obtidos para o alinhamento **2**.

Tabela 7. Medidas estatísticas, número de descritores de ocupação (GCOD) e de amostras atípicas obtidos para os 10 melhores modelos de QSAR (N = 38) para o alinhamento **2**

Modelo	n° GCOD	r^2	q^2	LSE	LOF	n° amostras atípicas (*outliers*)
1	7	0,90	0,83	0,06	0,20	0
2	7	0,90	0,83	0,06	0,21	1
3	7	0,90	0,83	0,07	0,21	1
4	6	0,88	0,83	0,07	0,19	1
5	6	0,88	0,83	0,07	0,20	1
6	6	0,88	0,83	0,07	0,19	1
7	7	0,88	0,83	0,07	0,19	1
8	7	0,90	0,83	0,06	0,21	1
9	7	0,90	0,83	0,06	0,21	1
10	6	0,88	0,83	0,07	0,19	1

Calculou-se, então, a matriz de correlação-cruzada de resíduos de ajuste entre os pares dos 10 melhores modelos obtidos para o alinhamento **2** (tabela 8) para verificar se os modelos gerados eram equivalentes ($R > 0,5$) ou distintos ($R < 0,5$). Os valores de R foram todos maiores de 0,5 ($R \approx 1$), indicativo que os modelos são equivalentes. Portanto, existe um único modelo de QSAR para este conjunto de dados.

Tabela 8. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os 10 melhores modelos (N = 38) do alinhamento **2**

Modelo	1	2	3	4	5	6	7	8	9	10
1	**1,00**									
2	0,92	**1,00**								
3	0,92	1,00	**1,00**							
4	0,85	0,93	0,93	**1,00**						
5	0,85	0,93	0,93	1,00	**1,00**					
6	0,85	0,93	0,93	1,00	1,00	**1,00**				
7	0,85	0,93	0,93	1,00	1,00	1,00	**1,00**			
8	0,90	0,97	0,97	0,88	0,88	0,88	0,88	**1,00**		
9	0,90	0,97	0,97	0,88	0,88	0,88	0,88	1,00	**1,00**	
10	0,86	0,94	0,94	1,00	1,00	1,00	1,00	0,89	0,89	**1,00**

Os descritores de ocupação das células da grade virtual (GCOD) selecionados para os modelos 1, 2, 3, 8 e 9 estavam altamente correlacionados entre si ($R > 0,5$). Entretanto, de acordo com a matriz de correlação-cruzada entre pares de descritores selecionados para o modelo **4**, apresentada na tabela 9, os valores de R encontrados foram menores que 0,5 (em módulo), ou seja, os descritores não são correlacionados entre si e contribuem com informação distinta ao modelo.

Tabela 9. Matriz de correlação-cruzada entre pares dos descritores de ocupação (GCOD) GC*i* (*x*,*y*,*z*;*IPE*)) do modelo **4** (N = 38), alinhamento **2**

GCOD	GC1 (3,1,-1; *np*)	GC2 (3,4,-1; *np*)	GC3 (-1,0,0; *any*)	GC4 (-2,-2,1; *any*)	GC5 (0,-3,0; *any*)	GC6 (1,-1,2; *any*)
GC1	**1,00**					
GC2	0,37	**1,00**				
GC3	-0,18	-0,13	**1,00**			
GC4	-0,16	-0,12	-0,15	**1,00**		
GC5	-0,19	-0,14	-0,18	-0,15	**1,00**	
GC6	0,48	0,47	-0,12	-0,27	-0,33	**1,00**

Além disso, o modelo **4** apresenta altos valores de q^2 (0,83) e r^2 (0,88), baixos valores de erro (LSE = 0,07 e LOF = 0,19) e possui seis descritores de ocupação. No entanto, o modelo **4** tem uma amostra atípica ou *outlier* (tabela 10), que corresponde ao composto **14b** ($pCI_{50} = 6,05$). O valor de resíduo (diferença entre a atividade experimental

e a calculada) obtido para o composto **14b** foi de 0,68, que excedeu o limite de dois desvios padrão da média do modelo (2DP = 0,53).

Tabela 10. Valores de resíduos obtidos para os inibidores de MMP-2 com o modelo **4** (N = 38) do alinhamento 2

Composto	pCI_{50exp}	pCI_{50calc}	Resíduo	DP	2DP
5a	7,21	6,78	-0,42	0,26	0,53
5b	6,85	7,11	0,26		
5c	6,29	6,63	0,34		
5f	6,23	6,19	-0,03		
5g	6,40	6,25	-0,14		
5h	6,46	6,82	0,36		
5j	6,61	6,86	0,25		
5l	6,86	6,83	-0,02		
5o	6,55	6,42	-0,12		
13a	7,33	7,01	-0,31		
13b	7,25	7,21	-0,03		
14b	6,05	6,73	**0,68**		
5k	7,45	6,94	-0,50		
7a	7,44	7,62	0,18		
8a	8,04	8,16	0,12		
$8b_1$	8,40	8,21	-0,18		
$8b_2$	6,75	6,78	0,03		
8c	7,96	8,00	0,04		
8d	8,00	7,84	-0,15		
8e	8,02	7,85	-0,16		
8f	7,77	7,75	-0,01		
8ir	7,62	7,40	-0,21		
8is	7,92	7,91	-0,01		
8k	7,19	7,39	0,20		
8n	7,57	7,39	-0,17		
8q	6,55	6,19	-0,35		
8r	6,58	6,23	-0,34		
8t	8,25	8,24	0,00		
8x	8,12	8,44	0,32		
8z	8,26	8,07	-0,18		
11a	6,17	6,25	0,08		
8aa	7,33	7,57	0,24		
8cc	8,26	8,25	0,00		
8dd	8,60	8,25	-0,34		
8ee	5,88	6,24	0,36		
9a	5,90	6,19	0,29		
9b	6,54	6,19	-0,34		
3b	7,14	7,40	0,26		

Do ponto de vista estrutural, o composto **14b** possui o grupo CH_2 ligado ao átomo de oxigênio na porção R_{1a}. Este tipo de substituição parece ser responsável pela redução de atividade inibitória frente à enzima MMP-2, pois confere maior flexibilidade à porção R_{1a} da molécula. Por sua vez, o composto **13b** que apresenta o grupo substituinte C≡C em R_{1a}, mas todos os demais substituintes similares ao composto **14b**, é potencialmente mais ativo (pCI_{50} = 7,25). É possível que uma amostra atípica apareça em uma população amostrada e o inibidor **14b** é a única molécula do conjuntode dados que apresenta um heteroátomo ligado a um grupo CH_2 em R_{1a}. Apesar disso, o coeficiente de correlação (*R*) entre os valores da atividade experimental e calculada de pCI_{50}, considerando o modelo **4** (N = 38) do alinhamento 2, foi 0,97 e o gráfico está apresentado na figura 21.

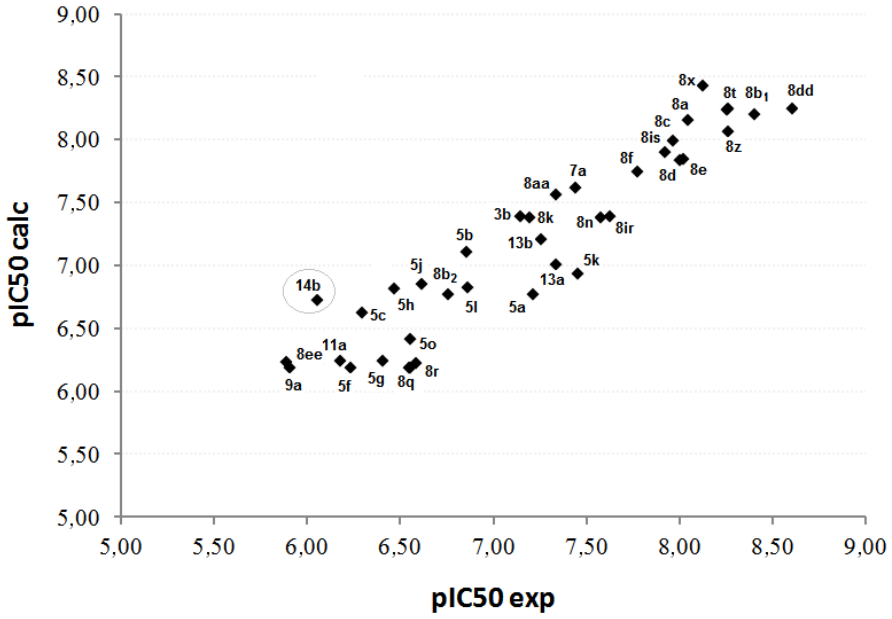


Figura 21. Gráfico com valores experimentais e calculados de pCI_{50} para o modelo **4** (N = 38) do alinhamento **2**. O coeficiente de correlação (*R*) entre a atividade calculada e experimental para este conjunto de dados foi 0,97. O composto **14b** foi a única amostra cujo valor de resíduo (0,68) excedeu 2 DP (confiança de 95 %) da média do modelo (2 DP = 0,53).

Após a obtenção de um modelo de QSAR estatisticamente significativo (modelo **4**) para o conjunto total de inibidores (N = 38), foram aplicados os procedimentos de validação cruzada LNO (triplicata) e aleatorização do vetor ***y***, conforme descrito nos itens **3.5.9** e **3.5.10** de Material e Métodos, para verificar a robustez do modelo **4** e se o mesmo não foi obtido ao acaso. Na validação cruzada LNO, o *N* variou de 2 a 16 (42 % das amostras) e o *N* crítico foi identificado quando 12 amostras foram removidas. Então, o valor de q^2_{LNO} foi mantido semelhante ao modelo original (q^2_{LOO}) até que 32 % das amostras fossem removidas do conjunto de dados (2DP < 0,1). Este dado indica que o

modelo **4** é robusto. A tabela 11 contêm os valores de q^2_{LNO} para cada réplica (triplicata), a média de q^2_{LNO} ($<q^2_{LNO}>$) e o valor de dois desvios padrão (2DP), onde N = 2 a 16.

Tabela 11. Validação cruzada LNO (triplicata) para o modelo **4** (N = 38) do alinhamento **2** obtido com o conjunto de compostos com atividade inibitória frente à enzima MMP-2

N	q^2_{LNO}	q^2_{LNO}	q^2_{LNO}	$<q^2_{LNO}>$	2SD
2	0,83	0,83	0,83	0,83	0,00
3	0,82	0,81	0,83	0,82	0,02
4	0,83	0,84	0,84	0,83	0,01
5	0,84	0,80	0,84	0,82	0,04
6	0,84	0,84	0,78	0,82	0,07
7	0,83	0,82	0,84	0,83	0,02
8	0,85	0,83	0,80	0,83	0,05
9	0,82	0,84	0,82	0,83	0,02
10	0,84	0,80	0,82	0,82	0,05
11	0,79	0,82	0,81	0,81	0,03
12	**0,84**	**0,79**	**0,83**	**0,82**	**0,04**
13	0,56	0,83	0,84	0,74	0,32
14	0,82	0,81	0,80	0,81	0,02
15	0,81	0,79	0,83	0,81	0,04
16	0,81	0,83	0,81	0,81	0,02

O modelo obtido para os 38 compostos foi, então, submetido ao procedimento de aleatorização do vetor **y**, com 10, 30 e 50 corridas. O embaralhamento de dados foi aplicado somente à coluna contendo os valores de atividade inibitória (**y**). O bloco **X** (variáveis independentes) não sofreu quaisquer modificações (ordem original da tabela de dados). Os modelos foram gerados utilizando a validação cruzada LOO em cada corrida, mesma condição utilizada para obter o modelo com os dados originais (***y*** não embaralhado). Os modelos obtidos apresentaram baixa qualidade e nenhum significado real. De acordo com a abordagem de Wold and Eriksson (1995), o modelo **4** do alinhamento **2** apresenta correlação ao acaso não significativa. Os valores de q^2_{yrand} foram iguais a/ou menores que zero (valores negativos) e os valores r^2_{yrand} variaram de 0,07 a 0,27, utilizando 10, 30 ou 50 aleatorizações.

Os gráficos de validação cruzada LNO e de aleatorização do ***y*** para o modelo **4** estão demonstrados na Figura 22. A tabela 2A que apresenta os valores de q^2_{yrand} e r^2_{yrand} para as 10 randomizações e as figuras 1A e 2A referentes aos resultados para 30 e 50 aleatorizações (tabelas e gráficos), respectivamente, podem ser visualizados no Anexo A.

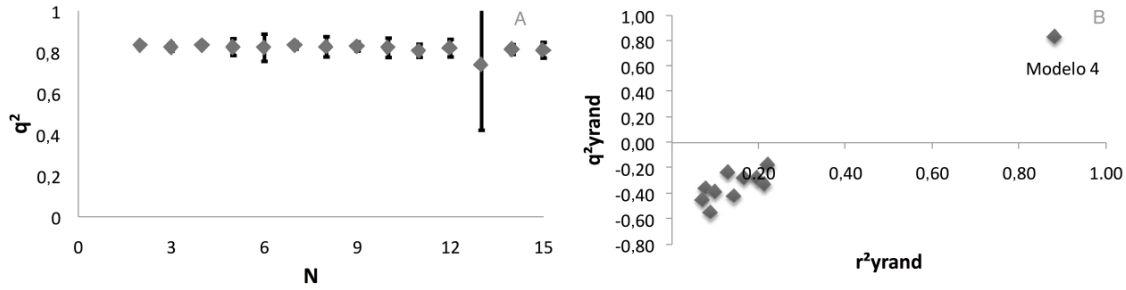


Figura 22. (A) Validação cruzada *leave-N-out* aplicada ao modelo final obtido para o conjunto de inibidores de MMP-2 (N = 38). Os losangos representam a média e as barras indicam o desvio padrão (triplicata) para cada valor de *N*. O *N* crítico para o modelo de 38 inibidores foi igual a 12 (32 % das amostras). (B) Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de 38 inibidores com 10 corridas ou aleatorizações do vetor **y** (valores de pCI_{50} frente à MMP-2).

Apesar do modelo de QSAR ter sido obtido a partir de conjunto de 38 compostos e ser estatisticamente confiável, em geral, é recomendado definir um conjunto de treinamento e um conjunto de avaliação. O conjunto de avaliação é importante para se desenvolver a validação externa do modelo obtido com o conjunto de treinamento. A validação externa permite verificar a capacidade de previsão do modelo QSAR para amostras que não participaram na etapa de construção do modelo.

Desta forma, os 38 inibidores foram divididos em dois conjuntos: de treinamento (N = 30), utilizado para a construção dos modelos de QSAR, e de avaliação (N = 8; 21 % do total de amostras), conforme descrito no item **3.5**. Os inibidores que compuseram o conjunto de avaliação foram: **5f**, **5h**, **14b**, **5k**, **8a**, **8n**, **8t** e **9a**.

O formalismo de QSAR-4D IR foi aplicado ao conjunto de treinamento (N = 30) utilizando as condições previamente otimizadas (70.000 operações genéticas e 10% de mutação em cada ciclo de recombinação (*crossover*) e na otimização GFA-MLR empregou-se fator de ajuste no valor de 2,4. As medidas estatísticas obtidas para os 10 melhores modelos de cada alinhamento estão descritas na tabela 12.

Tabela 12. Medidas estatísticas, número de descritores de ocupação (GCOD) e de amostras atípicas (*outliers*) para os 10 melhores modelos QSAR (N = 30) obtidos para cada alinhamento (**1**,**2**,**3**), utilizando fator de ajuste de 2,4

Alinhamento	n° GCOD	r^2	q^2	LSE	LOF	amostras atípicas (*outliers*)
1	7-8	0,96-0,97	0,95-0,96	0,09-0,10	0,01-0,02	1-2
2	**5-6**	**0,91-0,94**	**0,86-0,89**	**0,03-0,05**	**0,11-0,14**	**0-2**
3	8	0,98	0,97	0,06-0,07	0,08	2

Os três alinhamentos apresentaram modelos com valores elevados de q^2 (0,86 a 0,97) e de r^2 (0,91 a 0,98) e com valores baixos de erro (LSE = 0,03 a 0,10 e LOF = 0,01 a 0,14). No entanto, somente o alinhamento **2** gerou modelos sem amostras atípicas. Além disso, o número de descritores (GCOD) nos modelos do alinhamento **2** foi apropriado para o tamanho do conjunto de dados (N = 30), já que os modelos QSAR finais são otimizados com a abordagem GFA-MLR (relação de 5 ou 6 descritores para cada amostra). Então, o alinhamento **2** gerou a melhor representação espacial, considerando todos os PAC, para o conjunto de inibidores de MMP-2. Este alinhamento considera o carbono carbonílico da porção hidroxâmica, o grupo substituinte na porção R_3 e o átomo de nitrogênio da porção sulfonamida. As medidas estatísticas, o número de descritores (GCOD) e de amostras atípicas (*outliers*) para os 10 melhores modelos do alinhamento **2** estão apresentados na tabela 13.

Tabela 13. Medidas estatísticas, número de descritores de ocupação (GCOD) e de amostras atípicas obtidos para os 10 melhores modelos QSAR (N = 30) do alinhamento **2**

Alinhamento	n° GCOD	r^2	q^2	LSE	LOF	amostras atípicas (*outliers*)
1	5	0,92	0,89	0,04	0,11	1
2	6	0,94	0,89	0,03	0,12	1
3	6	0,93	0,88	0,04	0,14	1
4	**5**	**0,93**	**0,88**	**0,04**	**0,11**	**0**
5	5	0,92	0,88	0,04	0,12	2
6	5	0,92	0,87	0,04	0,12	1
7	5	0,93	0,86	0,04	0,11	1
8	5	0,92	0,86	0,04	0,12	1
9	5	0,91	0,86	0,05	0,13	2
10	5	0,91	0,86	0,05	0,13	2

O modelo **4** apresentou valores altos de q^2 (0,88) e de r^2 (0,93), valores baixos de erro (LSE = 0,04 e LOF = 0,11), 5 descritores de ocupação e nenhuma amostra atípica (tabela 14). Por isso, este modelo foi considerado para análises posteriores.

Tabela 14. Valores de resíduos obtidos para os inibidores de MMP-2 com o modelo **4** (N = 30) do alinhamento **2**

Composto	pCI_{50exp}	pCI_{50calc}	Resíduos	DP	2DP
5a$_1$	7,21	7,17	-0,03	0,20	0,41
5b	6,84	6,79	-0,04		
5c	6,29	6,66	0,38		
5g	6,39	6,66	0,27		
5j	6,60	6,68	0,09		
5l	6,85	6,68	-0,16		

Tabela 14 (cont.). Valores de resíduos obtidos para os inibidores de MMP-2 com o modelo **4** (N = 30) do alinhamento **2**

Composto	pCI_{50exp}	pCI_{50calc}	Resíduos	DP	2DP
5o	6,54	6,64	0,11		
13a	7,32	7,10	-0,22		
13b	7,25	6,94	-0,31		
7a	7,44	7,37	-0,06		
8b$_1$	8,39	7,99	-0,40		
8b$_2$	6,75	6,93	0,18		
8c	7,95	7,67	-0,27		
8d	8,00	8,19	0,19		
8e	8,02	7,76	-0,26		
8f	7,77	7,87	0,10		
8ir	7,62	7,68	0,07		
8is	7,92	8,31	0,40		
8k	7,19	7,05	-0,13		
8q	6,55	6,56	0,02		
8r	6,58	6,73	0,16		
8x	8,11	7,96	-0,14		
8z	8,26	8,19	-0,06		
11a	6,17	6,18	0,02		
8aa	7,32	7,14	-0,17		
8ee$_2$	5,88	5,93	0,06		
8cc	8,26	8,43	0,17		
8dd	8,60	8,76	0,16		
9b	6,54	6,69	0,16		
3b	7,13	6,86	-0,27		

A matriz de correlação-cruzada de resíduos de ajuste entre os pares dos 10 melhores modelos para o alinhamento **2** foi calculada para verificar se os modelos são equivalentes ($R > 0,5$) ou distintos ($R < 0,5$) (Tabela 15).

Tabela 15. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os 10 melhores modelos (N = 30) do alinhamento **2**

Modelo	1	2	3	4	5	6	7	8	9	10
1	**1,00**									
2	0,73	**1,00**								
3	0,79	0,82	**1,00**							
4	0,76	0,90	0,96	**1,00**						
5	0,91	0,78	0,82	0,81	**1,00**					
6	0,88	0,74	0,77	0,74	0,87	**1,00**				
7	0,81	0,73	0,96	0,88	0,83	0,77	**1,00**			
8	0,74	0,83	0,92	0,95	0,75	0,68	0,87	**1,00**		
9	0,79	0,77	0,86	0,85	0,86	0,73	0,90	0,87	**1,00**	
10	0,75	0,83	0,73	0,79	0,82	0,70	0,78	0,82	0,91	**1,00**

De acordo com a **Tabela 15**, os modelos gerados com o alinhamento **2** podem ser considerados equivalentes ($R \approx 1$), indicativo de que existe um único modelo de QSAR.

A próxima etapa foi verificar se os descritores presentes no modelo **4** fornecem informação independente ou distinta para a interpretação do modelo. Calculou-se, então, a matriz de correlação-cruzada entre pares de descritores do modelo **4** (Tabela 16). De acordo com os valores de *R* encontrados ($R < 0,5$; em módulo), os descritores de ocupação, GCOD, do modelo **4** não estão correlacionados entre si, ou seja, contribuem com informações distintas ao modelo QSAR obtido.

Tabela 16. Matriz de correlação-cruzada entre pares de descritores de ocupação (GCOD ou GC*i* (*x,y,z;IPE*)) do modelo **4** (N = 30), alinhamento **2**

GCOD	GC1 (-2,8,-4; *any*)	GC2 (-1,5,-2; *any*)	GC3 (-2,4,-1; *any*)	GC4 (0,-1,3; *np*)	GC5 (2,2,3; *any*)
GC1	**1,00**				
GC2	0,40	**1,00**			
GC3	-0,14	0,15	**1,00**		
GC4	-0,02	0,18	0,26	**1,00**	
GC5	0,04	-0,09	-0,23	-0,04	**1,00**

O modelo **4** é definido pela seguinte equação:

$$pCI_{50} = 8,91\ GC1_{(any)} - 0,77\ GC2_{(any)} - 0,78\ GC3_{(any)} + 2,09\ GC4_{(np)} + 1,11\ GC5_{(any)} + 6,70 \quad (\textbf{eq. 4})$$

$$N = 30 \quad r^2 = 0,93 \quad q^2 = 0,88 \quad LSE = 0,04 \quad LOF = 0,11 \quad \text{Amostras atípicas} = 0$$

De acordo com a equação **4**, o modelo **4** é composto por 5 descritores de ocupação (GC1 a GC5) pertencentes a duas classes de IPE, ocupação não-diferenciada ou referente a qualquer tipo de átomo na molécula (*any*) e átomo de caráter não polar ou apolar na molécula (*np*). Os descritores GC2 e GC3 desfavorecem a atividade inibitória frente à enzima MMP-2 (pCI_{50}), pois apresentam coeficientes de regressão negativos na equação **4**. Por outro lado, os descritores GC1, GC4 e GC5 contribuem positivamente à atividade inibitória da MMP-2.

Para verificar a robustez do modelo **4** e a chance deste modelo ter sido obtido por correlação ao acaso, os testes de validação cruzada LNO e de aleatorização do ***y***, respectivamente, foram aplicados ao conjunto de dados. Na validação cruzada LNO, o *N*

variou de 2 a 11 (36% das amostras). O valor de q^2_{LNO} permaneceu próximo ao valor de q^2_{LOO} até que 9 amostras (30%) fossem retiradas do conjunto de dados (2DP < 0,1). Com base neste dado, o modelo **4** foi considerado robusto. A tabela 17 contém os valores de q^2_{LNO} para cada réplica (triplicata), os valores médios de q^2_{LNO} ($<q^2_{LNO}>$) e os valores de dois desvios padrão (2DP), onde N = 2 a 11.

Tabela 17. Valores de q^2 obtidos na validação cruzada LNO (triplicata) para o modelo **4** (N = 30) do alinhamento **2**, com o conjunto de inibidores de MMP-2

N	q^2_{LNO}	q^2_{LNO}	q^2_{LNO}	$<q^2_{LNO}>$	2DP
2	0,88	0,88	0,88	0,88	0,00
3	0,87	0,86	0,85	0,86	0,03
4	0,87	0,87	0,83	0,86	0,04
5	0,87	0,88	0,82	0,86	0,07
6	0,87	0,84	0,88	0,87	0,04
7	0,90	0,85	0,88	0,88	0,05
8	0,86	0,87	0,88	0,87	0,02
9	**0,82**	**0,87**	**0,84**	**0,84**	**0,06**
10	0,76	0,56	0,82	0,71	0,28
11	0,75	0,89	0,86	0,83	0,16

Os gráficos de validação interna LNO e de aleatorização do ***y*** (10 corridas) para o modelo **4** (N = 30), do alinhamento **2**, estão apresentados na Figura 23:

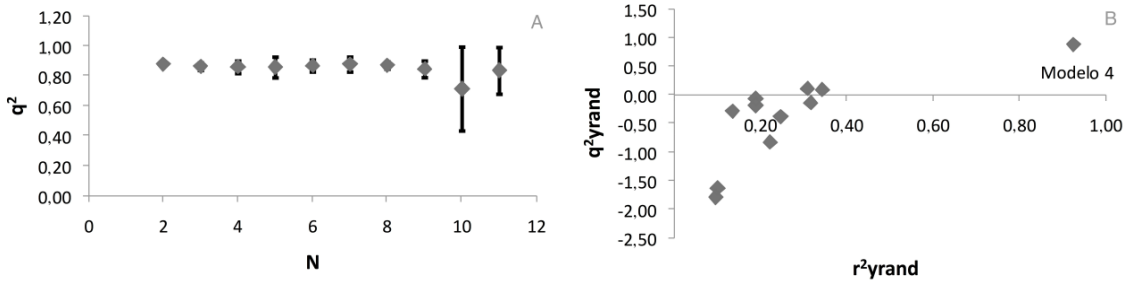


Figura 23. (A) Validação *leave-N-out* aplicada ao modelo QSAR final obtido para o conjunto de treinamento (N = 30). Os losangos representam a média e as barras indicam o desvio padrão (triplicata), para cada valor de *N*. O *N* crítico para o modelo de 30 inibidores foi correspondente à retirada de subconjuntos de 9 amostras (30 %). (B) Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de treinamento (N = 30) com 10 corridas ou aleatorizações do vetor ***y*** (pCI_{50} frente à enzima MMP-2).

O procedimento de aleatorização do ***y*** foi realizado com 10, 30 e 50 corridas e os modelos obtidos nestas condições apresentaram baixa qualidade e nenhum significado de acordo com as medidas estatísticas q^2_{yrand} e r^2_{yrand}. Considerando a abordagem de Wold and Eriksson (1995), o modelo **4** apresentou correlação ao acaso tolerável. Os valores de q^2_{yrand} variaram de 0,1 avalores negativos e os valores de r^2_{yrand} variaram de 0,1 a 0,39, em 10, 30 e 50 aleatorizações. A tabela 3A que apresenta os valores de q^2_{yrand} e r^2_{yrand} para as

10 randomizações e as figuras 3A e 4A referentes aos resultados para 30 e 50 aleatorizações (tabelas e gráficos), respectivamente, podem ser visualizados no Anexo A.

4.3.1.1. Validação externa

Para avaliar a capacidade de previsão externa do melhor modelo de QSAR obtido para o conjunto de treinamento (modelo **4**/alinhamento **2**), 8 inibidores (**5f**, **5h**, **14b**, **5k**, **8a**, **8n**, **8t**, **9a**) (tabela 2) foram utilizados como conjunto de avaliação. Os valores de pCI_{50} de cada um dos inibidores foram calculados utilizando a equação **4**, gerada com o conjunto de treinamento. Estes valores foram obtidos a partir da substituição dos valores dos descritores de ocupação, GCOD ou GC (listados na Tabela 18), para cada um dos inibidores do conjunto de avaliação.

Tabela 18. Valores dos descritores de ocupação (GCOD) calculados para os compostos do conjunto de avaliação com o programa 4D-QSAR 3.0 (HOPFINGER, 1999)

Composto	$GC1_{(any)}$ (-2,8,-4)	$GC2_{(any)}$ (-1,5,-2)	$GC3_{(any)}$ (-2,4,-1)	$GC4_{(np)}$ (0,-1,3)	$GC5_{(any)}$ (2,2,3)
5f	0	0,0026	0,0644	0	0,0018
5h	0	0	0	0	0,3672
14b	0	0	0	0	0,7376
5k	0	0	0,0612	0,3658	0,0008
8a	0	0	0	0	0,9402
8n	0	0	0	0	0,7374
8t	0,1274	0,2622	0	0,242	0,7086
9a	0	0	0	0	0,2198

A tabela 19 apresenta os valores experimentais e previstos/calculadosde pCI_{50} com o modelo **4**/alinhamento **2**, além dos valores de resíduos (diferença entre pCI_{50_exp} epCI_{50_calc}) para cada um dos inibidores, considerando um limite de confiança de 99 %.

Tabela 19. Valores de experimentais e calculados/previstos de pCI_{50} para o conjunto de avaliação de acordo com a equação **4** e valores de resíduo ($pCI_{50exp} - pCI_{50calc}$) considerando um limite de confiança de 99 % (um desvio padrão (DP) da média dos resíduos)

Composto	pCI_{50_exp}	pCI_{50_calc}	resíduos	DP
5f	6,23	6,95	0,72	0,73
5h	6,46	6,77	0,31	
14b	6,04	7,51	**1,46**	
5k	7,45	7,51	0,06	
8a	8,03	7,41	-0,62	
8n	7,56	7,10	-0,46	
8t	8,25	8,92	0,67	
9a	5,89	6,94	**1,04**	

pCI50_exp: valores de pCI50 experimental; pCI50_calc: valores de pCI50 calculados

Seis dos oito inibidores do conjunto de avaliação foram bem previstos pelo modelo, pois apresentaram valores de resíduo (pCI_{50_exp} -pCI_{50_calc}) menores ou iguais a um desvio-padrão da média dos resíduos do modelo (0,73). Então, a capacidade ou o poder de previsão do modelo **4**/alinhamento 2 foi de 75 %. O modelo **4** também apresentou valor alto de coeficiente de correlação entre os valores de atividade calculada e experimental dos compostos do conjunto de avaliação (R = 0,80), reforçando sua confiabilidade. O gráfico da atividade experimental *versus* a atividade calculada está apresentado na figura 24.

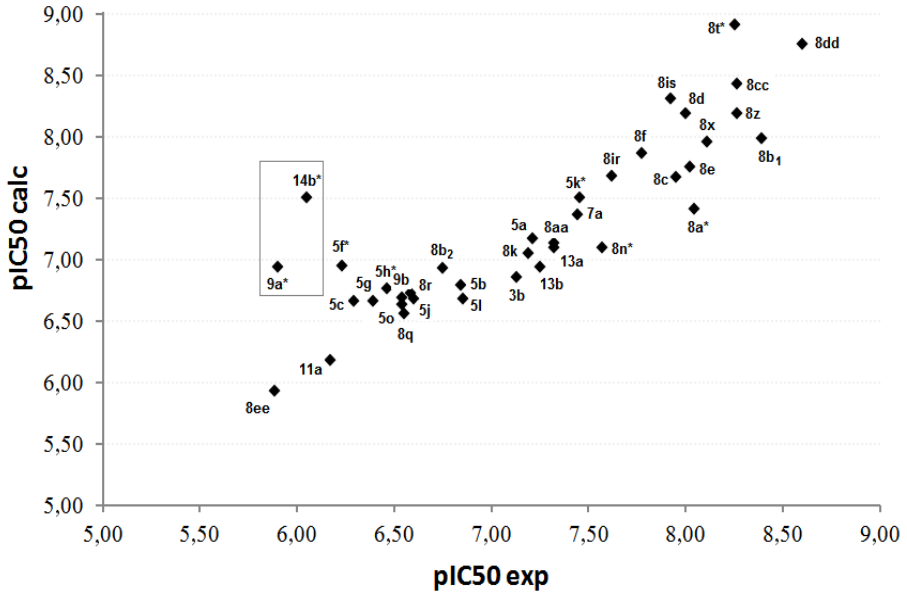


Figura 24. Gráfico com valores experimentais e calculados de pCI_{50} para o modelo **4** (N = 30) do alinhamento **2**. Os compostos do conjunto de avaliação estão representados com asterisco (*). Os compostos **9a*** e **14b*** não foram bem previstos pelo modelo **4** considerando o intervalo de confiança de 99 %.

Dentre os compostos que não foram bem previstos pelo modelo **4** está o composto **14b**. A razão pela qual ele pode ter sido considerado amostra atípica foi discutida anteriormente para o modelo obtido com o conjunto total (N = 38). Já o composto **9a** (pCI_{50} = 5,90) tem os mesmos grupos substituintes que o composto **5k** (pCI_{50} = 7,45), por exemplo, mas ao invés do grupo sulfonil ligado ao nitrogênio n°9, possui um grupo carbonila. O composto **9b** (pCI_{50} = 6,54)**,** do conjunto de treinamento, apresenta o mesmo padrão de substituição do **9a** no nitrogênio n°9, mas tem um grupo isopropil na posição R_2 ao invés de metil (**9a**).

4.3.1.2. Conformação bioativa prevista e interpretação dos descritores de ocupação do modelo QSAR (N = 30)

Após os procedimentos de validação, a hipótese de conformação bioativa para cada um dos inibidores do conjunto de treinamento foi calculada a partir do modelo **4**/alinhamento **2**. Os confôrmeros que apresentaram variação de energia (ΔE) de até 5,0 kcal/mol em relação ao confôrmero de energia mais baixa do PAC de cada inibidor foram considerados pelo programa 4D-QSAR 3.0 (HOPFINGER, 1999) nesta etapa. Os valores dos descritores da amostragem de confôrmeros, para cada ligante, foram substituídos na equação do modelo **4** (**eq. 4**) a fim de gerar os valores calculados/previstos de pCI_{50}. O confôrmero que apresentou o maior valor calculado de pCI_{50} foi selecionado pelo programa como a conformação bioativa prevista (HOPFINGER, 1999). A conformação bioativa prevista com os respectivos descritores de ocupação (GCOD ou GC), para o modelo **4**/alinhamento **2**, foi avaliada para cada inibidor do conjunto de treinamento. Nesta análise, foi possível observar um padrão de distribuição dos descritores, GCOD, na estrutura geral dos compostos, relacionado ao gradiente de atividade biológica (compostos mais e menos ativos). Para ilustrar, utilizou-se como exemplo de interpretação do modelo **4**/alinhamento **2**, a conformação bioativa prevista com os respectivos descritores de ocupação (GCOD ou GC) do inibidor mais (**8dd,** A) e menos (**8ee$_2$**, B) ativo do conjunto de compostos investigados (Figura 25). Estes estados conformacionais foram calculados utilizando uma grade virtual de 1 Å de resolução.

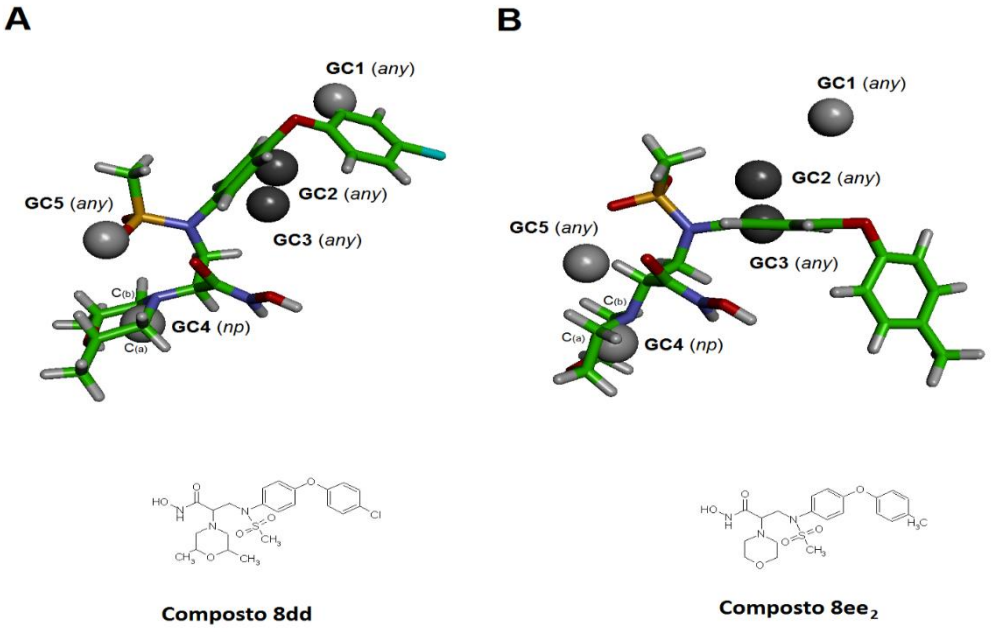


Figura 25. Conformações bioativas previstas com os respectivos GCOD para o composto (A) mais (**8dd**; enantiômero *S* em C_7) e (B) menos (**8ee$_2$**; enantiômero *R* em C_7) ativo do conjunto de

inibidores de MMP-2, utilizando o modelo **4**/alinhamento **2**. Os descritores de ocupação, GCOD ou GC, que contribuem de forma favorável e desfavorável à atividade inibitória estão representados por esferas na cor cinza mais claro e mais escuro, respectivamente. Os inibidores estão apresentados em modelo molecular de tubo ou palito e os carbonos estão representados na cor verde, os oxigênios em vermelho, hidrogênios em branco, enxofre em amarelo, cloro em mostarda e nitrogênio em azul.

Os descritores de ocupação (GCOD ou GC) que contribuem positivamente (**eq**. **4**) à atividade inibitória estão representados por esferas em cinza mais claro (GC1, GC4 e GC5), enquanto que os descritores que desfavorecem a ação inibitória (GC2 e GC3) correspondem às esferasemcinza mais escuro.

Para a conformação bioativa previstado inibidor mais ativo (**8dd**) com o modelo **4**/alinhamento **2** (equação **4**), a ocupação das células da grade virtual é mais favorecida nas regiões próximas ao substituinte na porção R_3 (GC4; 0,87 Å), à porção sulfonamida (GC5; 0,39 Å em relação ao átomo de oxigênio) e à região R_{1b} (GC1; 0,82 Å em relação ao átomo de carbono de um dos CH no anel aromático). A classe de IPE para o descritor GC4 é átomo de caráter não polar (*np*), ou apolar. Este GCOD está próximo ao átomo de carbono $C_{(a)}$ (figura 25) ligado ao nitrogênio do grupo dimetilmorfolino (0,87 Å). Este dado indica que grupos substituintes naquela região da molécula deveriam apresentar características preferencialmente apolares, a fim de contribuírem de forma favorável à atividade inibitória.

Os descritores de ocupação GC2 e GC3 não apresentam distinção em relação ao tipo de átomo (IPE = *any*). Sendo assim, substituições moleculares de qualquer natureza próximas à região do anel aromático ligado à porção sulfonamida deveriam ser evitadas (GC2 está a 1,21 Å de distância em relação ao carbono próximo à porção R_{1a}; GC3 está a 2,17 Å de distância em relação ao átomo de carbono próximo à porção sulfonamida).

Como esperado, o inibidor menos ativo do conjunto de dados, **8ee2**, apresentou padrão de distribuição diferente dos descritores de ocupação (GCOD ou GC), exceto para o descritor GC4 (0,94 Å próximo ao $C_{(a)}$). O descritor GC5 está mais próximo ao átomo de carbono $C_{(b)}$ ligado ao nitrogênio do grupo morfolino (2,36 Å; região R_3) ao invés do átomo de oxigênio da porção sulfonamida (2,56 Å). Os descritores GC2 e GC3 estão localizados na região do anel aromático ligado à porção sulfonamida, mas estão mais próximos ao carbono ligado diretamente ao átomo de nitrogênio (0,10 Å e 1,79 Å, respectivamente). A região do descritor GC1, por sua vez, não está ocupada pelo grupo R_{1b}. Este descritor está distante do átomo de carbono (CH, R_{1b}; 6,60 Å) no anel aromático. Todas as diferenças discutidas em relação à distribuição dos descritores de ocupação seriam responsáveis por reduzir a atividade inibitória frente à enzima MMP-2.

Para determinar o se o modelo **4** do alinhamento **2** se ajusta, do ponto de vista estereoquímico, ao sítio ativo e se há complementaridade no sítio de interação entre os resíduos de aminoácido importantes da enzima e os descritores de ocupação do ligante, as conformações bioativas previstas de cada composto do conjunto de treinamento e os respectivos GCOD (GC) foram ancorados manualmente no sítio catalítico do modelo molecular 3D da enzima MMP-2. Utilizou-se o alinhamento **2** (3 átomos) para orientar o ancoramento molecular dos ligantes. Este procedimento permite, também, a diferenciação ou identificação do tipo de átomo para os descritores de ocupação da classe de IPE definidos como *any* no modelo QSAR selecionado. A estrutura 3D do domínio catalítico da enzima MMP-2 complexada a um inibidor hidroxamato (SC-74020, figura 26), obtida por RMN (**PDB ID** 1HOV, FENG *et al*., 2002), foi empregada nesta investigação.

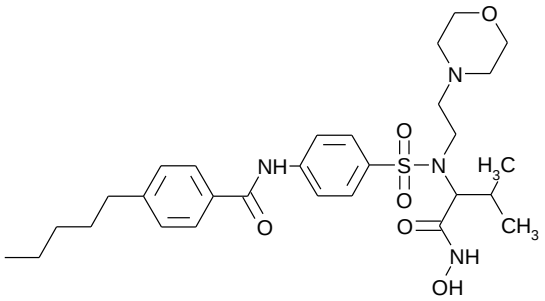


Figura 26. Estrutura química do inibidor de MMP-2, SC-74020 (FENG *et al*. 2002).

As coordenadas cartesianas do inibidor SC-74020 foram utilizadas como referência para ancorar as conformações bioativas previstas e os descritores de ocupação (modelo **4**/alinhamento **2**) dos compostos do conjunto de treinamento no sítio catalítico da MMP-2. O relaxamento dos complexos moleculares, para melhor acomodar os ligantes no sítio de interaçãofoi obtido a partir de simulações computacionais de DM de curta duração (50 ps) em 298 K, cada passo de 1 fs (MOLSIM 3.2; DOHERTY, 2002). O complexo de energia mais baixa para cada ligante foi selecionado da região de equilíbrio da simulação e, após minimização de energia, foi considerado para a avaliação dos descritores no sítio da MMP-2 (complementaridade).

A seguir estão apresentados os resultados obtidos para o processamento da proteína MMP-2 antes de ser utilizada na etapa de investigação da complementaridade no alvo do modelo QSAR-4D IR obtido para o conjunto de inibidores hidroxamatos.

4.3.1.3. Processamento da proteína para os estudos de ancoramento molecular

A estrutura da enzima MMP-2 complexada ao inibidor hidroxamato SC-74020, depositada no banco de dados **PDB** com o código 1HOV (FENG *et al*. 2002), foi utilizada como modelo para construção da geometria do receptor, conforme descrito em Materiais e Métodos (item **3.9.1**). A energia do modelo molecular 3D do complexo foi minimizada considerando o critério de convergência de 0,5 kcal/mol, com 30 iterações para o método de declive máximo e 112 iterações para o método de gradientes conjugados (MOLSIM 3.2; DOHERTY, 2002), à temperatura de 298 K. O modelo molecular do complexo MMP-2/SC-74020 otimizado foi o dado de entrada (*input*) para as simulações de DM no processo de preparação da proteína. O programa MOLSIM 3.2 (DOHERTY, 2002) foi também empregado para realizar as simulações de DM do modelo molecular 3D do complexo previamente processado. Empregou-se como protocolo para as simulações de DM um esquema de aquecimento lento *(warming-up*), com simulações curtas (50 ps) em T crescente (50 K, 100 K, 200 K, 300 K). O confôrmero de energia mais baixa selecionado da simulação a 300 K, após minimização de energia, foi utilizado como *input* para a simulação de 2 ns a 298 K (os gráficos dos PAC das simulações de 2ns estão apresentados no Anexo A - figura 5A). Este tipo de esquema de aquecimento é aplicado para gerar o estado de menor energia do complexo de forma mais rápida, permitindo que o sistema transponha barreiras energéticas (TOKARSKI, HOPFINGER, 1997). De acordo com os PAC provenientes das simulações de DM apresentados na figura 27, o modelo molecular do complexo MMP-2/SC74020 atingiu o equilíbrio gradativamente e estabilizou completamente apenas na simulação de DM de 2ns a 298 K. Uma amonstragem (100 ps) da região de equilíbrio do PAC, da simulação mais longa a 298 K, está apresentada na figura 27 (e).

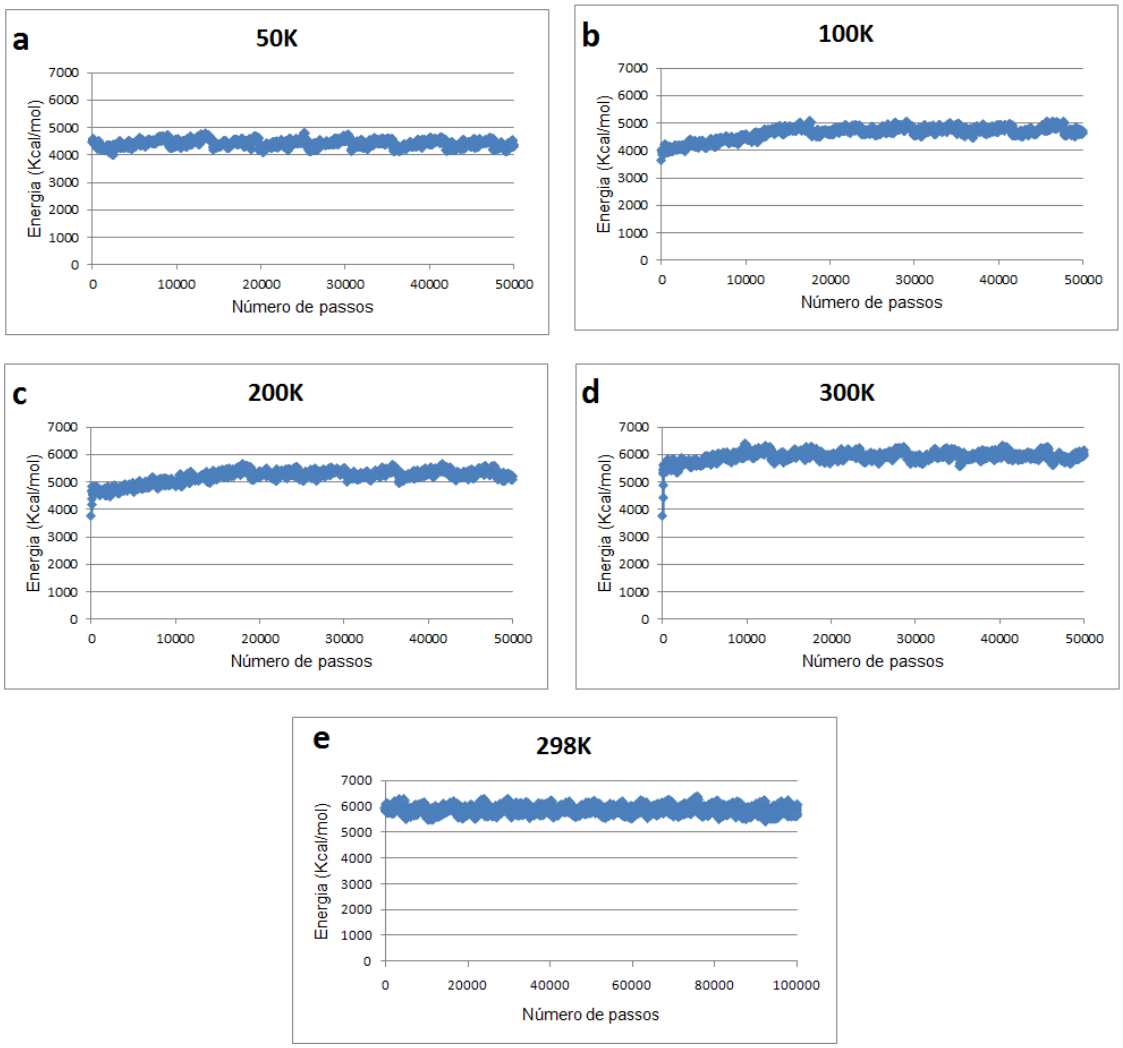


Figura 27. PAC obtido para as simulações de DM resultantes do esquema de aquecimento lento (50 ps) e para a simulação mais longa (2 ns) com o complexo MMP-2/SC-74020 (FENG *et al.*, 2002). (**a**) Simulação de DM a 50 K; (**b**) simulação de DM a 100 K; (**c**) simulação de DM a 200 K; (**d**) simulação de DM a 300 K; (**e**) simulação de DM a 298 K. Para a simulação de 2 ns (298 K) está representada apenas uma porção de 100 ps referente à região de equilíbrio do PAC. As contribuições de energia intramolecular de solvatação e de ligação de hidrogênio não foram consideradas no valor de energia (eixo y; kcal/mol).

O valor de energia potencial total (soma de todas as contribuições de energia, inclusive de solvatação e de ligação de hidrogênio) para o confôrmero de energia mais baixa do complexo MMP-2-SC-74020, selecionado em cada T de simulação, foi computado (Tabela 20) e verificou-se a manutenção de integridade estrutural a partir dos valores de desvio quadrático médio de posições atômicas (R*oot Mean Square Deviation*, RMSD) em relação à estrutura 3D de referência, obtida por RMN. Os resultados indicaram que, até 298 K, os modelos moleculares selecionados mantiveram a integridade estrutural em relação ao complexo original, uma vez que os valores de RMSD não ultrapassaram 1,8 Å (figura 28).

Tabela 20. Valores das contribuições de energia obtidos em cada T de simulação para o confôrmero de energia mínima do complexo MMP-2/SC-74020 (**PDB ID** 1HOV) selecionado da região de equilíbrio do PAC

T (K)	Eest	Eang	Etors	E1,4	EvdW	Eel	INTERV+C	Esolv	Ehb	Etot
50	1932,44	1259,52	1185,05	3956,13	-731,18	-3536,9	80,33	-178,13	-20778,15	-16810,9
100	2054,64	1430,02	1214,12	3957,51	-698,58	-3557,58	65,64	-169,86	-20882,21	-16586,3
200	2285,93	1653,43	1272,79	4033,64	-663,34	-3616,28	61,73	-158,59	-20705,88	-15836,6
300	2513,75	1792,46	1328,47	4088,79	-594,32	-3608,07	79,62	-189,68	-20062,40	-14651,4
298	2431,92	1867,57	1315,12	4084,95	-646,19	-3562,14	73,32	-180,73	-19857,08	-14473,3

* Eest = contribuição de energia intramolecular de deformação axial ou estiramento; Eang = contribuição de energia intramolecular de deformação angular; Etors = contribuição de energia intramolecular torsional; $E_{1,4}$ = contribuição de energia intramolecular de interação do tipo 1-4 ou Lennard-Jones; EvdW = contribuição de energia intramolecular de van der Waals; Eel = contribuição de energia intramolecular eletrostática; INTERV+C = soma das contribuições de energia intermolecular eletrostática e de van der Waals; Esolv = contribuição de energia intramolecular de solvatação; E_{Hb} = contribuição de energia intramolecular de ligação de hidrogênio; Etot = energia potencial total, que corresponde à soma de todas as contribuições de energia. As unidades de todas as contribuições de energia estão em kcal/mol.

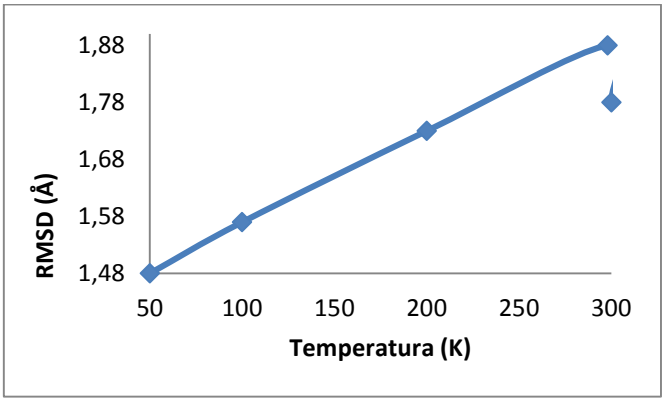


Figura 28. Valores de RMSD obtidos na comparação (sobreposição) entre os confôrmeros dos complexos selecionados em cada T de simulação, do esquema de aquecimento lento, e o complexo MMP-2/SC-74020 experimental (RMN; **PDB ID** 1HOV, FENG *et al*., 2002).

Então, utilizou-se o confôrmero de energia mais baixa do complexo MMP-2-SC-74020, selecionado da simulação de DM em 298 K (2 ns), após a minimização, para o ancoramento molecular das conformações bioativas previstas com os descritores de ocupação, GCOD ou GC. Após o ancoramento de cada conformação bioativa com os descritores de ocupação do modelo **4**/alinhamento **2**, o inibidor referência (SC-74020) foi deletado do sítio de interação e procedeu-se as etapas de minimização de energia e simulações de relaxamento (50 ps a 298 K) dos complexos, a fim de se obter uma melhor acomodação dos ligantes. O confôrmero de energia mais baixa de cada complexo selecionado do PAC da simulação de 50 ps, após a minimização de energia, foi utilizado para a análise de complementaridade ligante-enzima, já mencionada. Este tipo de avaliação permitiu a interpretação e, de certa forma, a validação dos descritores de ocupação gerados com a análise QSAR-4D IR. A elucidação dos tipos de interação

possíveis entre os inibidores + GCOD e os resíduos de aminoácidos do sítio catalítico da enzima alvo validaria o mapeamento do farmacóforo obtido com o modelo **4**/alinhamento **2** IR.

Avaliaram-se, então, as distâncias entre os resíduos de aminoácidos do sítio catalítico (Leu82, Leu83, Ala84, His85, Leu116, Val117, His130, Leu137, Met39, Pro140, Ile141, Tyr142, Thr143, Thr145, Lys146, Asn147, Leu150) e os descritores de ocupação (GCOD ou GC) associados às conformações bioativas previstas. Todos os possíveis conjuntos de interações entre inibidor-enzima (complementaridade), com base no modelo e alinhamento da análise QSAR-4D IR, podem ser identificados neste procedimento.

As interações encontradas para o inibidor mais ativo (**8dd**) do conjunto de treinamento foram comparadas às interações encontradas para o inibidor SC-74020 no sítio catalítico da enzima MMP-2 (1HOV, FENG *et al*., 2002). Apesar dos inibidores **8dd** e SC-74020 pertencerem a classes químicas relacionadas, algumas diferenças puderam ser observadas considerando principalmente as interações intermoleculares estabelecidas com os resíduos de aminoácidos do sítio catalítico da MMP-2 (figura 29).

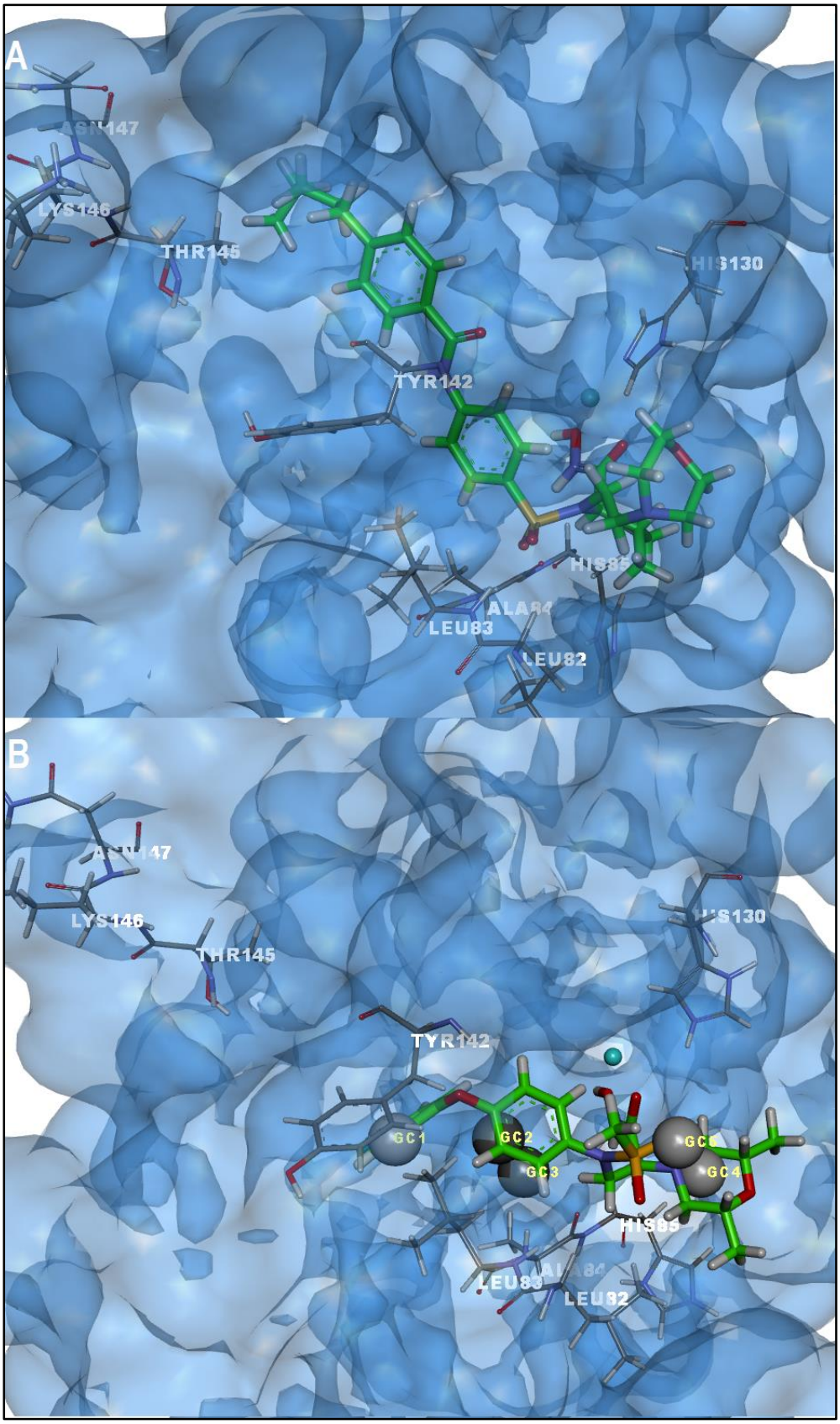


Figura 29. (A) Inibidor SC-74020 no sítio catalítico da enzima MMP-2. (B) Visualização da conformação bioativa prevista para o inibidor mais ativo (**8dd**) com os descritores de ocupação (GCOD, GC1-GC5). Os inibidores estão representados em modelo tubo ou palito e os átomos de carbono estão em verde, oxigênios em vermelho, nitrogênio em azul, enxofre em amarelo, cloro em azul ciano e hidrogênios em branco. Os íons de zinco (azul claro) e os descritores de ocupação

[GCOD ou GC; esferas em cinza claro (favoráveis) e cinza escuro (desfavoráveis)] estão representados em modelo CPK (preenchido). Os resíduos de aminoácido Leu82, Leu83, Ala84, His85, His130, Tyr142, Thr145, Lys146, Asn147 estão representados em modelo palito. Os átomos de carbono estão em cinza, oxigênio em vermelho e nitrogênio em azul. A superfície molecular da proteína MMP-2 está demonstrada em azul.

Os resíduos de aminoácido situados no fundo do bolso acessório S1' (Thr145, Lys146, Asn147) não estabelecem qualquer tipo de interação com porção alguma do inibidor mais ativo (**8dd**) do conjunto de treinamento. A estrutura deste inibidor é menor do que a do composto SC-74020 e se orienta no sítio catalítico de forma a acomodar a porção mais hidrofóbica (R_1) em um bolso adjacente ao sítio acessório S1', o que sugere um tipo de orientação diferente (modo de interação) para estes derivados no sítio catalítico da MMP-2. No entanto, a porção do ácido hidroxâmico em ambos os derivados, **8dd** e SC-74020, interage com o metal Zn no sítio catalítico. O carbono carbonílico e o oxigênio da hidroxila da porção hidroxâmica apresentam praticamente a mesma distância em relação ao átomo de Zn (2,1 Å e 2,0 Å), respectivamente.

O aumento de atividade inibitória (pCI_{50}) é previsto para a ocupação das células da grade virtual com átomos do ligante em GC1, GC4 e GC5 por tipos/classes apropriados de IPE. A situação oposta (diminuição de atividade) é verdadeira para a ocupação em GC2 e GC3 (**eq.4**). Os descritores de ocupação não apresentam qualquer especificidade em relação ao tipo de IPE (classe não diferenciada), exceto para GC4, que apresenta como classe de IPE átomo de caráter não polar ou apolar.

Considerando o ancoramento do inibidor **8dd** no sítio ativo da MMP-2 (Figura 29), o descritor GC4 está próximo à cadeia lateral do resíduo His85 (folha-β IV) (3,9 Å de distância do carbono ligado ao anel imidazol), da mesma forma que estáo grupo isopropil do inibidor SC-74020, reforçando o tipo de interação de G4 (interação hidrofóbica) no sítio catalítico da MMP-2.

O descritor GC1 está localizado próximo à cadeia lateral do resíduo de aminoácido Tyr142 (2,4 Å) e ao grupo metil do resíduo Leu83 (1,9 Å do átomo de carbono e 1,6 Å a 2,3 Å dos átomos de hidrogênio). A classe de IPE deste descritor é não diferenciada (*any*), ou seja, qualquer tipo de ocupação por qualquer tipo de átomo ou grupo substituinte. A não especificidade para o tipo de átomo em GC1 pode ser explicada pela sua localização no sítio catalítico. Este descritor de ocupação pode estabelecer ligações hidrofóbicas com o grupo metila da cadeia lateral da Leu83 ou com a cadeia lateral do aminoácido Tyr142.

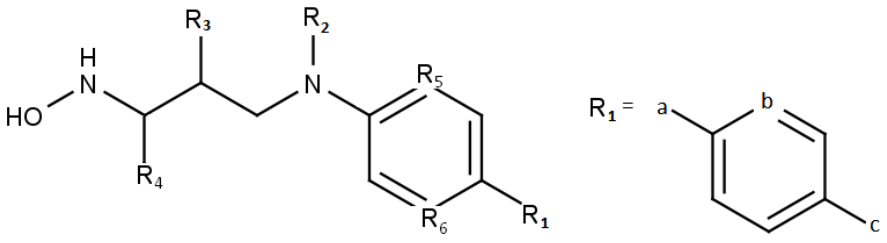
O descritor GC5 se encontra na vizinhança do carbono do grupo metil da cadeia lateral da Leu83 (5,8 Å) e do nitrogênio do anel imidazol da His130 (5,2 Å). Então, o

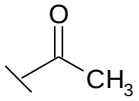
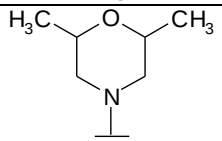
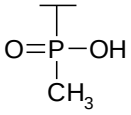
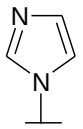
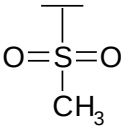
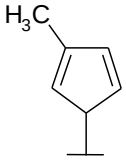
descritor GC5 poderia estabelecer interações do tipo hidrofóbica e polar com a cadeia lateral da Leu83 e His130, respectivamente. Cabe ressaltar que explorar as interações com o resíduo His130 poderia constituir em nova abordagem para o desenvolvimento de novos inibidores de MMP-2.

Por outro lado, grupos substituintes de qualquer tipo nas coordenadas/posições dos descritores GC2 e GC3 poderiam diminuir a atividade inibitória. Estes descritores de ocupação estão relacionados ao anel aromático ligado ao nitrogênio da porção sulfonamida, que deve estabelecer interações com os resíduos Leu82, Leu83, e Ala84 no sítio ativo da MMP-2.

O melhor modelo de QSAR (modelo **4**/alinhamento **2**) foi estatisticamente validado, robusto e, também, foi passível de interpretação tanto do ponto vista químico como biológico. Então, este modelo pode ser considerado confiável para prever novos derivados capazes de inibir a MMP-2. Assim, uma quimioteca (N = 11) foi construída e a atividade inibitória dos compostos foi calculada utilizando o modelo **4**/alinhamento **2**. Os compostos foram planejados a partir da estrutura do inibidor mais ativo do conjunto de dados (**8dd**), de acordo com as informações de substituições fornecidas pelo modelo. Ainda, variações na porção R_3 foram consideradas, mantendo-se grupos volumosos neste local. Os três inibidores mais promissores estão apresentados na Tabela 21.

Tabela 21. Ligantes propostos e respectivos valores atividade inibitórias frente à MMP-2 calculados com o modelo **4**/alinhamento 2



Ligantes	R_{1a}	R_{1b}	R_{1c}	R_2	R_3	R_4	PCI_{50cal}
L4	O	CH	Cl			H_2	8,12
L6	O	CH	Cl			O	7,33
L7	O	CH	Cl			O	7,03

O inibidor proposto que apresentou maior atividade foi o ligante **L4**, sugerindo que apenas um grupo para coordenar com o átomo de Zn seria suficiente para manter a atividade catalítica da MMP-2. A porção sulfonamida foi substituída por uma amida, mas se mantiveram os grupos R_1 e R_3 presentes no inibidor mais ativo do conjunto, **8dd**.

Então, informações importantes foram fornecidas a partir da análise QSAR-4D IR para direcionar e auxiliar no planejamento racional de novos inibidores potencialmente ativos contra MMP-2.

O estudo de QSAR-4D IR para os inibidores de MMP-2 foi apresentado no evento Globalization in Pharmaceutics Education Network (GPEN), 2012 e no 17° Simpósio Brasileiro de Química Teórica, 2013; e publicado no periódico *Molecular Informatics* em 2014 (**Anexo B**).

4.3.1.4. Previsão da atividade do 4-NC a partir do melhor modelo de QSAR (N = 30) obtido para o conjunto de inibidores de MMP-2

Com o intuito de verificar a capacidade de previsão do melhor modelo QSAR obtido com o conjunto de inibidores de MMP-2 para classes de moléculas não quimicamente relacionadas, calculou-se a atividade inibitória prevista para o modelo molecular tridimensional do 4-NC.

Os descritores de ocupação, GCOD, para o composto 4-NC foram calculados nas mesmas condições utilizadas para o conjunto de inibidores hidroxamatos e os valores obtidos para os descritores (Tabela 22) foram substituídos na equação do modelo **4**/alinhamento **2**. O valor de pCI_{50} calculado para o 4-NC com a **eq.4** foi de 6,70. De acordo com este modelo, a atividade inibitória do 4-NC frente à enzima MMP-2 seria considerada moderada.

Tabela 22. Valores dos descritores de ocupação, GCOD ou GC, para o 4-NC calculados com o modelo QSAR de inibidores de MMP-2 (modelo **4** do alinhamento **2**)

GC1	GC2	GC3	GC4	GC5
0	0	0	0,001	0

A conformação bioativa para o 4-NC também foi prevista (Figura 30, A) e ancorada no sítio catalítico da enzima MMP-2 (Figura 30, B) com os descritores de ocupação do modelo **4**/alinhamento **2**.

O padrão de distribuição dos descritores definiu a atividade moderada prevista para o 4-NC. O descritor GC1, por exemplo, de classe de átomo não diferenciada está bem distante da molécula e tem caráter favorável à atividade inibitória. O descritor GC4 (favorável à atividade inibitória), átomo de caráter não polar, está próximo ao oxigênio de um dos grupos hidroxila que provavelmente coordenaria com o Zn. A validação da distribuição de descritores (GCOD ou GC) e consequente gradiente de atividade só será possível com o ensaio experimental de inibição do 4-NC frente à MMP-2.

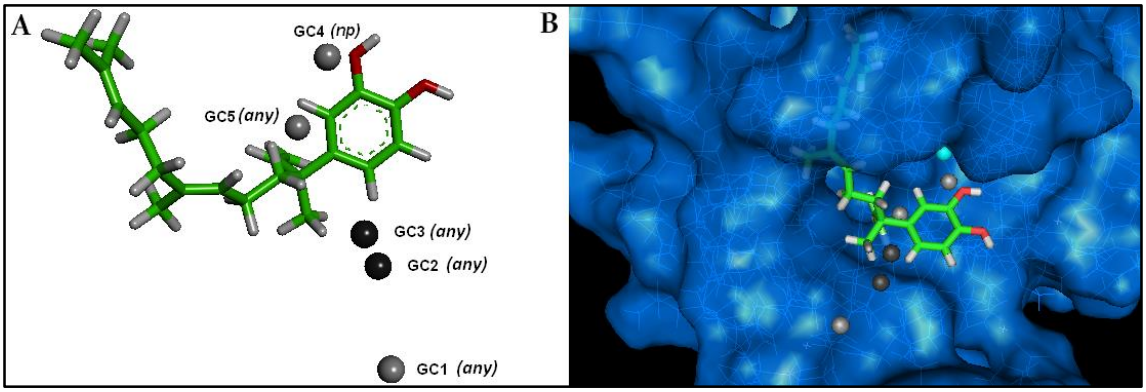


Figura 30. (A) Conformação bioativa prevista para o 4-NC com os descritores de ocupação, GCOD ou GC, utilizando o modelo **4**/alinhamento **2**. (B) Conformação bioativa prevista para o 4-NC com os descritores de ocupação no sítio catalítico da enzima MMP-2. O modelo molecular 3D do 4-NC está representado em tubo ou palito, onde os carbonos estão em verde, os oxigênios em vermelho e os hidrogênios em branco. Os descritores que desfavorecem a atividade estão representados em esferas cinza escuro e os que favorecem em esferas cinza claro (Pymol 1.5.0.1, DeLano, 2012).

4.3.2. Aplicação do formalismo QSAR-4D IR ao conjunto de hidroxamatos com atividade inibitória frente à enzima MMP-9

A análise de QSAR-4D IR foi aplicada ao conjunto inicial de dados constituído por 64 compostos (amostragem mais representativa), com atividade inibitória (baixa, média e alta) frente à enzima MMP-9 (a YANG, 2008; b YANG, 2008), para cada um

dos 3 alinhamentos propostos (tabela 4). Estes alinhamentos exploraram a estrutura comum às moléculas em toda sua extensão e foram considerados o grupo farmacofórico, representado pela porção hidroxamato (responsável por quelar o átomo de zinco das enzimas MMP) e a porção sulfonamida, que permite a interação dos inibidores com a cadeia polipeptídica das enzimas por meio de ligações de hidrogênio (a YANG, 2008). Os modelos QSAR gerados foram avaliados em relação às medidas estatísticas, número de descritores (GCOD/IPE) e número de amostras atípicas. Na tabela 23 estão apresentados os resultados obtidos para três alinhamentos investigados (**1**, **2** e **3**), para os 10 melhores modelos, utilizando 100.000 operações genéticas, 10% de mutação em cada ciclo de recombinação (crossover) e fator de ajuste de 2,6 na etapa de otimização empregando GFA-MLR.

Tabela 23. Medidas estatísticas, número de descritores de ocupação (GCOD) e de amostras atípicas para os 10 melhores modelos de QSAR (N = 64) obtidos para cada alinhamento (**1**,**2**,**3**) utilizando fator de ajuste de 2,6

Alinhamento	n° GCOD	r^2	q^2	LSE	LOF	amostras atípicas (*outliers*)
1	13-14	0,93-0,94	0,89-0,90	0,06	0,27-0,31	1-3
2	**10-11**	**0,86-0,88**	**0,82-0,83**	**0,12-0,13**	**0,37-0,42**	**0-2**
3	12-14	0,92-0,93	0,87-0,88	0,07-0,08	0,30-0,37	1

Todos os alinhamentos apresentaram modelos com valores altos de q^2 (0,82 a 0,90) e r^2 (0,86 a 0,94) e valores baixos de erro de LSE (0,06 a 0,13) e valores de LOF que variaram de 0,27 a 0,42, conforme apresentado na tabela 23. No entanto, somente para o alinhamento **2** foram gerados modelos de QSAR sem quaisquer amostras atípicas. Então, o alinhamento **2** foi considerado o melhor alinhamento. A tabela 24 apresenta os 10 melhores modelos obtidos com o alinhamento **2**.

Tabela 24. Medidas estatísticas, número de descritores de ocupação (GCOD) e de amostras atípicas (*outliers*) obtidos para os 10 melhores modelos de QSAR (N = 64) do alinhamento **2**

Modelo	n° GCOD	r^2	q^2	LSE	LOF	amostras atípicas
1	11	0,88	0,83	0,12	0,39	1
2	11	0,88	0,83	0,12	0,39	2
3	11	0,88	0,82	0,12	0,38	2
4	11	0,88	0,82	0,12	0,38	2
5	11	0,88	0,82	0,12	0,38	2
6	**10**	**0,86**	**0,82**	**0,13**	**0,38**	**0**
7	11	0,88	0,82	0,12	0,38	2
8	10	0,87	0,81	0,13	0,37	1
9	11	0,87	0,81	0,13	0,43	1
10	10	0,87	0,81	0,13	0,37	1

Os dez melhores modelos de QSAR do alinhamento **2** apresentarem valores altos de q^2 (0,81 a 0,83) e de r^2 (0,86 a 0,88) e valores baixos de LSE (0,12 a 0,13) e valores de LOF variando de 0,38 a 0,43. O único modelo que não apresentou amostras atípicas foi o modelo **6**. Então, a matriz de correlação-cruzada de resíduos de ajuste entre pares dos 10 melhores modelos obtidos com o alinhamento **2** foi calculada para verificar se todos os modelos são equivalentes (R > 0,5) ou distintos (R < 0,5). De acordo com os dados da tabela 25, os modelos gerados com o alinhamento **2** foram considerados equivalentes ($R \approx 1$), reforçando a escolha do modelo **6** como o melhor modelo para este alinhamento.

Tabela 25. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os 10 melhores modelos (N = 64) do alinhamento **2**

Modelos	1	2	3	4	5	6	7	8	9	10
1	**1,00**									
2	0,86	**1,00**								
3	0,99	0,85	**1,00**							
4	0,99	0,85	0,99	**1,00**						
5	0,98	0,84	0,99	0,99	**1,00**					
6	0,91	0,78	0,92	0,91	0,93	**1,00**				
7	0,85	0,98	0,86	0,86	0,86	0,80	**1,00**			
8	0,95	0,82	0,94	0,94	0,93	0,82	0,81	**1,00**		
9	0,78	0,91	0,78	0,78	0,78	0,86	0,91	0,70	**1,00**	
10	0,94	0,82	0,94	0,94	0,93	0,82	0,81	0,99	0,70	**1,00**

O modelo **6** apresentou 10 descritores de ocupação (GCOD), valor de q^2 igual a 0,82 e valor de r^2 de 0,86. Os valores de LSE e de LOF foram iguais a 0,13 e 0,38, respectivamente. Calculou-se, então, a matriz de correlação-cruzada entre pares de descritores do modelo **6** (tabela 26), que indicou que os descritores não estão correlacionados entre si (R < 0,5, em módulo) e que, portanto, contribuem com informação distinta ao modelo QSAR selecionado como o melhor modelo.

Tabela 26. Matriz de correlação-cruzada entre pares dos descritores de ocupação (GCOD; GC*i* $_{(x,y,z;IPE)}$) do modelo **6** (N = 64) utilizando o alinhamento **2**

GCOD	GC1 (5,4,4; *any*)	GC2 (6,0,2; *any*)	GC3 (1,2,5; *np*)	GC4 (5,1,0; *any*)	GC5 (3,5,-2; *ar*)	GC6 (0,-1,3; *np*)	GC7 (0,0,-3, *any*)	GC8 (-1,5,1; *hbd*)	GC9 (2,-3,4;*any*)	GC10 (1,3,1;*ar*)
GC1	**1,00**									
GC2	0,13	**1,00**								
GC3	0,07	-0,07	**1,00**							
GC4	0,13	0,28	-0,05	**1,00**						
GC5	0,06	-0,05	-0,02	-0,04	**1,00**					
GC6	0,12	0,35	-0,09	-0,05	-0,07	**1,00**				
GC7	0,34	-0,05	-0,06	-0,10	-0,03	-0,03	**1,00**			
GC8	0,10	0,30	-0,08	-0,07	-0,08	0,19	-0,15	**1,00**		
GC9	0,18	-0,15	-0,08	-0,12	-0,05	0,21	-0,15	-0,21	**1,00**	
GC10	0,17	0,24	-0,09	0,41	-0,07	0,07	-0,15	0,37	-0,20	**1,00**

De acordo com a tabela 27, o modelo **6**/alinhamento **2** não possui amostras atípicas. Os valores de resíduo das amostras (diferença entre a atividade experimental e calculada) não excederam 2DP da média do modelo (intervalo de confiança de 95 %; 2DP = 0,74).

Tabela 27. Valores de resíduos obtidos para os inibidores de MMP-9 com o modelo **6** (N = 64) do alinhamento **2**

Composto	pCI_{50exp}	pCI_{50calc}	Resíduos	DP	2DP
5c	6,91	7,22	0,31	0,37	0,74
12a	8,26	8,26	0,00		
13a	7,35	7,13	-0,22		
13b	7,32	7,07	-0,25		
13c	7,04	7,10	0,06		
14a	5,31	6,00	0,69		
14b	6,00	5,96	-0,04		
14c	5,82	5,79	-0,03		
3a	6,41	6,90	0,49		
$8b_2$	7,50	7,00	-0,50		
$5a_2$	7,92	7,41	-0,51		
$5a_1$	8,10	8,04	-0,06		
5b	7,55	7,80	0,25		
5c	6,90	6,89	-0,01		
5d	6,54	7,07	0,53		
5e	6,46	6,39	-0,07		
5f	7,03	6,59	-0,44		
5g	7,16	6,87	-0,29		
5h	7,26	7,43	0,17		
5j	7,40	7,50	0,10		
5k	8,32	8,04	-0,28		
5l	7,85	7,06	-0,79		
5m	6,32	7,07	0,75		
5n	7,26	7,84	0,58		
5o	6,98	6,91	-0,07		
$7a_2$	7,88	7,92	0,04		
$7a_1$	8,42	8,37	-0,05		
8	8,18	7,40	-0,78		
8aa	7,77	7,70	-0,07		
8a	8,62	8,67	0,05		
8bb	6,59	7,06	0,47		
$8b_2$	7,50	7,34	-0,16		
$8b_1$	9,28	8,64	-0,64		
8k	8,69	8,57	-0,12		
8c	8,88	8,99	0,11		
8dd	9,33	9,15	-0,18		
8d	8,77	9,24	0,47		
$8ee_2$	6,53	7,18	0,65		
$8ee_1$	8,72	8,68	-0,04		
8e	8,74	8,79	0,05		
8ff	9,16	9,27	0,11		
8s	8,55	8,06	-0,49		
8g	7,63	8,14	0,51		
8h	8,14	8,34	0,20		
$8b_1$	9,28	8,97	-0,31		
8c	8,88	8,99	0,11		

Tabela 27 (cont.). Valores de resíduos obtidos para os inibidores de MMP-9 com o modelo **6** (N = 64) do alinhamento **2**

Composto	pCI_{50exp}	pCI_{50calc}	Resíduos	DP	2DP
8l	9,05	8,71	-0,34	0,37	0,74
8m	7,69	8,29	0,60		
8n	9,20	8,82	-0,38		
8o	7,06	7,06	0,00		
8p	8,05	7,70	-0,35		
8q	7,16	7,14	-0,02		
8r	7,53	7,08	-0,45		
8s	8,53	8,72	0,19		
8t	8,95	9,26	0,31		
8u	8,49	8,60	0,11		
8v	8,31	7,66	-0,65		
8x	8,63	8,76	0,13		
8y	8,72	8,37	-0,35		
8z	8,60	8,34	-0,26		
9a	6,03	6,53	0,50		
9b	6,61	6,77	0,16		
8ir	8,82	8,84	0,02		
8is	9,26	9,73	0,47		

O modelo de regressão obtido com o conjunto de dados total (N = 64) foi submetido aos procedimentos de validação cruzada LNO (triplicata) e de aleatorização do vetor ***y*** para avaliar sua robustez e se foi obtido ao acaso, respectivamente. Retiraram-se até 31 % das amostras (20 compostos) do conjunto de 64 compostos. O valor de *N*crítico (LNO) para o modelo **6** foi identificado quando subconjuntos de 7 amostras (11%) foram removidos do conjunto de dados (2DP < 0,1). O modelo QSAR permanece estável com a retirada de um número pequeno de amostras. Entretanto, apenas o teste de validação interna LNO não é suficiente para excluir um modelo QSAR (KIRALJ, FERREIRA, 2009). A tabela 28 contém os valores de q^2_{LNO} para cada procedimento (triplicata), onde o *N* variou de 2 a 20, a média do q^2_{LNO} ($<q^2_{LNO}>$) e os valores de dois desvios padrão (2DP).

Tabela 28. Valores de q^2 obtidos no procedimento de validação cruzada LNO (triplicata) para o modelo **6** (N = 64) do alinhamento **2** com o conjunto de inibidores de MMP-9

N	q^2_{LNO}	q^2_{LNO}	q^2_{LNO}	$<q^2_{LNO}>$	2DP
2	0,82	0,82	0,82	0,82	0,00
3	0,81	0,81	0,81	0,81	0,00
4	0,81	0,82	0,82	0,82	0,00
5	0,81	0,81	0,81	0,81	0,00
6	0,81	0,80	0,82	0,81	0,01
7	0,82	0,79	0,82	0,81	0,02
8	**0,14**	**-2,22**	**0,80**	**-0,43**	**1,59**
9	0,74	0,80	0,82	0,79	0,04
10	0,82	0,80	0,75	0,79	0,03

Tabela 28 (cont.). Valores de q^2 obtidos no procedimento de validação cruzada LNO (triplicata) para o modelo **6** (N = 64) do alinhamento **2** com o conjunto de inibidores de MMP-9

N	q^2_{LNO}	q^2_{LNO}	q^2_{LNO}	$<q^2_{LNO}>$	2DP
11	0,82	0,66	0,70	0,73	0,08
12	0,75	-3,22	0,66	-0,60	2,27
13	0,77	-0,94	0,80	0,21	1,00
14	0,74	0,79	0,49	0,67	0,16
15	-1,30	0,24	0,00	-0,35	0,83
16	0,69	-1,79	0,78	-0,11	1,46
17	0,78	0,74	0,63	0,72	0,08
18	-1,99	0,62	0,51	-0,29	1,48
19	-1,79	0,54	0,62	-0,21	1,37
20	-1,99	-1,68	-0,71	-1,46	0,67

O modelo **6** (N = 64) foi submetido ao procedimento de aleatorização do vetor ***y***, com 10, 30 e 50 corridas. Neste processo, apenas a coluna de valores da atividade inibitória (***y***) é embaralhada e os dados do bloco X (variáveis independentes) permanecem sem quaisquer alterações. Os modelos foram gerados utilizando validação cruzada LOO, em cada corrida, que corresponde à mesma condição utilizada para gerar o modelo com os dados originais. De acordo com os critérios de Eriksson e Wold (1995) os modelos obtidos com o vetor ***y*** aleatorizado seriam classificados na categoria de correlação ao acaso não significativa ($0,2 < r^2_{yrand} < 0,3$; para 10 e 50 corridas), ou de correlação ao acaso tolerável ($0,3 < r^2_{yrand} < 0,4$; 30 corridas). O gráfico de aleatorização do vetor ***y*** (10 corridas) para o modelo **6** está disponível na Figura 31. As figuras 6A e 7A (Anexo A) apresentam as tabelas e os gráficos para 30 e 50 aleatorizações e a tabela 4A, apresenta os dados de aleatorização do vetor ***y*** com 10 corridas.

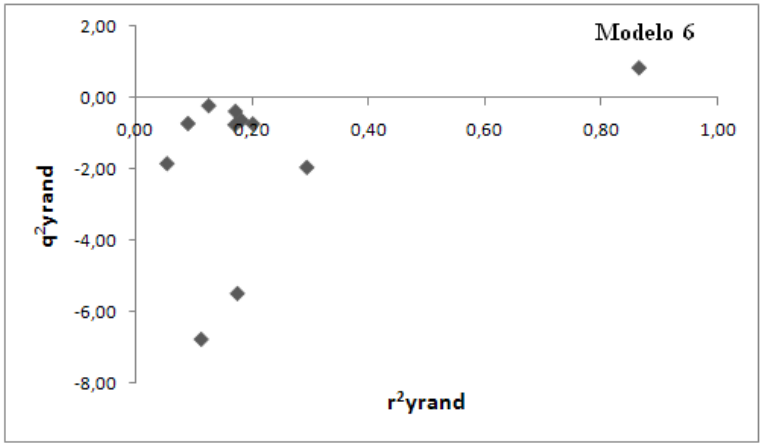


Figura 31. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de 64 inibidores com 10 aleatorizações do vetor ***y*** para o modelo 6 (atividade inibitória na MMP-9).

Apesar do modelo de QSAR ter sido obtido a partir de uma amostragem representativa (N = 64) e ser estatisticamente confiável, recomenda-se definir os conjuntos de treinamento e de avaliação, para desenvolver a validação externa. A validação externa, conforme já mencionado, é condição importante em uma análise de QSAR e tem a finalidade de verificar a capacidade de previsão do modelo QSAR obtido para amostras que não participaram na etapa de construção do modelo. Então, dividiram-se os dados originais em conjunto de treinamento (N = 47), utilizado para a construção dos modelos de QSAR, e conjunto de avaliação (N = 17; 26 % do total de amostras (KIRALJ, FERREIRA, 2009; GRAMATICA, 2007; GOLBRAIKH, TROPSHA, 2002). Nesta separação dos compostos, consideraram-se a diversidade estrutural e os gradientes de atividade inibitória (compostos altamente ativos, moderados e menos ativos). Os inibidores que compuseram o conjunto de avaliação foram os seguintes: **5e**, **5n**, **13a**, **13c**, **8d**, **8f**, **8g**, **8l**, **8q**, **8r**, **8s**, **8t**, **8u**, **8y**, **11a**, **8bb**, **3a** (figura 32; quadrados preenchidos em vermelho).

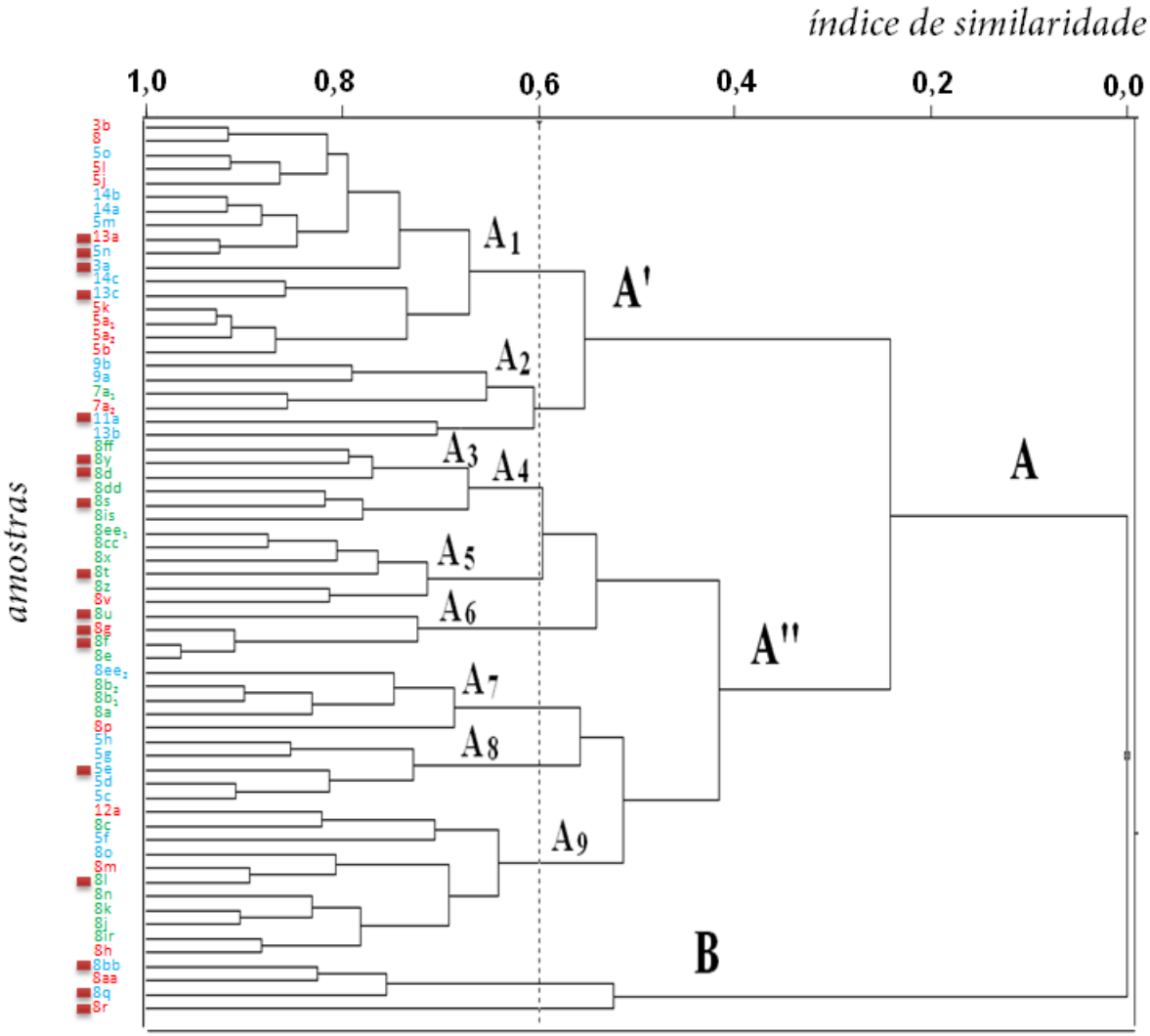


Figura 32. Dendrogramade amostras do conjunto de inibidores hidroxamatos com atividade inibitória frente à enzima MMP-9. Os quadrados preenchidos em vermelho (■) indicam os compostos que compõem o conjunto de avaliação (N = 17).

Os modelos de QSAR-4D IR foram construídos com o conjunto de treinamento para cada um dos três alinhamentos propostos, conforme apresentado na tabela 4. Os modelos gerados foram avaliados em relação às medidas estatísticas, número de descritores de ocupação (GCOD/IPE) e número de amostras atípicas (*outliers*). Os resultados obtidos com os três alinhamentos investigados (**1**, **2** e **3**), para os 10 melhores modelos, utilizando 100.000 operações genéticas, 10% de mutação em cada ciclo de recombinação (*crossover*) e fator de ajuste de 2,5 estão apresentados na Tabela 29.

Tabela 29. Medidas estatísticas, número de descritores de ocupação (GCOD) e de amostras atípicas obtidos para os 10 melhores modelos de QSAR (N = 47) para cada alinhamento (**1**, **2** e **3**), utilizando fator de ajuste 2,5

Alinhamento	n° GCOD	r^2	q^2	LSE	LOF	amostras atípicas
1	9-11	0,88-0,89	0,83	0,12-0,13	0,45-0,70	1
2	8-9	0,89-0,91	0,82-0,83	0,09-0,12	0,35-0,42	0-2
3	9-11	0,87-0,89	0,81-0,83	0,11-0,13	0,47-0,66	2-3

Todos os alinhamentos apresentaram modelos com altos valores de q^2 (0,81 a 0,83) e r^2 (0,89 a 0,91), valores baixos de erro de LSE (0,09 a 0,013) e valores de LOF variando de 0,35 a 0,70, conforme apresentado na tabela 29. No entanto, somente o alinhamento **2** apresentou modelos de QSAR sem quaisquer amostras atípicas. Além disso, o número de descritores de ocupação foi apropriado para o tamanho do conjunto de dados (N = 47; 8 a 9 descritores), uma vez que os modelos QSAR finais são otimizados com a abordagem GFA-MLR (relação de 5 ou 6 descritores para cada amostra). A tabela 30 apresenta os 10 melhores modelos obtidos a partir do alinhamento **2**.

Tabela 30. Medidas estatísticas, número de descritores de ocupação (GCOD) e de amostras atípicas para os 10 melhores modelos de QSAR (N = 47) do alinhamento **2**, com fator de ajuste 2,5

Alinhamento	n° GCOD	r^2	q^2	LSE	LOF	amostras atípicas
1	**9**	**0,91**	**0,83**	**0,09**	**0,35**	**0**
2	9	0,91	0,83	0,09	0,36	1
3	8	0,89	0,83	0,11	0,35	2
4	8	0,89	0,82	0,12	0,35	2
5	9	0,89	0,82	0,11	0,42	2
6	9	0,91	0,82	0,09	0,36	1
7	8	0,89	0,82	0,12	0,36	2
8	8	0,89	0,82	0,12	0,36	2
9	8	0,88	0,82	0,12	0,36	2
10	8	0,88	0,82	0,12	0,36	2

O modelo **1** apresenta valores elevados de q^2 (0,83) e de r^2 (0,91), valores de LSE e de LOF iguais a 0,09 e 0,35, respectivamente, 9 descritores de ocupação e nenhuma amostra atípica (tabela 31). Este modelo foi, então, selecionado como o melhor modelo para o alinhamento **2** e considerado para análises posteriores.

Tabela 31. Valores de resíduos obtidos para os inibidores de MMP-9 com o modelo **1** (N = 47) do alinhamento **2**

Composto	pCI_{50exp}	pCI_{50calc}	Resíduos	DP	2DP
8	8,18	8,05	-0,12	0,31	0,62
$5a_1$	8,10	7,88	-0,22		
$5a_2$	7,92	8,08	0,15		
5b	7,55	7,33	-0,22		
5c	6,90	7,43	0,53		
5d	6,54	6,52	-0,01		
5f	7,03	6,41	-0,61		
5g	7,16	7,72	0,56		
5h	7,26	7,62	0,36		
5j	7,40	7,38	-0,01		
5k	8,32	7,79	-0,52		
5l	7,85	7,50	-0,34		
5m	6,32	6,79	0,47		
5o	6,98	7,31	0,33		
13b	7,32	7,34	0,01		
14a	5,31	5,64	0,33		
14b	6,00	5,64	-0,35		
14c	5,82	5,58	-0,23		
$7a_2$	7,88	7,35	-0,52		
$7a_1$	8,42	8,21	-0,20		
8a	8,62	8,88	0,26		
8b	9,28	8,93	-0,34		
$8b_2$	7,50	7,29	-0,20		
8c	8,88	8,99	0,11		
8e	8,74	8,85	0,11		
8h	8,14	8,63	0,49		
8ir	8,82	8,74	-0,07		
8is	9,26	9,10	-0,16		
8j	9,28	8,65	-0,62		
8k	8,88	8,72	-0,15		
8m	7,69	8,08	0,38		
8n	9,20	9,12	-0,08		
8o	7,06	7,37	0,30		
8p	8,05	7,91	-0,14		
8v	8,31	8,08	-0,23		
8x	8,63	8,86	0,23		
8z	8,60	8,64	0,04		
12a	8,26	8,37	0,11		
8aa	7,77	7,58	-0,18		
8cc	8,69	8,97	0,27		
8dd	9,33	9,01	-0,32		
$8ee_1$	8,72	8,97	0,24		
$8ee_2$	6,53	6,72	0,18		
8ff	9,16	9,06	-0,09		
9a	6,03	6,52	0,48		
9b	6,61	6,60	-0,01		
3b	7,50	7,53	0,03		

A matriz de correlação-cruzada de resíduos de ajuste entre os pares dos 10 melhores modelos para o alinhamento **2** foi calculada para verificar se todos os modelos são equivalentes ($R > 0,5$) ou distintos ($R < 0,5$) e o resultado está apresentado na tabela

32. Os modelos gerados com o alinhamento **2** foram equivalentes ($R \approx 1$), indicando que existe apenas um único modelo de QSAR.

Tabela 32. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os 10 melhores modelos de QSAR (N = 47) do alinhamento **2**

Model	1	2	3	4	5	6	7	8	9	10
1	**1,00**									
2	0,98	**1,00**								
3	0,89	0,90	**1,00**							
4	0,89	0,90	1,00	**1,00**						
5	0,91	0,89	0,97	0,97	**1,00**					
6	0,99	0,97	0,88	0,88	0,90	**1,00**				
7	0,90	0,88	0,98	0,98	0,99	0,89	**1,00**			
8	0,88	0,90	0,99	0,99	0,97	0,88	0,98	**1,00**		
9	0,88	0,89	0,88	0,88	0,87	0,87	0,86	0,87	**1,00**	
10	0,89	0,87	0,87	0,87	0,89	0,89	0,88	0,86	0,98	**1,00**

Verificou-se, então, se os descritores que aparecem no modelo **1** fornecem informação independente ao modelo. A tabela 33 apresenta a matriz de correlação-cruzada entre pares de descritores de ocupação presentes no modelo **1**. Pode-se observar que os descritores não estão correlacionados entre si ($R < 0,5$, em módulo), ou seja, contribuem com informações distintas ao modelo **1** (N = 47).

Tabela 33. Matriz de correlação-cruzada entre pares de descritores de ocupação (GCOD ou GC*i* $_{(x,y,z;IPE)}$) encontrados no modelo **1** (N = 47) para o alinhamento **2**

GCOD	GC1 (-1,3,-2; *hbd*)	GC2 (4,3,6; *any*)	GC3 (-2,5,1; *hbd*)	GC4 (-1,0,5; *any*)	GC5 (1,5,-3; *any*)	GC6 (3,5,-2; *np*)	GC7 (1,2,1; *any*)	GC8 (1,-1,5; *np*)	GC9 (0,0,2; *np*)
GC1	**1,00**								
GC2	0,42	**1,00**							
GC3	-0,17	-0,20	**1,00**						
GC4	-0,10	-0,11	-0,09	**1,00**					
GC5	-0,09	0,01	-0,14	-0,05	**1,00**				
GC6	-0,08	0,10	-0,12	-0,02	0,01	**1,00**			
GC7	-0,15	-0,17	0,06	-0,09	-0,12	0,38	**1,00**		
GC8	-0,22	-0,22	-0,23	0,18	0,14	-0,13	-0,23	**1,00**	
GC9	0,41	0,37	-0,17	-0,24	-0,17	0,07	0,20	-0,54	**1,00**

O modelo **1** é definido pela seguinte equação:

$$pCI_{50} = 3,02\ GC1_{(hbd)} - 4,04\ GC2_{(any)} + 4,05\ GC3_{(hbd)} - 2,43\ GC4_{(any)} - 1,58\ GC5_{(any)} + 1,24\ GC6_{(np)} - 1,69\ GC7_{(any)} + 1,84\ GC8_{(np)} - 0,75\ GC9_{(np)} + 8,09 \quad (\textbf{eq. 5})$$

N = 47 $r^2 = 0,91$ $q^2 = 0,83$ LSE = 0,09 LOF = 0,35 Amostras atípicas = 0

De acordo com a equação 5, o modelo **1** é composto por 9 descritores (GC1 a GC9) pertencentes a três classes de IPE: classe não diferenciada ou qualquer tipo de átomo (*any*), átomo de caráter não polar ou apolar (*np*) e átomos de caráter doador de ligação de hidrogênio (*hbd*). Considerando os coeficientes de regressão dos descritores de ocupação, GC2, GC4, GC5, GC7 e GC9 desfavorecem a atividade inibitória em MMP-9 (pCI_{50}) (coeficientes negativos), enquanto que os descritores GC1, GC3, GC6, e GC8 contribuem favoravelmente à atividade inibitória em MMP-9 (coeficientes positivos).

O modelo de regressão obtido com o conjunto de treinamento (N = 47) foi submetido ao procedimento de validação cruzada LNO (triplicata) e aleatorização do vetor ***y*** para verificar sua robustez e se o modelo foi obtido ao acaso, respectivamente. Para a validação LNO, o N crítico foi obtido com a retirada de subconjuntos de até 15 amostras (32 %), pois quando se retiraram 16 amostras o valor de dois desvios padrão foi maior que 0,1. O valor de q^2_{LNO} foi semelhante ao do modelo original (q^2_{LOO}), ou seja, permaneceu robusto, até a retirada de subconjuntos de 15 amostras. A tabela 34 contém os valores de q^2_{LNO} para cada réplica (triplicata), onde *N* varia de 2 a 16, a média do q^2_{LNO} ($<q^2_{LNO}>$) e os valores de dois desvios padrão (2DP).

Tabela 34. Valores de q^2 obtidos no procedimento de validação cruzada LNO (triplicata) para o modelo **1** (N = 47) do alinhamento 2 com o conjunto de inibidores de MMP-9

N	q^2_{LNO}	q^2_{LNO}	q^2_{LNO}	$<q^2_{LNO}>$	2DP
2	0,83	0,83	0,83	0,83	2,71E-16
3	0,81	0,82	0,82	0,82	0,01
4	0,82	0,83	0,82	0,82	0,01
5	0,83	0,82	0,82	0,82	0,01
6	0,84	0,83	0,82	0,83	0,02
7	0,78	0,83	0,78	0,80	0,05
8	0,81	0,80	0,81	0,81	0,01
9	0,75	0,78	0,82	0,79	0,06
10	0,81	0,83	0,85	0,83	0,03
11	0,80	0,83	0,82	0,82	0,03
12	0,79	0,82	0,78	0,80	0,04
13	0,78	0,80	0,79	0,79	0,02
14	0,85	0,84	0,81	0,83	0,04
15	**0,73**	**0,82**	**0,80**	**0,78**	**0,08**
16	0,84	0,57	0,73	0,71	0,26

O modelo **1** foi, então, submetido ao procedimento de aleatorização do vetor ***y***, com 10, 30 e 50 corridas. Os resultados demonstraram que os modelos obtidos nestas condições apresentaram baixa qualidade e nenhum significado. Os valores de q^2_{yrand} obtidos variaram de 0,1 a valores negativos e os valores de r^2_{yrand} variaram de 0,40 a 0,04 considerando as 10, 30 e 50 aleatorizações. De acordo com as medidas estatísticas q^2_{yrand} e r^2_{yrand} (Eriksson, Wold;1995), o modelo **1** apresentou correlação ao acaso tolerável (0,3

< r^2_{yrand} < 0,4). Os gráficos de LNO e de aleatorização do ve*tor* ***y*** (10 aleatorizações), para o modelo **1** (N = 47) do alinhamento **2**, estão apresentados na Figura 33. A tabela 5A, que contém os valores de q^2_{yrand} e r^2_{yrand} para as 10 randomizações, e as figuras 8A e 9A que apresentam os resultados para 30 e 50 corridas, respectivamente, estão no Anexo A.

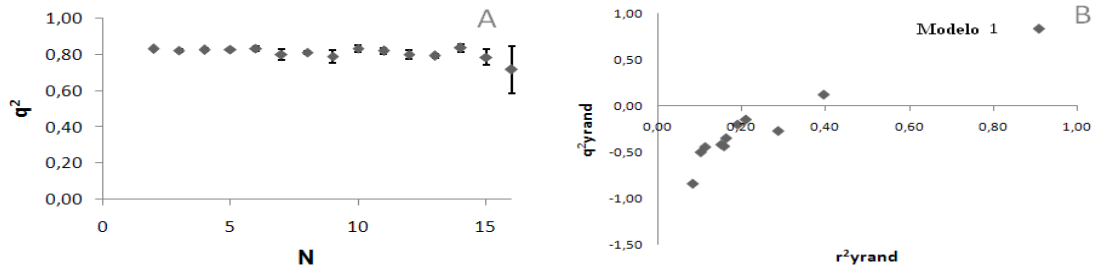


Figura 33. (A) Validação *leave-N-out* aplicada ao modelo final obtido para o conjunto de treinamento (N = 47). Os losangos representam a média e as barras indicam o desvio padrão (triplicata) para cada valor de *N*. O *N* crítico para o modelo de 47 inibidores foi igual a 15 (32 % das amostras). (B) Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de treinamento (N = 47) com 10 corridas, ou aleatorizações, do vetor ***y*** (atividade inibitória frente à MMP-9).

4.3.2.1. Validação externa

Para avaliar a capacidade ou poder de previsão do melhor modelo de QSAR obtido para o conjunto de treinamento (modelo **1**/alinhamento **2**), utilizaram-se os 17 inibidores (**5e**, **5n**, **13a**, **13c**, **8d**, **8f**, **8g**, **8l**, **8q**, **8r**, **8s**, **8t**, **8u**, **8y**, **11a**, **8bb**, **3a**) contidos no conjunto de avaliação. Os valores de pCI_{50} para cada um destes inibidores foram calculados a partir da equação **5**, gerada com o conjunto de treinamento, com a substituição dos valores de cada descritor de ocupação, GCOD ou GC, obtidos com o programa 4D-QSAR 3.0 (HOPFINGER, 1999) (tabela 35).

Tabela 35. Valores dos descritores de ocupação (GCOD ou GC) calculados para os compostos do conjunto de avaliação com o programa 4D-QSAR 3.0 (HOPFINGER, 1999)

Inibidor	GC1	GC2	GC3	GC4	GC5	GC6	GC7	GC8	GC9
5e	0,00	0,01	0,00	0,00	0,0018	0,0002	0,00	0,00	1,00
5n	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	1,00
13a	0,00	0,00	0,01	0,00	0,00	0,00	1,00	0,00	1,00
13c	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	1,00
8d	0,00	0,00	0,07	0,04	0,00	0,00	0,10	0,02	0,00
8f	0,00	0,00	0,23	0,00	0,00	0,00	0,48	0,00	0,00
8g	0,00	0,00	0,28	0,00	0,00	0,00	0,58	0,00	0,00
8l	0,00	0,00	0,06	0,01	0,00	0,00	0,00	0,00	0,00
8q	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00
8r	0,00	0,00	0,00	0,54	0,00	0,00	0,00	0,03	0,00
8s	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
8t	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
8u	0,00	0,16	0,02	0,00	0,00	0,00	0,00	0,00	0,00
8y	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,28	0,00
11a	0,00	0,00	0,41	0,00	0,00	0,23	0,62	0,00	0,00
8bb	0,00	0,00	0,00	0,04	0,66	0,00	0,00	0,00	0,00
3a	0,00	0,26	0,00	0,00	0,00	0,00	0,00	0,00	0,00

A tabela 36 apresenta os valores experimentais de pCI_{50} e previstos com o modelo **1**/alinhamento **2** para cada um dos inibidores de MMP-9. Quatorze dos dezessete inibidores da MMP-9 do conjunto de avaliação foram bem previstos pelo modelo, pois apresentaram valores de resíduo (pCI_{50_exp} -pCI_{50_calc}) menores ou iguais a um desvio-padrão da média dos resíduos do modelo (DP = 1,07) (intervalo de confiança de 99%). Portanto, a capacidade ou o poder de previsão do modelo **1** (N = 47), alinhamento **2**, foi de 82 %.

Tabela 36. Valores de pCI_{50} experimentais e calculados para o conjunto de avaliação (equação **5)**, considerando um intervalo de confiança de 99 %

Composto	pCI_{50_exp}	pCI_{50_calc}	resíduos	DP
5e	6,46	7,30	0,85	1,07
5n	7,26	7,41	0,16	
13a	**7,35**	**5,71**	**-1,63**	
13c	**7,04**	**5,64**	**-1,40**	
8d	8,77	8,16	-0,61	
8f	8,55	8,21	-0,33	
8g	7,63	8,25	0,63	
8l	9,05	8,34	-0,71	
8q	7,16	8,05	0,89	
8r	7,53	6,80	-0,73	
8s	8,53	8,06	-0,46	
8t	8,95	8,06	-0,89	
8u	8,49	7,53	-0,95	
8y	8,72	8,57	-0,14	
11a	**6,91**	**9,02**	**2,12**	
8bb	6,59	6,93	0,34	
3a	6,41	7,04	0,63	

Entre os compostos que não foram bem previstos pelo modelo **1**/alinhamento **2** estão os compostos **13a**, **13c** e **11a**. Do ponto de vista estrutural, os compostos **13a** e **13c** apresentam o grupo C≡C como substituinte em R_{1a}, no lugar do átomo de oxigênio na porção R_{1a}, que os diferem da maioria do conjunto de dados. Este tipo de substituição parece ser responsável por modificar interação dos inibidores com o bolso S1'da MMP-9, pois a introdução de um grupo acetileno, ou etino, na molécula proporciona o alongamento da estrutura, característica que prejudicaria a interação com o bolso S1'. Os inibidores **13a e 13c** são os únicos compostos do conjunto de avaliação que apresentam C≡C em R_{1a}.

O composto **11a** (pCI_{50} = 6,92) tem os mesmos grupos substituintes que o composto **12a** (pCI_{50} = 8,27), do conjunto de treinamento, mas ao invés do grupo isobutiramida na porção R_3 possui uma amina primária. Desta forma, grupos mais

volumosos na posição R_3 parecem melhorar a atividade inibitória dos compostos frente à MMP-9.

4.3.2.2. Conformação bioativa prevista e interpretação dos descritores de ocupação do modelo QSAR (N = 47)

Após os procedimentos de validação, a hipótese de conformação bioativa para cada um dos inibidores do conjunto de treinamento foi calculada a partir do modelo **1**/alinhamento **2**. Os confôrmeros que apresentaram variação de energia (ΔE) de até 5,0 kcal/mol em relação ao confôrmero de energia mais baixa do PAC de cada inibidor foram considerados pelo programa 4D-QSAR 3.0 (HOPFINGER, 1999) nesta etapa. Os valores dos descritores da amostragem de confôrmeros, para cada ligante, foram substituídos na equação do modelo **1** (**eq. 5**) a fim de gerar os valores calculados/previstos de pCI_{50}. O confôrmero que apresentou o maior valor calculado de pCI_{50} foi selecionado pelo programa como a conformação bioativa prevista (HOPFINGER, 1999). As conformações bioativas previstas e os seus respectivos descritores de ocupação (GCOD ou GC) foram avaliados para os inibidores do conjunto de treinamento. Nesta análise, observou-se um padrão de distribuição dos descritores de ocupação ao longo da estrutura geral dos compostos, de acordo com os gradientes de atividade biológica (compostos mais e menos ativos). Este padrão de distribuição estaria relacionado a possíveis substituições nos compostos de forma a favorecer ou desfavorecer as interações com a enzima alvo (MMP-9). Para exemplificar o processo de interpretação do modelo QSAR-4D IR obtido (modelo 1/alinhamento **2**) utilizaram-se, então, as conformações bioativas previstas com os respectivos descritores de ocupação (GCOD ou GC) para os inibidores mais (**8dd** (A)) e menos (**14a** (B)) ativos do conjunto de dados (Figura 34). Estes estados conformacionais também foram calculados utilizando uma grade virtual de 1 Å de resolução. Os descritores de ocupação (GC) que contribuem favoravelmente (**eq. 5**) à atividade inibitória de MMP-9 estão representados por esferas em cinza claro (GC1, GC3, GC6 e GC8), enquanto que os descritores que desfavorecem a ação inibitória (GC2, GC4, GC5, GC7 e GC9) estão em esferas cinza escuro.

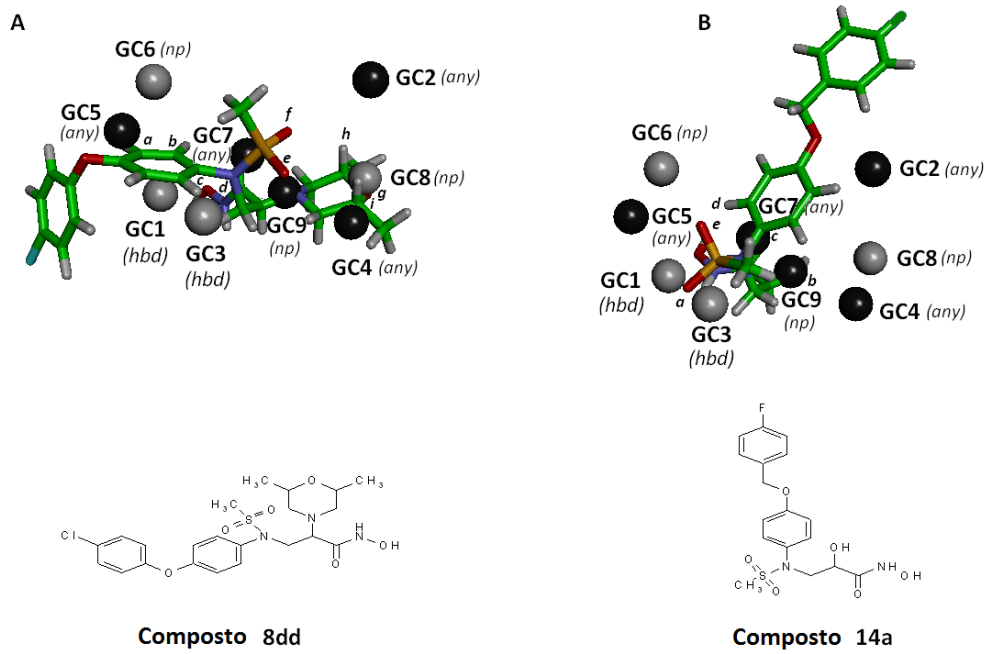


Figura 34. Conformações bioativas previstas com os respectivos GCOD para o composto (A) mais (**8dd**) e (B) menos (**14a**) ativo do conjunto de inibidores de MMP-9, utilizando o modelo **1**/alinhamento **2**. Os descritores de ocupação, GCOD ou GC, que contribuem de forma favorável e desfavorável à atividade inibitória estão representados por esferas na cor cinza mais claro e mais escuro, respectivamente. Os inibidores estão apresentados em modelo molecular de tubo ou palito e os carbonos estão representados na cor verde, os oxigênios em vermelho, hidrogênios em branco, enxofre em amarelo, cloro em azul ciano, nitrogênio em azul e flúor em verde escuro.

Na conformação bioativa prevista para o inibidor mais ativo (**8dd**), as regiões importantes para a atividade inibitória correspondem àquelas próximas ao anel aromático ligado à porção R_1 (GC1, GC3 e GC6) e à porção R_3 (GC8). A classe de IPE do descritor de ocupação GC8 é átomo de caráter não polar ou apolar (*np*) e este descritor está próximo ao átomo de oxigênio (g) do grupo dimetilmorfolina (0,77 Å). Isto indica que átomos nesta posição devem apresentar preferencialmente características apolares para que favoreçam a atividade inibitória em MMP-9. Os descritores GC1e GC3 são da classe de IPE correspondente a átomo doador de ligação de hidrogênio (*hbd*) e estão próximos ao carbono (b) e hidrogênio (c) do anel aromático ligados à porção R_1 (1,64 Å e 0,95 Å, respectivamente). Substituições nestas porções da molécula por átomos que possuam caráter de doador de ligação de hidrogênios favoreceriam a atividade inibitória frente à MMP-9. O descritor de ocupação GC6 apresenta classe de IPE correspondente a átomo de caráter não polar (*np*) e está próximo ao anel aromático ligado à porção R_1 (2,60 Å do carbono (a)). Os descritores de ocupação GC2, GC4, GC5 e GC7 pertencem à classe não diferenciada de IPE (*any*), indicando que substituições moleculares de qualquer tipo deveriam ser evitadas próximas ao nitrogênio (d) da porção sulfonamida (GC7; 1,30 Å),

ao anel aromático ligado à porção R_1 (GC5; 1,23 Å do carbono a), do anel dimetilmorfolina (GC4; 0,86 Å do carbono (i)) e entre a região da porção sulfonamida (GC2; 4,01 Å em relação ao oxigênio (f)) e R_3 (GC2; 4,83 Å em relação ao hidrogênio (h)).

Cabe ressaltar que o inibidor **14a** (menos ativo do conjunto de dados) apresentou padrão de distribuição diferente dos descritores de ocupação, GCOD ou GC, ao longo da estrutura, com exceção do descritor GC9, que também se encontra próximo à R_3 (0,4 Å). GC1 e GC3 não estão relacionados ao anel aromático ligado à porção R_1, mas ao átomo de oxigênio (a) da porção sulfonamida (2,51 Å e 1,34, respectivamente). Os descritores GC8 e GC4 estão mais distantes da porção R_3(GC8: 2,95 Å e GC4: 2,77 Å em relação ao oxigênio (b) da hidroxila em R_3, respectivamente), no composto menos ativo, e isto reduziria os pontos de interação no bolso S1da enzima MMP-9. O descritor GC7 encontra-se próximo ao anel aromático ligado à porção sulfonamida (0,74 Å do carbono (c)). Os descritores GC5 e GC6, por sua vez, se encontram distantes do anel aromático ligado à porção R_1 (GC6 está 3,62 Å em relação ao hidrogênio (d) deste anel e GC5 encontra-se a 3,14 Å do átomo de oxigênio (e) da porção sulfonamida). Estas diferenças seriam, provavelmente, responsáveis na redução de atividade inibitória frente à MMP-9 do composto **14a**.

Para determinar o se o modelo **1** do alinhamento **2** se ajusta, do ponto de vista estereoquímico, ao sítio ativo e se há complementaridade no sítio de interação entre os resíduos de aminoácido importantes da enzima e os descritores de ocupação do ligante, as conformações bioativas previstas de cada composto do conjunto de treinamento e os respectivos GCOD (GC) foram ancorados manualmente no sítio catalítico do modelo molecular 3D da enzima MMP-9. Utilizou-se o alinhamento **2** (3 átomos) para orientar o ancoramento molecular dos ligantes. Este procedimento permite, também, a diferenciação ou identificação do tipo de átomo para os descritores de ocupação da classe de IPE definidos como *any* (GC2, GC4, GC5, GC7) no modelo QSAR selecionado. A estrutura 3D do domínio catalítico da enzima MMP-9 complexada a um inibidor hidroxamato (7-MR, figura 35), obtida por difração de raios-X (**PDB ID** 2OW1 TOCHOWICZ *et al*., 2007), foi empregada nesta investigação.

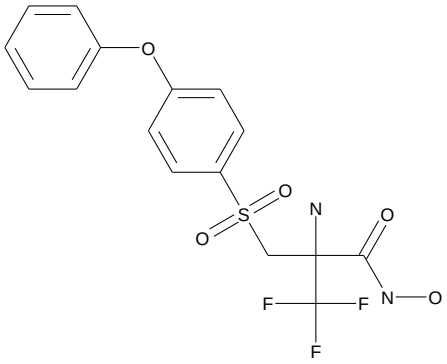


Figura 35. Estrutura molecular do inibidor de MMP-9, 7-MR (**PDB ID** 2OW1, TOCHOWICZ *et al*., 2007).

As coordenadas cartesianas do inibidor 7-MR foram utilizadas como referência para ancorar as conformações bioativas previstas e os descritores de ocupação (modelo **1**/alinhamento **2**) dos compostos do conjunto de treinamento no sítio catalítico da MMP-9. O relaxamento dos complexos moleculares, para melhor acomodar os ligantes no sítio de interação foi obtido a partir de simulações computacionais de DM de curta duração (50 ps) em 298 K, cada passo de 1 fs (MOLSIM 3.2; DOHERTY, 2002). O complexo de energia mais baixa para cada ligante foi selecionado da região de equilíbrio da simulação e, após minimização de energia, foi considerado para a avaliação dos descritores no sítio da MMP-9 (complementaridade).

A seguir estão apresentados os resultados obtidos para o processamento da proteína MMP-9 antes de ser utilizada na etapa de investigação da complementaridade no alvo do modelo QSAR-4D IR obtido para o conjunto de inibidores hidroxamatos.

4.3.2.3. Processamento da proteína para os estudos de ancoramento das conformações bioativas/descritores, GCOD

A estrutura da enzima MMP-9 co-cristalografada com o inibidor hidroxamato 7-MR, depositada no banco de dados **PDB** com o código 2OW1 (TOCHOWICZ *et al*., 2007), foi utilizada como modelo para construção da geometria do receptor, conforme descrito em Materiais e Métodos (item **3.9.1).** A energia do modelo molecular 3D do complexo foi minimizada considerando o critério de convergência de 0,5 kcal/mol, com 462 iterações para o método de declive máximo e 111 iterações para o método de gradientes conjugados (MOLSIM 3.2; DOHERTY, 2002), à temperatura de 298 K. O modelo molecular do complexo MMP-9/7-MR otimizado foi o dado de entrada (*input*) para as simulações de DM no processo de preparação da proteína.O programa MOLSIM 3.2 (DOHERTY, 2002) foi também empregado para realizar as simulações de DM do modelo

molecular 3D do complexo previamente processado. Empregou-se como protocolo para as simulações de DM um esquema de aquecimento lento (*warming-up*), com simulações curtas (50 ps) em T crescente (50 K, 100 K, 200 K, 300 K). O confôrmero de energia mais baixa selecionado da simulação a 300 K, após minimização de energia, foi utilizado como *input* para a simulação de 2 ns a 298 K (os gráficos dos PAC das simulações de 2ns estão apresentados no Anexo A - figura 10A). Este tipo de esquema de aquecimento é aplicado para gerar o estado de menor energia do complexo de forma mais rápida, permitindo que o sistema transponha barreiras energéticas (TOKARSKI, HOPFINGER, 1997). De acordo com os PAC provenientes das simulações de DM apresentados na figura 36, o modelo molecular do complexo MMP-9/7-MR atingiu o equilíbrio gradativamente e estabilizou completamente apenas na simulação de DM de 2ns a 298 K. Uma amonstragem (100 ps) da região de equilíbrio do PAC, da simulação mais longa a 298 K, está apresentada na figura 36 (e).

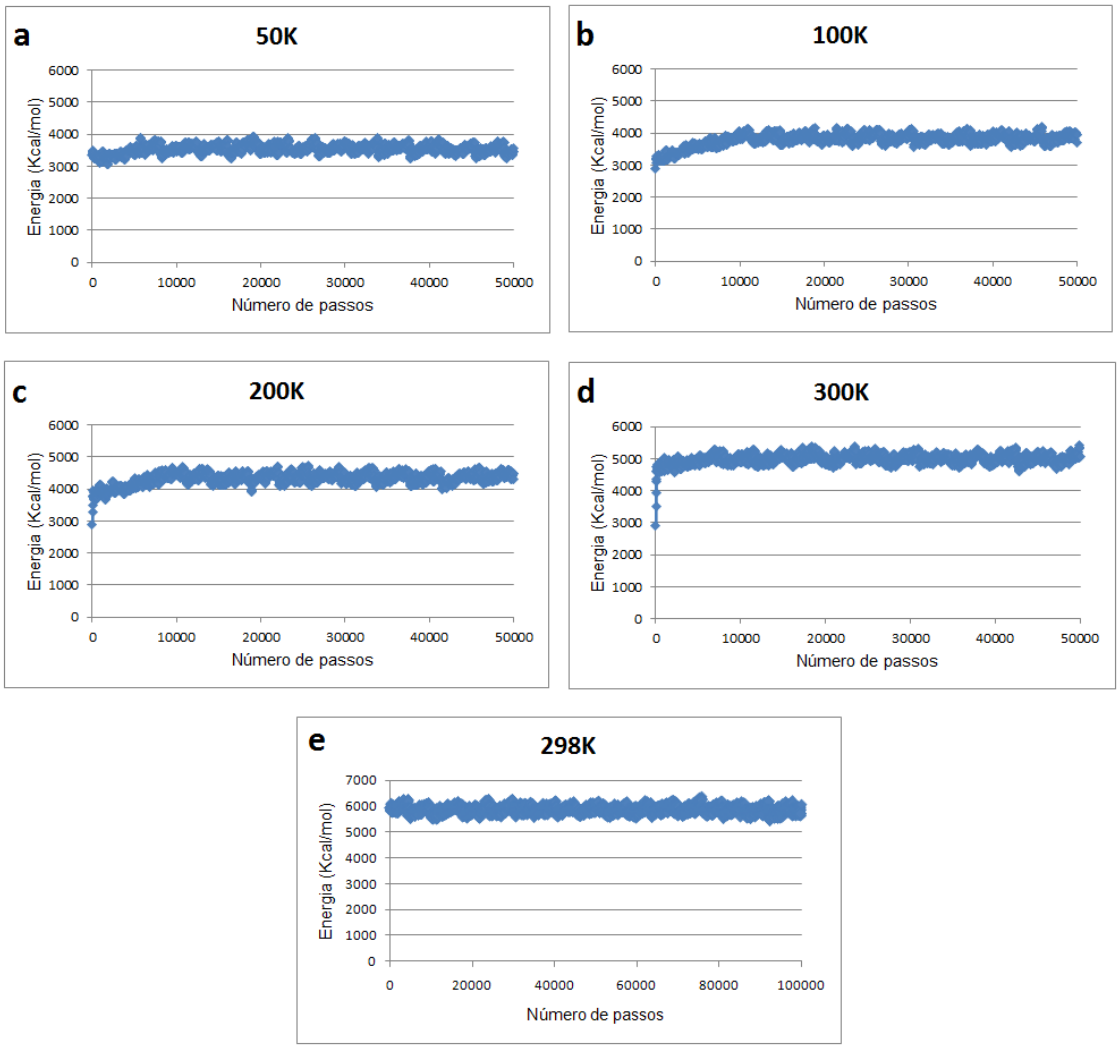


Figura 36. PAC obtido para as simulações de DM resultantes do esquema de aquecimento lento (50 ps) e para a simulação mais longa (2 ns) com o complexo MMP-9/7-MR (TOCHOWICZ *et*

al., 2007). (**a**) Simulação de DM a 50 K; (**b**) simulação de DM a 100 K; (**c**) simulação de DM a 200 K; (**d**) simulação de DM a 300 K; (**e**) simulação de DM a 298 K. Para a simulação de 2 ns (298 K) está representada apenas uma porção de 100 ps referente à região de equilíbrio do PAC. As contribuições de energia intramolecular de solvatação e de ligação de hidrogênio não foram consideradas no valor de energia (eixo y; kcal/mol).

O valor de energia potencial total (soma de todas as contribuições de energia, inclusive de solvatação e de ligação de hidrogênio) para o confôrmero deenergia mais baixa do complexo MMP-9/7-MR, selecionado em cada T de simulação, foi computado (Tabela 37) e verificou-se a manutenção de integridade estrutural a partir dos valores de desvio quadrático médio de posições atômicas (*Root Mean Square Deviation*, RMSD) em relação à estrutura 3D de referência, obtida por difração de raios-X. Os resultados indicaram que, até 298 K, os modelos moleculares selecionados mantiveram a integridade estrutural em relação ao complexo original, uma vez que os valores de RMSD não ultrapassaram 1,5 Å (Figura 37).

Tabela 37. Valores das contribuições de energia obtidos em cada T de simulação para o confôrmero de energia mínima do complexo MMP-9/7-MR **(PDB ID** 2OW1) selecionado da região de equilíbrio do PAC

T (K)	Eest	Eang	Etors	E1,4	EvdW	Eel	INTERV+C	Esolv	Ehb	Etot
50	1250,96	1264,49	1030,38	3910,48	-824,79	-3589,18	277,34	-192,44	-19134,22	-16007
100	1307,62	1378,09	1041,21	3953,76	-812,83	-3536	284,01	-193,91	-19226,61	-15804,7
200	1534,24	1629,38	1122,08	3991,54	-783,68	-3643,84	266,76	-164,80	-19556,48	-15604,8
300	1745,92	1840,18	1203,21	4020,86	-724,8	-3638,84	289,13	-147,95	-19298,26	-14710,6
298	1737,15	1780,84	1180,4	4011,63	-725,89	-3559,07	301,07	-157,73	-19945,40	-15377

* Eest = contribuição de energia intramolecular de deformação axial ou estiramento; Eang = contribuição de energia intramolecular de deformação angular; Etors = contribuição de energia intramolecular torsional; $E_{1,4}$ = contribuição de energia intramolecular de interação do tipo 1-4 ou Lennard-Jones; EvdW = contribuição de energia intramolecular de van der Waals; Eel = contribuição de energia intramolecular eletrostática; INTERV+C = soma das contribuições de energia intermolecular eletrostática e de van der Waals; Esolv = contribuição de energia intramolecular de solvatação; E_{Hb} = contribuição de energia intramolecular de ligação de hidrogênio; Etot = energia potencial total, que corresponde à soma de todas as contribuições de energia. As unidades de todas as contribuições de energia estão em kcal/mol.

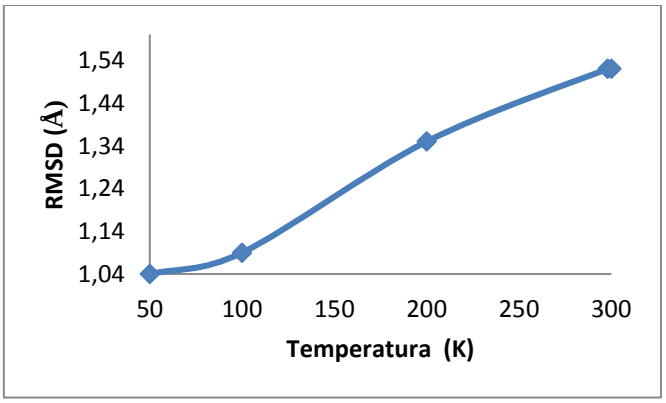


Figura 37. Valores de RMSD obtidos na comparação (sobreposição) entre os confôrmeros dos complexos selecionados em cada T de simulação, do esquema de aquecimento lento, e o complexo MMP-9/7-MR experimental (**PDB ID** 2OW1, TOCHOWICZ *et al*., 2007).

Então, utilizou-se o confôrmero de energia mais baixa do complexo MMP-9/7-MR, selecionado da simulação de DM em 298 K (2 ns), após a minimização, para o ancoramento molecular das conformações bioativas previstas com os descritores de ocupação, GCOD ou GC. Após o ancoramento de cada conformação bioativa com os descritores de ocupação do modelo **1**/alinhamento **2**, o inibidor referência (7-MR) foi deletado do sítio de interação e procedeu-se as etapas de minimização de energia e simulações de relaxamento (50 ps a 298 K) dos complexos, a fim de se obter uma melhor acomodação dos ligantes. O confôrmero de energia mais baixa de cada complexo selecionado do PAC da simulação de 50 ps, após a minimização de energia, foi utilizado para a análise de complementaridade ligante-enzima, já mencionada. Este tipo de avaliação permitiu a interpretação e, de certa forma, a validação dos descritores de ocupação gerados com a análise QSAR-4D IR. A elucidação dos tipos de interação possíveis entre os inibidores + GCOD e os resíduos de aminoácidos do sítio catalítico da enzima alvo validaria o mapeamento do farmacóforo obtido com o modelo **1**/alinhamento **2** IR.

Avaliaram-se, então, as distâncias entre os resíduos de aminoácidos do sítio catalítico (Leu187, Leu188, Ala189, His190, Val398, Pro421, Met422 e Tyr423) e os descritores de ocupação (GCOD ou GC) associados às conformações bioativas previstas. Todos os possíveis conjuntos de interações entre inibidor-enzima (complementaridade), com base no modelo e alinhamento da análise QSAR-4D IR, podem ser identificados neste procedimento.

As interações encontradas para o inibidor mais ativo (**8dd**) do conjunto de treinamento (figura 38, A) foram comparadas às encontradas para o inibidor 7-MR (figura 38, B) no sítio catalítico da enzima MMP-9 (2OW1, TOCHOWICZ *et al*., 2007). Apesar dos inibidores **8dd** e 7-MR pertencerem a classes químicas relacionadas, algumas diferenças puderam ser observadas considerando principalmente a orientação e as interações intermoleculares estabelecidas com os resíduos de aminoácidos no sítio catalítico da MMP-9.

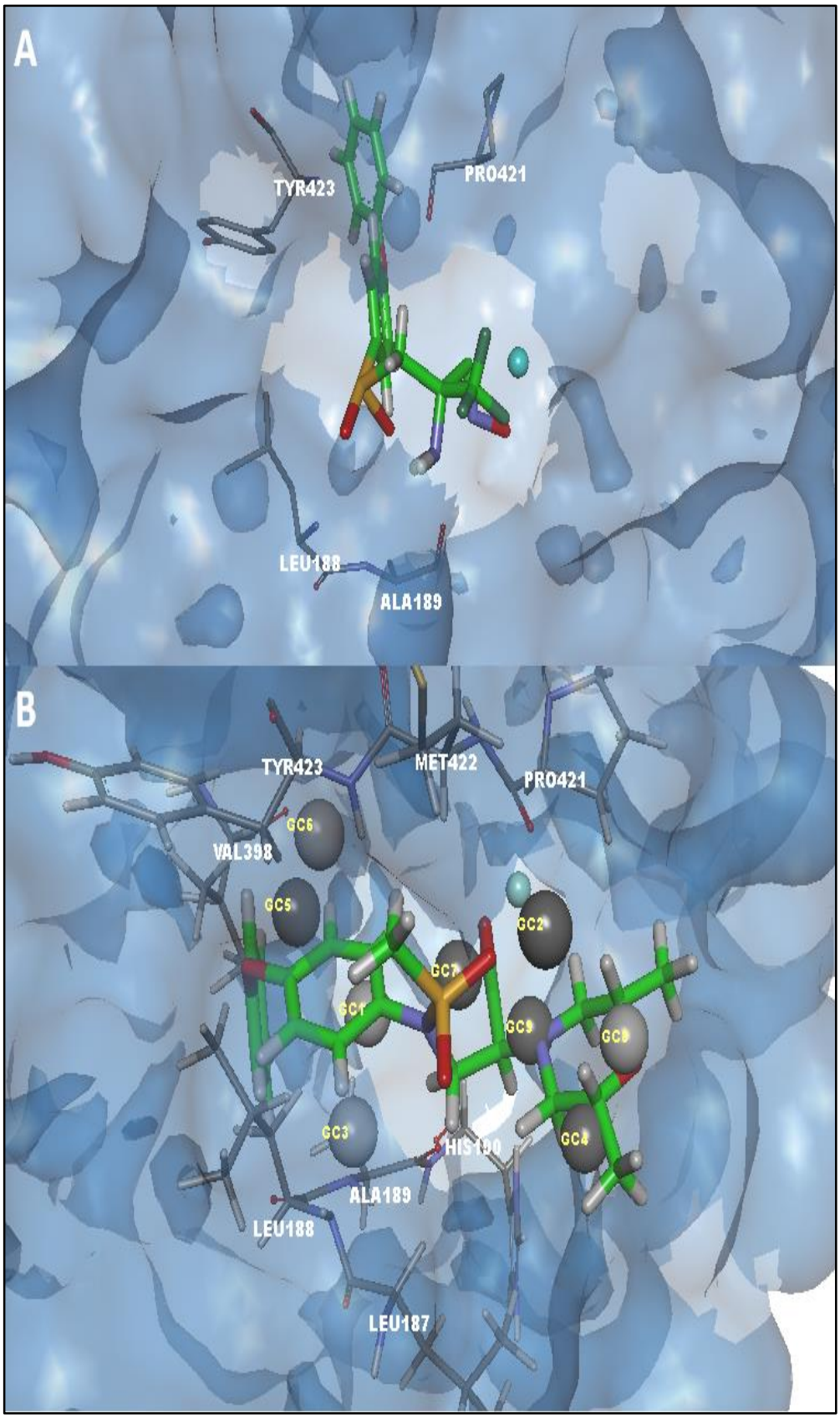


Figura 38. (A) Inibidor 7-MR no sítio catalítico da enzima MMP-9. (B) Visualização da conformação bioativa prevista para o inibidor mais ativo (**8dd**) com os descritores de ocupação (GCOD, GC1-GC9). Os inibidores estão representados em modelo tubo ou palito e os átomos de carbono estão em verde, oxigênios em vermelho, nitrogênio em azul, enxofre em amarelo, flúor em vede escuro e hidrogênios em branco. Os íons de zinco (azul claro) e os descritores de

ocupação [GCOD ou GC; esferas em cinza claro (favoráveis) e cinza escuro (desfavoráveis)] estão representados em modelo CPK (preenchido). Os resíduos de aminoácido Leu187, Leu188, Ala189, His190, Val390, Thr423, Pro421 e Met422 estão representados em modelo palito. Os átomos de carbono estão em cinza, oxigênio em vermelho e nitrogênio em azul. A superfície molecular da proteína MMP-9 está demonstrada em azul.

A conformação bioativa do inibidor mais ativo (**8dd**) participa de interações envolvendo os resíduos de aminoácido Leu187, Leu188, Ala189, His190, Val398, Pro421, Met422 e Tyr423 do sítio catalítico da MMP-9.

O descritor de ocupação GC1 está próximo ao resíduo de aminoácido Ala189 (GC1; 3,88 Å distante do N). De acordo com a classe de IPE deste descritor (*hbd*), grupos substituintes doadores de ligação hidrogênio nesta posição poderiam estabelecer interações com o nitrogênio do resíduo Ala189 no sítio da MMP-9, reforçando o caráter de GC1 descrito pelo modelo **1**/alinhamento **2**. Da mesma forma, grupos substituintes nas coordenadas do descritor GC3, que apresenta a mesma natureza de IPE (*hbd*), poderiam estabelecer interações do tipo doador de ligação de hidrogênio com o nitrogênio do resíduo Leu188 (GC3; 1,94 Å distante do N).

O descritor GC6, de natureza não polar, está 3,10 Å distante do átomo de carbono da cadeia lateral do resíduo de aminoácido Tyr423 no sítio catalítico da MMP-9, sugerindo que a presença de grupos substituintes apolares nesta posição contribuiriam favoravelmente à atividade inibitória. O descritor GC8, também de caráter não polar (porção R_3), está próximo ao carbono do anel do resíduo Pro421 (GC8; 4,85 Å), enfatizando o tipo de interação descrita pelo modelo QSAR-4D IR.

Modificações moleculares considerando as coordenadas cartesianas dos GCOD descritos anteriormente, bem como a natureza destes descritores (IPE), corresponderiam à estratégia interessante visando incrementar a atividade inibitória frente à MMP-9.

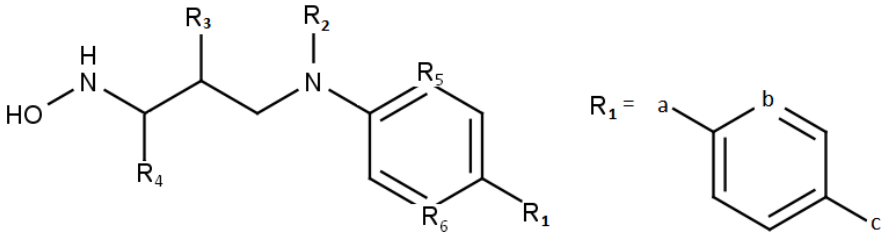
Os descritores de ocupação GC2, GC4, GC5 e GC7 são de classe não diferenciada (*any*) e apresentam distâncias favoráveis para estabelecerem interações com os resíduos de aminoácidos Met422, Leu187, Val398 e Pro421, respectivamente, do sítio catalítico da MMP-9. GC2 poderia estabelecer, por exemplo, interações do tipo não polar com o carbono da cadeia lateral do resíduo Met422 (4,27Å), o que definiria a natureza do grupo substituinte a ser considerado nesta coordenada. Este descritor se encontra mais distante da estrutura do inibidor mais ativo do conjunto de dados (**8dd**) (GC2; 4,01 Å em relação ao oxigênio (f) da porção sulfonamida (figura 34)) e poderia ser considerado para auxiliar o estabelecimento de pontos adicionais de interação nesta região da MMP-9. O descritor GC4 está próximo à cadeia lateral do resíduo Leu187 (4,50 Ådo grupo metil) e, portanto,

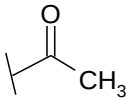
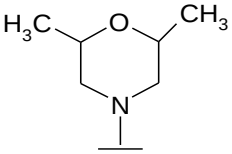
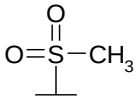
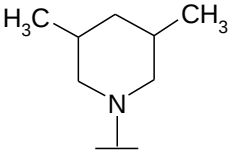
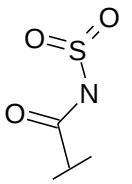
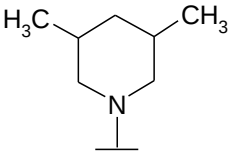
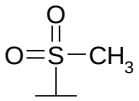
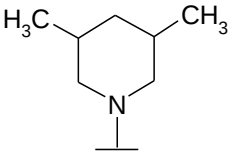
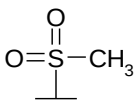
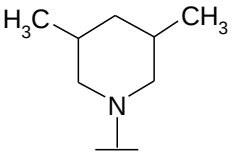
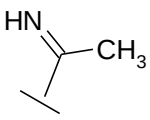
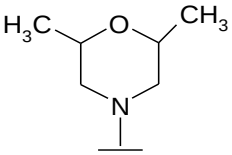
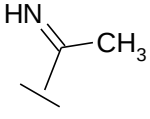
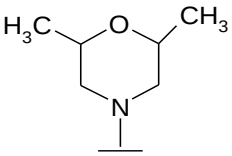
poderia estabelecer interações do tipo não polar com tal resíduo. GC5 está próximo ao grupo metil do resíduo Val398 (3,01 Å) e poderia estabelecer interações hidrofóbicas (tipo não polar ou apolar) coma cadeia lateral daquele resíduo. O descritor GC7 poderia ser definidocomo polar, já que está próximo ao oxigênio da carbonila do resíduo Pro421 (4,29 Å).

GC9, que possui coeficiente de regressão negativo no modelo QSAR, foi definido como átomo de caráter não polar (*np*), ou seja, grupos desta natureza naquela coordenada desfavoreceriam a atividade inibitória frente à enzima MMP-9. Este descritor está próximo ao resíduo His190 (4,9 Å do nitrogênio do anel imidazol), reforçando que a natureza deste átomo não deveria ser apolar, mas de doador ou aceitador de ligações de hidrogênio, por exemplo.

O melhor modelo de QSAR (modelo **1**/alinhamento **2**) foi estatisticamente validado, robusto e, também, foi passível de interpretação tanto do ponto vista químico como biológico. Então, este modelo pode ser considerado confiável para prever novos derivados capazes de inibir a MMP-9. Assim, uma quimioteca (N = 7) foi construída e a atividade inibitória dos compostos foi calculada utilizando o modelo **1**/alinhamento **2** (Tabela 38). Cinco compostos foram considerados bem promissores, mas o ligante **L1** apresentou a maior atividade prevista para MMP-9 (pCI_{50} = 9,94). Este composto possui somente um átomo de oxigênio, do grupo hidroxila, para interagir com o Zn no sítio ativo da MMP-9 e apresenta volume grande na porção R_3, para estabelecer interações no sítio acessório S1', adjacente ao sítio catalítico. A porção sulfonamida foi substituída por uma amida, mas o grupo R_1 do composto **8dd** foi mantido. Então, informações importantes foram fornecidas a partir da análise QSAR-4D IR para direcionar e auxiliar no planejamento racional de novos inibidores potencialmente ativos contra MMP-9.

Tabela 38. Ligantes propostos e respectivos valores atividade inibitórias frente à MMP-9 calculados com o modelo **1**/alinhamento **2**



Ligantes	R_{1a}	R_{1b}	R_{1c}	R_2	R_3	R_4	R_5	R_6	pCI_{50cal}
L1	O	CH	Cl			H_2	CH	CH	9,94
L2	O	CH	Cl			O	COH	COH	9,44
L3	O	CH	Cl			O	N	N	7,91
L4	O	CH	Cl			O	N	N	9,88
L5	O	CH	Cl			H_2	N	N	8,08
L6	O	CH	Cl			H_2	N	N	9,12
L7	O	CH	Cl			O	N	N	9,86

4.3.2.4. Previsão da atividade do 4-NC a partir do melhor modelo de QSAR (N = 47) obtido para o conjunto de inibidores de MMP-9

Com o intuito de verificar a capacidade de previsão do melhor modelo QSAR obtido com o conjunto de inibidores de MMP-9 para classes de moléculas não quimicamente relacionadas, calculou-se a atividade inibitória prevista para o modelo molecular tridimensional do 4-NC.

Os descritores de ocupação, GCOD, para o composto 4-NC foram calculados nas mesmas condições utilizadas para o conjunto de inibidores hidroxamatos e os valores obtidos para os descritores (Tabela 39) foram substituídos na equação do modelo **1**/alinhamento **2**. O valor de pCI_{50} calculado para o 4-NC com a **eq. 5** foi de 7,93. De acordo com este modelo, a atividade inibitória do 4-NC frente à enzima MMP-9 seria considerada moderada.

Tabela 39. Valores dos descritores de ocupação, GCOD ou GC, para o 4-NC calculados com o modelo QSAR de inibidores de MMP-9 (modelo 1 do alinhamento **2**)

GC1	GC2	GC3	GC4	GC5	GC6	GC7	GC8	GC9
0	0	0	0	0	0	0	0	0,0634

A conformação bioativa para o 4-NC também foi prevista (Figura 39, A) e ancorada no sítio catalítico da enzima MMP-9 (Figura 39, B) com os descritores de ocupação do modelo **1**/alinhamento **2**.

O padrão de distribuição dos descritores definiu a atividade moderada prevista para o 4-NC. Os descritores de ocupação se encontram próximos à conformação postulada como biotiva para o 4-NC. O modelo QSAR para MMP-9 foi construído com um número maior número de compostos (N = 47) que o modelo para MMP-2 (N = 30), envolvendo maior diversidade estrutural, o que permite gerar um conjunto maior de descritores para mapear o farmacófoco do conjunto de dados, facilitando a proposição de novas estruturas. O modelo molecular 3D do composto 4-NC foi ancorado no sítio catalítico da enzima MMP-9 de acordo com o alinhamento **2**. É possível observar que, assim como o inibidor mais ativo do conjunto de dados (**8dd**), o composto 4-NC poderia estabelecer interação com os resíduos de aminoácidos His104, Leu187, Leu188, His190 e Tyr423.

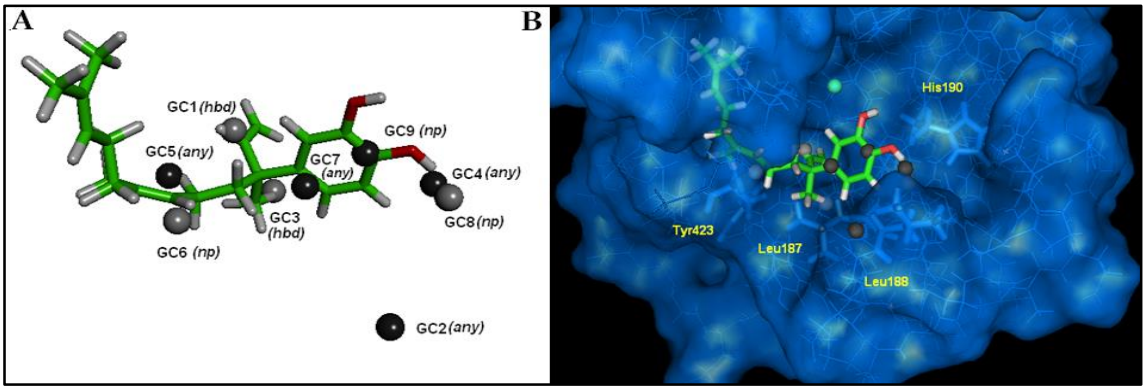


Figura 39. (**A**) Conformação bioativa prevista para o 4-NC com os descritores de ocupação, GCOD ou GC, utilizando o modelo **1**/alinhamento **2**. (**B**) Conformação bioativa prevista para o 4-NC com os descritores de ocupação no sítio catalítico da enzima MMP-9. O modelo molecular 3D do 4-NC está representado em tubo ou palito, onde os carbonos estão em verde, os oxigênios em vermelho e os hidrogênios em branco. Os descritores que desfavorecem a atividade estão representados em esferas cinza escuro e os que favorecem em esferas cinza claro (Pymol 1.5.0.1, DeLano, 2012).

Então, de acordo com as informações obtidas com o modelo **1**/alinhamento **2**, modificações na estrutura molecular do 4-NC foram propostas. Os modelos moleculares 3D foram construídos, otimizados e o PAC calculado, conforme descrito nos itens **3.6** e **3.3** para desenvolver o cálculo dos descritores de ocupação.

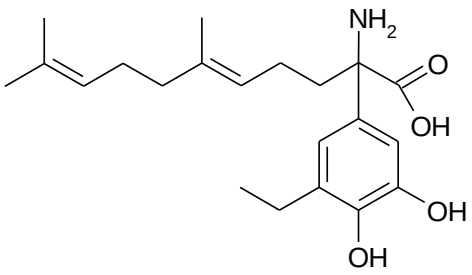


Figura 40. Modificações propostas para a estrutura do 4-NC com base na interpretação dos descritores de ocupação do modelo **1**/alinhamento **2**.

Tabela 40. Valores dos descritores de ocupação (GCOD ou GC) calculados para a estrutura do 4-NC com modificações com o modelo **1**/alinhamento **2** (programa 4D-QSAR 3.0, HOPFINGER, 1999)

GC1	GC2	GC3	GC4	GC5	GC6	GC7	GC8	GC9
0	0	0	0	0	0	0	0	0,21320

Após a substituição dos valores de GCOD na equação 5, do modelo **1**/alinhamento **2**, o valor de pCI_{50} calculado foi de 8,09. Este novo composto seria, também, classificado como moderadamente ativo.

O estudo de QSAR-4D IR para os inibidores de MMP-9 foi apresentado no evento Strasbourg Summer School in Chemoinformatics, 2014.

4.4. Determinação do estado de protonação do 4-NC

Para a determinação do estado de protonação do composto 4-NC, realizou–se uma varredura nos valores de pH de 0 a 14, utilizando o programa computacional Marvin Beans 5.8.0 (ChemAxon Ltd., 2009). Esta análise permitiu identificar as possíveis espécies ionizadas, referentes à estrutura do 4-NC, em diferentes valores de pH, relacionados aos compartimentos do organismo. Durante a varredura, geraram-se quatro espécies moleculares (Figura 41).

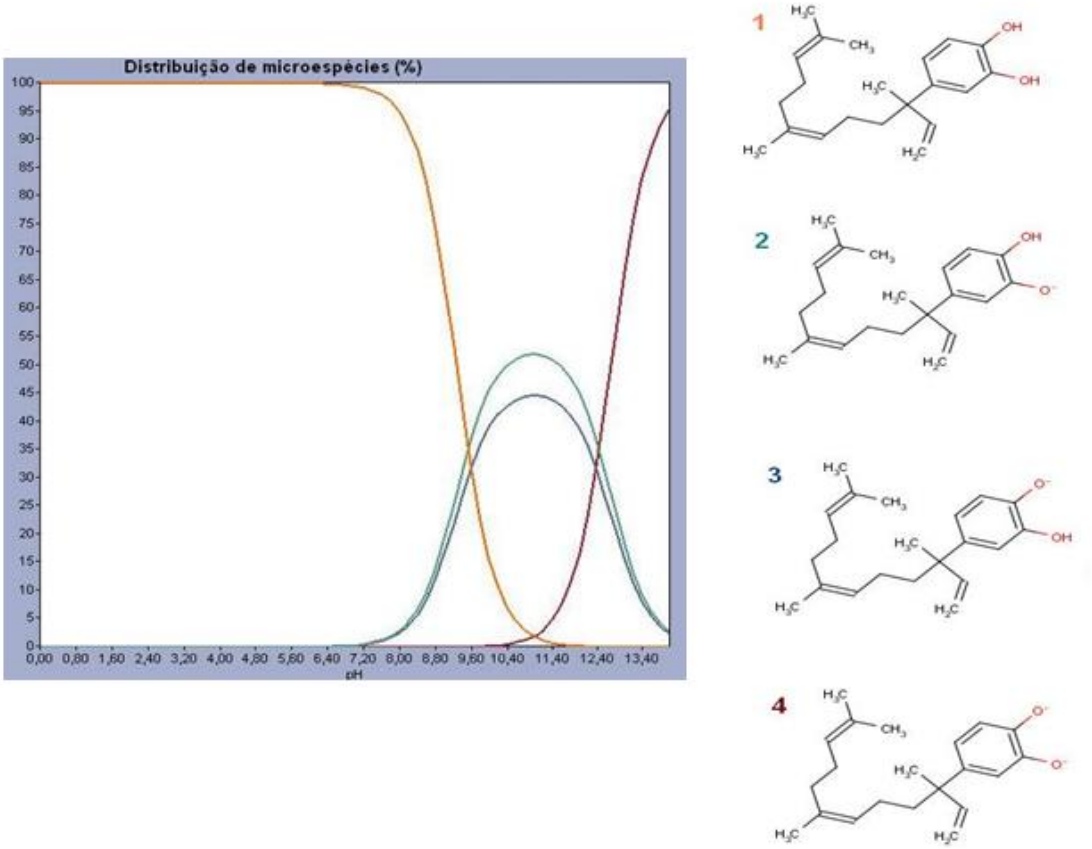


Figura 41. Análise das espécies iônicas do composto 4-NC encontradas em diferentes valores de pH.

A espécie que apresenta dois grupos hidroxila protonados (número **1**, Figura 41), foi a de maior concentração na variação de pH entre 0,0 a 10,40. Como esta espécie também se encontra em maior concentração no pH fisiológico (7,4), foi considerada em estudos posteriores de ancoramento e dinâmica molecular com as enzimas MMP-2 e MMP-9.

Com o aumento do valor de pH, as espécies **2** e **3**, que apresentam apenas um dos grupos hidroxila desprotonado, parecem ser predominantes e poderiam, também, estar presentes nas porções finais do intestino delgado (pH 8,0 a 9,4). A espécie **4**, que apresenta os dois grupos hidroxila desprotonados, estaria em maior concentração em um ambiente com pH de 12,0 a 13,40. Estes dados fornecem uma previsão sobre o comportamento referente à ionização do 4-NC, caso fosse administrado por via oral. Porém, estudos de farmacocinética devem ser desenvolvidos a fim de verificar a melhor via de administração para este composto.

4.5. Estudos de *redocking*

4.5.1. Estudos de *redocking* do complexo inibidor/MMP-2

Com o objetivo de verificar se os parâmetros utilizados no processo de ancoramentomolecular do 4-NC na enzima MMP-2 seriam capazes de reproduzir dados de complexos obtidos experimentalmente, o inibidor hidroxamato (SC-74020) foi extraído do sítio catalítico do modelo molecular 3D obtido por RMN (**PDB ID** 1HOV, FENG *et al*., 2002) e ancorado novamente no sítio catalítico da MMP-2, empregando-se o software Autodock 4 (HUEY *et al.*, 2007). Os parâmetros utilizados nestes procedimentos estão descritos no item **3.9.2** de Material e Métodos.

Dos três diferentes tipos de algoritmo de busca oferecidos no programa Autodock 4 (HUEY *et al*., 2007) para explorar o sítio catalítico da proteína, selecionou-se o algoritmo genético Lamarckiano (AGL), uma vez que estudos preliminares utilizando os dois outros métodos presentes neste software (arrefecimento simulado (*Simulated Annealing*) e algoritmo genético), demostraram menor eficiência quando comparados ao AGL para ligantes altamente flexíveis (DADRAS *et al*., 2004).

Cabe ressaltar que, em geral, os campos de força utilizados nos estudos de ancoramento molecular apresentam deficiências na parametrização de metais, como no caso de metaloproteinases. Por este motivo, os valores de r e ε utilizados foram baseados em estudos prévios desenvolvidos Hu e Shelver (2003) (r = 0,87 Å, ε = 0,35 kcal/mol, carga dos íons de zinco = +0,95). Hu e Shelver (2003) avaliaram a otimização destes parâmetros para 14 complexos envolvendo metaloproteinases. Os parâmetros otimizados proporcionaram melhor acurácia no ancoramento molecular de ligantes no sítio de ligação contendo zinco e melhor previsão dos valores de energia livre de ligação (afinidade), utilizando o programa Autodock 4 (HUEY *et al*, 2007).

A conformação de menor energia obtida nos estudos de ancoramento molecular foi comparada à estrutura do ligante experimental (RMN), utilizando o valor de RMSD como referência (figura 42). O valor de RMSD obtido foi de 2,94 Å, que pode ser considerado bom de acordo com os critérios adotados (DADRAS *et al*., 2004). Por este motivo, consideraram-se os parâmetros/condições utilizados no processo de ancoramento molecular satisfatórios.

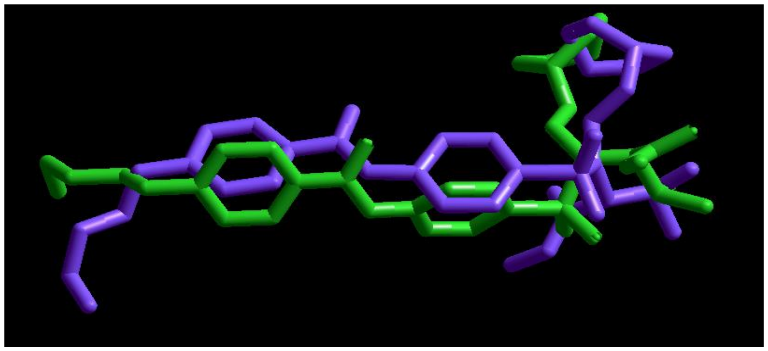


Figura 42. Sobreposição da estrutura do inibidor hidroxamato (SC-74020) obtido no processo de *redocking* quando comparado à estrutura obtida por RMN (FENG *et al*., 2002). O confôrmero na cor verde corresponde à estrutura resultante do processo de *redocking* e a estrutura em violeta corresponde ao ligante na conformação obtida experimentalmente (Hyperchem 8.0; HyperChem, 2007)

4.5.2. Estudos de *redocking* do complexo inibidor/MMP-9

Para verificar se os parâmetros utilizados no processo de ancoramento molecular do 4-NC no sítio catalítico da enzima MMP-9 seriam capazes de reproduzir os dados de complexo obtido experimentalmente (difração de raios-X), efetuou-se o *redocking* do derivado hidroxamato (7-MR) presente na estrutura cristalografada da MMP-9 (TOCHOWICZ *et al*., 2007), novamente no sítio catalítico. Neste estudo, o confôrmero de menor energia proveniente do processo de *redocking* apresentou valor de RMSD de 1,69 Å quando comparado ao ligante experimental no sítio catalítico da MMP-9 (figura 43). O resultado pode ser considerado satisfatório. Assim, utilizaram-se os mesmos parâmetros definidos neste estudo para o ancoramento molecular do 4-NC ao sítio catalítico da MMP-9.

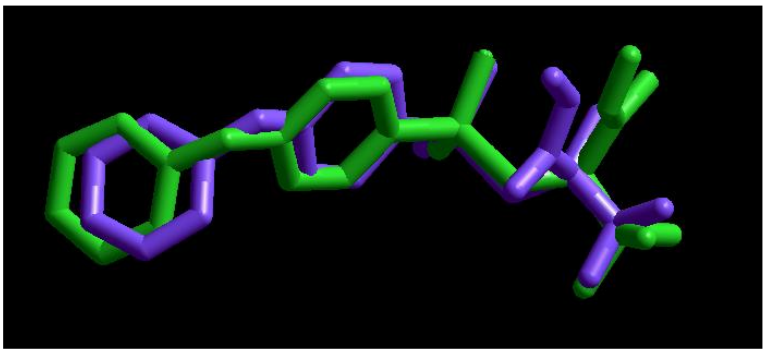


Figura 43. Sobreposição da estrutura do derivado hidroxamato (7-MR) obtido no processo de *redocking* quando comparado à estrutura cristalografada (TOCHOWICZ *et al*., 2007). O confôrmero na cor verde é a estrutura proveniente do processo de *redocking* e a estrutura em violeta é o ligante na conformação do cristal (Hyperchem 8.0; HyperChem, 2007).

4.6. Estudos de ancoramento

4.6.1. Estudos de ancoramento do 4-NC no sítio catalítico da enzima MMP-2

O modelo molecular 3D do 4-NC foi construído, em sua forma protonada (Hyperchem 8.0 (Hypecube, 2007) e otimizado em campo de força empírico MM+ (Allinger, 1977). Calcularam-se as cargas atômicas parciais com o método semi-empírico AM1 (DEWAR *et al*., 1985). Após processado no software Autodocktools 4 (MORRIS *et al*., 2009), o modelo molecular 3D do 4-NC foi ancorado à estrutura da enzima MMP-2 por trinta vezes, sob condições idênticas. Cada corrida gerou, em média, 50 confôrmeros classificados em ordem crescente de energia total de interação do complexo (ΔG). Desta forma, 1500 confôrmeros foram analisados. Na tabela 6A (Anexo A) estão apresentados os valores de energia obtidos para os confôrmeros com menor e maior energia de cada corrida. A classificação, ou ranqueamento, dos ancoramentos obtidos com o programa computacional é obtida pela função denominada *score*, que consiste na somatória de diferentes contribuições energéticas, a saber:

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{ele} + \Delta G_{conf} + \Delta G_{sol}$$

Onde: ΔG_{vdw} é a contribuição de energia de van der Waals, ΔG_{hbond} corresponde às contribuições de energia das ligações de hidrogênio, ΔG_{ele} é a contribuição de energia eletrostática, ΔG_{conf} corresponde às contribuições de energia de arranjos conformacionais e ΔG_{sol} é a contribuição de energia de solvatação do complexo (MORRIS *et al*., 2009).

O ranqueamento dos complexos obtidos (*score*) considerando os valores de energia foi considerado na escolha dos confôrmeros, uma vez que os confôrmeros melhores classificados seriam energeticamente mais favoráveis, traduzindo a melhor afinidade do ligante ao sítio de interação. Para a seleção dos confôrmeros, também se

considerou a frequência de orientação das porções da cadeia lateral e do núcleo catecólico da molécula do 4-NC no sítio catalítico da MMP-2.

A análise dos resultados de ancoramento molecular sugeriu que, uma vez ancorados no sítio ativo da MMP-2, 90% dos confôrmeros de menor energia tendem a se orientar de forma a acomodarem a porção da cadeia lateral no bolso hidrofóbico (S1'), adjacente ao sítio catalítico (Figura 44, A). Outros confôrmeros, que apareceram com bastante frequência, mas não correspondiam aos confôrmeros de menor energia, se orientaram de forma a acomodar o núcleo catecólico da estrutura do 4-NC dentro do bolso S1´ (Figura 44, B). Por este motivo, dois confôrmeros foram selecionados como padrão de orientação para as simulações de DM: um com o anel catecólico voltado para o bolso S1'e outro com a porção da cadeia lateral voltada para o bolso S1'.

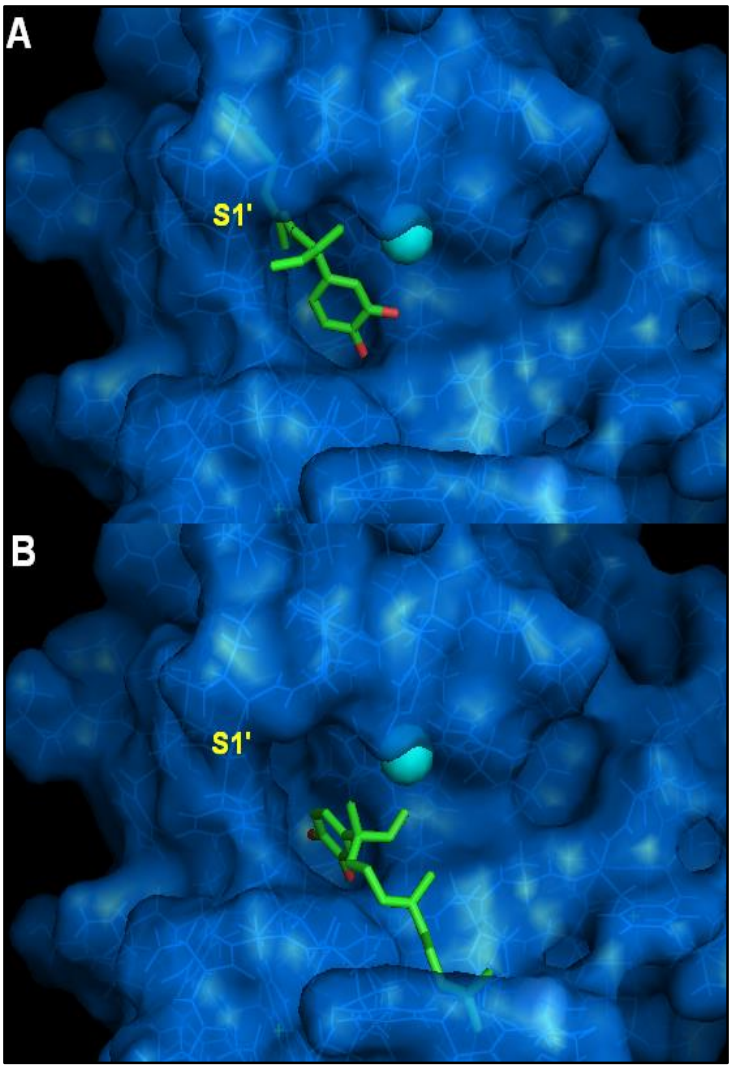


Figura 44. Confôrmero do 4-NC orientado no sítio catalítico da enzima MMP-2 com a cadeia lateral voltada para o bolso S1' (**A**) e com o núcleo catecólico voltado para o bolso S1' (**B**). Os átomos de carbono estão representados em verde, oxigênios em vermelho e os átomos de hidrogênios não estão explícitos. O átomo de zinco do sítio catalítico está representado em modelo CPK (preenchido) em azul ciano (Pymol 1.5.0.1, DeLano, 2012).

Os confôrmeros cujas cadeias laterais estavam voltadas para o bolso S1' foram muito semelhantes. O confôrmero que apresentou a menor energia entre os demais (-8,13 kcal/mol) foi selecionado como *input* para as simulações computacionais de dinâmica molecular.

Entre os confôrmeros que acomadaram o anel catecólico dentro do bolso S1', aquele que apresentou a menor energia (-6,64 kcal/mol) foi também selecionado para os estudos de dinâmica molecular.

O estudo de ancoramento molecular foi empregado com o intuito de encontrar o confôrmero do 4-NC com melhor ajuste (orientação) no sítio ativo da MMP-2. Entretanto, cabe ressaltar que uma das limitações desta ferramenta está relacionada à flexibilidade do receptor durante as análises e, também, à aproximação das variações entrópicas conformacionais das moléculas que participam do processo de interação (MORGON, COUTINHO, 2007; ALONSO, BLIZNYUK, GREADY, 2006). Por isso, os complexos selecionados nos estudos de ancoramento molecularserão utilizados como estruturas de entrada (*input*) para simulações computacionais de DM. A combinação dos dois métodos (ancoramento molecular e simulação de DM) permite compreender melhor o comportamento do sistema, auxiliando na obtenção de informações mais detalhadas e de maior confiabilidade sobre o mecanismo de inibição do 4-NC frente à enzima MMP-2.

4.6.2. Estudos de ancoramento do 4-NC no sítio catalítico da enzima MMP-9

A estrutura tridimensional do 4-NC previamente construída e otimizada foi ancorada ao sítio catalítico da enzima MMP-9 por trinta vezes, sob condições idênticas. Os confôrmeros gerados foram analisados de acordo com os critérios já descritos para a enzima MMP-2 (no item **4.6.1**). Avaliaram-se 1500 confôrmeros para o 4-NC no sítio catalítico da MMP-9. Na tabela 7A (Anexo A) estão apresentados os valores de energia obtidos para os confôrmeros de menor e de maior energia em cada corrida. A análise dos resultados de ancoramento molecular sugeriu que, uma vez ancorado no sítio ativo da MMP-9, a maioria dos confôrmeros de menor energia (73%) tende a se orientar de forma a acomodar a porção da cadeia lateral em um bolso hidrofóbico (S1'), adjacente ao sítio catalítico (Figura 45, A). Os outros confôrmeros (27%) acomodaram o núcleo catecólico da estrutura do 4-NC no bolso S1´ (Figura 45, B).

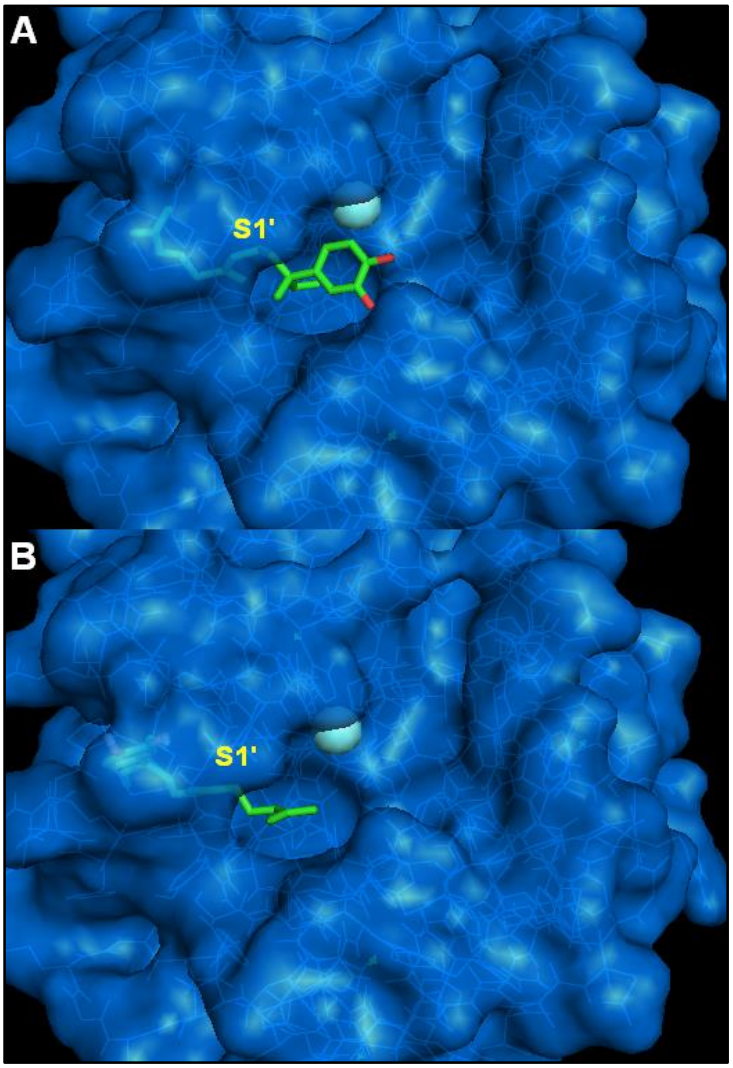


Figura 45. Confôrmero do 4-NC orientado no sítio catalítico da enzima MMP-9 com a cadeia lateral voltada para o bolso S1' (**A**) e com o núcleo catecólico voltado para o bolso S1' (**B**). Átomos de carbono estão representados em verde, oxigênios em vermelho e os átomos de hidrogênios não estão explícitos. O átomo de zinco do sítio catalítico está representado em modelo CPK (preenchido) na cor azul ciano (Pymol 1. 5.0.1, DeLano, 2012).

O confôrmero energeticamente mais favorável, cuja cadeia lateral estava voltada para o bolso S1´ (-9,0 kcal/mol), foi selecionado como *input* para as simulações computacionais de DM. Selecionou-se ainda uma conformação que acomodou o núcleo catecólico no bolso S1´para avaliação (-8,79 kcal/mol). O estudo ancoramento foi empregado com o intuito de encontrar o confôrmero do 4-NC com melhor ajuste no sítio ativo da MMP-9. Os complexos obtidos foram utilizados como estrutura de entrada (*input*) nas simulações de DM, a fim de se avaliar melhor o comportamento do sistema molecular e para obterinformações mais detalhadas e mais confiáveis sobre o mecanismo de inibição do 4-NC frente à enzima MMP-9.

4.7. Simulações de dinâmica molecular dos complexos com o 4-NC

Para as simulações de dinâmica molecular (DM) empregou-se o software MOLSIM 3.2 (DOHERTY, 2002). Os modelos tridimensionais dos complexos MMP-2/4-NC e MMP-9/4-NC, previamente obtidos no software Autodock 4 (HUEY *et al*., 2007), foram minimizados com os métodos de declive máximo e de gradientes conjugados e, após este procedimento, foram submetidos a simulações curtas de DM com esquema de aquecimento (*warming-up)*. Esta abordagem teve como objetivo explorar diferentes estados conformacionais dos complexos, em diferentes condições térmicas, para que barreiras de energia fossem transpostas e se encontrasse o encaixe do ligante mais aproximado do ideal no sítio ativo das enzimas MMP-2 e MMP-9 (TOKARSKI, HOPFINGER, 1997). Na figura 11A (Anexo A) estão apresentados os PAC dos complexos para cada uma das temperaturas do esquema de aquecimento. Nota-se que os complexos atingiram o equilíbrio gradativamente e estabilizaram completamente apenas na simulação de 5ns à temperatura de 310 K. As contribuições de energia computadas nas simulações de DM foram as seguintes: deformação axial ou de estiramento, deformação angular, deformação torsional, van der Waals, eletrostática, ligação de hidrogênio, Lennard-Jones e solvatação. Estas energias foram calculadas para cada um dos confôrmeros de energia mais baixa selecionados da simulação à 310 K e estão apresentadas na Tabela 8A, Anexo A. A energia potencial total doscomplexos também foi calculada e estão listadas na tabela 41. Todos os valores podem ser considerados estáveis, mas não estão diretamente relacionados à afinidade. Para avaliar a afinidade do 4-NC em relação à MMP-2 e MMP-9, teríamos que obter os valores de energia relativos ao estado ligado (complexo) e não ligado (ligante isolado e enzima isolada). Estes estudos ainda estão em desenvolvimento. Além disso, a contribuição do meio solvente é muito importante no processo de reconhecimento em nível molecular e deve ser considerada. Desta forma, os dados obtidos até o momento são preliminares e não conclusivos em relação à afinidade do 4-NC em MMP-2 e MMP-9.

Tabela 41. Valores de energia potencial total dos complexos selecionados das simulações de dinâmica molecular à 310K

Complexos	Energia total (kcal/mol)
MMP-2 (4-NC com anel voltado para bolso S1´)	-14330,7
MMP-2 (4-NC com a cauda voltada para bolso S1´)	-14863,9
MMP-9 (4-NC com a cauda voltada para bolso S1´)	-15071,6
MMP-9 (4-NC com anel voltado para bolso S1´)	-14470,5

Os confôrmeros obtidos para os complexos após as simulações de DM (dados preliminares) foram similares àqueles obtidos no estudo de ancoramento molecular (figura 46).

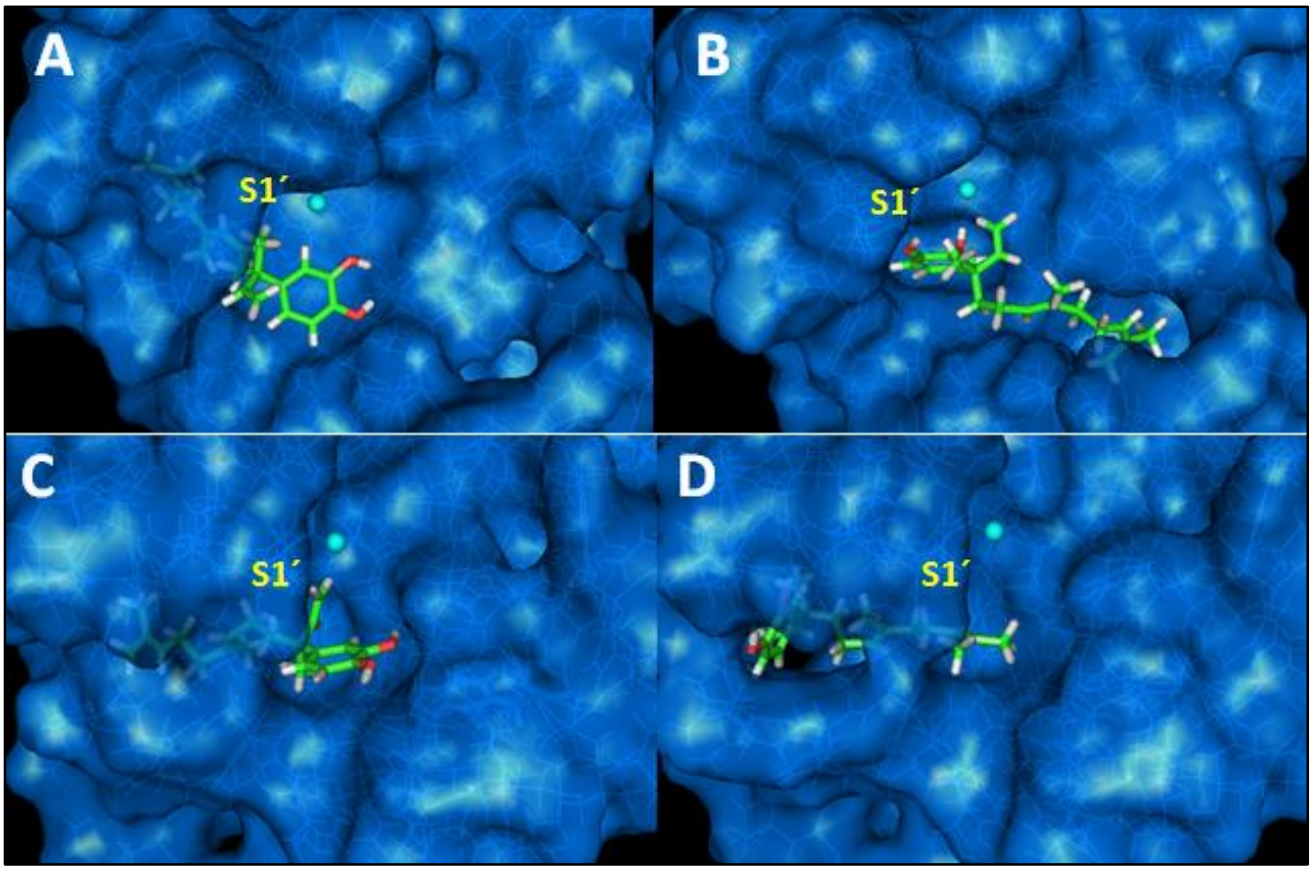


Figura 46. Conformações selecionadas para os complexos MMP-2/4-NC após as simulações de DM (**A**: 4-NC com cauda voltada para bolso S1´; **B**: 4-NC com anel catecol voltado para bolso S1´). Conformações selecionadas para os complexos MMP-9/4-NC após as simulações de DM (**C**: 4-NC com cauda voltada para bolso S1´; **D**: 4-NC com anel catecol voltado para bolso S1´). O modelo molecular do 4-NC está representado em palito ou tubo, onde os átomos de carbono são representados em verde, oxigênio em vermelho e os hidrogênios e branco. A superfície molecular da enzima MMP-2 e da enzima MMP-9 está apresentada em azul. Os átomos de zinco do sítio catalítico estão representados como modelo CPK (preenchido) na cor azul ciano (Pymol 1.5.0.1, DeLano, 2012).

Os valores de energia total demonstraram que os complexos onde a cadeia lateral do 4-NC estava voltada para o bolso S1´se apresentaram termodinamicamente mais estáveis em comparação aos complexos onde o anel catecólico do 4-NC estava voltado para este mesmo bolso. Estes resultados indicam que existe uma tendência de orientação do 4-NC no sítio catalítico das enzimas MMP-2 e MMP-9 e que estes dados estão acordo com os aqueles obtidos anteriormente nos estudos de QSAR-4D IR. Apesar dos modelos de QSAR terem sido construídos a partir de inibidores de classes químicas diferentes do 4-NC, eles foram capazes de prever a orientação desta molécula no sítio ativo destas enzimas, enfatizando a robustez dos mesmos.

Cabe ressaltar que a identidade total considerando o alinhamento das sequências primárias de resíduos de aminoácidos entre as enzimas MMP-2 e MMP-9 é alta (59,7 %; domínio catalítico). Entretanto, os bolsos S1' adjacentes ao sítio, considerados os bolsos da especificidade (TAURO, LYNCH, 2014), apresentam pequenas diferenças. Tais diferenças poderiam explicar os valores de atividade inibitória distintos, descritos para o 4-NC, considerando as duas enzimas (BROHEM, 2009; ROPKE *et al*., 2006).

A sequência de resíduos de aminoácidos da alça, ou *loop*, S1´ na família das metaloproteinases não apresenta similaridade e difere em comprimento. A variabilidade estrutural, a sequência de aminoácidos deste *loop* e a natureza do ligante contribuem para o tamanho total e o formato dos bolsos S1´ nas diferentes MMP (LOVEJOY *et al*., 1999). A principal característica que determina o formato deste bolso é a orientação dos resíduos localizados na porção final (MMP-2: Thr426; MMP-9: Arg 424 – aminoácidos enumerados de acordo com a estrutura das enzimas intactas (MORGUNOVA *et al*., 1999)). Na estrutura da MMP-2, o resíduo Thr426 está longe do bolso S1´, formando um túnel que direciona para o solvente (ROSWELL *et al.* 2002). Por outro lado, na MMP-9, o aminoácido Arg 424 está orientado para dentro do bolso S1´ formando um bolso anexo muito menor. Desta forma, o bolso S1´da enzima MMP-2 pode interagir melhor com inibidores que possuam a porção P1´ mais longas (YANG *et al*., 2008).

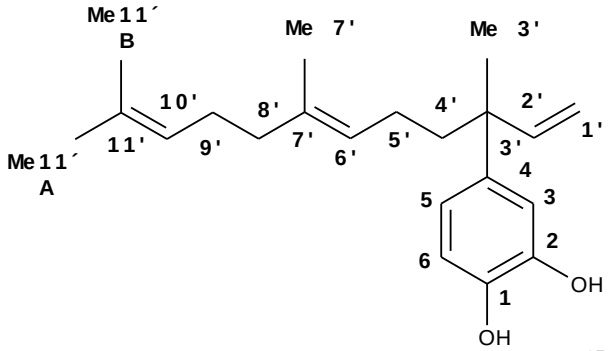
O complexo MMP-9/4-NC, onde a cadeia lateral ddo 4-NC estava voltada para o bolso S1´, apresentou menor valor de energia total (-15071,6 Kcal/mol) quando comparado ao complexo MMP-2/4-NC na mesma orientação (-14863,9 Kcal/mol). Os valores de energia total não consideraram os valores de entropia do sistema, o que dificulta os cálculos para estimar a energia de interação dos complexos (ΔG). Ainda, como já mencionado, as energias do estado não-ligado (ligante isolado com o meio solvente e enzima isolada com o meio solvente) não foram consideradas até o momento.

Estes cálculos estão em desenvolvimento (programa GROMACS) (HESS *et al*., 2008; VAN DER SPOEL *et al*., 2005; LINDAHL *et al*., 2001; BERENDENSEN *et al*., 1995; BEKKER *et al*., 1993) e os resultados serão publicados em momento oportuno. Por este motivo, ainda não é possível afirmar por qual das enzimas o 4-NC teria maior afinidade.

Os estudos de ancoramento e simulações de dinâmica molecular foram apresentados no evento 3rd Medicinal Chemistry and Computer Aided Drug Designing, 2014.

4.8. Extração do 4-Nerolidilcatecol

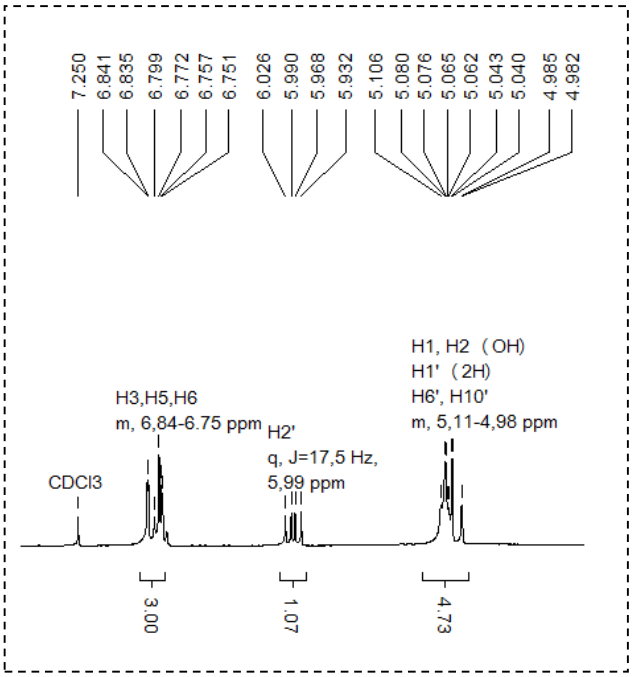
Após preparo do extrato concentrado e liofilizado de *Pothomorphe umbellata*, o isolamento do 4-NC resultou em 131,62 mg (3,29 %) do composto. O composto foi caracterizado por ressonância magnética nuclear de hidrogênio (RMN-^{1}H) e carbono (RMN-^{13}C). Os espectros obtidos são apresentados nas Figuras 47 e 48:

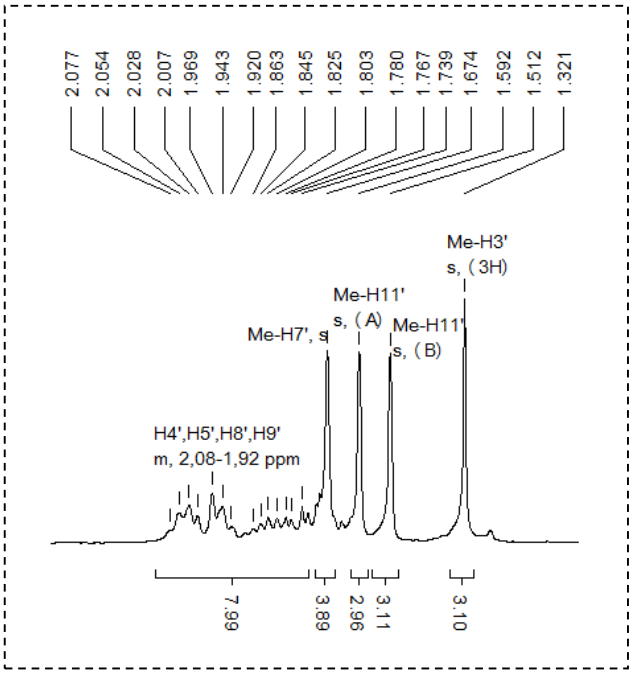


Parâmetro processo:
300,1299945 MHz

Solvente:
Clorofórmio - $CDCl_3$

Padrão de referência interna:
TMS





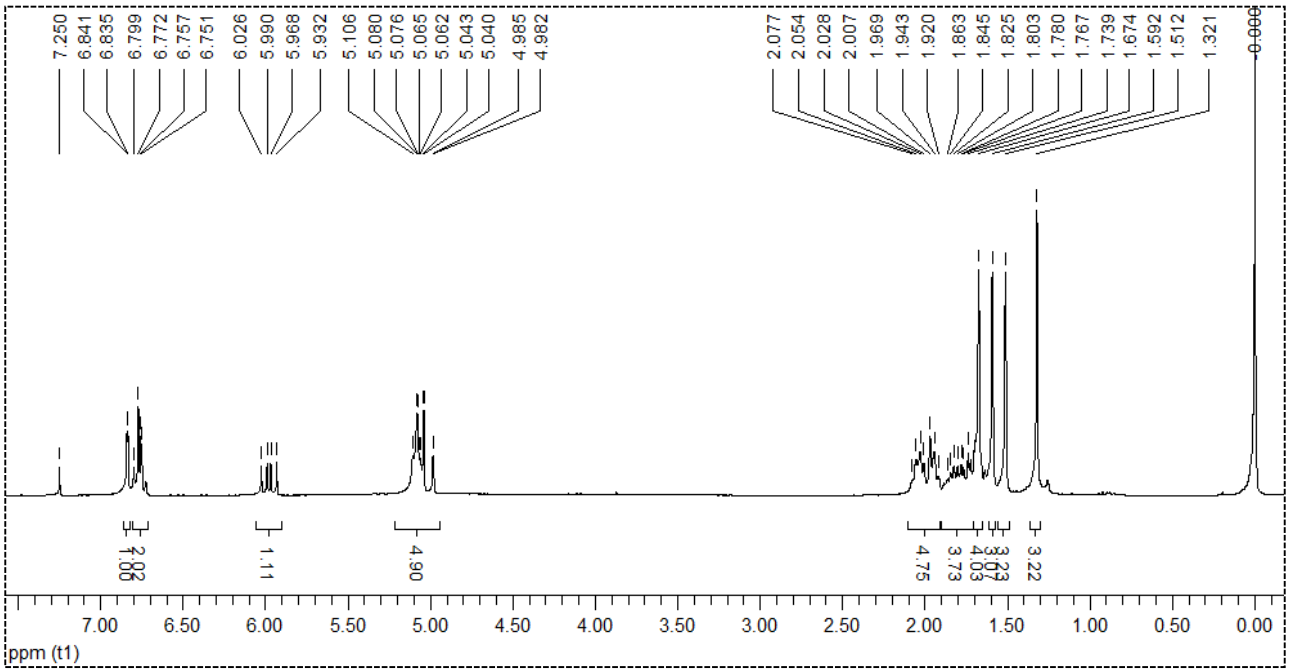
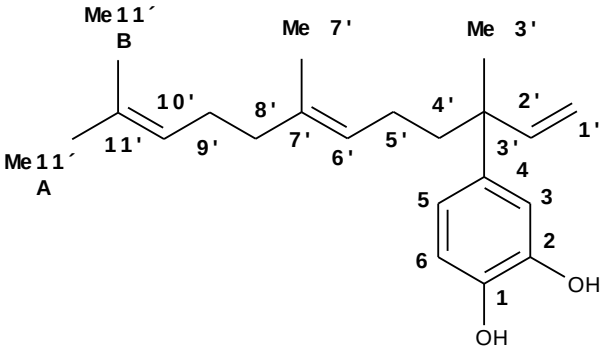


Figura 47. Espectro de RMN ^{1}H do 4-nerolidilcatecol.



Parâmetro processo:
75,4677867 MHz

Solvente:
Clorofórmio - $CDCl_3$

Padrão de referência interna:
TMS

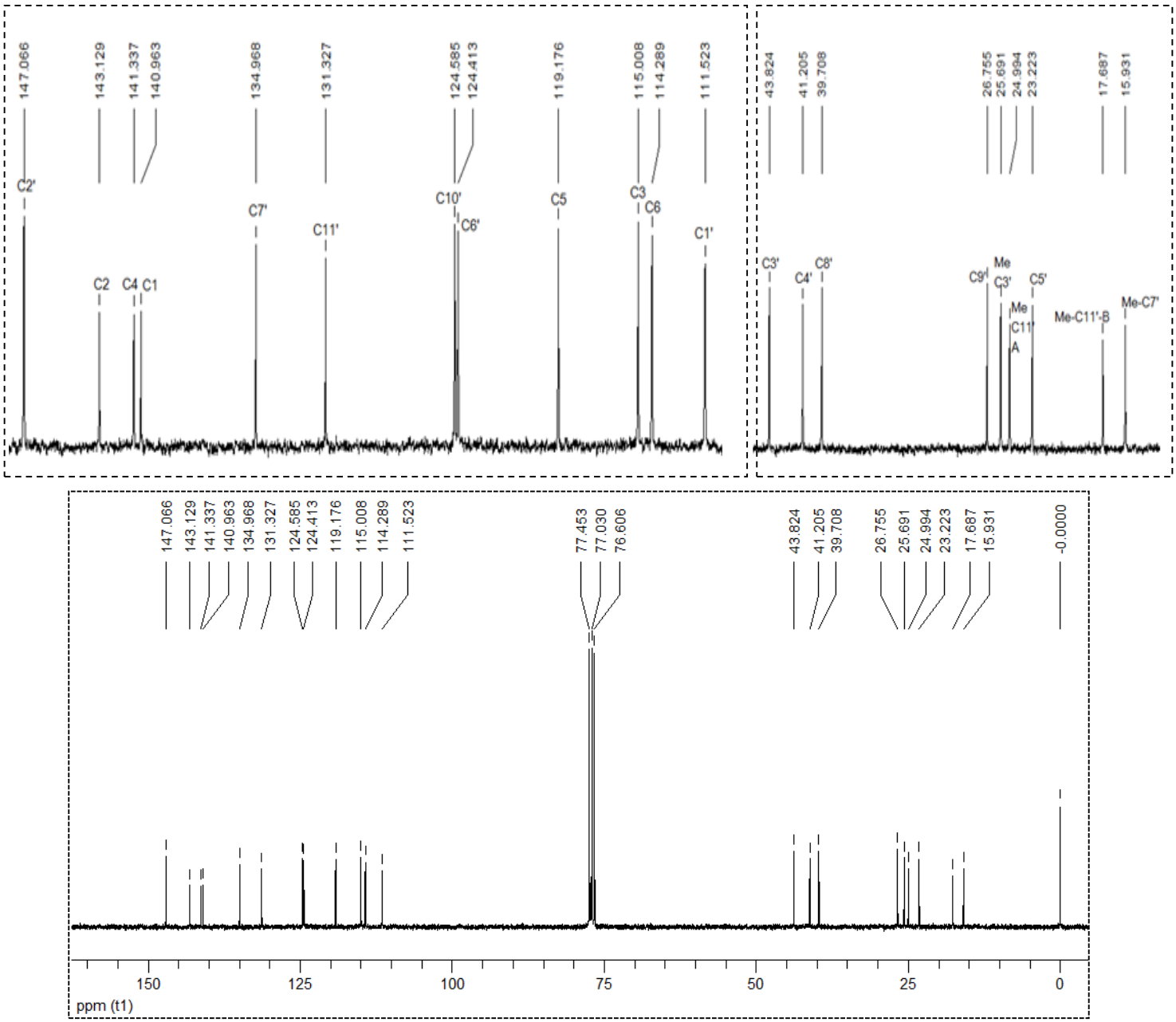


Figura 48. Espectro de RMN ^{13}C do 4-nerolidilcatecol.

Os sinais referentes aos hidrogênios e carbonos da estrutura do 4-NC foram condizentes aos descritos por Kijjoa e colaboradores (1980).

4.9. Ensaios de zimografia

A atividade inibitória *in vitro* dos compostos 4-nerolidilcatecol, nerolidol, catecol e 1,2- dimetoxibenzeno frente às metaloproteinases MMP-2 e MMP-9 (fração total: pró + ativa) foi avaliada por ensaios de zimografia.

Estas moléculas foram escolhidas para que se pudesse avaliar a contribuição da cadeia lateral e do núcleo catecólico da estrutura química do 4-NC na atividade inibitória frente às enzimas em estudo. Acredita-se que as hidroxilas presentes no núcleo catecólico do 4-NC sejam responsáveis por quelar os átomos de zinco das metaloproteinases, íons fundamentais para sua atividade enzimática. Por essa razão, avaliou-se a capacidade inibitória do catecol isolado e também do composto 1,2-dimetoxibenzeno, que possuias duas hidroxilas bloqueadas por grupos metila (figura 8).

Como se pode verificar na figura 49, A, os valores obtidos sugerem um padrão de atividade inibitória dos compostos para ambas enzimas. As áreas de digestão das bandas referentes às frações total da MMP-2 e MMP-9 dos géis incubados com 4-NC e catecol apresentam áreas significativamente menores que o tampão (Figura 49, B).

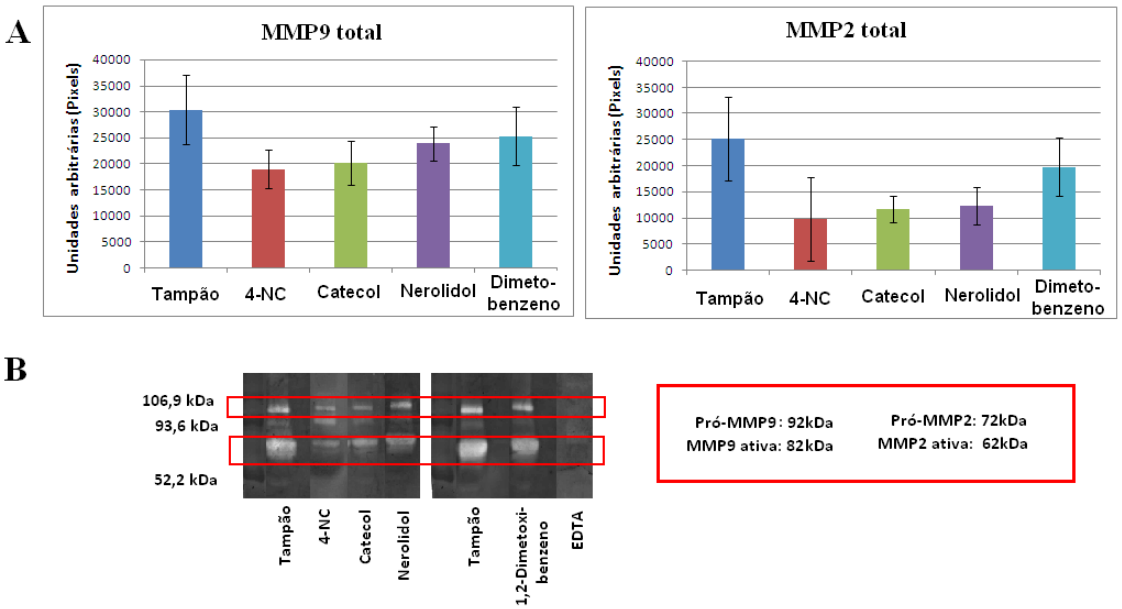


Figura 49. (A) Gráficos referentes aos valores de pixels correspondentes à fração total (pró + ativa) das MMP-2 e 9 nos géis de zimografia. Os resultados foram expressos como média de pixels ± erro padrão (n = 4-5). (B) Géis obtidos nos ensaios de zimografia.

O 4-NC apresentou atividade inibitória frente às MMP-2 e MMP-9 quando comparado ao tampão, e a porção catecol apresentou atividade semelhante ao 4-NC para ambas enzimas. Por outro lado, o nerolidol apresentou atividade inibitória apenas frente à MMP-2.

Os resultados sugerem que o núcleo catecólico da estrutura do 4-NC (hidroxilas livres) seja oresponsável pela sua atividade, já que o composto 1,2-dimetoxibenzeno, que possui as hidroxilas bloqueadas por grupos metil, não foi capaz de exercer atividade inibitória significante para nenhuma das enzimas. Ainda, vale ressaltar que a ausência das bandas nos géis de zimografia pode ter ocorrido devido à propriedade quelante do 4-NC e do catecol frente aos íons Zn^{2+}. A ausência destes íons no tampão de incubação impediu que as enzimas fossem ativadas corretamente e, por este motivo, não degradaram o gel. Este fato reforça a hipótese de que o 4-NC possua a capacidade de quelar os íons de Zn^{2+}.

4.10. Ensaios de voltametria

A voltametria cíclica é uma técnica eletroquímica muito utilizada na caracterização de processos eletroquímicos que ocorre na superfície de um eletrodo de trabalho, neste casocarbono vítreo; enquanto a voltametria de pulso diferencial, por ser uma técnica bem mais sensível, tem sido mais utilizada em estudos de interação do composto alvo (4-NC) com eventuais espécies presentes no meio e, por esse motivo foi aqui utilizada. Em ambas as técnicas os processos de oxidação e redução são exatamente os mesmos. A utilização das técnicas eletroquímicas no estudo de sistemas biológicos tem sido cada vez mais frequente, principalmente porque podem simular, *in vitro*, reações que ocorrem *in vivo* e, muitas vezes assistidas por sistemas enzimáticos.

O voltamograma cíclico obtido com eletrodo de carbono vítreo em solução contendo (5 mmol/L) de $CaCl_2$ + (1,59 mmol/L) de 4-NC em tampão Tris-HCl (Figura 50, A), apresentou um par redox reversível com o potencial de pico de oxidação (Epa) em (0,100V) e potencial de pico de redução (Epc) em 0,075V, de modo que nestas condições experimentais o potencial redox (E^0) do par $4\text{-}NC_{oxidado}/4\text{-}NC_{reduzido}$ foi de (0,0875V) e corresponde ao valor de (Epa + Epc)/2. A corrente de pico de oxidação (Ip, a), corresponde à altura do pico em relação à linha base e este valor é proporcional à concentração de 4-NC em solução. Este conceito vale também para a voltametria de pulso diferencial (Figura 50, B). O processo de oxidação e redução observados no voltamograma cíclico apresentado na Figura 50, A,

correspondem, respectivamente, à oxidação de grupos OH da molécula de 4-NC e redução da respectiva quinona formada no eletrodo após o processo de oxidação.

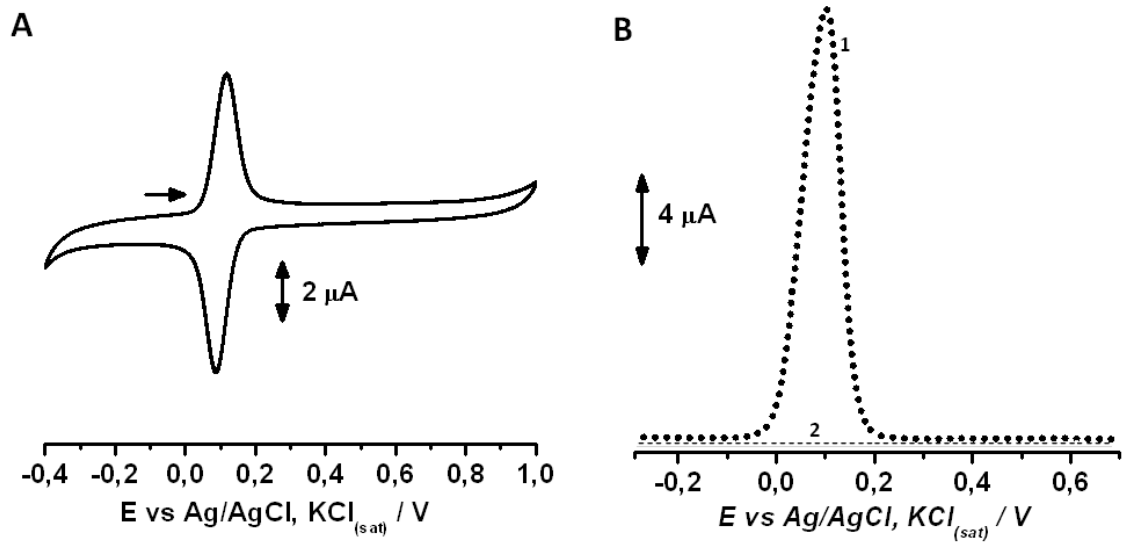


Figura 50. **(A)** Voltamograma cíclico obtido em uma solução contendo 5 mmol/L de $CaCl_2$, e 1,59 mmol/L de 4-NC em tampão Tris-HCl, pH 8. **(B)** Voltamograma de pulso diferencial obtido em solução contendo 5 mmol/L de $CaCl_2$ e 1,59 mmol/L de 4-NC em tampão Tris-HCl, pH 8,0 **(curva 1)** e em solução contendo 5 mmol/L de $CaCl_2$, 5 μmol/L de $ZnCl_2$ em tampão Tris-HCl, pH 8,0 **(curva 2)**.

Para avaliar a eventual possibilidade de complexação de íons Zn^{2+} pelo 4-NC obtiveram-se voltamogramas de pulso diferencial na solução contendo 5 mmol/L de $CaCl_2$ e 1,59 mmol/L de 4-NC em tampão Tris-HCl, pH 8,0. Em seguida, várias alíquotas de íons Zn^{2+} foram adicionadas à solução contida na célula eletroquímica sem que houvesse variação apreciável de volume. Novo voltamograma de pulso diferencial foi obtido após a adição de cada alíquota de solução do metal. Os dados são apresentados na Figura 51.

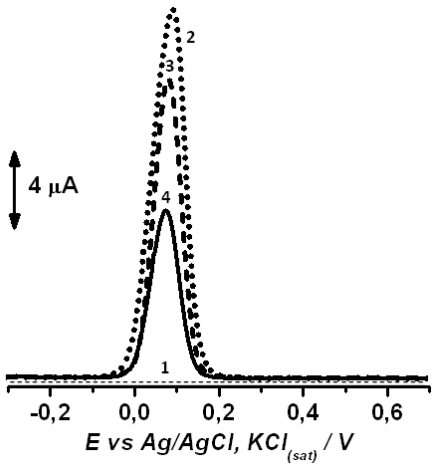


Figura 51. Voltamogramas de pulso diferencial obtidos em soluções contendo: **(1)** 5 mmol/L de $CaCl_2$, 5 μM de $ZnCl_2$ em tampão Tris-HCl, pH 8,0; **(2)** 5 mmol/L de $CaCl_2$ + 1,59 mM de 4-NC em tampão Tris-HCl, pH 8,0; **(3)** 5 mmol/L de $CaCl_2$ + 1,59 mM de 4-NC em tampão Tris-HCl,

pH 8,0 + 5 μM $ZnCl_2$ e **(4)** 5 mmol/L de $CaCl_2$ + 1,59 mM de 4-NC em tampão Tris-HCl, pH 8,0 + 50 μM $ZnCl_2$.

De acordo com os dados apresentados na Figura 51, a corrente de oxidação observada em Epa1 diminuiu com o aumento da concentração de íons Zn^{2+} em solução, indicando que de fato, o 4-NC complexa o metal. Como o 4-NC sofre oxidação eletroquímica em um potencial similar ao observado para o catecol (ENACHE, OLIVEIRA-BRETT, 2011), sugere-se a interação e complexação entre os íons Zn^{2+} ocorra envolvendo os grupos eletroativos fenol/quinona, já que este é o sistema eletroquímico ativo. Deste modo, por meio das técnicas eletroquímicas foi possível ter fortes indícios que ocorra a coordenação entre o íon metálico Zn^{2+} e o 4-NC.

5. CONCLUSÃO

A análise exploratória de dados gerou resultados complementares para os métodos HCA e PCA considerando os dois conjuntos de inibidores de MMP, MMP-2 e MMP-9. As propriedades moleculares de natureza estérica/intrínseca, estérica/hidrofóbica e estérica/eletrônica foram importantes para a descrição e discriminação dos compostos altamente ativos. A presença de substituintes volumosos em R_3 se mostrou crucial para este conjunto de inibidores, considerando o estabelecimento de interações moleculares com a cavidade S1 das duas enzimas. Entretanto, há um limite de volume a ser considerado.

Em relação aos estudos de QSAR-4D IR, obtiveram-se o mapeamento do farmacóforo e a hipótese de conformação bioativa prevista pelos modelos para cada inibidor, com a avaliação das possíveis interações frente às enzimas MMP (complementaridade no sítio de ação), explorando diferentes alinhamentos (orientações no sítio ativo). Obtiveram-se modelos robustos, sem correlação ao acaso, com boa capacidade de previsão (inclusive para classes químicas não relacionadas) e que podem auxiliar na proposição de novos inibidores potenciais para MMP-2 e MMP-9.

Os estudos de *redocking*, empregados para verificar se os parâmetros utilizados no processo de ancoramento do 4-NC nas enzimas MMP-2 e MMP-9 estavam adequados, apresentaram resultados satisfatórios e delinearam as condições para os estudos de ancoramento molecular. Os resultados de ancoramento molecular desenvolvidos para determinar o modelo de interação do 4-NC com as biomacromoléculas (MMP-2 e MMP-9) apresentaram duas conformações favoráveis no processo de inibição dos complexos. A associação de simulações computacionais de DM aos estudos de ancoramento forneceram informações mais detalhadas e de maior confiabilidade, indicando que o 4-NC tende a se orientar no sítio catalítico de forma a acomodar a cadeia lateral no bolso S1´ nas duas enzimas.

Ensaios de zimografia também foram desenvolvidos com intuito de auxiliar na compreensão da possível contribuição do núcleo catecólico da estrutura química do 4-NC em sua atividade inibitória frente às enzimas em estudo. Os resultados sugerem que este anel seja o responsável por sua atividade quelante. Esta hipótese foi confirmada por ensaios de voltametria.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ABDI, H.; WILLIAMS, L.J. Principal component analysis. **WIREs Computational Statistics**. New Jersey, v. 2, p. 434-459, 2010.

Accelrys Inc, WebLab Viewer Lite, version 5.0, Molecular Simulations Inc, (San Diego, CA), 2002.

Accelrys Software Inc., Discovery Studio Modeling Environment, Release 4.0, San Diego: Accelrys Software Inc., 2013.

ADMET/MI-QSAR Version 1.0 User's Guide; The Chem21 Group Inc. Chicago, 2001.

ALLINGER, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. **Journal of the American Chemical Society**, Washington, v. 99, n. 25, p. 8127-8134, 1977.

ALONSO, H.; BLIZNYUK, A. A.; GREADY, J.E. Combining docking and molecular dynamic simulations in drug design. **Medicinal Research Reviews**, New York, v. 26, n. 5, p. 531-68, 2006.

ANDRADE, C.H.; PASQUALOTO, K.F.M.P.; FERREIRA, E.I.; HOPFINGER, A.I. Rational Design and 3D-Pharmacophore Mapping of 5'- Thiourea-Substituted α-Thymidine Analogues as Mycobacterial TMPK Inhibitors. **Journal of Chemical Information and Modeling**, Washington, DC, v. 49, p. 1070-1078, 2009.

ANDREI, C.C.; FERREIRA, D.T.; FACCIONE, M.; FARIA, T.J. Da química medicinal à química combinatória e modelagem molecular. In: **Modelagem Molecular**. São Paulo: Manole, 2003. p.111-129.

AURELI, L.; GIOIA, M.; CERBARA, I.; MONACO, S.; FASCIGLIONE, G.F.; MARINI, S.; ASCENZI, P.; TOPAI, A.; COLETTA, M. Structural Bases for Substrate and Inhibitor Recognition by Matrix Metalloproteinases. **Current Medicinal Chemistry**, v. 15, p. 2192-2222, 2008.

BADKE, G.L.; PANAGOPOULOS, A.T. AGUIAR, G.B.; VEIGA, J.C.E. Glioblastoma multiforme em idosos: uma revisão sobre seu tratamento com ênfase na abordagem cirúrgica. **Arquivos Brasileiros de Neurocirurgia**., v. 33, n°1, p. 45-51, 2014.

BALABAN, A.T. Highly discriminating distance-based topological index. **Chemical Physics Letters**, Amsterdam, v.89, n.5, p.399-404, 1982.

BALCH, C.M.; GERSHENWALD, J.E.; SOONG, S.J.; THOMPSON, J.F.; ATKINS, M.B.; BYRD, D.R.; BUZAID, A.C.; COCHRAN, A.J.; COIT, D.G.; DING, S.; EGGERMONT, A.M.; FLAHERTY, K.T.; GIMOTTY, P.A.; KIRKWOOD, J.M.; MCMASTERS, K.M.; MIHM, M.C. JR.; MORTON, D.L.; ROSS, M.I.; SOBER, A.J.; SONDAK, V.K. Final version of 2009 AJCC melanoma staging and classification. **Journal of Clinical Oncology**, New York, v. 27, n°36, p.6199-206, 2009.

BARROS, S.B.M.; TEIXEIRA, D.S.; AZNAR, A. E.; MOREIRA, JR., J. A.; ISHII, I.; FREITAS, P. C. D. Antioxidant activity of ethanolic extracts of *Pothomorphe umbellata* L. Miq. **Ciência e Cultura**, Campinas, v. 48, p. 114-116, 1996.

BAS, D.C.; ROGERS, D.M.; JENSEN, J.H. Very Fast Prediction and Rationalization of pKa Values for Protein-Ligand Complexes. **Proteins**, v. 73, p. 765-783, 2008.

BAUMANN, K.**Trends in Analytical Chemistry**.*v. 22, p. 395-406,* 2003.

BAUMANN, K.; STIEFL, N. **Journal of Computer-Aided Molecular Design.** *v. 18, p. 549-562,* 2004.

BEKKER, H.; BERENDSEN, H.J.C.; DIJKSTRA, E.J.; ACHTEROP, S.; VAN DRUNEN, R.;VAN DERSPOEL, D.; SIJBERS, A.;KEEGSTRA, H.; REITSMA, B.; RENARDUS, M.K.R.Gromacs: A parallel computer for molecular dynamics simulations. In: GROOT, R.A.; NADRCHAL, J. Physics Computing 92, Singapore. World Scientific. p. 252-256, 1993.

BERENDSEN, H.J.C.; POTSMA, J.P.M.; VAN GUNSTEREN, W.F.; DI NOLA, A.D.; HAAK, J.R. Molecular dynamics with coupling to an external bath. **Journal of Chemical Physics,** v. 81, p. 3684–3690, 1984.

BERENDSEN, H.J.C.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: A message-passing parallel molecular dynamics implementation. **Computer Physics Communications**., v. 91, p. 43–56, 1995.

BERMAN, H.M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T.N.; WEISSIG, H.; SHINDYALOV, I.N.; BOURNE, P.E. The Protein Data Bank. **Nucleic acids research**, Oxford**,** v. 28, n. 1, p. 235-242.

BODOR, N.; GABANYI, Z.; WONG, C.K. A New Method for the Estimation of Partition-Coefficient. **Journal of the American Chemical Society**, Washington, v.111, n.11, p.3783-3786, 1989.

BRENEMAN, C.M.; WIBERG, E.K.B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis. **Journal of Computational Chemistry**, New York, v.11, n.3, p.361-373, 1990.

BROHEM, C.A. **Mecanismo de ação do 4-nerolidilcatecol na indução da morte celular e contenção da invasão em linhagens de melanoma humano e modelo de pele artificial**. 2010. (Doutorado). Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo.

BROHEM, C.A.; SAWADA, T.C.; MASSARO, R.R.; ALMEIDA, R.L.; RIVELLI, D.P.; ROPKE, C.D.; DA SILVA, V.V.; DE LIMA, T.M.; CURI, R.; BARROS, S.B.; MARIA-ENGLER, S.S. Apoptosis induction by 4-nerolidylcatechol in melanoma cell lines. **Toxicology in Vitro**, New York, v. 23, n. 1, p. 111-9, 2009.

BULTINCK, P.; WINTER, H.; LANGEBAEKER, W.; TOLLENAERE, J.P. **Computational medicinal chemistry for drug discovery**. New York: Marcel Dekker, 2004.

CARVALHO, I.; PUPO, M.T.; BORGES, A.D.; BERNARDES, L. Introdução a modelagem molecular de fármacos no curso experimental de química farmacêutica. **Química Nova**, São Paulo, v. 26, p. 428-438, 2003.

CHAMBERS, J.M.; CLEVELAND, W.S.; KLEINER, B.; TUKEY, P.A. **Graphical methods for data analysis**. 1°Edição. London: CRC Press, 1983.

CHOE, G., PARK, J.K., JOUBEN-STEELE, L., KREMEN, T.J., LIAU, L.M,, VINTERS, H.V., CLOUGHESY, T.F., MISCHEL, P.S. Active matrix metalloproteinase 9 expression is associated with primary glioblastoma subtype. **Clinical Cancer Research**, v.8, n. 9, p.2894-901, 2002.

CLARK, R.D.*;* FOX, P.C. **Journal of Computer-Aided Molecular Design. Des.** *v.18, p.*563–576, 2004.

CORREIA, P.R.M.; FERREIRA, M.M.C. Reconhecimento de padrões por métodos não supervisionados: Explorando procedimentos quimiométricos para tratamento de dados analíticos. **Química Nova**, v. 30, n. 2, p. 481-487, 2007.

CORNELL, W.D.; CIEPLAK P.; BAYLY, C.I.; GOULD, I.R.; MERZ, K. M., JR.; FERGUSON, D.M.; SPELLMEYER, D.C.; FOX, T.; CALDWELL, J.W.; KOLLMAN, P.A. A second generation force field for the simulation of proteins and nucleic acids. **Journal of the American Chemical Society**, Washington, v. 117, p. 5179-5197, 1995.

DADRASS, O.G.; SOBHANI A.M.; SHAFIEE, A.; MAHMOUDIAN, M. Flexible Ligand Doking Studies of Matrix Metalloproteinase Inhibitors Using Lamarkian Genetic Algorithm. **DARU Journal of Pharmaceutical Sciences,** v. 12, n. 1, p. 1-10, 2004.

DAPPRICH, S.; DANIELS, A. D.; STRAINS, M.C.; FARKAS, O.; MALICK, D. K.; RABUCK, A.D.; RAGHAVACHARI, K.; FORESMAN, J.B.; ORTIZ, J. V.; CUI, Q.; BABOUL, A.G.; CLIFFORD, S.; CLOSLOWSKI, J.; STEFANOV, B.B.; LIU, G.; LIASHENKO, A.; PISKORZ, P.; KOMAROMI, I.; MARTIN, R. L.; FOX, D.J.; KEITH, T.; AL-LAHAM, M. A.; PENG, C. Y.; NANAYAKKARA, A.; CHALLACOMBE, M.; GILL, P. M. W.; JOHNSON, B.; CHEN, W.; WONG, M.W.; GONZALEZ, C.; POPLE, J. A. **Gaussian**. Wallingford CT, 2004.

DE LA FUENTE-GARCIA, A.; OCAMPO-CANDIANI, J. Cutaneous melanoma. **Gaceta medica de Mexico**, México, v. 146, n. 2, p. 126-35.

DELANO, W.L. **The PyMOL Molecular Graphics System.** Palo Alto: DeLano Scientific LLC. 2002.

DEWAR, M.J.S.E., ZOEBISCH, G., HEALY, E.F., STEWART, J.J.P. AM1: A New General Purpose Quantum Mechanical Molecular Model. **Journal American of Chemical Society**, Washington, v. 107, p. 3903-3909, 1985.

DOHERTY, D. MOLSIM: Molecular Mechanics and Dynamics Simulation Software, version 3.2 – User's Guide. The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA, 2002.

DUNBRACK Jr, R.L. Rotamer libraries in the 21st century. **Current Opinion in Structural Biology**., v. 12, n. 4, p. 431-440, 2002.

ENACHE, T.A.; OLIVEIRA-BRET, A.M. Phenol and *para*-substituted phenols electrochemical oxidation pathways. **Journal of Electroanalytical Chemistry**,v. 655, n. 1, p. 9-16, 2011.

ERIKSSON L.*;* JAWORSKA, J.; WORTH, A.P.; CRONIN, M.T.D.; MCDOWELLR M.; GRAMATICA, P. **Environmental Health Perspectives**.*, v.*111, *p.*1361–1375, 2003.

ERTL, P.; ROHDE, B.; SELZER, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. **Journal of Medicinal Chemistry**, Amsterdam, v.43, n.20, p.3714-3717, 2000.

ESSMAN, V.; PERERA, L.; BERKOWITZ, M.L.; DARDEN, T.; LEE, H.A smooth particle mesh Ewald method. **Journal of Chemical Physics,** v. 103, p. 8577–8593, 1995.

FENG, Y.; LIKOS, J.J.; ZHU, L.; WOODWARD, H.; MUNIE, G.; MCDONALD, J.J.; STEVENS, A.M.; HOWARD, C.P.; DE CRESCENZO, G. A.; WELSCH, D.; SHIEH, H. S.; STALLINGS, W. C. Solution structure and backbone dynamics of the catalytic domain of matriz metalloproteinase-2 complexed with a hydroxamic acid inhibitor. **Biochimica et Biophysica Acta**, New York, v. 1598, n. 1-2, p. 10-23, 2002.

FERRARA, P.; APOSTOLAKIS, J.; CAFLISCH, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. **Proteins**, v. 46, p. 24–33, 2002.

FERREIRA, M.M.C.; ANTUNES, A.M.; MELGO, M.S.; VOLPE, P.L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, São Paulo, v. 22, n.5, p. 724-731, 1999.

FERREIRA, M.M.C; MONTANARI, C.A.; GAUDIO, A.C. Seleção de variáveis em QSAR. **Química Nova**, São Paulo, v. 25, n. 3, p. 439-448, 2002.

FERREIRA, M.M.C. Multivariate QSAR. **Journal of the Brazilian Chemical Society**. São Paulo, v. 13, p. 742-753, 2002.

FLAHERTY, K.T.; PUZANOV, I.; KIM,K. B. , RIBAS, A.; MCARTHUR, G.A.; SOSMAN, J.A.; O'DWYER, P. J.; LEE, R. J.; GRIPPO, J. F.; NOLOP, K.; CHAPMAN, P. B. Inhibition of Mutated, Activated BRAF in Metastatic Melanoma. **The New England Journal of Medicine**, United States, v. 363, n. 9, p.809-819, 2010.

FOCK, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. **Zeitschrift Fur Physik**, Berlin, v.61, n.1-2, p.126-148, 1930.

FORSYTHE, K.H.; HOPFINGER, A.J. Influence of Solvent on Secondary Structures of Poly (L-Alanine) and Poly (L-Proline). **Macromolecules**, Washington, v.6, n.3, p.423-437, 1973.

FRIEDMAN, J.H. Multivariate adaptive regressionsplines, **Annals of Statistics,** Filadélfia, v.19, p.1-141, 1991.

Gaussian 03, Revision B.03, FRISCH, M. J.; TRUCKS, G. W.; SCHLEGEL, H. B.; SCUSERIA, G. E.; ROBB, M. A.; CHEESEMAN, J. R.; MONTGOMERY, J. A.; VREVEN, JR. T.; KUDIN, K. N.; BURANT, J. C.; MILLAM, J. M.; IYENGAR, S. S.; TOMASI, J.; BARONE, V.; MENNUCCI, B.; COSSI, M.; SCALMANI, G.; REGA, N. PETERSSON, G. A.; NAKATSUJI, H.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA,J.; ISHIDA, M.; NAKAJIMA,T.; HONDA, Y.; KITAO, O.; NAKAI, H.; KLENE, M.; LI, X.; KNOX, J. E.; HRATCHIAN, H. P.; CROSS, J. B.; ADAMO, C.; JARAMILLO, J.; GOMPERTS, R.;STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; AYALA, P. Y.; MOROKUMA, K.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; ZAKRZEWSKI, V. G.; DAPPRICH, S.; DANIELS, A. D.; STRAIN, M. C.; FARKAS, O.; MALICK, D. K.; RABUCK, A. D.; RAGHAVACHARI, K.; FORESMAN, J. B.; ORTIZ, J. V.; CUI, Q.; BABOUL, A. G.; CLIFFORD, S.;CIOSLOWSKI, J.;STEFANOV, B. B.; LIU, G.; LIASHENKO, A.; PISKORZ, P.; KOMAROMI, I.; MARTIN, R. L.; FOX, D. J.; KEITH,T.; AL-LAHAM, M. A.; PENG, C. Y.; NANAYAKKARA, A.; CHALLACOMBE, M.; GILL, P. M. W.;

JOHNSON, B.; CHEN, W.; WONG, W., GONZALEZ C. and POPLE, J. A. Gaussian, Inc., Pittsburgh PA, 2003.

GEMPERLINE, P.J. **Practical Guide To Chemometrics.** 2°Edição. London: CRC Press, 2006.

GRAMATICA, P. QSAR & Combinatorial Science**.** *v.*26, p.694–701, 2007.

GREINERT, R. Skin cancer: new markers for better prevention. **Pathobiology**, Basel, v. 76, n. 2, p. 64-81, 2009.

GHOSE, A.K.; CRIPPEN, G.M. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. **Journal of Chemical Information and Computer Sciences**, Washington, v.27, n.1, p.21-35, 1987.

GIALELI, C.; THEOCHARIS, A.D.; KARAMANOS, N.K. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. **FEBS Journal**, v. 278, n.1, p. 16–27, 2011.

GOLBRAIKH, A.; TROPSHA, A. **Journal of Molecular Graphics and Modelling**. New York, v. 20, p. 269 – 276, 2002.

GUPTA, S.P.; PATIL, V.M. Specificity of Binding with Matrix Metalloproteinases. Experientia Supplementum, v.103, p. 35-56, 2012.

GraphPadInstat - GraphPad software, San Diego, Califórnia, USA

GRAVES, A. P.; BRENK, R.; SHOICHET, B. K. Decoys for docking. **Journal of Medicinal Chemistry.**, v. 48, n. 11, p. 3714-3728, 2005.

HARDING, M.M. The Geometry of Metal ligand Interactions in Proteins. **Acta Crystallographica Section D57, p.** 401-411, 2001.

HASEL, W.; HENDRICKSON, F.T.; STILL, W.C. A rapid approximation to the solvent accessible surface areas of atoms. **Tetrahedron Computer Methodology**, Amsterdam, v.1, n.2, p.103-116. 1988.

HEHRE, W.J.; DITCHFIELD, R.; POPLE, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules **Journal of Chemical Physics**, Philadelphia, v.56, n.5, p.2257-2261, 1972.

HERSEY, P.; ZHANG, X.D. How melanoma cells evade trail-induced apoptosis. **Nature Reviews-Cancer**, New York, v. 1, n. 2, p. 142-50, 2001.

HESS, B.; BEKKER, H.; BERENDSEN, H.J.C.; FRAJIE, J.G.E M. LINCS: A Linear Constraint Solver for molecular simulations. **Journal of Computational Chemistry**.v.18, p. 1463–1472, 1997.

HESS, B.; KUTZNER, C.; VAN DER SPOEL, D.; LINDAHL, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. **Journal of Chemical Theory and Computation**., v. 4, n. 3, p. 435–447, 2008.

HOFMANN, U.B.; WESTPHAL, J.R.; WAAS, E.T. ;ZENDMAN, A.J.; CORNELISSEN, I. M.; RUITER, D. J.; VAN MUIJEN, G. N. Matrix metalloproteinases in human melanoma cell lines and xenografts: increased expression of activated matrix metalloproteinase-2 (MMP-2) correlates with melanoma progression. **British Journal of Cancer.**, v. 81, n. 5, p. 774-882, 1999.

HOLLAND, J.H. (1975/1992), Adaptation in Natural and Artificial Systems. 2ªedição. Cambridge: MIT, 1992. 217 p.

HOPFINGER, A.J. 4D-QSAR Software, v. 3.0 – User's Manual. The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA, 1999.

HOPFINGER, A.J.; WANG, S.; TOKARSKI, J.S.; JIN, B.; ALBUQUERQUE, M.; MADHAV, P.J.; DURAISWAMI, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. **Journal of the American Chemical Society**, Washington, v. 119, p. 10509-10524, 1997.

HU, X.; SHELVER W.H. Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. **Journal of Molecular Graphics and Modelling.,** v. 22, n. 2, p. 115-126, 2003.

HUEY, R.; MORRIS, G.M.; OLSON, A.J.; GOODSELL, D.S. A Semi-empirical Free Energy Force Field with Charge-Based Desolvation. **Journal of Computational Chemistry.**, v. **28**, p. 1145-1152, 2007.

HUI, L.; ROBERTSON, A.D.; JENSEN, J.H. Very Fast Empirical Prediction and Interpretation of Protein pKa Values. **Proteins**, v. 61, p. 704-721, 2005.

HyperChem Program Release 8 for Windows, Hypercube, Inc.: Gainesville, FL, 2007. (Evaluation copy).

INCA, Instituto Nacional do Câncer. Disponível em: http://www2.inca.gov.br/wps/wcm/connect/inca/portal/home. Acesso em 23 de outubro de 2010.

INCA, Instituto Nacional do Câncer. Distribuição das taxas de incidência, ajustadas por idade*, segundo o RCBP e período de referência Disponível em: http://www.inca.gov.br/cancernobrasil/2010/docs/taxa_neoplasias/Parte458_registro_de_base_populacional_completo.pdf. Acesso em: maio de 2015

INCA, Instituto Nacional do Câncer. Rio de Janeiro, 1996-2014. Disponível em: http://www.inca.gov.br/. Acesso em: 27 de novembro de 2014.

ImageJ 1.44p - Wayne Rasband, Natiol Institute of Health, USA

INOMATA, S.; MATSUNAGA, Y.; AMANO, S.; TAKADA, K.; KOBAYASHI, K.; TSUNENAGA, M.; NISHIYAMA, T.; KOHNO, Y.; FUKUDA, M. Possible involvment of gelatinases in basement membrane damage and wrinkle formation in chronically ultraviolet B –exposed hairless mouse. **Journal of Investigative Dermatolology**, v.120, p. 1-7, 2003.

JOHNSON, S.C. Hierarchical Clustering Schemes. **PyschometriKa**. New York, v.32 n. 3, p. 241-254, 1967.

JORGENSEN, W.L.; MADURA, J.D. Solvation and Conformation of Methanol in Water. **Journal of the American Chemical Society**, v. 105, p. 1407, 1983.

KASPER, B.; D'HONDT, V.; VEREECKEN, P.; AWADA, A. Novel treatment strategies for malignant melanoma: a new beginning? **Critical Reviews in Oncology Hematology**, New York, v. 62, n. 1, p. 16-22, 2007.

KATO, Y.; YAMASHITA, T.; ISHIKAWA, M. Relationship between expression of matriz metalloproteinase-2 and matriz metalloproteinase-9 and invasion ability of cervical cancer cells. **Oncology Reports**, Athens, v. 9, n. 3, p. 565-9, 2002.

KHADIKAR, P.V.; DESHPANDE, N.V.; KALE, P.P.; DOBRYNIN, A.; GUTMAN, I.; DOMOTOR, G. The Szeged Index and an Analogy with the Wiener Index. **Journal of Chemical Information and Computer Sciences**, Washington, v.35, n.3, p.547-550, 1995.

KIJJOA, A.; GIESBRETCH, A.M.; AKISUE, M.K.; GOTTLIEB, O.R.; GOTTLIEB, H.E. 4-Nerolidylcatechol from *Pothomorphe umbellata*. **Planta Medica**, Stuttgart, v. 39, p. 85-87, 1980.

KIRALJ, R.; FERREIRA, M.M.C. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application, **Journal of Brazilian Chemical Society**., São Paulo, v.20, n.4, p. 770-787, 2009.

KLAHOLZ, B.P.; RENAUD, J.P.; MITSCHLER, A.; ZUSI, C.; CHAMBON, P.; GRONEMEYER, H.; MORAS, D. Conformational adaptation of agonists to the human nuclear receptor RAR gamma. **Nature Structural Biology**, New York, v. 5, n. 3, p. 199-202, 1998.

KLUSÁK, V.; BAŘINKA, C.; PLECHANOVOVÁ, A.; MLČOCHOVÁ, P.; KONVALINKA, J.; RULÍŠEK, L.; LUBKOWSKI, J. Reaction mechanism of glutamate carboxypeptidase II revealed by mutagenesis, x-ray crystallography, and computational method. **Biochemistry,** v. 48, p. 4126–4138, 2009.

KLOPMAN, G.; LI, J.Y.; WANG, S.M.; DIMAYUGA, M. Computer Automated Log P Calculations Based on an Extended Group-Contribution Approach. **Journal of Chemical Information and Computer Sciences**, Washington, v.34, n.4, p.752-781, 1994.

KUBINYI, H. **QSAR: Hanch Analysis and Related Approache**s. Weinheim:, 1993. p.240

LA PORTA, C.A. Drug resistance in melanoma: new perspectives. **Current Medicinal Chemistry**, Sharjah, v. 14, n. 4, p. 387-91, 2007.

LEACH, A.R. **Molecular Modelling: Principles and Applications**. 2 ed. London: Pearson Education, 2001.p. 768

LEVICAR, N.; NUTTALL, R.K.; LAH, T.T. Proteases in brain tumour progression. **Acta Neurochirurgica (Wien)**, v. 145, n. 9, p. 835-838, 2003.

LINDAHL, E.; HESS, B.; VAN DER SPOEL, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. **Journal of Molecular Modeling**., v. 7, p. 306–317, 2001.

LINDORFF-LARSEN K., PIANA S., PALMO K., MARAGAKIS P., KLEPEIS J. L., DROR R. O., SHAw D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. **Proteins,** v. 78, p. 1950-1958, 2010.

LILJEFORS, T.; GUNDERTOFTE, K.; NORRBY, P.; PETTERSSON, I. Molecular mechanics and comparison of force fields. In: **Computational Medicinal Chemistry for Drug Discovery.** New York: Marcel Dekker, 2004. cap. 1°

a LOUIS, D.N.; OHGAKI, H.; WIESTLER, O.D.; CAVENEE, W.K. WHO Classification of tumors of the central nervous system, 3ªed. IARC Press, Lyon, França, 2007.

b LOUIS, D.N.; OHGAKI, H.; WIESTLER, O.D.; CAVENEE, W.K.; BURGER, P.C.; JOUVET, A.; SCHEITHAUER, B.W.; KLEIHUES, P. The 2007 WHO classification of tumors of the central nervous system. **Review Acta Neuropathology**, v. 114, n. 2, p. 97-109, 2007. Errata em: **Review Acta Neuropathology**, v. 114, n. 5, p. 547, 2007.

LOVEJOY, B.; WELCH,A.R.; CARR, S.; LUONG, C.; BROKA, C.; HENDRICKS, R.T.; CAMPBELL, J.A.; WALKER, K.A.; MARTIN, R.; VAN WART, H.; BROWNER, M.F.Crystal structures of MMP-1 and 13 reveal the structural basis for selectivity of collagenase inhibitors. **Nature Structural & Molecular Biology**. v. 6, n. 3, p. 217-221, 1999.

LOWRY, O.H.; ROSEBROUGH, N.J.; FARR, A.; RANDALL, R.J. Protein measurement with Folin-phenol reagent. Journal of Biological Chemistry, v.193, p. 265-275, 1951.

LUCIC, B.; LUKOVITS, I.; NIKOLIC, S.; TRINAJSTIC, N. Distance-related indexes in the quantitative structure-property relationship modeling. **Journal of Chemical Information and Computer Sciences,** Washington, v. 41, p. 527-535, 2001.

LUKOVITS, I.; LINERT, W. A Novel Definition of the Hyper-Wiener Index for Cycles. **Journal of Chemical Information and Computer Sciences**, Washington, v.34, n.4, p.899-902, 1994.

MABALANOBIS, P.C. On tests and measures of group divergence. **Journal of the Asiatic Society of Bengal**, v. 26, p. 541-588, 1930.

MARTENS, H.; NAES, T. Multivariate Calibration. 1ª edição. New York: John Wiley and Sons, 1989. 422 p.

Marvin Beans, versão 5.8.0, Chemaxon, Inc.: Budapeste, HU. 2012

MARTINS, J.A.M.; FERREIRA, M.M C. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR, **Química Nova,** v. 36, n. 4, p. 554-560, 2013.

MASSARO, R.R. Avaliação do efeito inibitório de metaloproteases do extrato de Pothomorphe de linhagens de glioma humano**. 2009. (Mestrado). Faculdade de Ciências Farmacêuticas (FCF). Universidade de São Paulo, São Paulo.**

MILLER, K.J. Additivity methods in molecular polarizability. **Journal of the American Chemical Society**, Washington, v.112, n.23, p.8533-8542, 1990.

MOITA NETO, J.M.; MOITA, G.C. Uma introdução à análise exploratória de dados multivariados. **Química Nova**, São Paulo, v. 21, n. 4, 1998.

MONGELLI, E.; ROMANO, A.; DESMARCHELIER, C.; COUSSIO, J.; CICCIA, G. Cytotoxic 4-nerolidylcatechol from Pothomorphe peltata inhibits topoisomerase I activity. **Planta medica**, Bloomfield, v. 65, n. 4, p. 376-8, 1999.

MORGON, N.H.; COUTINHO, K. **Métodos de química teórica e modelagem molecular**. São Paulo: Livraria da física, 2007. p.489-497.

MORGUNOVA, E.; TUUTTILA, A.; BERGMANN, U.; ISUPOV, M.; LINDQVIST, Y.; SCHNEIDER, G.; TRYGGVASON, K. Structure of human pro-matrix metalloproteinase-2: activation mechanism revealed. Science, v. 284, n. 5420, p. 1667-1670, 1999.

MORRIS, G. M.; GOODSELL, D. S.; HALLIDAY, R. S.; HUEY, R.; HART, W. E.; BELEW, R. K.; OLSON, A. J. Automated docking using a Lamarckian genetic algorithm mandand empirical binding free energy function. **Journal of Computational Chemistry**, New York, v. 19, p. 1639–1662, 1998.

MORRIS, G.M.; HUEY, R.; LINDSTROM, W.; SANNER, M.F.; BELEW, R.K.; GOODSELL, D.S.; OLSON, A.J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. **Journal of Computational Chemistry.**, v. 30, n. 16, p. 2785–2791, 2009.

NAGASE, H.; VISSE, R.; MURPHY, G. Structure and function of matrix metalloproteinases and TIMPs, **Cardiovascular Research,** Oxford, v.69,p.562 – 573, 2006.

NAJMAN, H.; GADELHA, M. I.P. Temozolamida. **Revista Brasileira de Cancerologia**, v. 48, n°3, p. 439-445, 2002.

NAMBA, A.M.; SILVA, V.B.; SILVA, C.H.T.P. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. **Eletica quimica**, v. 33, n. 4, 2008.

NANTASENAMAT, C.; ISARANKURA-NA-AYUDHYA, C.; NAENNA, T.; PRACHAYASITTIKUL, V. A practical overview of quantitative structure-activity relationship. **Experimental and Clinical Sciences**, Dortmund, v. 8, p. 77-88, 2009.

NATIONAL CANCER INSTITUTE, Bethesda. Disponível em: http://www.cancer.gov/. Acesso em: 27 de novembro de 2014.

NATSUME, A. IFN-beta down-regulates the expression of DNA repair gene MGMT and sensitizes resistant glioma cells to temozolomide. **Cancer Research,** v. 65, n. 17, p.7573-7579, 2005.

NELSON, K.K.; MELENDEZ, J.A. Mitochondrial redox control of matriz metalloproteinases. **Free Radical Biology& Medicine**, New York, v. 37, n. 6, p. 768-84, 2004.

Nova software 1.10.4. Metrohm Autolab B. V. Disponível em: http://www.metrohm-autolab.com/

NÚÑEZ, V.; CASTRO, V.; MURILLO, R; PONCE-SOTO, L. A.; MERFORT, I.; LOMONTE, B. Inhibitory effects of *Piper umbellatum and Piper peltatum* extracts towards myotoxic phospholipases A2 from Bothrops snake venoms: isolation of 4-nerolidylcatechol as active principle. **Phytochemistry**, v. 66, n. 9, p. 1017-25, 2005.

ODIA, Y.; SHIH, J.H.; KREISL, T.N.; FINE, H.A. Bevacizumab-related toxicities in the National Cancer Institute malignant glioma trial cohort. J**ournal of Neuro-Oncology**., v. 120, n°2, p. 431-440, 2014.

OH, J.; TAKAHASHI, R.; KONDO, S.; MIZOGUCHI, A.; ADACHI, E.; SASAHARA, R.M.; NISHIMURA, S.; IMAMURA, Y.; KITAYAMA, H.; ALEXANDER, D.B.; IDE, C.; HORAN, T.P.; ARAKAWA, T.; YOSHIDA, H.; NISHIKAWA, S.; ITOH, Y.; SEIKI, M.; ITOHARA, S.; TAKAHASHI, C.; NODA, M. The membrane-anchored MMP inhibitor

RECK is a key regulator of extracellular matriz integrity and angiogenesis. **Cell,** Cambridge, v. 107, n. 6, p. 789-800, 2001.

OLSEN, J.G.; KADZIOLA, A.; VON WETTSTEIN-KNOWLES, P.; SIGGAARD-ANDERSEN, M.; LARSEN, S. Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. **Structure,** v. 9, n. 3, p. 233-243, 2001.

OLSSON, M.H.M.; SØNDERGARD, C.R.; ROSTKOWSKI, M.; JENSEN, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa predictions. **Journal of Chemical Theory and Computation**, v.7, n°2, p. 525-537, 2001.

PARK, K.; GOPALSAMY, A.; APLASCA, A.; ELLINGBOE, J.W.; XU, W.; ZHANG, Y.; LEVIN, J.I. Synthesis and activity of tryptophan sulfonamide derivatives as novel non-hydroxamate TNF-alpha converting enzyme (TACE) inhibitors. **Bioorganic & Medicinal Chemistry**, Amsterdam**,** v. 17, n. 11, p. 3857-65, 2009.

PASQUALOTO, K.F.; FERREIRA, E.I.; SANTOS-FILHO, O.A.; HOPFINGER, A.J. Rational design of new antituberculosis agents: receptor-independent four-dimensional quantitative structure-activity relationship analysis of a set of isoniazid derivatives. **Journal of Medicinal Chemistry**,v. 47, n° 15, p.3755-3764, 2004.

PEARSON, K.; LEE, A.On the generalized probable error in multiple normal correlation, *Biometrika*, Oxford, v.6, p.59-68, 1908.

PEREIRA, C.P.; CARMO, E.D.; SILVEIRA, V.A.S.; AMADEI, S.U.; ROSA, L.E.B. O papel das MMP-2 e -9 no desenvolvimento do carcinoma epidermóide. **Revista Brasileira de Cancerologia**, Rio de Janeiro, v. 52, n. 3, p. 257-262, 2006.

PETTERSEN, E.F., GODDARD T. D., HUANG, C.C., COUCH, G.S., GREENBLATT, D.M., MENG, E.C., FERRIN, T. E.UCSF Chimera--a visualization system for exploratory research and analysis. **Journal of Computational Chemistry.**, *v.* 25, n.13, p. 1605-12, 2004.

PINTO, A.C.; SILVA, L.F.; CAVALCANTI, B.C.; MELO, M.R.; CHAVES, F.C.; LOTUFO, L. V.; DE MORAES, M.O.; DE ANDRADE-NETO, V.F.; TADEI, W.P.; PESSOA, C.O.; VIEIRA, P.P.; POHLIT, A.M. New antimalarial and cytotoxic 4-nerolidylcatechol derivatives. **European Journal of Medicinal Chemistry**, Amsterdam, v. 44, n. 6, p. 2731-5, 2009.

Pirouette, version 3.11, Woodinville: Infometrix Inc., 1990-2003.

PLAVSIC, D.; NIKOLIC, S.; TRINAJSTIC, N.; MIHALIC, Z. On the Harary Index for the Characterization of Chemical Graphs. **Journal of Mathematical Chemistry**, Berlin, v.12, n.1-4, p.235-250, 1993.

PLATT, J.R. Influence of Neighbor Bonds on Additive Bond Properties in Paraffins. **Journal of Chemical Physics**, Philadelphia, v.15, n.6, p.419-420, 1947.

PLATT, J.R. Prediction of Isomeric Differences in Paraffin Properties. **Journal of Physical Chemistry**, Washington, v.56, n.3, p.328-336, 1952.

PUBCHEM. The PubChem Compound Database. Dispoínvel em: https://pubchem.ncbi.nlm.nih.gov/ Acesso: 21 de Outubro de 2014.

PUTNAM, C.D.; ARVAI, A.S.; BOURNE, Y.; TAINER, J.A. Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism. **Journal of Molecular Biology**, Sydney,v. 296, n. 1, p. 295-309, 2000.

RANDIC, M. Novel Molecular Descriptor for Structure-Property Studies. **Chemical Physics Letters**, Amsterdam, v.211, n.4-5, Aug 20, p.478-483, 1993.

REDONDO, P.; LLORET, P.; IDOATE, M.; INOGES, S. Expression and serum levels of MMP-2 and MMP-9 during human melanoma progression. **Clinical and Experimental Dermatology.**,v. 30, n. 5, p. 541-545, 2005.

RENCHER, A.C.; CHRISTENSEN, W. F. **Methods of Multivariate Analysis (Wiley Series in Probability and Statistics)**. New York: John Wiley & Sons, 2012. p. 503, 508.

RIGEL, D.S.; RUSSAK, J.; FRIEDMAN, R. The evolution of melanoma diagnosis: 25 years beyond the ABCDs. **CA: a cancer journal for clinicians**, New York**,** v. 60, n. 5, p. 301-316, 2010.

ROGERS, D. WOLF Genetic Function Approximation. Reference Manual, version 5.5; Molecular Simulation Inc.: Burlignton, MA, 1994.

ROGERS, D.; HOPFINGER, A.J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. **Journal of Chemical Information and Computer Sciences**, Washington, v. 34, p. 854-866, 1994.

ROGERS, D. G/SPLINES: A Hybrid of Friedman's Multivariate Adaptive Regression Splines (MARS) Algorithm with Holland's Genetic Algorithm, *ICGA*, San Diego, p.384-391, 1991.

ROPKE, C.D.; MEIRELLES, R.R.; DA SILVA, V.V.; SAWADA, T.C.; BARROS, S.B. Pothomorphe umbellata extract prevents alpha-tocopherol depletion after UV-irradiation. **Photochemistry and photobiology**, Hoboken,v. 78, n. 5, p. 436-9, 2003.

ROPKE, C.D.; SILVA,V.V.; KERA, C.Z.; MIRANDA, D.V.; ALMEIDA, R.L.; SAWADA, T. C.H.; BARROS, S.B. M .*In vitro* and *in vivo* inhibition of skin matrix metalloproteinases by Pothomorphe umbellata root extract. **Photochemistry and Photobiology**., v. 82, n. 2, p. 439-442, 2006.

ROWSELL, S.; HAWTIN, P.; MINSHULL, C.A.; JEPSON, H.; BROCKBANK, S.M.; BARRATT, D. G.; SLATER, A.M.; MCPHEAT, W. L.; WATERSON, D.; HENNEY, A.M.; PAUPTIT, R.A. Crystal structure of human mmp9 in complex with a reverse hydroxamate inhibitor. **Journal of Molecular Biology**, Sydney, v. 319, n. 1, p. 173-81, 2002.

RÜCKER, C.; RÜCKER, G.; MERINGER, M. **Journal of Chemical Information and Modeling.** *v.*47, *p.*2345–2357, 2007.

RUSH T.S.; POWERS R. The application of x-ray, NMR, and molecular modeling in the design of MMP inhibitors. **Current Topics in Medicinal Chemistry,** Doylestown, v. 4, p.1311-1327, 2004.

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

SANT'ANNA, C.M.R. Métodos de modelagem molecular para estudo e planejamento de compostos bioativos: Uma introdução. **Revista Virtual de Química**, Rio de Janeiro, v. 1, n.

1, p. 49-57, 2009.

SANTOS-FILHO,O.A.; MISHRA, R.K.; HOPFINGER, A.J. Free energy force field (FEFF) 3D-QSAR analysis of a set of Plasmodium falciparumdihydrofolatereductase inhibitors. **Journal of Computer-Aided Molecular Design**, Berlim, v.15, p.787–810, 2001.

SCHNEIDER, T.; MAWRIN, C.; SCHERLACH, C.; SKALEJ, M.; FIRSCHING, R. Gliomas in adults. **DtschArztebl Int.,** v. 107, n. 45, p. 799-807, 2010.

SHEU, B.C.; LIEN, H.C.; HO, H.N.; LIN, H.H.; CHOW, S.N.; HUANG, S.C.; HSU, S.M. Increased expression and activation of gelatinolytic matriz metalloproteinases is associated with the progression and recurrence of human cervical cancer. **Cancer Research**, Philadelphia, v. 63, n. 19, p. 6537-42, 2003.

SILVA, T.H.A. Modelagem Molecular. In: Da Química Medicinal à Química Combinatória e Modelagem Molecular, um curso prático, 1a ed., Ed. Manole, p. 111-139, 2002.

SØNDERGAARD, C.R.; OLSSON, M.H.M.; ROSTKOWSKI, M.; JENSEN, J.H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. **Journal of Chemical Theory and Computation**, v.7, n.7, p. 2284-2295, 2011.

SOUZA, A.M. de; POPPI, R.J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte l. **Química Nova**, v. 35, n. 1, p. 223-229, 2012.

STILL, W.C.; TEMPCZYK, A.; HAWLEY, R.C.; HENDRICKSON, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. **Journal of the American Chemical Society**, Washington, v.112, n.16, Aug 1, p.6127-6129, 1990.

STOTE, R.H.; KARPLUS. M. Zinc Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation, **Proteins: Structure, Function, and Genetics**., v. 23, p. 12-31, 1995.

TAURO, M.; MCGUIRE, J.; LYNCH, C.C. New approaches to selectively target cancer-associated matrix metalloproteinase activity. **Cancerand Metastasis Reviews.** (artigo não impresso até o momento)

TAVARES, L. C. QSAR: A abordagem de Hansch. **Química Nova**, São Paulo, v. 27, n.4, p. 631-639, 2004.

The PROPKA Web Interface. Disponível em: http://propka.ki.ku.dk/. Acesso: 20 de Outubro de 2014.

TIAN, F.; LEE, S.W. X-linked inhibitor of apoptosis protein as a therapeutic target in metastatic melanoma. **Journal of Investigative Dermatology**, New York, v. 130, p. 2169-2172, 2010.

TOCHOWICZ, A.; MASKOS, K.; HUBER, R.; OLTENFREITER, R.; DIVE, V.; YIOTAKIS, A.; ZANDA, M.; BODE, W.;GOETTIG, P.Crystal Structures of MMP-9 Complexes with Five Inhibitors: Contribution of the Flexible Arg424 Side-chain to Selectivity.**Journal of Molecular Biology**, Amsterdã, v.371, p.989–1006,2007.

TOKARSKI, J.S.;HOPFINGER, A.J. Constructing protein models for ligand-receptor binding thermodynamic simulations: an application to a set of peptidometicrenin inhibitors, **Journal of Chemical Information and Computer Sciences**,Washington,v.37, p.779–791, 1997.

TOKARSKI, J.S.; HOPFINGER, A.J. Prediction of ligand-receptor binding thermodynamics by free energy force field (FEFF) 3D-QSAR analysis: application to a set of peptidometicrenin inhibitors. **Journal of Chemical Information and Computer Sciences**, Washington, v. 37, n. 4, p. 792-811, 1997.

VALADARES, M.C.; REZENDE, K.R.; PEREIRA, E.R.; SOUSA, M.C.; GONCALVES, B.; DE ASSIS, J.C.; KATO, M.J. Protective effects of 4-nerolidylcatechol against genotoxicity induced by cyclophosphamide. **Food and Chemical Toxicology**, Amsterdam, v. 45, n. 10, p. 1975-1978, 2007.

VAN DERSPOEL, D.; LINDAHL, E.; HESS, B.; GROENHOF, G.; MARK, A.E.; BERENDSEN, H.J. GROMACS: fast, flexible, and free. **Journal of Computational Chemistry**, New York, v. 26, n. 16, p. 1701-18, 2005.

VERMA, R.P.; HANSCH, C. Matriz metalloproteinases (MMPs): chemical-biological functions and (Q) SARs. **Bioorganic & Medicinal Chemistry**, Amsterdam, v. 15, n. 6, p. 2223-68, 2007.

VISWANADHAN, V.N.; GHOSE, A.K.; REVANKAR, G.R.; ROBINS, R.K. Atomic Physicochemical Parameters for three Dimensional Structure Directed Quantitative Structure - Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally-Occurring Nucleoside Antibiotics. **Journal of Chemical Information and Computer Sciences**, Washington, v.29, n.3, p.163-172, 1989.

WANG, M., WANG, T., LIU, S., YOSHIDA, D., TERAMOTO, A.The expression of matrix metalloproteinase-2 and -9 in human gliomas of different pathological grades. **Brain Tumor Patholology**., v20, n°2, p.65-72, 2003.

WANG, X.; ZHAO, D.; MAO, H.; WANG, X. Hybrid algorithm in phase diversity wavefront sensing. **Proceedings of SPIE**, Bellingham, v. 6624, 2009.

WEINER,S.J.; KOLLMAN, P.A.; CASE, D.A.; SINGH, U.C.; GHIO, C.; ALAGONA, G.; PROFETA, S.; WEINERL, P.A. New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins.**Journal of the American Chemical Society**, Washington, v.106, p.765-784, 1984.

WEINER, S.J.; KOLLMAN, P.A.; NGUYEN, D.T. An all atom force field for simulations of proteins and nucleic acids. **Journal of Computational Chemistry,** New York**,** v. 5, p. 230-252, 1986.

WERMUTH, C.G. **The practice of medicinal chemistry**. 2°Edição. London: Academic Press, 2008.

WERMUTH, C.G.; GANELLIN, C.R.; LINDBERG, P.; MITSCHER, L.A. Glossary of terms used in Medicinal Chemistry. **Annual Reports in Medicinal Chemistry**, Bloomfield, v. 33, p. 385-395, 1998.

WIENER, H. Structural Determination of Paraffin Boiling Points. **Journal of the American Chemical Society**, Washington, v.69, n.1, p.17-20, 1947.

WOLD, S.; ERIKSSON, L. Statistical Validation of QSAR Results. In: VAN DE WATERBEEMD, H. (Ed.) Chemometric Methods in Molecular Design, Weinheim, p. 309-318, 1995.

XI, L.; DU, J.; LI, S.; LI, J.; LIU, H.; YAO, X.A combined molecular modeling study on gelatinases and their potent inhibitors. **Journal of Computational Chemistry**, New York, v. 31, n. 1, p. 24-42, 2010.

a YANG, S.M.; SCANNEVIN, R.H.; WANG, B.; BURKE, S.L.; HUANG, Z.; KARNACHI, P.; WILSON, L.J.; RHODES, K.J.; LAGU, B.; MURRAY, W.V. Beta-N-biaryl ether sulfonamide hydroxamates as potent gelatinase inhibitors: part 2. Optimization of alpha-amino substituents. **Bioorganic & medicinal chemistry letters**, Amsterdam, v. 18, n. 3, p. 1140-5, 2008.

b YANG, S.M.; SCANNEVIN, R.H.; WANG, B.; BURKE, S.L.; WILSON, L.J.; KARNACHI, P.; RHODES, K.J.; LAGU, B.; MURRAY, W.V. Beta-N-biaryl ether sulfonamide hydroxamates as potent gelatinase inhibitors: part 1. Design, synthesis, and lead identification. **Bioorganic & medicinal chemistry letters**, Amsterdam, v. 18, n. 3, p. 1135-9, 2008.

ANEXO A

Tabela 1A. Descritores calculados para o conjunto de dados β-N-biaril eter sulfonamida hidroxamatos

Composto	**E_{est}** (kcal/mol)*	**E_{ang}** (kcal/mol)*	**E_{tors}** (kcal/mol)*	**$E_{1,4}$** (kcal/mol)*	**E_{el}** (kcal/mol)*	**E_{vdW}** (kcal/mol)*	**E_{LH}** (kcal/mol)*
8	8,77	17,21	-5,20	-46,07	7,02	-6,68	-32,17
$5a_1$	8,94	14,38	-2,67	-7,09	-14,57	-7,00	-50,21
$5a_2$	10,31	13,06	-4,15	-7,77	-16,35	-6,61	-56,66
5b	11,82	16,83	-6,09	41,40	-29,96	-6,27	-51,69
5c	12,03	18,14	-3,37	6,75	-3,56	-8,35	-51,98
5d	14,27	17,52	5,31	89,67	-41,64	-8,24	-60,05
5e	11,41	17,70	-5,40	35,75	-20,34	-8,38	-56,32
5f	14,23	11,27	-4,18	-10,01	5,94	-7,01	-57,10
5g	13,52	16,91	-6,36	8,78	-6,98	-10,80	-58,54
5h	10,97	18,11	-2,40	-23,94	10,55	-7,97	-59,70
5j	12,82	12,15	-1,07	-8,22	-12,07	-7,26	-58,39
5k	8,460	16,97	-3,30	-5,50	-19,21	-6,67	-57,19
5l	12,00	12,88	-1,94	-2,69	-12,3	-8,18	-59,35
5m	11,61	14,30	1,86	-7,57	-17,28	-7,30	-58,07
5n	10,94	12,36	-2,00	-7,21	-8,44	-7,23	-51,97
5o	9,25	13,07	-1,00	-5,65	-10,04	-7,56	-52,40
13a	7,12	10,47	-10,10	-9,26	-7,81	-6,53	-48,70
13b	8,54	9,200	-10,49	-7,58	-11,39	-7,08	-56,74
13c	12,83	12,82	-10,28	-4,25	-13,65	-4,95	-54,82
14a	12,24	12,13	-7,10	-3,19	-11,38	-5,61	-57,86
14b	10,09	14,36	-3,62	-2,39	-12,69	-6,85	-57,22
14c	10,10	15,15	-1,73	11,53	-15,2	-5,23	-42,88
$7a_2$	13,12	12,19	-5,24	-11,91	-4,19	-5,95	-25,30
$7a_1$	11,05	16,59	-2,50	-11,17	-1,98	-7,27	-26,07
8a	13,37	20,84	-0,88	-13,06	-7,73	-10,20	-52,8
$8b_1$	14,46	19,80	-4,08	-16,61	-7,8	-7,72	-52,66
$8b_2$	14,27	18,76	-4,75	-17,47	0,46	-7,71	-41,87
8c	45,65	19,25	-1,46	-29,41	2,25	-7,64	-35,16
8d	15,31	22,22	1,67	-7,31	-13,2	-8,88	-50,24
8e	13,90	21,97	-6,94	-4,55	-16,17	-10,64	-51,84
8f	18,83	20,61	-6,86	-5,01	-12,52	-10,52	-50,05
8g	17,07	18,78	-8,30	-4,23	-17,71	-8,95	-52,19
8h	14,48	22,48	-10,70	-10,77	-7,61	-10,33	-48,92
8ir	20,14	25,26	-4,82	-11,28	-11,65	-11,24	-46,42
8is	20,04	17,83	-8,69	-3,37	-19,38	-10,72	-52,33
8j	11,50	17,63	-6,86	-4,91	-9,000	-11,04	-50,25
8k	12,18	19,51	-4,57	-1,79	-18,89	-9,95	-53,85
8l	10,92	16,84	-7,76	-11,40	-5,21	-8,82	-49,97
8m	13,75	16,95	-4,82	-7,53	-11,67	-9,25	-49,38
8n	16,41	18,89	-2,04	-10,75	-11,14	-11,53	-55,44
8o	16,39	22,58	-13,52	12,98	-20,15	-10,18	-62,83
8p	15,86	26,28	-0,68	-13,83	-5,13	-4,62	-32,64
8q	21,12	30,30	-3,45	-8,41	-10,61	-5,27	-32,89
8r	58,55	45,28	2,31	-2,59	1,73	-3,43	-52,49
8s	16,89	28,78	-0,78	-20,61	-1,69	-8,54	-22,31
8t	17,61	24,55	-1,71	-21,34	6,66	-5,11	-22,21
8u	22,74	30,89	-0,46	-17,46	-77,67	-9,66	-22,91
8v	14,49	26,44	9,25	-17,87	2,07	-1,80	-24,54
8x	12,27	23,60	0,79	-14,89	5,14	-4,00	-19,76
8y	15,77	24,10	3,12	-19,95	3,12	-9,36	-47,45
8z	15,20	33,11	5,35	-3,95	-9,62	-5,65	-27,64
11a	12,63	16,07	-5,72	-13,45	-6,99	-6,23	-45,98
12a	14,28	20,72	5,77	-0,22	-23,43	-7,23	-48,32
8aa	22,58	27,32	-5,04	-10,44	-7,42	-6,68	-19,50
8bb	22,60	31,20	-0,86	-15,00	2,45	-8,00	-24,94
8cc	19,61	21,46	-0,82	-23,05	-0,74	-4,96	-22,08
8dd	12,69	25,33	-1,52	-23,73	8,08	-7,50	-22,10
$8ee_1$	13,55	23,96	-3,06	-21,18	0,63	-4,47	-22,17
$8ee_2$	15,40	20,98	-1,31	-22,54	2,73	-5,01	-28,88
8ff	20,75	29,22	1,04	-25,06	14,71	-6,75	-22,20
9a	9,57	10,06	-0,13	23,24	-10,65	-5,97	-25,60
9b	11,74	13,34	1,59	38,24	-16,73	-6,29	-26,50
3a	36,30	9,97	-0,79	22,28	-36,95	-5,69	-33,66
3b	37,79	12,41	-2,24	22,10	-40,25	-5,99	-33,23

* Para maiores informações sobre metodologia utilizada nos cálculos, verificar o item **3.4**.

Tabela 1A (cont.). Descritores calculados para o conjunto de dados β-N-biaril eter sulfonamida hidroxamatos

Composto	E_{solv} (kcal/mol)*	E_T (kcal/mol)*	$ChelpG_1$ (eV)*	$ChelpG_2$ (eV)*	$ChelpG_3$ (eV)*	$ChelpG_4$ (eV)*	$ChelpG_5$ (eV)*
8	-10,32	-74,47	0,48	-0,48	-0,28	0,40	0,60
$5a_1$	-10,43	-54,08	0,49	-0,50	-0,24	0,30	0,56
$5a_2$	-12,01	-63,83	0,48	-0,49	-0,21	0,31	0,50
5b	-10,01	-3,97	0,50	-0,49	-0,27	0,33	0,60
5c	-8,63	-35,42	0,49	-0,51	-0,17	0,29	0,47
5d	-6,50	51,97	0,50	-0,51	-0,14	0,27	0,44
5e	-7,41	-12,70	0,47	-0,46	-0,25	0,34	0,47
5f	-5,95	-58,76	0,50	-0,51	-0,16	0,29	0,43
5g	-7,63	-44,13	0,50	-0,50	-0,13	0,28	0,38
5h	-7,39	-72,33	0,49	-0,50	-0,17	0,29	0,46
5j	-10,23	-60,23	0,50	-0,52	-0,11	0,27	0,40
5k	-11,38	-58,62	0,50	-0,51	-0,17	0,31	0,48
5l	-12,02	-59,31	0,50	-0,51	-0,10	0,25	0,38
5m	-12,43	-57,61	0,48	-0,50	-0,17	0,31	0,46
5n	-9,52	-54,69	0,49	-0,50	-0,24	0,32	0,55
5o	-10,80	-55,15	0,49	-0,49	-0,20	0,30	0,49
13a	-10,34	-68,06	0,49	-0,50	-0,25	0,32	0,58
13b	-12,03	-76,18	0,44	-0,44	-0,21	0,31	0,46
13c	-12,02	-60,67	0,47	-0,46	-0,30	0,40	0,54
14a	-11,64	-61,03	0,47	-0,45	-0,31	0,41	0,53
14b	-12,29	-57,92	0,48	-0,46	-0,273	0,40	0,47
14c	-14,92	-27,98	0,49	-0,53	-0,16	0,31	0,59
$7a_2$	-10,60	-33,70	0,48	-0,54	-0,05	0,27	0,44
$7a_1$	-8,69	-28,06	0,49	-0,49	-0,28	0,30	0,54
8a	-9,05	-51,85	0,46	-0,45	-0,32	0,40	0,55
$8b_1$	-6,94	-53,68	0,46	-0,45	-0,33	0,40	0,53
$8b_2$	-8,76	-47,54	0,49	-0,52	-0,13	0,33	0,41
8c	-6,54	-15,31	0,45	-0,47	-0,49	0,42	0,83
8d	-7,14	-34,38	0,48	-0,47	-0,32	0,39	0,50
8e	-9,30	-47,40	0,48	-0,50	-0,19	0,36	0,46
8f	-9,14	-42,15	0,48	-0,50	-0,20	0,36	0,40
8g	-8,85	-46,68	0,48	-0,49	-0,25	0,38	0,50
8h	-7,90	-51,66	0,49	-0,51	-0,18	0,36	0,36
8ir	-5,91	-34,27	0,46	-0,48	-0,25	0,35	0,47
8is	-7,63	-44,87	0,48	-0,48	-0,23	0,38	0,40
8j	-4,97	-48,90	0,42	-0,47	-0,18	0,32	0,38
8k	-6,52	-44,99	0,47	-0,50	-0,14	0,31	0,39
8l	-6,09	-56,28	0,46	-0,48	-0,22	0,34	0,41
8m	-6,06	-46,35	0,49	-0,50	-0,19	0,32	0,43
8n	-6,41	-50,88	0,45	-0,49	-0,14	0,32	0,38
8o	-5,05	-39,63	0,49	-0,55	-0,18	0,34	0,32
8p	-5,76	-15,40	0,49	-0,55	-0,09	0,25	0,34
8q	-4,91	-3,51	0,52	-0,63	-0,04	0,22	0,47
8r	-9,41	38,20	0,46	-0,50	-0,38	0,43	0,69
8s	-7,58	-14,15	0,50	-0,53	-0,23	0,31	0,61
8t	-4,70	-12,91	0,51	-0,54	-0,23	0,30	0,60
8u	-12,71	-9,57	0,51	-0,53	-0,24	0,31	0,68
8v	-7,89	-1,92	0,51	-0,54	-0,22	0,31	0,58
8x	-6,15	-8,14	0,51	-0,54	-0,23	0,32	0,58
8y	-5,32	-39,09	0,50	-0,53	-0,21	0,32	0,54
8z	-7,45	8,96	0,50	-0,51	-0,22	0,31	0,59
11a	-9,64	-52,33	0,48	-0,49	-0,23	0,32	0,53
12a	-7,79	-22,79	0,48	-0,51	-0,17	0,30	0,46
8aa	-5,95	2,27	0,48	-0,50	-0,22	0,31	0,54
8bb	-8,71	-3,62	0,48	-0,49	-0,23	0,32	0,50
8cc	-6,92	-16,76	0,50	-0,54	-0,23	0,31	0,63
8dd	-4,03	-20,87	0,51	-0,53	-0,20	0,31	0,50
$8ee_1$	-4,71	-18,08	0,50	-0,52	-0,28	0,32	0,70
$8ee_2$	-8,38	-29,74	0,47	-0,52	-0,14	0,28	0,58
8ff	-2,72	-5,73	0,50	-0,54	-0,21	0,31	0,59
9a	-6,35	4,80	0,49	-0,51	-0,19	0,31	0,46
9b	-4,61	27,50	0,49	-0,51	-0,19	0,31	0,51
3a	-11,97	16,43	0,49	-0,52	-0,39	0,32	0,85
3b	-11,32	19,51	0,47	-0,51	-0,36	0,27	0,83

* Para maiores informações sobre metodologia utilizada nos cálculos, verificar o item **3.4**.

Tabela 1A (cont.). Descritores calculados para o conjunto de dados β-N-biaril eter sulfonamida hidroxamatos

Composto	**ChelpG$_6$**(eV)*	**ChelpG$_7$**(eV)*	**ChelpG$_8$**(eV)*	**ChelpG$_9$** (eV)*	**ChelpG$_{10}$** (eV)*	**ChelpG$_{11}$** (eV)*	**ChelpG$_{12}$** (eV)*
8	-0,69	0,25	0,45	-0,60	0,21	0,93	-0,56
5a$_1$	-0,64	0,38	0,39	-0,52	0,35	0,89	-0,56
5a$_2$	-0,64	0,02	0,19	-0,45	0,20	0,94	-0,55
5b	-0,64	0,29	0,26	-0,42	0,35	0,83	-0,54
5c	-0,63	0,30	0,43	-0,50	0,12	0,91	-0,58
5d	-0,63	0,14	0,26	-0,34	0,19	0,78	-0,54
5e	-0,62	0,34	0,18	-0,23	-0,01	0,62	-0,46
5f	-0,63	0,16	0,20	-0,33	0,02	1,02	-0,59
5g	-0,62	0,07	0,14	-0,17	-0,20	0,75	-0,53
5h	-0,64	0,18	0,19	-0,31	0,12	1,00	-0,58
5j	-0,62	0,03	0,07	-0,25	0,11	0,97	-0,57
5k	-0,63	0,48	0,19	-0,29	0,01	0,91	-0,59
5l	-0,62	0,08	0,13	-0,20	0,02	0,91	-0,55
5m	-0,64	0,16	0,21	-0,21	-0,06	0,87	-0,57
5n	-0,64	0,25	0,48	-0,50	0,21	0,89	-0,57
5o	-0,64	0,14	0,28	-0,38	0,17	0,92	-0,55
13a	-0,64	0,17	0,42	-0,48	0,07	0,89	-0,54
13b	-0,69	0,00	0,47	-0,74	0,66	1,26	-0,66
13c	-0,67	0,18	0,50	-0,60	0,08	0,95	-0,57
14a	-0,67	0,27	0,20	-0,47	0,24	0,93	-0,57
14b	-0,67	0,31	0,21	-0,45	0,09	0,93	-0,56
14c	-0,65	0,24	0,18	-0,22	0,05	0,88	-0,54
7a$_2$	-0,64	0,18	0,26	-0,58	0,21	1,01	-0,58
7a$_1$	-0,60	0,32	0,30	-0,51	0,12	1,08	-0,59
8a	-0,67	0,47	0,52	-0,62	0,10	0,95	-0,57
8b$_1$	-0,65	0,23	0,45	-0,64	0,24	0,95	-0,57
8b$_2$	-0,65	0,51	0,18	-0,37	0,15	1,00	-0,58
8c	-0,6	0,19	0,47	-0,66	0,43	0,94	-0,56
8d	-0,64	0,57	0,35	-0,59	0,11	1,12	-0,59
8e	-0,68	0,60	0,21	-0,45	0,01	0,97	-0,58
8f	-0,68	0,52	0,22	-0,37	-0,19	0,97	-0,57
8g	-0,66	0,63	0,25	-0,36	-0,06	0,94	-0,58
8h	-0,67	0,56	0,20	-0,37	-0,07	1,04	-0,61
8ir	-0,69	0,58	0,09	-0,30	0,03	1,01	-0,58
8is	-0,6	0,56	0,27	-0,49	0,01	1,03	-0,59
8j	-0,65	0,64	0,39	-0,34	-0,15	0,90	-0,57
8k	-0,65	0,87	0,23	-0,27	-0,01	0,90	-0,57
8l	-0,66	0,51	0,30	-0,33	-0,10	0,91	-0,56
8m	-0,66	0,18	0,14	-0,29	0,04	0,88	-0,56
8n	-0,66	0,06	0,23	-0,34	-0,26	1,01	-0,59
8o	-0,64	-0,09	0,06	-0,23	-0,23	0,92	-0,57
8p	-0,64	0,03	0,27	-0,33	0,08	0,82	-0,56
8q	-0,69	-0,20	0,16	-0,23	0,08	0,89	-0,58
8r	-0,62	-0,07	0,21	-0,38	0,27	0,97	-0,56
8s	-0,61	0,07	0,18	-0,34	0,14	0,97	-0,61
8t	-0,61	-0,19	0,20	-0,36	0,17	1,01	-0,62
8u	-0,64	0,08	0,31	-0,47	0,25	1,01	-0,62
8v	-0,61	0,15	0,18	-0,32	0,17	0,96	-0,61
8x	-0,60	0,21	0,14	-0,25	0,08	0,97	-0,60
8y	-0,59	0,10	0,15	-0,41	0,38	1,05	-0,62
8z	-0,65	0,29	0,21	-0,26	-0,01	1,03	-0,62
11a	-0,65	-0,12	0,16	-0,52	0,25	1,13	-0,60
12a	-0,64	0,26	0,27	-0,52	0,31	0,99	-0,59
8aa	-0,63	-0,16	0,20	-0,28	-0,05	0,94	-0,60
8bb	-0,61	-0,13	0,17	-0,27	-0,04	0,98	-0,59
8cc	-0,61	-0,06	0,15	-0,32	0,17	0,96	-0,60
8dd	-0,60	0,32	0,12	-0,25	0,09	0,96	-0,62
8ee$_1$	-0,62	0,26	0,20	-0,41	0,24	1,04	-0,62
8ee$_2$	-0,65	-0,31	0,22	-0,53	0,28	1,04	-0,61
8ff	-0,61	-0,25	0,24	-0,42	0,27	1,06	-0,63
9a	-0,63	0,34	0,13	-0,42	0,11	0,77	0,00
9b	-0,64	0,31	0,22	-0,50	0,14	0,67	0,00
3a	-0,60	----	-0,31	-0,24	-0,31	0,97	-0,59
3b	-0,61	----	-0,25	-0,24	-0,25	0,92	-0,55

* Para maiores informações sobre metodologia utilizada nos cálculos, verificar o ite **3.4**.

Tabela 1A (cont.). Descritores calculados para o conjunto de dados β-N-biaril eter sulfonamida hidroxamatos

Composto	**ChelpG$_{13}$** (eV)*	**ChelpG$_{14}$** (kcal/mol)*	**E$_{homo}$**(kcal/mol)*	**E$_{lumo}$** (kcal/mol)*	**μ$_x$** (Debye)*	**μ$_y$** (Debye)*	**μ$_z$**(Debye)*
8	-0,57	-0,21	-7,41	2,36	-3,01	0,21	1,92
5a$_1$	-0,56	-0,25	-7,19	2,55	-0,10	2,42	3,37
5a$_2$	-0,57	-0,25	-7,46	2,09	-8,75	4,44	-3,10
5b	-0,57	-0,01	-7,07	2,61	-1,43	5,99	4,60
5c	-0,55	-0,23	-7,15	2,51	-0,57	-0,33	2,14
5d	-0,54	0,11	-6,92	2,65	5,44	-1,87	6,20
5e	-0,51	0,09	-7,34	1,59	6,30	2,79	5,70
5f	-0,61	-0,28	-7,22	2,37	0,08	-1,10	1,90
5g	-0,56	-0,04	-7,47	2,25	3,36	0,63	2,96
5h	-0,56	-0,48	-7,14	2,06	2,45	4,29	-7,23
5j	-0,58	-0,37	-7,46	2,55	0,09	-4,69	8,68
5k	-0,57	-0,23	-7,60	2,32	-0,44	2,22	-0,86
5l	-0,56	-0,38	-7,70	2,20	3,25	-3,94	8,21
5m	-0,55	-0,36	-7,85	1,92	-2,20	-2,07	2,39
5n	-0,56	-0,23	-7,08	2,41	3,16	-0,49	4,69
5o	-0,57	-0,33	-7,24	2,19	5,78	-2,34	4,55
13a	-0,57	-0,27	-6,87	1,96	1,39	-0,95	4,56
13b	-0,63	-0,18	-6,73	1,42	1,49	-1,60	5,24
13c	-0,57	-0,26	-6,68	1,76	4,64	1,39	0,73
14a	-0,56	-0,25	-7,33	2,57	1,78	0,02	3,52
14b	-0,55	-0,26	-7,44	2,47	1,80	0,21	6,77
14c	-0,56	-0,34	-7,93	2,00	4,91	1,46	-7,09
7a$_2$	-0,59	-0,28	-7,11	2,61	3,08	5,51	-0,54
7a$_1$	-0,59	-0,49	-6,84	2,71	4,94	9,96	0,58
8a	-0,57	-0,26	-7,81	2,09	1,16	-2,40	1,56
8b$_1$	-0,56	-0,31	-7,27	2,44	3,39	0,98	1,95
8b$_2$	-0,57	-0,36	-7,35	2,52	-2,34	3,61	7,80
8c	-0,57	-0,27	-7,26	2,39	4,67	-0,25	1,87
8d	-0,62	-0,50	-7,08	2,57	3,13	2,84	4,79
8e	-0,59	-0,24	-7,31	2,39	-4,51	-0,94	0,29
8f	-0,59	-0,33	-6,99	2,46	1,48	0,38	0,89
8g	-0,58	-0,24	-7,22	2,54	-0,98	-0,12	1,23
8h	-0,58	-0,38	-7,01	2,60	5,56	-1,68	-4,02
8ir	-0,59	-0,42	-6,96	2,67	1,18	-3,37	7,51
8is	-0,58	-0,33	-7,06	2,58	-1,50	7,89	-1,82
8j	-0,57	-0,28	-7,16	2,44	-1,78	-2,35	-0,46
8k	-0,57	-0,26	-7,21	2,24	4,64	-0,39	-0,01
8l	-0,57	-0,32	-7,04	2,47	3,58	2,32	-0,01
8m	-0,55	-0,24	-7,11	2,50	2,46	2,91	6,53
8n	-0,59	-0,29	-6,94	2,37	3,31	-4,13	2,48
8o	-0,57	-0,27	-5,74	2,20	7,36	2,43	-4,80
8p	-0,54	-0,18	-7,69	1,76	3,23	-2,20	-1,52
8q	-0,55	-0,30	-6,79	1,90	-2,92	2,83	-0,67
8r	-0,58	-0,37	-7,78	2,24	-6,97	-3,75	-6,65
8s	-0,57	-0,27	-7,87	2,12	1,17	8,81	5,60
8t	-0,59	-0,34	-7,75	2,09	6,78	1,98	3,36
8u	-0,59	-0,23	-7,93	2,11	5,31	7,61	9,66
8v	-0,60	-0,23	-7,68	2,11	0,98	9,20	3,32
8x	-0,57	-0,40	-7,65	2,34	0,32	9,41	-2,54
8y	-0,63	-0,26	-7,73	2,10	0,85	9,78	6,21
8z	-0,59	-0,43	-7,52	2,40	2,22	10,63	-4,52
11a	-0,63	-0,40	-6,91	2,92	4,02	4,67	4,28
12a	-0,58	-0,27	-7,20	2,71	-0,31	-6,00	-5,12
8aa	-0,57	-0,25	-7,07	2,24	-1,62	6,67	3,92
8bb	-0,57	-0,35	-7,50	2,36	3,85	4,67	-7,87
8cc	-0,59	-0,25	-7,61	2,34	2,63	7,95	4,76
8dd	-0,57	-0,34	-7,39	2,66	4,17	9,67	-4,02
8ee$_1$	-0,61	-0,28	-7,27	2,58	-4,49	6,42	-3,80
8ee$_2$	-0,59	-0,27	-7,50	2,28	8,91	-5,19	-3,90
8ff	-0,63	-0,32	-7,37	2,59	7,63	6,51	4,48
9a	-0,64	-0,28	-7,37	2,60	-1,52	3,27	4,31
9b	-0,60	0,25	-7,39	2,58	1,81	3,06	-4,34
3a	-0,56	-0,35	-7,71	2,05	9,12	-4,02	1,24
3b	-0,57	-0,28	-7,61	2,12	11,26	-0,42	-4,86

* Para maiores informações sobre metodologia utilizada nos cálculos, verificar o item **3.4**.

Tabela 1A (cont.). Descritores calculados para o conjunto de dados β-N-biaril eter sulfonamida hidroxamatos

Composto	**μ_T** (Debye)*	**V_{vdW}**(Å^2)*	**V** (Å^3)*	**ClogP** *	**MR** (Å^3)*	**α**(Å^3)*	**PSA** (Å^2)*
8	3,581	382,00	315,25	2,28	92,50	34,60	104,32
5a$_1$	4,156	390,06	321,75	1,85	93,84	35,23	124,55
5a$_2$	10,296	392,50	321,97	1,85	93,84	35,23	124,55
5b	7,697	412,99	339,20	2,19	98,59	37,07	124,55
5c	2,251	429,50	355,74	2,66	103,11	38,90	124,55
5d	8,467	432,88	356,32	2,60	103,01	38,90	124,55
5e	8,943	446,75	378,56	3,53	113,62	43,06	124,55
5f	2,206	470,56	395,65	3,63	118,45	44,89	124,55
5g	4,527	457,08	389,44	1,51	119,10	44,18	137,44
5h	8,763	464,91	389,53	1,39	119,02	44,18	137,44
5j	9,873	371,18	305,51	1,38	88,80	33,40	124,55
5k	2,435	383,77	319,33	1,90	93,61	35,32	124,55
5l	9,676	402,53	329,59	2,26	94,77	34,96	124,55
5m	3,860	395,91	324,87	2,12	92,90	34,25	137,44
5n	5,679	402,84	333,07	2,19	100,28	37,60	140,62
5o	7,730	388,30	319,00	1,86	95,46	35,67	140,62
13a	4,864	405,04	336,63	2,41	101,30	37,16	115,32
13b	5,683	405,15	335,47	2,46	101,06	37,25	115,32
13c	4,906	413,85	345,00	1,69	102,72	37,80	124,55
14a	3,951	394,18	324,56	1,62	93,85	35,14	124,55
14b	7,015	407,88	337,00	2,00	98,44	37,16	124,55
14c	8,751	427,51	347,30	2,36	99,61	36,79	124,55
7a$_2$	6,339	414,89	340,40	2,13	98,59	37,07	113,55
7a$_1$	11,141	415,37	340,58	2,13	98,59	37,07	113,55
8a	3,099	467,18	385,53	3,08	110,37	41,18	116,35
8b$_1$	4,038	455,61	377,95	2,67	109,44	41,45	116,35
8b$_2$	8,913	458,21	378,42	2,67	109,44	41,45	116,35
8c	5,041	460,78	382,22	3,96	110,83	41,94	104,32
8d	6,395	479,66	401,35	3,10	116,69	44,35	116,35
8e	4,618	488,28	411,66	1,45	125,46	46,73	129,24
8f	1,772	487,42	411,63	1,57	125,53	46,73	129,24
8g	1,585	487,25	411,56	2,70	125,58	46,73	129,24
8h	7,071	495,78	418,24	3,69	124,89	47,44	116,35
8ir	8,316	519,05	435,63	4,10	129,30	49,28	116,35
8is	8,242	517,40	435,10	4,10	129,30	49,28	116,35
8j	2,997	478,61	401,80	3,74	121,83	45,61	116,35
8k	4,657	479,97	404,15	3,87	122,05	45,52	116,35
8l	4,270	481,61	404,43	3,87	122,05	45,52	116,35
8m	7,564	495,23	416,41	4,25	126,64	47,53	116,35
8n	5,853	507,47	426,44	3,48	128,30	48,08	125,58
8o	9,120	517,50	442,47	2,48	143,06	51,79	116,35
8p	4,197	485,59	403,38	3,44	115,67	43,01	107,56
8q	4,123	560,06	476,74	5,22	140,28	52,67	107,56
8r	10,345	659,19	547,69	4,88	159,30	60,25	145,89
8s	10,509	485,63	434,50	2,34	115,58	42,88	116,79
8t	7,826	525,49	400,61	3,17	124,41	46,55	116,79
8u	13,402	547,63	453,54	1,71	130,79	46,56	153,32
8v	9,833	529,23	439,76	1,79	126,82	47,35	127,87
8x	9,753	493,74	409,34	3,41	118,64	44,07	107,56
8y	11,623	496,09	415,03	2,24	120,22	44,71	127,79
8z	11,774	550,37	455,64	2,12	131,60	49,18	136,66
11a	7,511	404,70	327,26	1,50	95,50	35,95	130,34
12a	7,903	486,75	396,80	2,52	114,14	43,37	133,42
8aa	7,910	572,84	483,56	4,96	143,29	53,73	107,56
8bb	9,937	586,96	494,41	5,16	147,85	55,18	110,80
8cc	9,635	469,02	390,55	1,98	114,41	43,24	116,79
8dd	11,278	513,73	424,44	2,81	123,24	46,91	116,79
8ee$_1$	8,716	470,70	392,50	1,93	114,64	43,15	116,79
8ee$_2$	11,031	471,94	392,26	1,93	114,64	43,15	116,79
8ff	10,995	512,87	426,69	2,76	123,48	46,82	116,79
9a	5,626	370,41	306,56	1,98	89,63	36,11	99,10
9b	5,615	410,39	340,67	3,17	98,83	39,78	99,10
3a	10,048	342,82	280,88	1,63	82,76	30,92	104,32
3b	12,278	363,06	297,81	2,10	87,80	32,76	104,32

* Para maiores informações sobre metodologia utilizada nos cálculos, verificar o item **3.4**.

Tabela 1A (cont.). Descritores calculados para o conjunto de dados β-N-biaril eter sulfonamida hidroxamatos

Composto	SAM (Å^2)*	ClogD_1.5*	ClogD_5.0*	ClogD_6.5*	ClogD_7.4*	Platt*	Randic*
8	526,40	1,49	1,49	1,49	1,47	154,00	20,03
5a$_1$	533,44	0,77	0,77	0,77	0,75	156,00	20,59
5a$_2$	534,00	0,77	0,77	0,77	0,75	156,00	20,59
5b	566,90	1,28	1,28	1,28	1,26	168,00	21,84
5c	595,30	1,81	1,80	1,80	1,78	180,00	23,09
5d	593,14	1,85	1,85	1,85	1,83	180,00	23,09
5e	607,29	2,89	2,89	2,88	2,86	180,00	24,01
5f	637,28	2,65	2,65	2,65	2,62	192,00	25,26
5g	627,84	0,45	1,25	1,42	1,41	188,00	24,84
5h	627,28	0,45	1,17	1,42	1,40	188,00	24,84
5j	499,91	0,26	0,26	0,26	0,23	144,00	19,38
5k	515,74	0,86	0,86	0,86	0,84	144,00	19,38
5l	549,24	1,14	1,14	1,13	1,11	156,00	20,59
5m	541,52	0,44	0,51	0,51	0,49	152,00	20,16
5n	542,62	1,46	1,46	1,46	1,44	156,00	20,59
5o	516,93	1,09	1,09	1,09	1,07	144,00	19,38
13a	537,23	1,39	1,39	1,39	1,37	158,00	21,09
13b	521,73	1,48	1,48	1,48	1,46	146,00	19,88
13c	552,60	0,72	0,72	0,72	0,70	160,00	21,56
14a	537,22	0,47	0,47	0,47	0,44	156,00	20,61
14b	547,76	0,93	0,93	0,93	0,90	156,00	20,61
14c	583,58	1,20	1,20	1,20	1,18	168,00	21,82
7a$_2$	572,66	1,42	1,42	1,41	1,39	168,00	21,73
7a$_1$	572,11	1,42	1,42	1,41	1,39	168,00	21,73
8a	651,97	-1,00	0,26	1,62	2,07	196,00	24,68
8b$_1$	633,35	-1,37	-0,11	1,25	1,70	196,00	24,68
8b$_2$	637,15	-1,37	-0,11	1,25	1,70	196,00	24,68
8c	643,19	3,21	0,26	3,21	3,19	202,00	25,03
8d	667,17	-0,79	3,21	1,64	2,20	220,00	26,43
8e	671,28	-2,48	0,20	1,35	1,54	204,00	26,43
8f	670,84	-2,52	0,28	1,39	1,55	204,00	26,43
8g	670,63	-1,83	1,03	1,64	1,66	204,00	26,43
8h	681,11	-0,41	1,16	2,43	2,73	208,00	26,86
8ir	711,19	0,00	1,51	2,80	3,13	220,00	28,11
8is	710,20	0,00	1,51	2,80	3,13	220,00	28,11
8j	652,05	2,33	2,74	2,74	2,72	196,00	25,61
8k	658,28	2,42	2,88	2,88	2,86	196,00	25,61
8l	659,35	2,78	2,88	2,88	2,86	196,00	25,61
8m	669,12	3,17	3,34	3,34	3,32	196,00	25,61
8n	699,23	2,38	2,58	2,58	2,56	210,00	27,30
8o	704,45	3,46	3,73	3,72	3,71	220,00	28,43
8p	686,63	-0,82	1,72	2,53	2,59	208,00	25,89
8q	788,64	0,89	3,31	4,22	4,31	244,00	30,57
8r	914,43	0,70	3,57	3,96	3,96	282,00	35,36
8s	675,67	-0,47	1,61	1,62	1,61	210,00	25,60
8t	734,44	-0,20	2,39	2,46	2,44	234,00	28,10
8u	776,21	-1,50	0,40	0,41	0,40	238,00	28,51
8v	739,25	-1,41	0,89	0,92	0,90	232,00	28,17
8x	686,29	-0,73	1,88	2,62	2,67	220,00	26,39
8y	692,58	-2,16	0,68	1,14	1,14	222,00	26,95
8z	763,55	-2,09	0,57	1,27	1,31	244,00	29,46
11a	540,82	-2,36	-1,16	0,15	0,53	160,00	20,97
12a	665,00	1,73	1,73	1,73	1,71	202,00	25,59
8aa	794,33	0,65	2,95	3,96	4,07	256,00	31,07
8bb	806,81	1,07	3,96	4,18	4,17	250,00	30,69
8cc	642,46	-0,74	1,34	1,35	1,33	198,00	24,39
8dd	699,06	-0,48	2,12	2,18	2,17	222,00	26,89
8ee$_1$	656,90	-0,83	1,25	1,26	1,24	210,00	25,60
8ee$_2$	658,93	-0,83	1,25	1,26	1,24	210,00	25,60
8ff	716,55	-0,57	2,03	2,09	2,08	234,00	28,10
9a	482,91	1,37	1,37	1,37	1,34	138,00	19,04
9b	543,54	2,61	2,61	2,61	2,59	162,00	21,54
3a	463,62	0,74	0,74	0,74	0,72	130,00	17,57
3b	495,46	1,25	1,25	1,25	1,23	142,00	18,78

* Para maiores informações sobre metodologia utilizada nos cálculos, verificar o item **3.4**.

Tabela 1A (cont.). Descritores calculados para o conjunto de dados β-N-biaril eter sulfonamida hidroxamatos

Composto	Balaban*	Harary*	Hyper Wiener*	Szeged*	Wiener*	Wiener polarity*	a_1(grau)
8	1,65	87,11	7390,00	2416,00	1702,00	35,00	118,80
$5a_1$	1,84	93,11	8082,00	2615,00	1856,00	38,00	115,52
$5a_2$	1,84	93,11	8082,00	2615,00	1856,00	38,00	114,55
5b	1,82	98,71	8865,00	2832,00	2028,00	41,00	115,11
5c	1,78	103,96	9847,00	3076,00	2227,00	42,00	115,24
5d	1,80	104,81	9651,00	3051,00	2202,00	44,00	115,85
5e	1,56	123,00	12843,00	4074,00	2838,00	49,00	115,62
5f	1,49	127,31	14734,00	4439,00	3149,00	48,00	116,81
5g	1,49	127,31	14734,00	4439,00	3149,00	48,00	116,81
5h	1,49	127,31	14734,00	4439,00	3149,00	48,00	116,20
5j	1,71	88,20	6833,00	2290,00	1642,00	36,00	115,75
5k	1,84	93,11	8082,00	2615,00	1856,00	38,00	115,92
5l	1,82	108,57	12558,00	3674,00	2582,00	44,00	116,17
5m	1,82	108,57	12558,00	3674,00	2582,00	44,00	115,48
5n	1,84	93,11	8082,00	2615,00	1856,00	38,00	115,27
5o	1,84	93,11	8082,00	2615,00	1856,00	38,00	115,44
13a	1,72	97,07	9905,00	2948,00	2132,00	39,00	115,58
13b	1,72	97,07	9905,00	2948,00	2132,00	39,00	117,22
13c	1,64	101,73	11652,00	3327,00	2397,00	41,00	117,54
14a	1,72	97,07	9905,00	2948,00	2132,00	39,00	117,27
14b	1,72	97,07	9905,00	2948,00	2132,00	39,00	117,22
14c	1,71	112,54	15155,00	4091,00	2933,00	45,00	119,53
$7a_2$	1,81	98,58	8954,00	2840,00	2036,00	40,00	120,29
$7a_1$	1,81	98,58	8954,00	2840,00	2036,00	40,00	121,92
8a	1,89	125,43	16767,00	4600,00	3319,00	48,00	119,97
$8b_1$	1,92	109,40	11115,00	3346,00	2452,00	42,00	121,21
$8b_2$	1,92	109,40	11115,00	3346,00	2452,00	42,00	119,86
8c	1,92	109,40	11115,00	3346,00	2452,00	42,00	122,80
8d	1,50	121,32	13750,00	3968,00	2927,00	44,00	121,30
8e	1,40	131,20	17711,00	4892,00	3548,00	48,00	118,55
8f	1,40	131,20	17711,00	4892,00	3548,00	48,00	119,08
8g	1,40	131,20	17711,00	4892,00	3548,00	48,00	118,09
8h	1,40	131,20	17711,00	4892,00	3548,00	48,00	119,40
8ir	1,56	138,23	18840,00	5174,00	3776,00	51,00	120,30
8is	1,56	138,23	18840,00	5174,00	3776,00	51,00	118,95
8j	1,46	126,81	15340,00	4487,00	3197,00	47,00	121,09
8k	1,42	132,37	17221,00	4918,00	3496,00	49,00	119,89
8l	1,43	132,72	16937,00	4866,00	3470,00	49,00	121,24
8m	1,43	132,72	16937,00	4866,00	3470,00	49,00	122,26
8n	1,57	138,23	18840,00	5278,00	3776,00	51,00	120,80
8o	1,19	151,37	22947,00	6538,00	4415,00	56,00	121,52
8p	1,88	132,27	18014,00	4896,00	3552,00	52,00	122,86
8q	1,57	169,57	29990,00	7518,00	5472,00	61,00	124,53
8r	1,440	200,790	56398,00	11164,00	8686,00	68,00	116,63
8s	1,54	138,90	19791,00	5527,00	3841,00	53,00	118,73
8t	1,49	151,59	23385,00	6379,00	4435,00	57,00	118,48
8u	1,56	162,64	29274,00	7554,00	5244,00	61,00	117,81
8v	1,43	156,17	26810,00	7036,00	4882,00	59,00	118,79
8x	1,54	138,90	19791,00	5527,00	3841,00	53,00	121,30
8y	1,50	144,73	21891,00	6006,00	4164,00	55,00	118,31
8z	1,55	161,77	29668,00	7589,00	5279,00	61,00	118,89
11a	1,84	93,11	8082,00	2615,00	1856,00	38,00	119,70
12a	1,83	120,48	13758,00	3914,00	2930,00	46,00	121,21
8aa	1,21	174,49	35816,00	9067,00	6112,00	64,00	119,24
8bb	1,18	180,35	39577,00	9808,00	6586,00	66,00	122,27
8cc	1,53	122,49	13320,00	4114,00	2878,00	47,00	118,59
8dd	1,47	134,83	15996,00	4816,00	3370,00	51,00	120,86
$8ee_1$	1,53	122,49	13320,00	4114,00	2878,00	47,00	119,16
$8ee_2$	1,53	122,49	13320,00	4114,00	2878,00	47,00	118,91
8ff	1,47	134,83	15996,00	4816,00	3370,00	51,00	120,07
9a	1,64	86,89	7471,00	2424,00	1710,00	36,00	116,73
9b	1,80	97,92	9028,00	2854,00	2050,00	40,00	116,27
3a	1,82	78,15	5184,00	1878,00	1308,00	32,00	117,51
3b	1,74	82,92	6208,00	2162,00	1493,00	34,00	118,19

* Para maiores informações sobre metodologia utilizada nos cálculos, verificar o item **3.4**.

Tabela 2A. Valores de r^2_{yrand} e q^2_{yrand} das 10 aleatorizações obtidas para o modelo **4** (N = 38) do alinhamento 2

Aleatorização	r^2_{yrand}	q^2_{yrand}
1	0,07	-0,45
2	0,22	-0,17
3	0,12	-0,23
4	0,07	-0,35
5	0,19	-0,27
6	0,21	-0,32
7	0,16	-0,27
8	0,08	-0,54
9	0,09	-0,38
10	0,14	-0,42
Modelo 4	**0,88**	**0,83**

Corrida	r^2_{yrand}	q^2_{yrand}
1	0,25	0,00
2	0,13	-0,27
3	0,14	-0,38
4	0,23	-0,12
5	0,18	-0,34
6	0,08	-0,35
7	0,11	-0,45
8	0,16	-0,22
9	0,19	-0,25
10	0,09	-0,29
11	0,20	-0,13
12	0,18	-0,33
13	0,25	-0,16
14	0,12	-0,30
15	0,26	-0,10
16	0,13	-0,32
17	0,15	-0,44
18	0,14	-0,32
19	0,02	-0,51
20	0,03	-0,53
21	0,22	-0,16
22	0,17	-0,42
23	0,23	-0,16
24	0,21	-0,21
25	0,19	-0,25
26	0,21	-0,20
27	0,11	-0,32
28	0,15	-0,29
29	0,20	-0,18
30	0,15	-0,27
Modelo 4	0,88	0,83

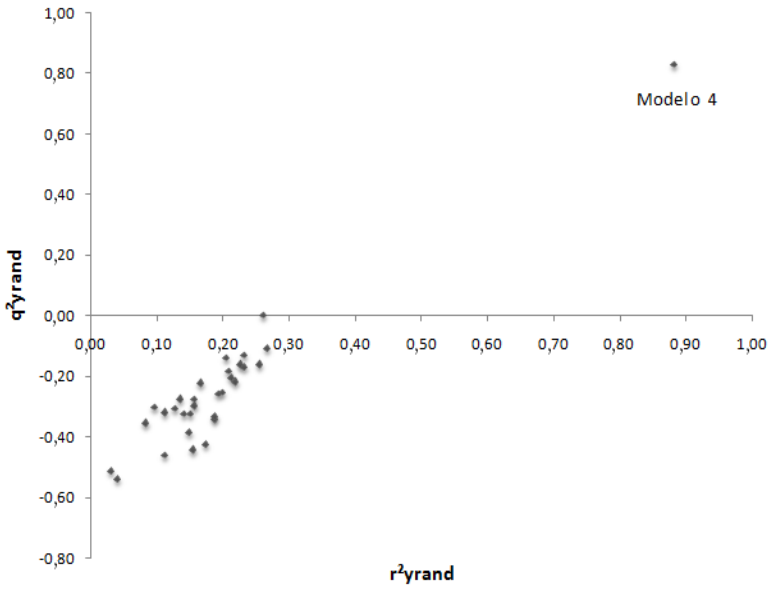


Figura 1A. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de inibidores (N = 38) com 30 corridas ou aleatorizações do vetor **y** (atividade inibitória na MMP-2).

Corrida	r^2_{yrand}	q^2_{yrand}
1	0,16	-0,24
2	0,16	-0,33
3	0,16	-0,19
4	0,12	-0,34
5	0,03	-0,46
6	0,26	-0,14
7	0,13	-0,33
8	0,14	-0,25
9	0,16	-0,28
10	0,18	-0,18
11	0,12	-0,26
12	0,26	-0,10
13	0,13	-0,31
14	0,26	-0,08
15	0,11	-0,44
16	0,14	-0,24
17	0,15	-0,30
18	0,16	-0,23
19	0,16	-0,28
20	0,12	-0,34
21	0,16	-0,42
22	0,07	-0,56
23	0,27	-0,09
24	0,12	-0,26
25	0,07	-0,45
26	0,18	-0,23
27	0,12	-0,39
28	0,02	-0,53
29	0,09	-0,30
30	0,24	-0,13
31	0,20	-0,18
32	0,24	-0,12
33	0,16	-0,20
34	0,27	-0,16
35	0,12	-0,29
36	0,25	-0,14
37	0,17	-0,27
38	0,26	-0,05
39	0,03	-0,55
40	0,05	-0,60
41	0,02	-0,48
42	0,14	-0,21
43	0,14	-0,21
44	0,27	-0,10
45	0,08	-0,46
46	0,25	-0,32
47	0,21	-0,14
48	0,25	-0,13
49	0,09	-0,35
50	0,11	-0,37
Modelo 4	0,88	0,83

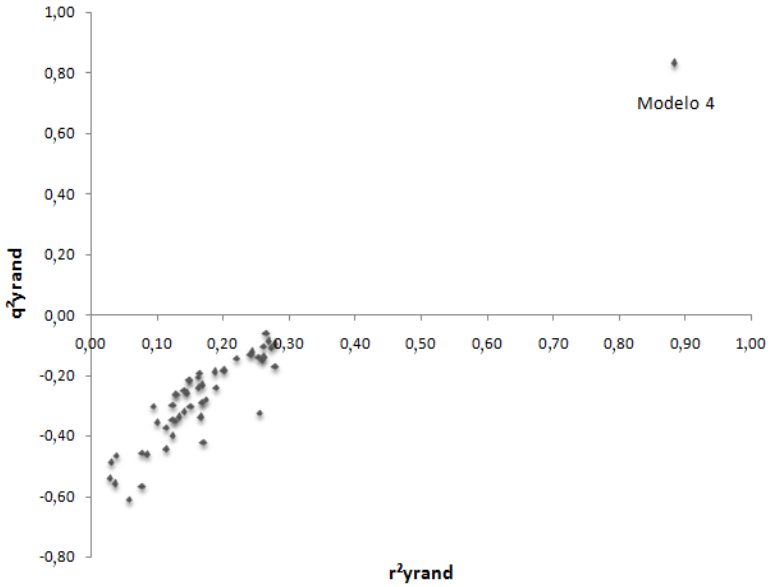


Figura 2A. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de inibidores (N = 38) com 50 corridas ou aleatorizações do vetor **y** (atividade inibitória na MMP-2).

Tabela 3A. Valores de r^2_{yrand} e q^2_{yrand} para as 10 aleatorizações do modelo **4** (N = 30) para o alinhamento 2

Aleatorização	r^2_{yrand}	q^2_{yrand}
1	0,34	0,1
2	0,25	-0,4
3	0,19	-0,2
4	0,22	-0,8
5	0,32	-0,1
6	0,10	-1,8
7	0,14	-0,3
8	0,10	-1,6
9	0,31	0,1
10	0,19	-0,1
Modelo 4	**0,93**	**0,88**

Corrida	r^2_{yrand}	q^2_{yrand}
1	0,12	-0,31
2	0,12	-0,59
3	0,20	-0,28
4	0,13	-0,17
5	0,39	0,07
6	0,13	-0,55
7	0,03	-0,60
8	0.12	-1.00
9	0,19	-0,29
10	0,16	-0,14
11	0,14	-0,45
12	0,26	-0,12
13	0,10	-0,44
14	0,22	-0,38
15	0,17	-0,17
16	0,15	-0,59
17	0,12	-0,50
18	0,22	-0,23
19	0,29	-0,07
20	0,26	-0,27
21	0,20	-0,11
22	0,13	-1,48
23	0,24	-0.15
24	0,10	-0,59
25	0,27	0,03
26	0,18	-0,42
27	0,20	-0,42
28	0,04	-1,44
29	0,13	-0,31
30	0,07	-0,65
Modelo 4	0,92	0,88

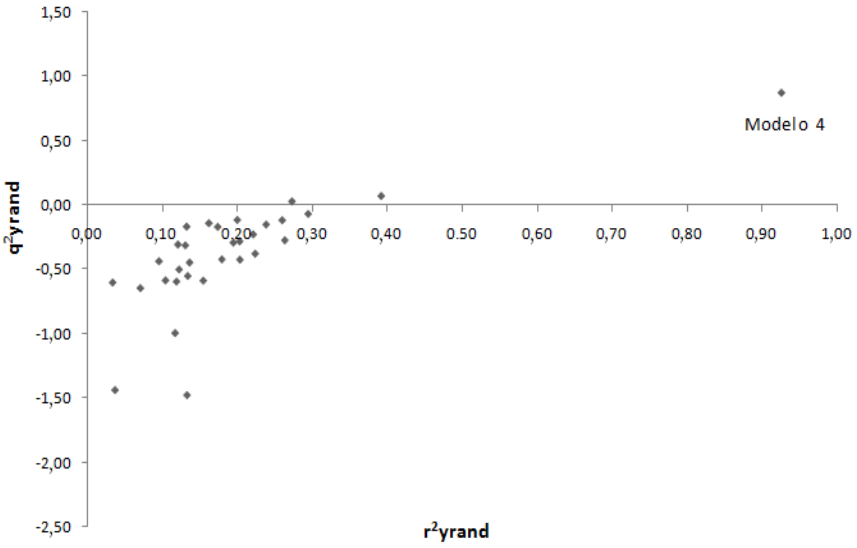


Figura 3A. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de inibidores (N = 30) com 30 corridas ou aleatorizações do vetor **y** (atividade inibitória frente à MMP-2).

Corrida	r^2_{yrand}	q^2_{yrand}
1	0,10	-0,50
2	0,12	-0,71
3	0,06	-0,88
4	0,24	-0,09
5	0,19	-0,21
6	0,23	-0,16
7	0,11	-0,71
8	0,05	-0,65
9	0,32	-0,72
10	0,25	0,05
11	0,22	-0,64
12	0,18	-1,30
13	0,33	0,11
14	0,04	-0,53
15	0,19	-0,48
16	0,22	-0,21
17	0,05	-0,69
18	0,17	-0,12
19	0,17	-0,28
20	0,19	-0,39
21	0,36	-0,03
22	0,26	-0,19
23	0,09	-0,49
24	0,29	0,07
25	0,36	0,13
26	0,25	0,02
27	0,16	-0,26
28	0,15	-0,27
29	0,03	-0,58
30	0,23	0,01
31	0,34	-0,39
32	0,10	-0,51
33	0,13	-0,22
34	0,11	-2,15
35	0,25	0,03
36	0,09	-0,40
37	0,16	-0,64
38	0,15	-0,27
39	0,30	-0,03
40	0,07	-1,21
41	0,10	-0,96
42	0,37	-0,46
43	0,25	-0,04
44	0,30	-0,09
45	0,13	-0,29
46	0,12	-0,40
47	0,06	-0,95
48	0,12	-0,66
49	0,09	-1,14
50	0,21	-0,12
Modelo 4	0,92	0,88

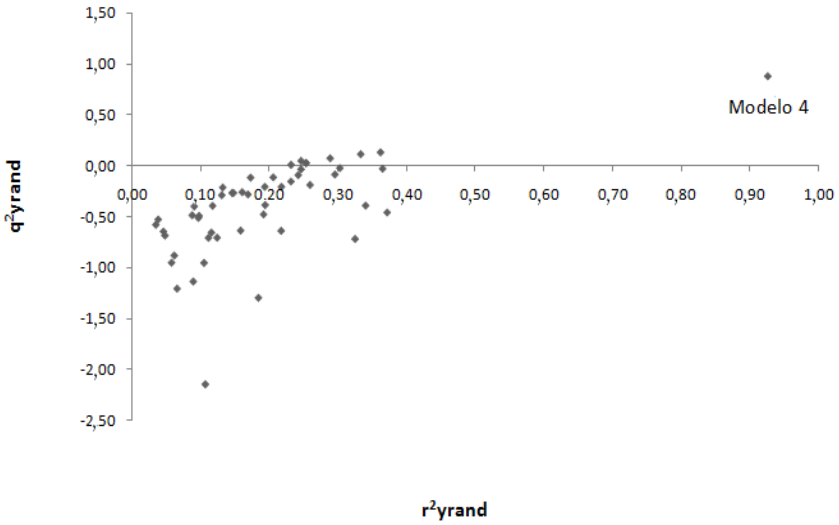


Figura 4A. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de inibidores (N = 30) com 50 corridas ou aleatorizações do vetor **y** (atividade inibitória frente à MMP-2).

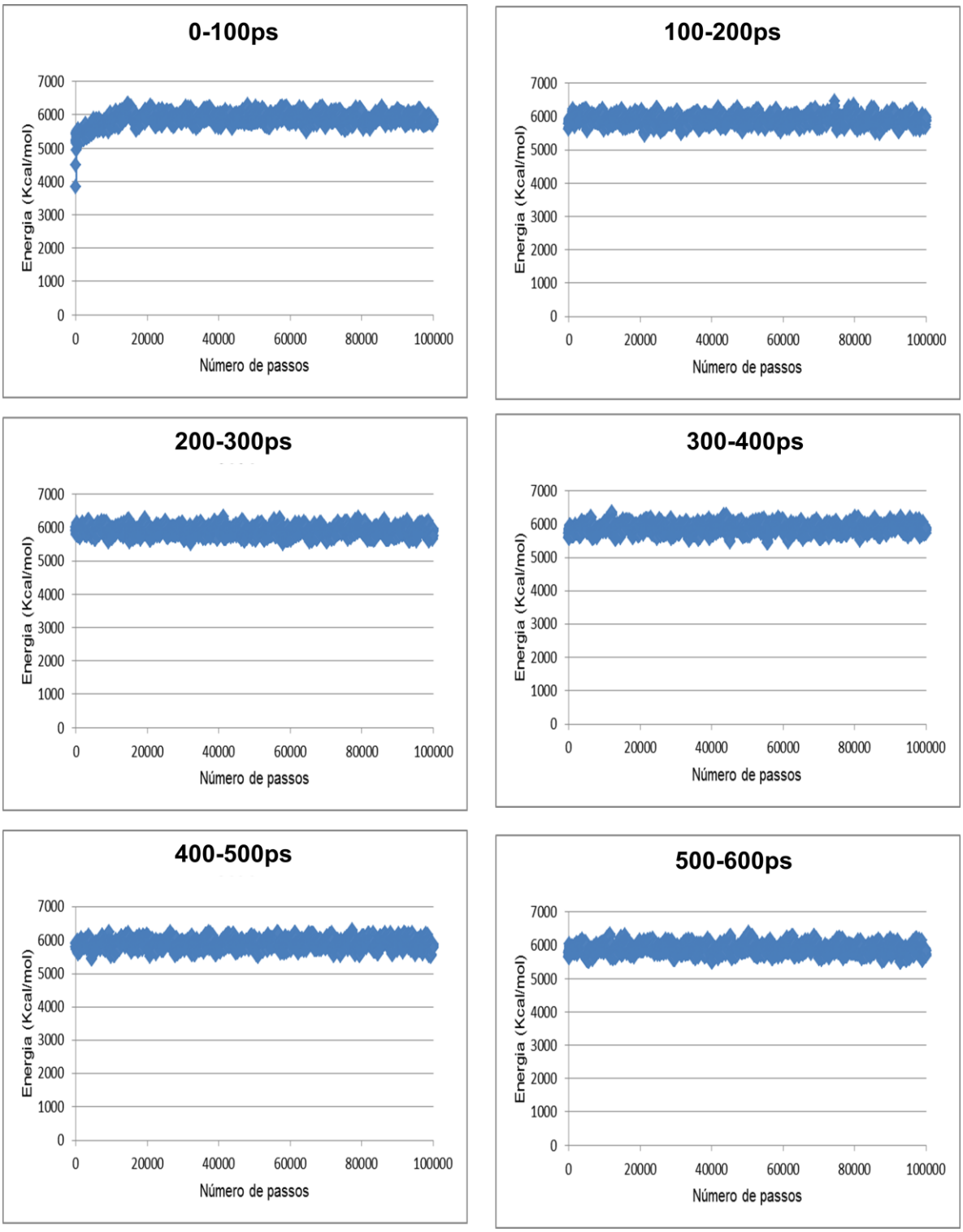


Figura 5A. PAC do complexo MMP-2-co-cristalografado com inibidor hidroxamato (**PDB ID** 1HOV) a uma temperatura de 298 K.

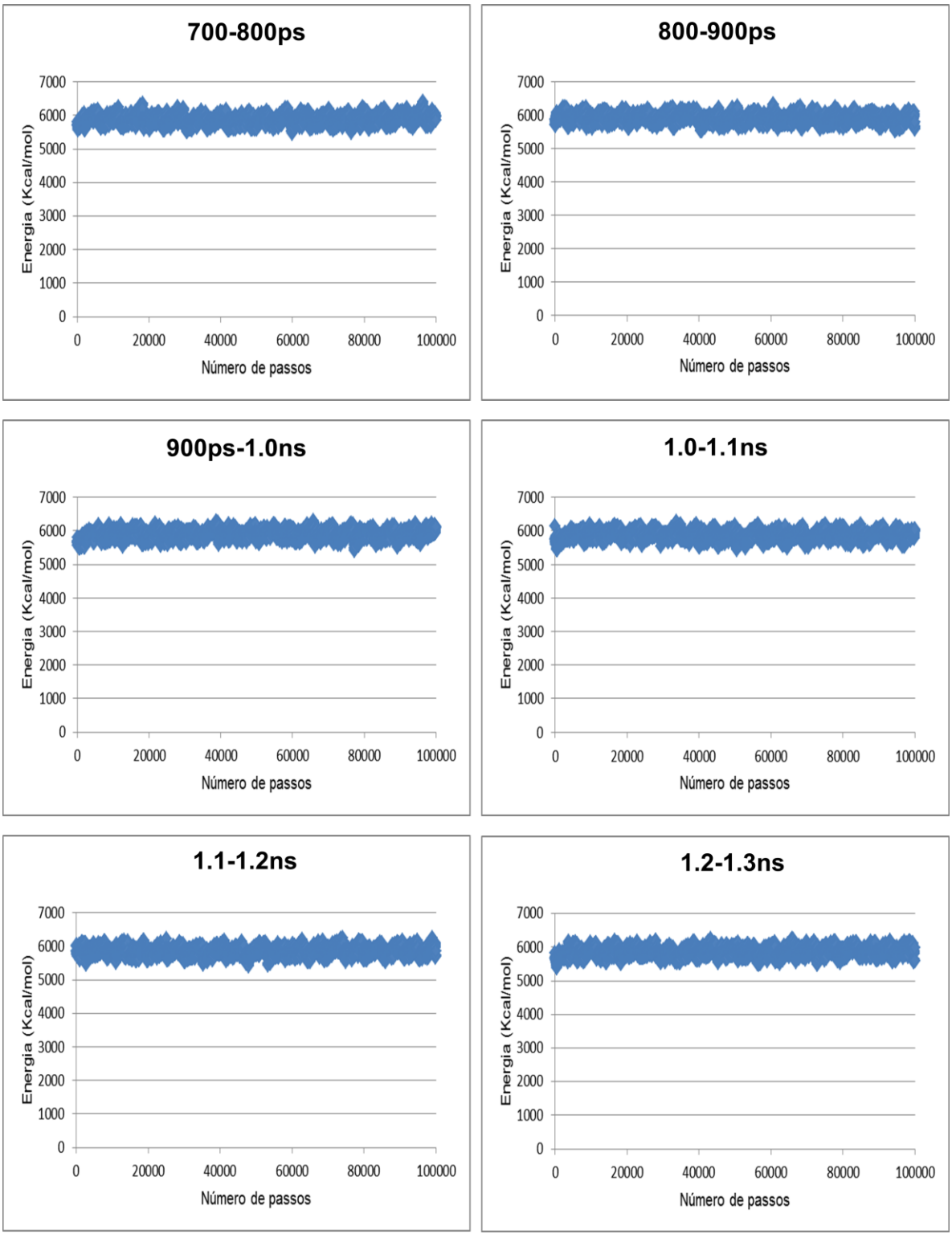


Figura 5A (cont). PAC do complexo MMP-2-co-cristalografado com inibidor hidroxamato (**PDB ID** 1HOV) a uma temperatura de 298 K.

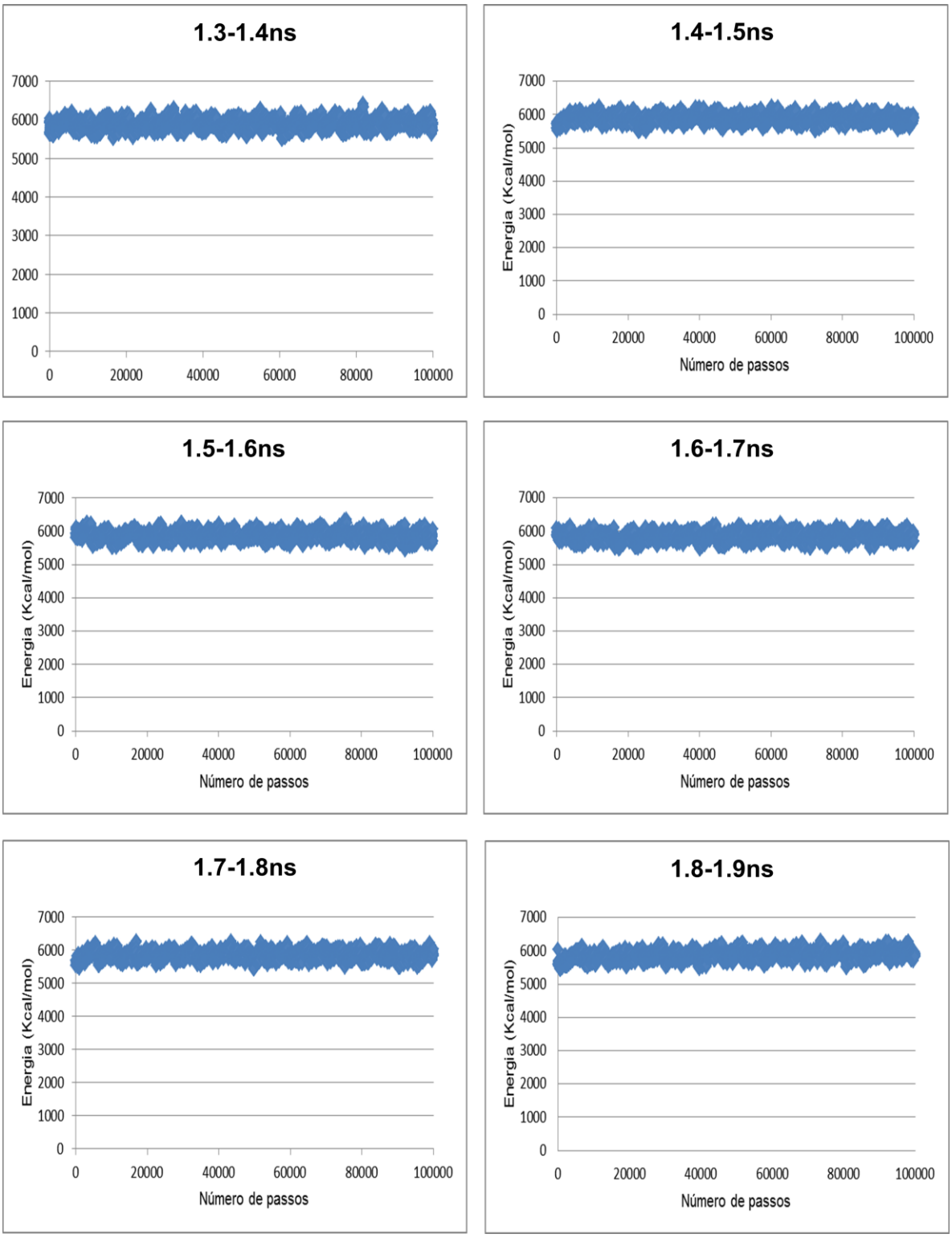


Figura 5A (cont). PAC do complexo MMP-2-co-cristalografado com inibidor hidroxamato (**PDB ID** 1HOV) a uma temperatura de 298 K.

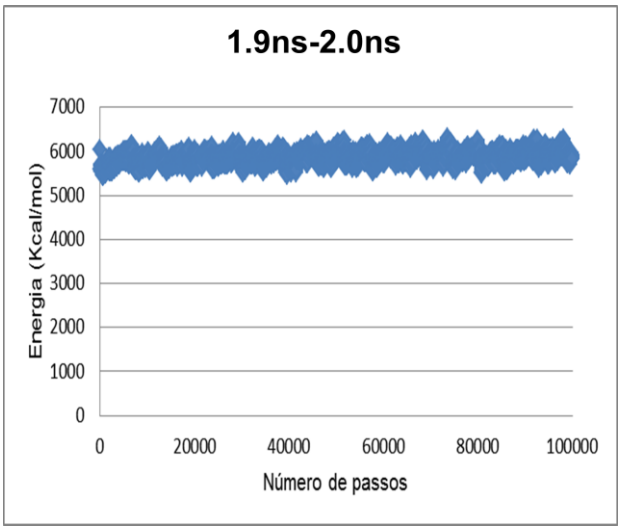


Figura 5A (cont). PAC do complexo MMP-2-co-cristalografado com inibidor hidroxamato (**PDB ID** 1HOV) a uma temperatura de 298 K.

Corrida	r^2_{yrand}	q^2_{yrand}
1	0,16	-5,24
2	0,20	-0,94
3	0,21	-3,34
4	0,11	-2,04
5	0,09	-7,59
6	0,08	-0,50
7	0,34	-1,22
8	0,17	-1,03
9	0,11	-1,96
10	0,16	-1,22
11	0,16	-0,26
12	0,15	-0,92
13	0,07	-1,62
14	0,11	-5,27
15	0,16	-1,04
16	0,13	-0,52
17	0,30	-0,40
18	0,15	-2,70
19	0,16	-1,60
20	0,17	-0,63
21	0,23	-2,35
22	0,13	-1,83
23	0,18	-5,70
24	0,11	-7,54
25	0,17	-3,38
26	0,15	-6,69
27	0,22	-2,21
28	0,30	-1,14
29	0,20	-9,42
30	0,14	-1,63
Modelo 6	0,86	0,82

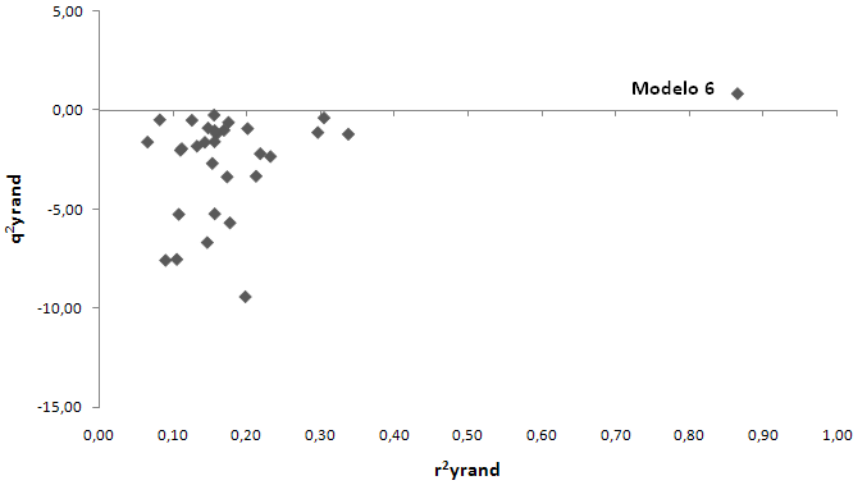


Figura 6A. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de inibidores (N = 64) com 30 corridas ou aleatorizações do vetor **y** (atividade inibitória na MMP-9).

Corrida	r^2_{yrand}	q^2_{yrand}
1	0,17	-13,46
2	0,17	-9,93
3	0,10	-1,11
4	0,08	-16,03
5	0,07	-4,32
6	0,15	-1,23
7	0,14	-2,13
8	0,09	-9,80
9	0,19	-5,61
10	0,13	-0,58
11	0,06	-0,72
12	0,16	-4,95
13	0,13	-4,55
14	0,15	-1,57
15	0,24	-14,52
16	0,27	-5,75
17	0,22	-4,09
18	0,07	-1,70
19	0,12	-2,87
20	0,23	-3,44
21	0,18	-10,45
22	0,23	-0,26
23	0,18	-2,92
24	0,17	-11,96
25	0,08	-11,01
26	0,07	-2,97
27	0,14	-3,90
28	0,19	-0,56
29	0,19	-1,00
30	0,15	-15,30
31	0,14	-1,25
32	0,16	-1,34
33	0,10	-2,56
34	0,07	-13,18
35	0,15	-1,12
36	0,13	-0,68
37	0,14	-2,23
38	0,14	-1,82
39	0,15	-5,09
40	0,07	-4,88
41	0,23	-0,78
42	0,10	-0,54
43	0,21	-10,03
44	0,20	-7,46
45	0,13	-3,57
46	0,12	-1,78
47	0,16	-0,34
48	0,14	-1,33
49	0,12	-1,17
50	0,19	-2,59
Modelo 6	0,86	0,82

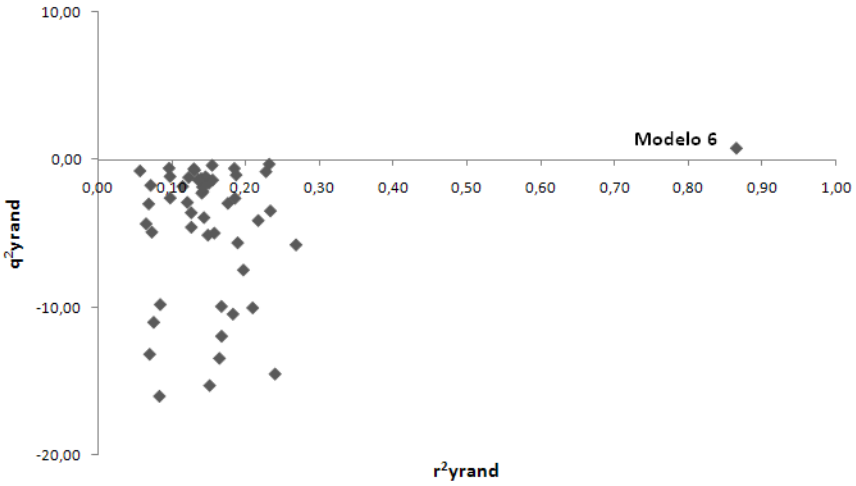


Figura 7A. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de inibidores (N = 64) com 50 corridas ou aleatorizações do vetor **y** (atividade inibitória na MMP-9).

Tabela 4A. Valores de r^2_{yrand} e q^2_{yrand} paraas 10 aleatorizações obtidas para o modelo 6 (N = 64) do alinhamento 2

Aleatorização	r^2_{yrand}	q^2_{yrand}
1	0,11	-6,78
2	0,17	-0,78
3	0,18	-0,65
4	0,06	-1,86
5	0,18	-5,50
6	0,20	-0,75
7	0,09	-0,74
8	0,29	-1,97
9	0,17	-0,40
10	0,13	-0,24
Modelo 6	**0,86**	**0,82**

Tabela 5A. Valores de r^2_{yrand} e q^2_{yrand} para as 10 aleatorizações do modelo **1** (N = 47) para o alinhamento 2

Aleatorização	r^2_{yrand}	q^2_{yrand}
1	0,21	-0,15
2	0,16	-0,43
3	0,19	-0,20
4	0,11	-0,44
5	0,15	-0,42
6	0,10	-0,50
7	0,08	-0,84
8	0,40	0,12
9	0,29	-0,27
10	0,16	-0,35
Modelo 1	**0,91**	**0,83**

Corrida	r^2_{yrand}	q^2_{yrand}
1	0,18	-0,30
2	0,25	-0,20
3	0,21	-0,19
4	0,17	-0,41
5	0,13	-0,34
6	0,15	-0,41
7	0,10	-0,60
8	0,28	-0,45
9	0,20	-0,43
10	0,18	-0,38
11	0,32	-0,19
12	0,28	-0,08
13	0,16	-0,93
14	0,09	-0,39
15	0,08	-0,75
16	0,07	-0,65
17	0,30	-0,57
18	0,18	-0,55
19	0,10	-0,49
20	0,15	-0,47
21	0,20	-0,73
22	0,19	-0,19
23	0,21	-0,21
24	0,13	-0,40
25	0,28	-0,07
26	0,20	-0,35
27	0,22	-0,34
28	0,22	-0,29
29	0,12	-0,33
30	0,18	-0,33
Modelo 1	0,91	0,83

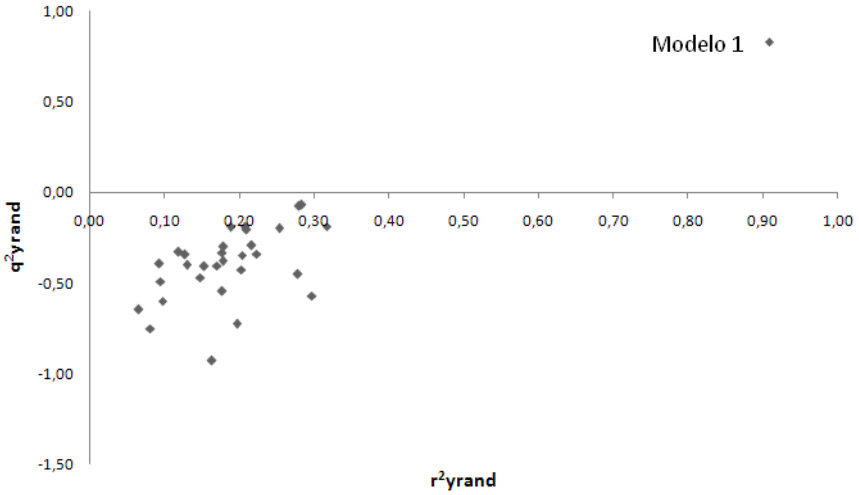


Figura 8A. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de inibidores (N = 47) com 30 corridas ou aleatorizações do vetor **y** (atividade inibitória na MMP-9).

Corrida	r^2_{yrand}	q^2_{yrand}
1	0,22	-0,32
2	0,13	-0,54
3	0,27	-0,36
4	0,09	-0,63
5	0,23	-0,76
6	0,16	-0,44
7	0,17	-0,32
8	0,33	-0,17
9	0,17	-0,43
10	0,16	-0,43
11	0,14	-0,32
12	0,26	-0,54
13	0,18	-0,36
14	0,12	-0,32
15	0,18	-0,47
16	0,20	-0,44
17	0,26	-0,15
18	0,28	-0,48
19	0,30	-0,31
20	0,11	-0,52
21	0,30	-0,30
22	0,14	-0,79
23	0,17	-0,47
24	0,06	-0,65
25	0,19	-0,31
26	0,26	-0,60
27	0,10	-0,61
28	0,20	-0,46
29	0,29	-0,23
30	0,15	-0,44
31	0,24	-0,19
32	0,15	-0,39
33	0,26	-0,17
34	0,17	-0,29
35	0,12	-0,62
36	0,22	-0,70
37	0,27	-0,28
38	0,07	-0,64
39	0,06	-0,88
40	0,24	-0,23
41	0,17	-0,87
42	0,06	-1,01
43	0,38	-0,09
44	0,10	-0,60
45	0,09	-0,55
46	0,14	-0,71
47	0,36	-0,04
48	0,28	-0,10
49	0,09	-0,70
50	0,04	-0,58
Modelo 1	0,91	0,83

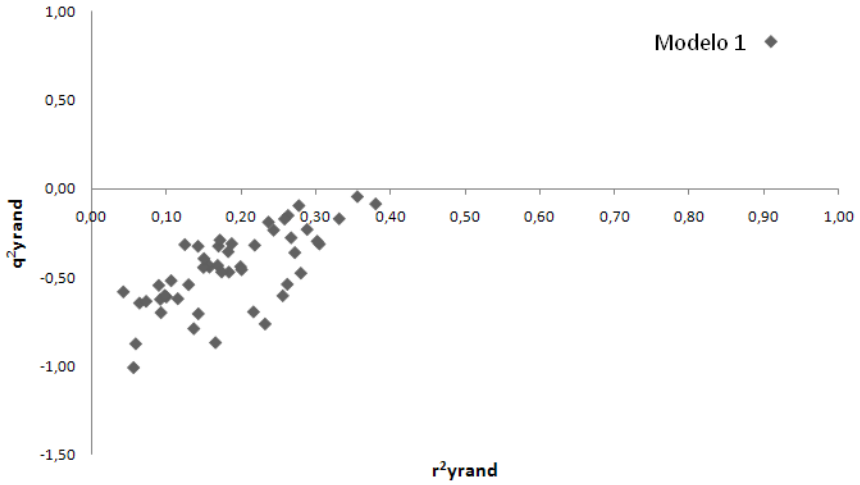


Figura 9A. Valores de q^2_{yrand} e r^2_{yrand} obtidos para o conjunto de inibidores (N = 47) com 50 corridas ou aleatorizações do vetor **y** (atividade inibitória na MMP-9).

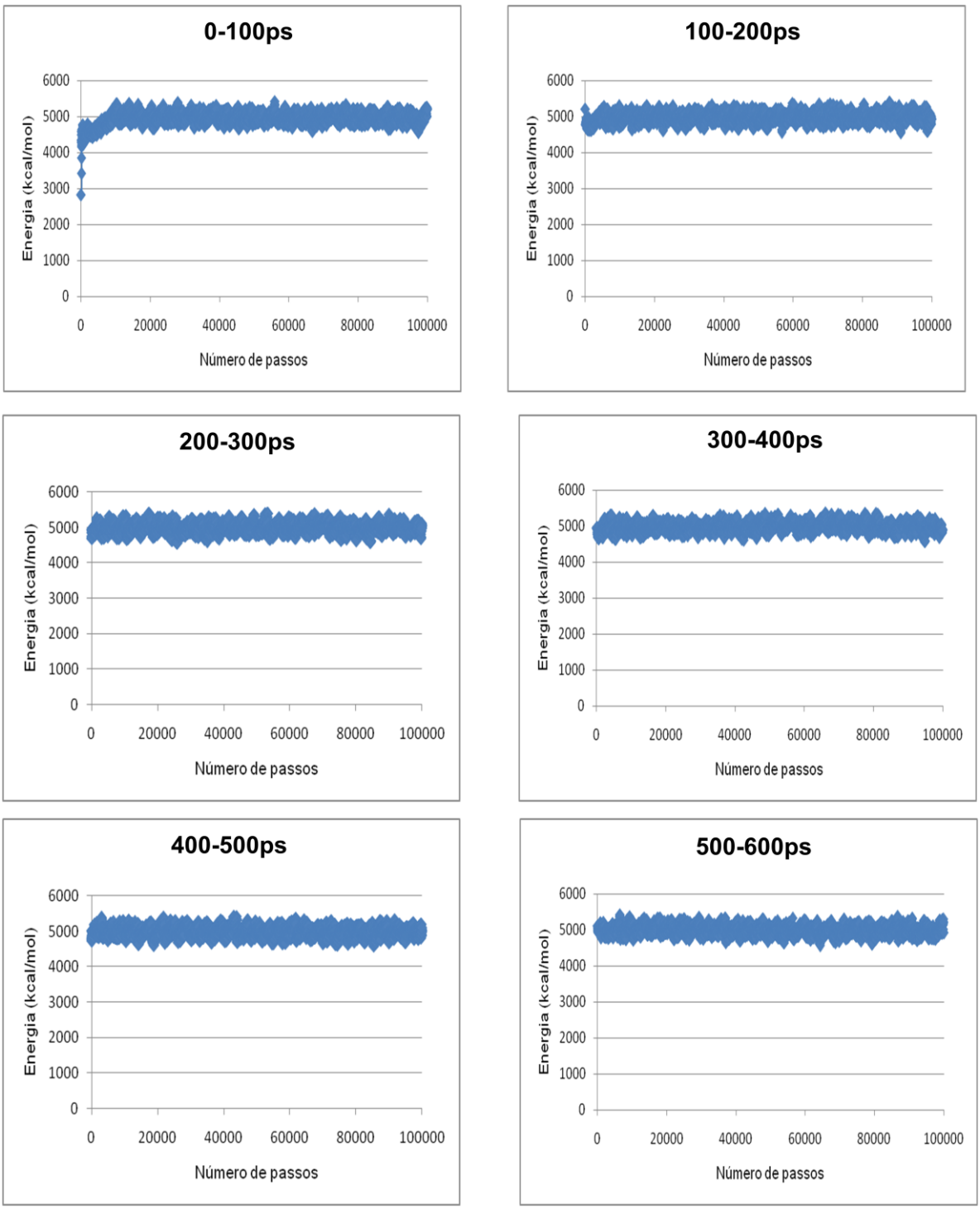


Figura 10A. PAC do complexo MMP-9-co-cristalografado com inibidor hidroxamato (**PDB ID** 2OW1) a uma temperatura de 298 K.

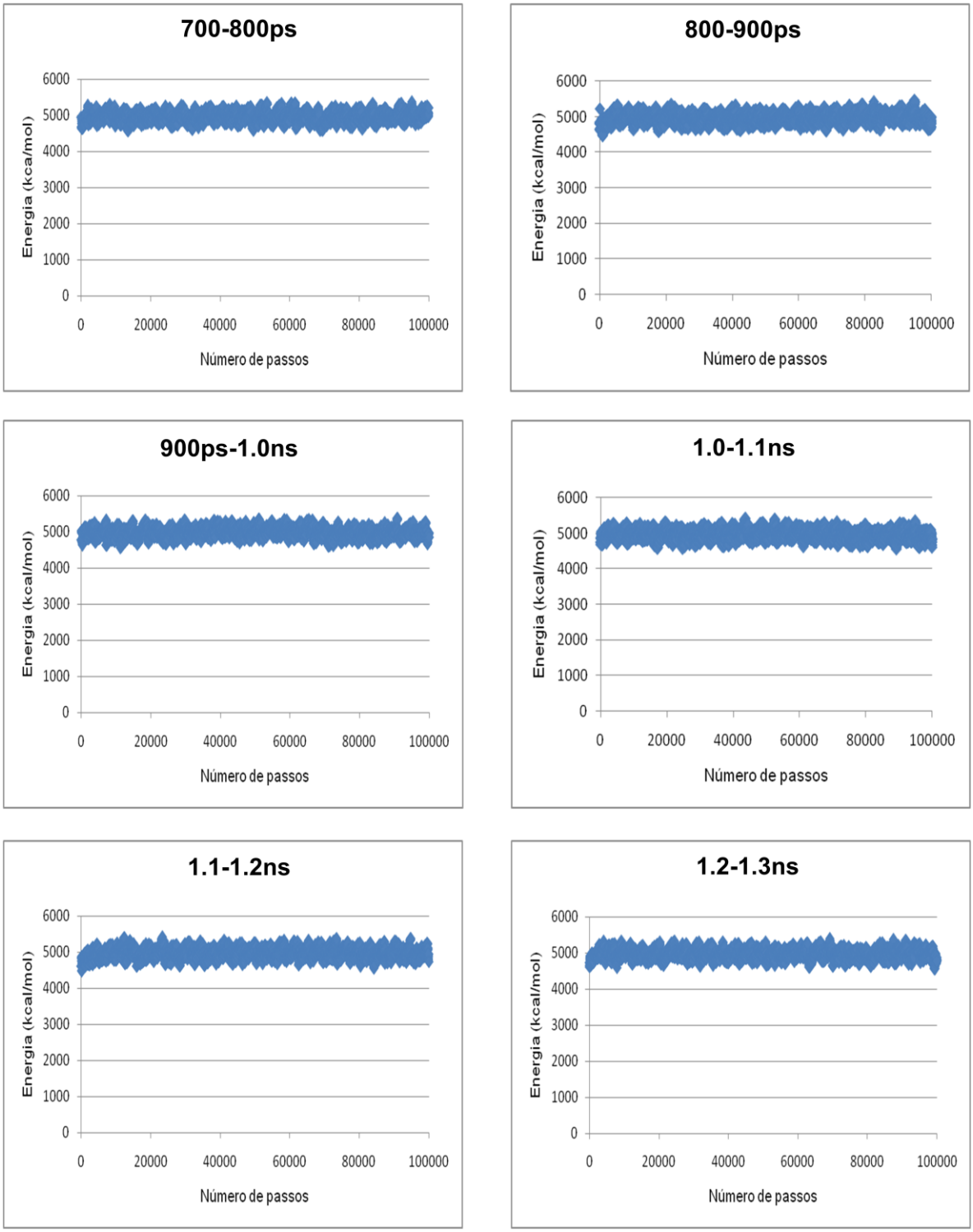


Figura 10A (cont.). PAC do complexo MMP-9-co-cristalografado com inibidor hidroxamato (**PDB ID** 2OW1) a uma temperatura de 298 K.

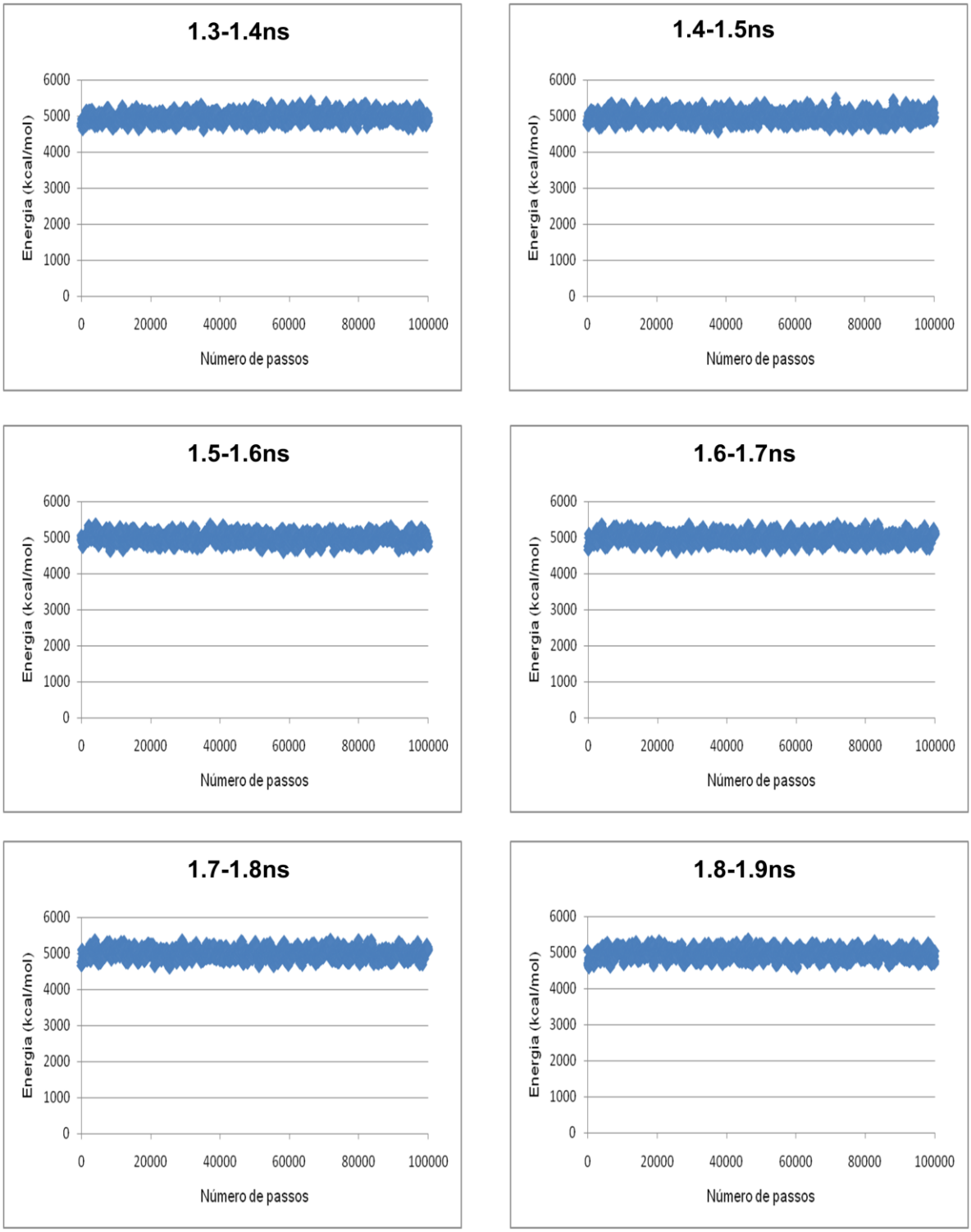


Figura 10A (cont.). PAC do complexo MMP-9-co-cristalografado com inibidor hidroxamato (**PDB ID** 2OW1) a uma temperatura de 298 K.

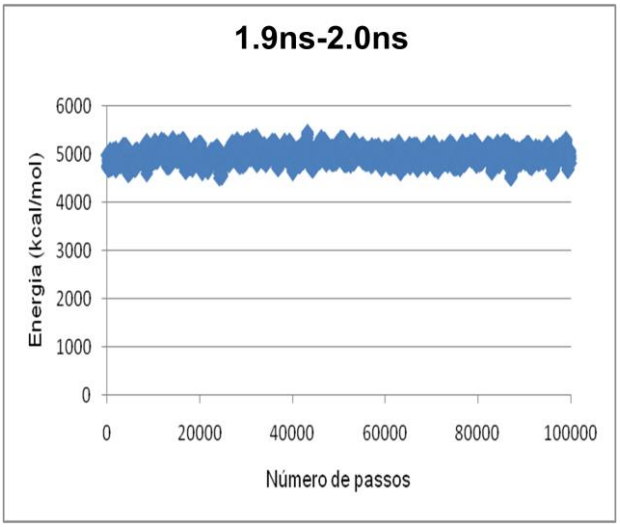


Figura 10A (cont.). PAC do complexo MMP-9-co-cristalografado com inibidor hidroxamato (**PDB ID**-2OW1) a uma temperatura de 298 K.

Tabela 6A. Valores de energia obtidos para os confôrmeros de menor e maior energia em cada corrida nos estudos de ancoramento do 4-NC na MMP-2

Corrida	**Valor obtido para o confôrmero de menor energia** (kcal/mol)	**Valor obtido para o confôrmero de maior energia** (kcal/mol)
1	-7,18	-4,34
2	-7,3	-4,56
3	-7,44	-4,63
4	-7,13	-4,71
5	-7,85	-4,01
6	-7,47	-4,5
7	-6,52	-4,63
8	-7,11	-4,62
9	-7,11	-4,62
10	-7,2	-4,27
11	-7,74	-4,6
12	-6,92	-3,99
13	-8,08	-4,55
14	-7,37	-4,57
15	-6,86	-4,1
16	-7,67	-4,77
17	-7,05	-4,34
18	-7,26	-4,58
19	-6,62	-4,49
20	-6,7	-4,48
21	-7,47	-4,11
22	-7,78	-4,25
23	-7,56	-4,52

Tabela 6A (cont.). Valores de energia obtidos para os confôrmeros de menor e maior energia em cada corrida nos estudos de ancoramento do 4-NC na MMP-2

Corrida	**Valor obtido para o confôrmero de menor energia** (kcal/mol)	**Valor obtido para o confôrmero de maior energia** (kcal/mol)
24	-8,13	-4,11
25	-7,23	-4,83
26	-7,53	-4,69
27	-7,12	-4,23
28	-7,2	-4,13
29	-7,2	-4,55
30	-7,79	-4,3

Tabela 7A. Valores de energia obtidos para os confôrmeros de menor e maior energia em cada corrida nos estudos de ancoramento do 4-NC na MMP-9

Corrida	**Valor obtido para o confôrmero de menor energia** (kcal/mol)	**Valor obtido para o confôrmero de maior energia** (kcal/mol)
1	-8,67	-5,99
2	-8,94	-4,61
3	-8,8	-6,03
4	-8,69	-5,45
5	-8,82	-5,43
6	-9,02	-5,59
7	-8,8	-5,73
8	-8,66	-5,91
9	-8,61	-5,75
10	-8,66	-5,91
11	-8,91	-5,89
12	-8,56	-5,47
13	-8,54	-5,51
14	-9,17	-6,31
15	-8,86	-6,07
16	-9,25	-5,91
17	-8,86	-5,47
18	-8,85	-5,96
19	-8,53	-5,66
20	-8,45	-5,99
21	-8,71	-5,54
22	-8,94	-5,26
23	-9,12	-6,22
24	-9,33	-6,12
25	-8,63	-5,4
26	-9,3	-5,84
27	-9,05	-6,18
28	-8,69	-5,17
29	-8,86	-5,15
30	-8,79	-5,59

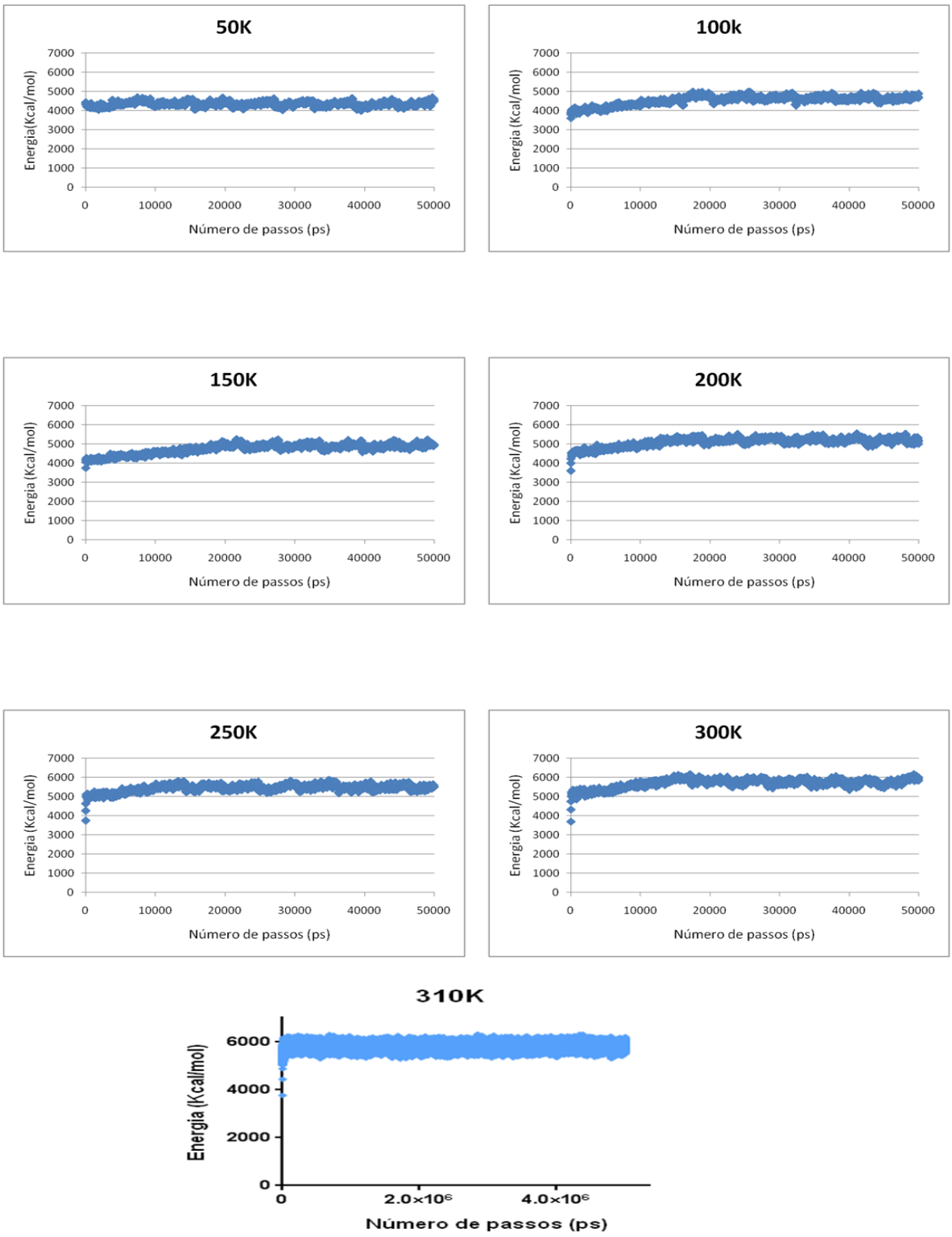


Figura 11A. PAC das simulações de DM do esquema de aquecimento lento e da simulação mais longa para o complexo 4-NC-MMP-2 (anel catecol voltado para o bolso S1'). Simulação do complexo a 50 K (50 ps), 100 K (50 ps), 200 K (50 ps), 300 K (50 ps) e 310 K (5ns). As contribuições de energia de solvatação e de ligação de hidrogênio não foram calculadas nos PAC.

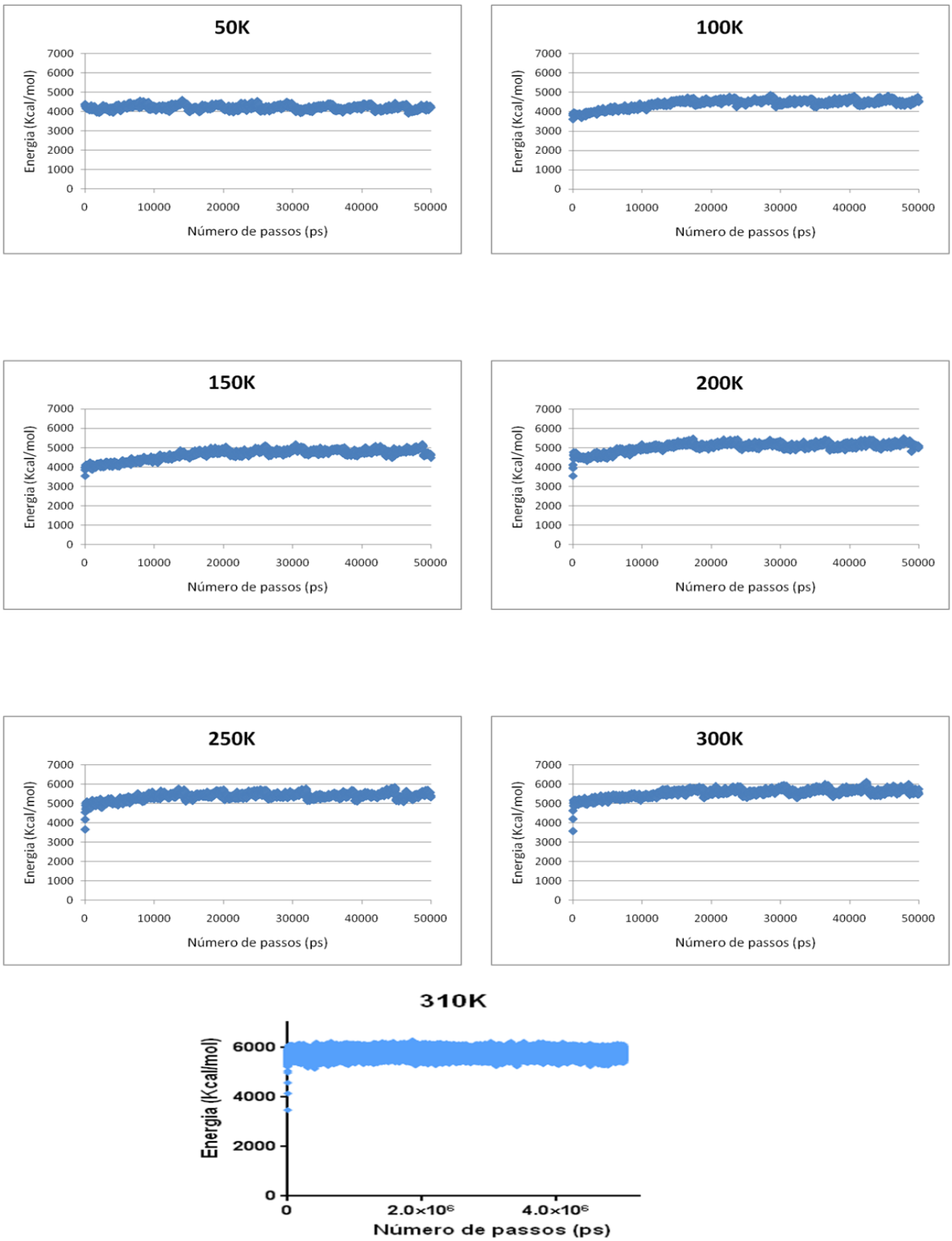


Figura 11A (cont.). PAC das simulações de DM do esquema de aquecimento lento e da simulação mais longa para o complexo 4-NC-MMP-2 (cadeia lateral voltada para o bolso S1'). Simulação do complexo a 50 K (50 ps), 100 K (50 ps), 200 K (50 ps), 300 K (50 ps) e 310 K (5ns). As contribuições de energia de solvatação e de ligação de hidrogênio não foram calculadas nos PAC.

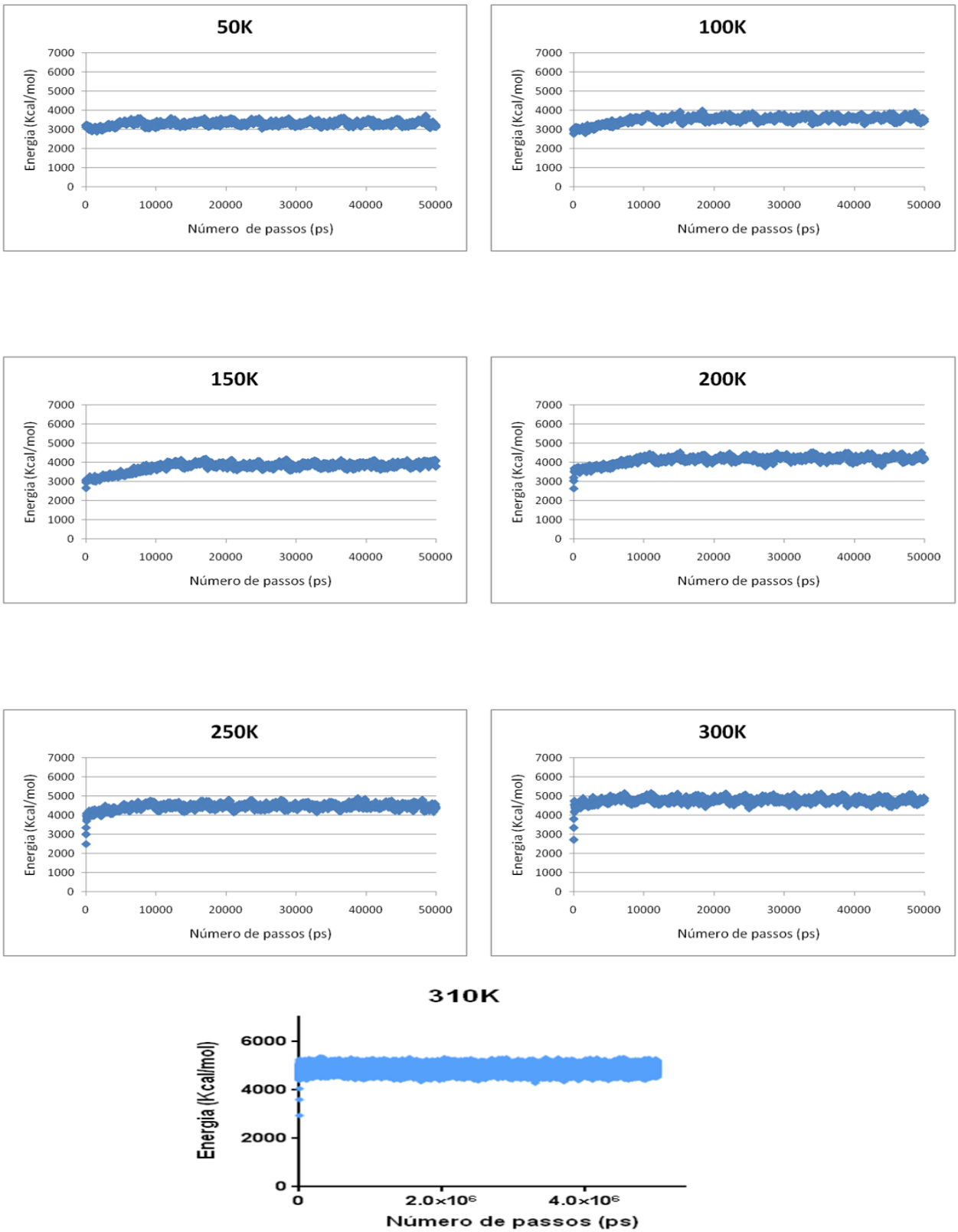


Figura 11A (cont.). PAC das simulações de DM do esquema de aquecimento lento e da simulação mais longa para o complexo 4-NC-MMP-9 (anel catecol voltado para o boso S1'). Simulação do complexo a 50 K (50 ps), 100 K (50 ps), 200 K (50 ps), 300 K (50 ps) e 310 K (5ns). As contribuições de energia de solvatação e de ligação de hidrogênio não foram calculadas nos PAC.

Figura 11A (cont). PAC das simulações de DM do esquema de aquecimento lento e da simulação mais longa para o complexo 4-NC-MMP-9 (cadeia lateral voltada para o bolso S1'). Simulação do complexo a 50 K (50 ps), 100 K (50 ps), 200 K (50 ps), 300 K (50 ps) e 310 K (5ns). As contribuições de energia de solvatação e de ligação de hidrogênio não foram calculadas nos PAC.

Tabela 8A. Valores de energia obtidos para os confôrmeros selecionados nas simulações de DM a 310 K (Valores expressos em kcal/mol)

Complexos	Eest	Eang	Etors	E1,4	EvdW	Eel	INTERV+C	Esolv	Ehb	Etot
MMP-2 (anel voltado para bolso S1´)	2372,11	1753,71	1294,22	4002,51	-657,29	-3621,18	30,45	-163,476	-19341,7	-14330,7
MMP-2 (cauda voltada para bolso S1´)	2313,87	1794,47	1279	3978,6	-692,59	-3531,73	3,22	-113,705	-19895,1	-14863,9
MMP-9 (cauda voltada para bolso S1´)	1777,28	1811,23	1197,59	3986,11	-816,25	-3781,44	144,9	-146,332	-19244,7	-15071,6
MMP-9 (anel voltado para bolso S1´)	1704,7	1734,8	1193,33	3997,52	-818,15	-3678,19	149,76	-122,981	-18631,3	-14470,5

* Eest = energia de deformação axial ou estiramento; Eang = energia de deformação angular; Etors = energia torsional; $E_{1,4}$ = energia de interação do tipo 1-4 ou Lennard-Jones; EvdW = energia de van der Waals; Eel = energia eletrostática; INTERV+C=soma das contribuições de energia eletrostática e de van der Waals intermoleculares; Esolv = energia de solvatação; EHb = energia de ligação de hidrogênio; Etot = energia potencial total, corresponde à soma de todas as contribuições de energia (kcal/mol).

ANEXO B

International Journal of QUANTUM CHEMISTRY

A Novel Set of β-*N*-biaryl Ether Sulfonamide Hydroxamates as Potential MMPs Inhibitors: Molecular Dynamics Simulations and Molecular Properties Evaluation

Kely Medeiros Turra,[a] Kerly Fernanda Mesquita Pasqualoto,*[b] and Silvia Berlanga de Moraes Barros[a]

Matrix metalloproteinases (MMPs) constitute a family of zinc-dependent proteases involved in the extracellular matrix degradation. MMP-2 and MMP–9 are overexpressed in several human cancer types, including melanoma, thus the development of new compounds to inhibit MMPs' activity is desirable. Molecular dynamic simulation and molecular properties calculations were performed on a set of novel β-*N*-biaryl ether sulfonamide-based hydroxamates, reported as MMP-2 and MMP-9 inhibitors, for providing data to develop an exploratory analysis. Thermodynamic, electronic, and steric descriptors have significantly discriminated highly active from moderately and less active inhibitors of MMP-2 whereas apparent partition coefficient at pH 1.5 was also significant for the MMP-9 data set. Compound 47 was considered an outlier in all analysis, indicating the presence of a bulky substituent group in R_3 is crucial to this set of inhibitors for the establishment of molecular interactions with the S1 subsite of both enzymes, but there is a limit. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/qua.24222

Introduction

Matrix metalloproteinases (MMPs) are a family of zinc-dependent proteases involved in tissue remodeling and degradation of the extracellular matrix, 26 distinct kinds of MMPs have been discussed for humans. The up-regulation of MMP-2 and MMP-9, for instance, has been implicated in several types of human cancer such as breast and lung cancer, hepatocarcinoma and, notably, in melanoma.[1] Therefore, the development of new potential compounds to inhibit the activity of those enzymes is desirable.

The strategy for designing MMP inhibitors is generally based upon two key features: a functional group able to chelate the active zinc site, which is known as zinc binding group, and at least one functional group able to interact with the enzyme backbone by establishing a hydrogen bonding.[2] Furthermore, substituent groups that provide additional interactions with adjacent subsites to the catalytic site can also improve the inhibitors potency. Recently, a novel set of β-*N*-biaryl ether sulfonamide hydroxamates was reported as potential MMP inhibitors.[3,4] These compounds exhibited a significant inhibitory activity against both proteases, MMP-2 and MMP-9, at nanomolar concentration. Distinct substituent groups were added at three different positions (R_1, R_2, and R_3), as showed in Figure 1, to allow additional interactions with subsites that have been previously neglected.

The application of quantitative structure-activity relationship (QSAR) formalism as well as investigation of the binding mode of those compounds could generate useful findings for the prediction and development of promising new inhibitors. Initially, chemometric methods such as hierarchical cluster analysis (HCA) and principal component analysis (PCA) can be used as initial tools to investigate this set of compounds and provide insights regarding their structure-activity relationships in a qualitative fashion. This methodology is suitable particularly when the investigated set is depending upon many variables.[5–7]

HCA is a multivariate method for calculating and comparing the distances between pairs of samples or variables, and it groups data into clusters having similar attributes. The findings are displayed as a dendrogram. The primary purpose of the method is to present data in a manner that emphasizes natural groupings and patterns. It is considered an excellent tool for a preliminary data analysis to better understand similarities and differences regarding the molecular system being investigated.[6,7]

PCA is a data compression method based upon the correlation among variables or descriptors. Its purpose is to reduce the dimensionality of a data set by grouping the correlated variables and replacing the original descriptors with new set of descriptors called the principal components (PCs), onto which the data is projected. PCs are completely uncorrelated variables built up as simple linear combinations of the original variables, containing most of the variability in the initial data set even though it is in a much lower dimensional space.[6,7] Thus, PCA and HCA can be used as complementary techniques for efficiently exploring a data set.

[a] *K. M. Turra, S. B. M. Barros*
Laboratory of Pathology, Department of Clinical Chemistry and Toxicology, Faculty of Pharmaceutical Sciences, University of Sao Paulo, Sao Paulo, SP 05508-900, Brazil

[b] *K. F. M. Pasqualoto*
Laboratory of Biochemistry and Biophysics, Butantan Institute, Sao Paulo, SP 05503-900, Brazil
E-mail: kely.turra@gmail.com

© 2012 Wiley Periodicals, Inc.

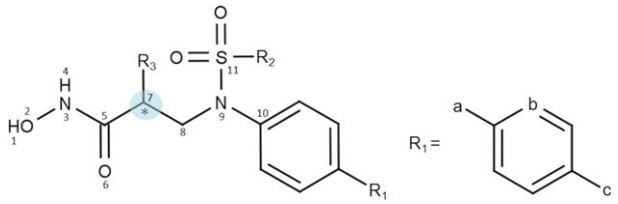


Figure 1. General chemical structure of the *β-N*-biaryl ether sulfonamide hydroxamate derivatives (Yang et al., 2008). The carbon at the 7-position (blue circle) is a chiral center (*) and can confer *S* or *R* configuration to the molecule. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

In this study, an exploratory data analysis of a set of novel *β-N*-biaryl ether sulfonamide-based hydroxamates,[3,4] formerly reported as potential MMP-2 and MMP-9 inhibitors, was performed to select the most relevant descriptors or molecular properties (thermodynamics, structural, electronic, lipophilic, and steric) and further establish correlations with the biological data.

Methods

Biological data set

A data set of 67 *β-N*-biaryl ether sulfonamide hydroxamate derivatives was selected from Yang et al.[3,4] (Fig. 1 and Table 1). Biological activity was calculated based upon the inhibitor concentration capable of reducing the MMP-2 and MMP-9 binding capacity by half (IC_{50}, nM). The experimental inhibitory activities (the dependent variables) were converted into their corresponding negative logarithmic units, pIC_{50} (-log IC_{50}) (see Table 1), and the values ranged from 5.53 to 8.60 and 5.32 to 9.34 for MMP-2 and MMP-9, respectively.

Building up the three-dimensional molecular models

The three-dimensional (3D) representations of the 67 inhibitors, in their neutral forms, were constructed using the HyperChem 8.0 software.[8] A solution structure of a hydroxamic acid inhibitor in complex with a catalytic domain of MMP-2[9] (PDB code 1HOV, solution NMR) was retrieved from the Protein Data Bank[10] and employed as a geometry reference for constructing all ligands. Furthermore, the chirality of the α-carbon center (at 7-position, Fig. 1) was assigned as shown in Table 1. The energy minimization of each ligand was carried out using the MM+ force field, without any constrains, and partial atomic charges were calculated using the AM1[11] semiempirical method, implemented in the HyperChem 8.0 program.[8] The MOLSIM 3.2 program[12] was also used for the geometry optimization procedure applying the steepest descent and, subsequently, conjugated gradient method with the convergence criterion of 0.01 kcal/mol. The energy-minimized structures were used as input for the molecular dynamics (MD) simulations performed with the MOLSIM 3.2 software.

MD simulation and descriptors calculation/selection

MD simulations were carried out to generate the conformational ensemble profile (CEP) of each inhibitor. The MD simulation protocol for each inhibitor was the following: 1 ns [1,000,000 steps, 1 fs step size] at 298 K, the same temperature of the biological assay.[3,4] Output trajectory files were saved at every 20 simulation steps, resulting in 50,000 conformations for each inhibitor. The lowest-energy conformation was selected from CEP, and the hydration shell model proposed by Forsythe and Hopfinger[13] was used to estimate the solvation energy contribution. The hydrogen bonding intramolecular energy contribution was also computed for the lowest-energy conformation of each ligand. At this point, thermodynamic properties were obtained for the entire data set and the selected conformers were energy-minimized before calculating the molecular properties (independent variables or descriptors) using the appropriate software package. Descriptors related to electronic, hydrophobic, steric, and structural contributions were calculated employing the following softwares: Gaussian 03W,[14] Marvin 5.8.0,[15] Hyperchem 8.0.[8] Detailed information regarding all descriptors, methods, and respective software used to perform these calculations are listed in Table 2.

After descriptors calculations, two data matrices for each target enzyme and its inhibition values (pIC_{50}) were constructed. Matrix **X1** included MMP-2 inhibition values as dependent variable, and ligands were classified as less/moderately (5.53–7.56) and highly active inhibitors (7.62–8.60). Matrix **X2** was composed of MMP-9 inhibition values and the low/moderate activity range was from 5.32 to 8.40 while the highly active inhibitors have pIC_{50} higher than 8.40 (8.42–9.34). A filter using the Pearson correlation coefficient value of 0.20 was formerly used as cutoff for selecting the calculated independent variables. In addition, a visual inspection of scatter plots for the biological data (pIC_{50} values) versus each calculated descriptors was performed (see chart 1 and 2 of the Supporting Information section), and only variables presenting a linear correlation tendency were selected for composing the final matrices that are the initial data sets for exploratory analysis (HCA and PCA). The final matrices, **X1** (MMP-2 data set) and **X2** (MMP-9 data set), exhibited the same number of rows (67 compounds or samples) but different number of columns (variables or descriptors). Matrix **X1** was composed of nine columns (eight descriptors selected plus the biological activity) whereas **X2** consisted of 12 columns (11 descriptors plus the biological activity).

Exploratory data analysis →PCA and HCA methods

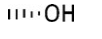
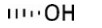
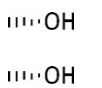
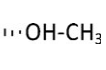
The exploratory analysis (HCA and PCA) was carried out using the Pirouette 3.11 software.[35] The autoscaling procedure was applied as a preprocessing method due to the distinct magnitude orders among the calculated variables.[36]

The complete linkage method and Euclidean distance were considered in HCA. The distances between samples or variables were calculated according the Eq. (1).

$$d_{kl} = \left[\sum_{i}^{m} \left(x_{kj} - x_{lj}\right)^{M}\right]^{1/M} \quad (1)$$

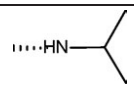
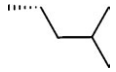
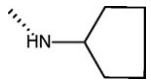
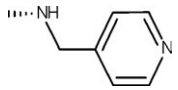
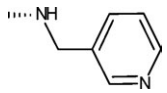
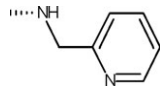
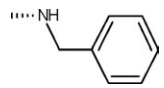
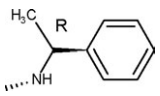
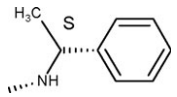
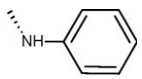
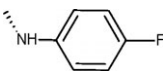
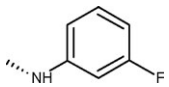
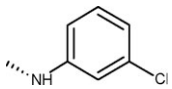
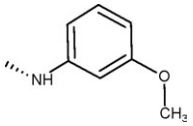
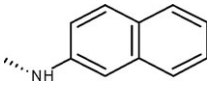
The multivariate distance d_{kl} between two samples vectors, *k* and *l*, is determined by computing differences at each of the m variables. *M* is the order of the distance, and here corresponds to the Euclidean distance ($M = 2$). The distance values were transformed into a similarity matrix whose elements correspond to the similarity indices [$similarity_{kl} = 1 - d_{kl}/d_{max}$ where d_{max} is the largest distance in the data set]. The

Table 1. Chemical notation and biological activity of the investigated inhibitors.

Compounds	C_7***R***/***S***[a]	pIC_{50} MMP-2	pIC_{50} MMP-9	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
1	–	7.21	8.18	O	CH	Me	Me	–
2	*S*	7.21	8.11	O	CH	Me	Me	
3	*R*	7.17	7.92	O	CH	Me	Me	
4	*S*	6.85	7.55	O	CH	Me	Et	
5	*S*	6.29	6.91	O	CH	Me	*n*-Pr	
6	*S*	5.91	6.54	O	CH	Me	*i*-Pr	
7	*S*	5.62	6.46	O	CH	Me	Ph	
8	*S*	6.23	7.03	O	CH	Me	Bn	
9	*S*	6.40	7.16	O	CH	Me	(pyridin-3-yl)CH2	
10	*S*	6.46	7.27	O	CH	Me	((pyridin-4-l)CH2	
11	*S*	6.61	7.41	O	CH	H	Me	
12	*S*	7.45	8.33	O	CH	Cl	Me	
13	*S*	6.86	7.85	O	CH	CF3	Me	
14	*S*	5.60	6.33	O	N	CF3	Me	
15	*S*	6.53	7.27	S	CH	Me	Me	
16	*S*	6.55	6.98	S	CH	F	Me	
17	*S*	7.33	7.36	C2	CH	Me	Me	
18	*S*	7.25	7.33	C2	CH	Cl	Me	
19	*S*	7.49	7.04	C2	CH	OMe	Me	
20	*S*	5.63	5.32	OCH2	CH	F	Me	
21	*S*	6.05	6.01	OCH2	CH	Cl	Me	
22	*S*	5.53	5.83	OCH2	CH	CF3	Me	
23	*S*	7.21	8.11	O	CH	Me	Me	
24	*S*	7.45	8.33	O	CH	Cl	Me	
25	*S*	6.86	7.85	O	CH	CF3	Me	
26	*R*	6.99	7.89	O	CH	Me	Me	
27	*S*	7.44	8.42	O	CH	Me	Me	
28	*S*	8.04	8.62	O	CH	CF3	Me	
29	*S*	8.40	9.28	O	CH	Me	Me	

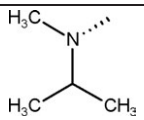
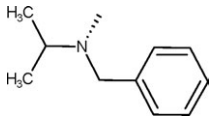
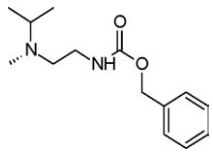
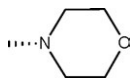
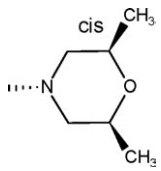
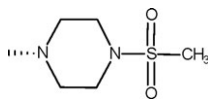
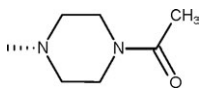
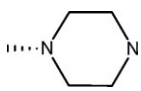
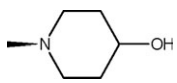
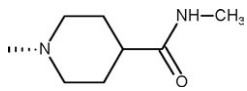
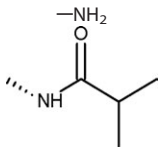
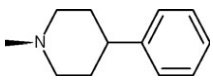
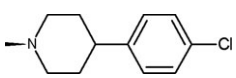
(Continued)

Table 1 (Continued)

Compounds	C_7*R*/*S*[a]	pIC_{50} MMP-2	pIC_{50} MMP-9	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
30	*R*	6.75	7.51	O	CH	Me	Me	
31	*S*	7.96	8.89	O	CH	Me	Me	
32	*S*	8	8.77	O	CH	Me	Me	
33	*S*	8.02	8.74	O	CH	Me	Me	
34	*S*	7.77	8.55	O	CH	Me	Me	
35	*S*	7.08	7.64	O	CH	Me	Me	
36	*S*	6.69	8.15	O	CH	Me	Me	
37	*S*	7.62	8.82	O	CH	Me	Me	
38	*S*	7.92	9.27	O	CH	Me	Me	
39	*S*	7.89	9.28	O	CH	Me	Me	
40	*S*	7.19	8.89	O	CH	Me	Me	
41	*S*	7.27	9.05	O	CH	Me	Me	
42	*S*	6.48	7.70	O	CH	Me	Me	
43	*S*	7.57	9.21	O	CH	Me	Me	
44	*S*	5.85	7.06	O	CH	Me	Me	

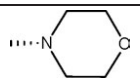
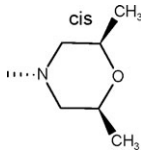
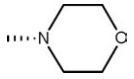
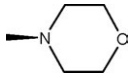
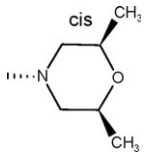
(Continued)

Table 1 (Continued)

Compounds	C_7*R*/*S*[a]	pIC_{50} MMP-2	pIC_{50} MMP-9	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
45	*S*	7.62	8.05	O	CH	CF3	Me	
46	*S*	6.55	7.17	O	CH	CF3	Me	
47	*S*	6.58	7.54	O	CH	CF3	Me	
48	*S*	8.03	8.54	O	CH	CF3	Me	
49	*S*	8.25	8.96	O	CH	CF3	Me	
50	*S*	8.00	8.49	O	CH	CF3	Me	
51	*S*	8.06	8.32	O	CH	CF3	Me	
52	*S*	8.12	8.64	O	CH	CF3	Me	
53	*S*	8.24	8.72	O	CH	CF3	Me	
54	*S*	8.26	8.60	O	CH	CF3	Me	
55	*S*	6.17	6.92	O	CH	Me	Me	
56	*S*	7.15	8.27	O	CH	Me	Me	
57	*S*	7.33	7.77	O	CH	CF3	Me	
58	*S*	6.92	6.59	O	CH	CF3	Me	

(Continued)

Table 1 (Continued)

Compounds	C_7*R*/*S*[a]	pIC_{50} MMP-2	pIC_{50} MMP-9	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
59	*S*	8.26	8.70	O	CH	Cl	Me	
60	*S*	8.60	9.34	O	CH	Cl	Me	
61	*S*	8.28	8.72	O	CH	Me	Me	
62	*R*	5.88	6.53	O	CH	Me	Me	
63	*S*	8.55	9.17	O	CH	Me	Me	
64	*S*	5.90	6.03	O	CH	Cl	–	—OH
65	*S*	6.54	6.62	O	CH	Cl	–	—OH
66	–	6.83	6.42	O	CH	H	Me	–
67	–	7.14	7.51	O	CH	Me	Me	–

[a] Chirality *R*/*S* related to α-carbon (7-position). Compounds **66** and **67** do not present the chiral carbon atom at 7-position.

similarity scale ranges from zero (dissimilar samples or variables) to one (identical samples or variables), and the larger the similarity index the smaller the distance between any pair of samples or variables.[5–7,35,36] Results were expressed as a dendrogram, which is a tree-shaped map constructed from the distances data.

Most chemical applications of data analysis are by nature multivariate and for such cases PCA is a suitable method. The PCs contain most of the variability in the data set, albeit in a much lower dimensional space. PC1, or factor 1, is defined in the direction of maximum variance of the whole data set whereas PC2 or factor 2 is the direction that describes the maximum variance in the orthogonal subspace to PC1. The subsequent components are taken orthogonal to those previously chosen and describe the maximum of the remaining variance. Once the redundancy is removed, only the first few PCs are required to describe most of the information contained in the original data set. The data matrix $\mathbf{X}(I \times J)$ corresponding to *I* molecules and *J* descriptors, is decomposed into two matrices, **T** and **L**, such that $\mathbf{X} = \mathbf{TL}^{T}$. The **T** matrix, known as the score matrix, represents the positions (classification) of the compounds in the new coordinate system where the PCs are the axes. Scores are integral to exploratory analysis because they show intersample relationships. In cases of a single category classification, for instance, the scores should not cluster strongly, but if the ultimate goal is multicategory classification, sample groupings corresponding to known categories suggest that a good classification model can be constructed. **L** is the loadings' matrix whose columns describe how the new axes (the PCs) are built from the old axes and indicates the variables importance, meaning which variables contribute more or less to each PC.[5–7,35,36] In this exploratory data analysis, PCA included up to seven factors or PCs. The Pirouette 3.11 software[35] displays many entities that can help to explore the relationships between samples, find samples outliers, choose the optimal number of factors and help make decisions about excluding variables. These entities were considered for investigating the two data sets in the present study. The outliers' diagnosis, for instance, can be made through the Mahalanobis distance,[35,37] which is a distance computed from its *k* factor (PC) score:

$$MD_i = (t_i - \bar{t})^T S_k^{-1} (ti - \bar{t}) \quad (2)$$

In Eq. (2), **S** is the scores covariance matrix and $\bar{t}$ is the mean score vector. Assuming that Mahalanobis distance is normally distributed, and a critical value (MD_{crit}) can be determined from a chi squared distribution with k degrees of freedom. If a sample's Mahalanobis distance exceeds MD_{crit} that sample may be an outlier.

Results and Discussion

The MD simulation findings regarding an equilibration region from part of the CEP of each ligand (100 ps of 1 ns) are presented in chart 3 of the Supporting Information section. The energy values did not include the hydrogen bonding and solvation energy contributions that were computed computed only for the lowest-energy conformation of each inhibitor selected from those CEP regions. The total potential energy

Table 2. Descriptors, methods and software used to perform calculations for the set of *β-N-* biaryl ether sulfonamide hydroxamate inhibitors.

Descriptors	Description	Nature	Methods	Software
$E_{stretch}$	Stretching energy contribution	Thermodynamic	MD simulations (see Method section)	*Molsim* 3.2
E_{bend}	Bending energy contribution			
E_{tors}	Torsional energy contribution			
$E_{1,4}$	Lennard-Jones energy contribution			
E_{vdW}	van der Waals energy contribution			
E_{charge}	Electrostatic energy contribution			
E_{solv}	Solvation energy contribution			
E_{Hb}	Hydrogen bonding energy contribution			
E_T	Total energy (sum of all energy contributions)			
$ChelpG_1$	H_1 electrostatic potential charge	Electronic	HF/6-31G (d,p) [16,17]	*Gaussian* 03W for Windows
$ChelpG_2$	O_2 electrostatic potential charge			
$ChelpG_3$	N_3 electrostatic potential charge			
$ChelpG_4$	H_4 electrostatic potential charge			
$ChelpG_5$	C_5 electrostatic potential charge			
$ChelpG_6$	O_6 electrostatic potential charge			
$ChelpG_7$	C_7 electrostatic potential charge			
$ChelpG_8$	C_8 electrostatic potential charge			
$ChelpG_9$	N_9 electrostatic potential charge			
$ChelpG_{10}$	C_{10} electrostatic potential charge			
$ChelpG_{11}$	S_{11} electrostatic potential charge			
$ChelpG_{12}$	O_{12} electrostatic potential charge			
$ChelpG_{13}$	O_{13} electrostatic potential charge			
$ChelpG_{14}$	Electrostatic potential charge of the first atom linked to S_{11} (R_2)			
E_{lumo}	Energy of the lowest unoccupied molecular orbital (LUMO)			
E_{homo}	Energy of the highest occupied molecular orbital (HOMO)			
μ_x	*X* component of dipole moment			
μ_y	*Y* component of dipole moment			
μ_z	*Z* component of dipole moment			
μ_{total}	Total dipole moment			
Platt	Platt index	Topologic	[18,19]	Marvin 5.8.0
Randic	Randic index		[20]	
Balaban	Balaban index		[21]	
Harary	Harary index		[22]	
Hyper Wiener	HyperWiener index		[20,23]	
Szeged	Szeged index		[24]	
Wiener	Wiener index		[25]	
Wiener polarity	Wiener polarity index		[25,26]	
ClogD_1.5	Apparent partition coefficient pH 1.5	Hydrofobic	[27]	
ClogD_5.0	Apparent partition coefficient pH 5.0			
ClogD_6.5	Apparent partition coefficient pH 6.5			
ClogD_7.4	Apparent partition coefficient pH 7.4			
ClogP	Calculated partition coefficient (water/n-octanol)		[28]	Hyperchem 8.0
MSA	Molecular surface area	Steric	[29]	Marvin 5.8.0
PSA	Polar surface area		[30]	
V_{vdW}	van der Waals volume		[31,32]	
V	Total volume (intrinsic property)		[33]	Hyperchem 8.0
MR	Molar refractivity	Steric/hydrofobic	[28]	
A	Polarizability	Steric/electronic	[34]	
A_1	O_2—N_3—C_5 angle	Structural		
a_2	N_3—C_5—C_7 angle			
a_3	O_6—C_5—C_7 angle			

(E_T) value (kcal/mol) found for each lowest-energy conformation is showed in Figure 2 and corresponds to the summation of the following intramolecular energy contributions: stretching ($E_{stretch}$), bending (E_{bend}), torsional (E_{tors}), Lennard-Jones or 1-4 interactions (E_{1-4}), electrostatic (E_{charge}), van der Waals (E_{vdW}), hydrogen bonding (E_{Hb}), and solvation (E_{solv}). These calculated thermodynamic properties are also displayed in table III of the Supporting Information section. The E_T values were mostly zero kcal/mol except for eight inhibitors (**6**, **47**, **54**, **57**, **64**–**67**) (see Fig. 2). Regarding the highly active compounds (inhibitors of MMP-2 and MMP-9), only the lowest-energy conformation of compound **54** presented a more positive E_T value (8.96 kcal/mol).

The conformers selected from MD simulations, after energy-minimized, were used as input to calculate other descriptors or molecular properties of distinct nature, as electronic, hydrophobic, steric, topological, and structural. All 51 calculated descriptors are listed in table III (Supporting Information section).

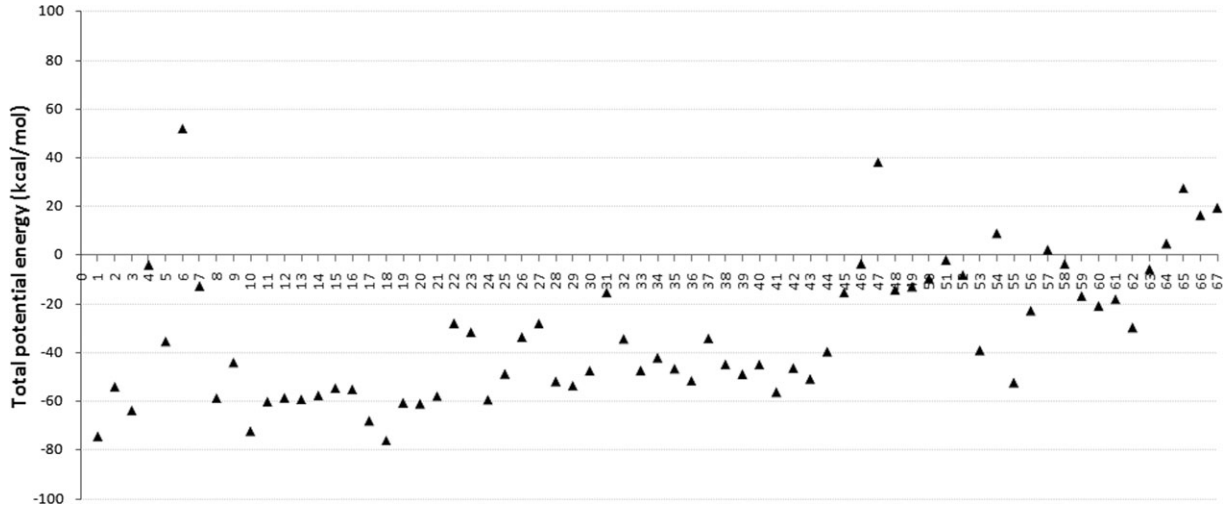


Figure 2. Total potential energy (E_T) value found for the lowest-energy conformation of each inhibitor (67 compounds).

Two filters were applied to select the calculated independent variables or descriptors: (i) Pearson linear correlation coefficient between descriptors and inhibitory activities (cutoff value = 0.2); and (ii) a visual inspection of data sets through scatter plots (inhibitory activities versus each calculated descriptor). These findings are presented in chart 1 and 2 (Supporting Information section). Although the descriptors volume (V), Van der Waals volume (V_{vdW}), molar refractivity (MR), and

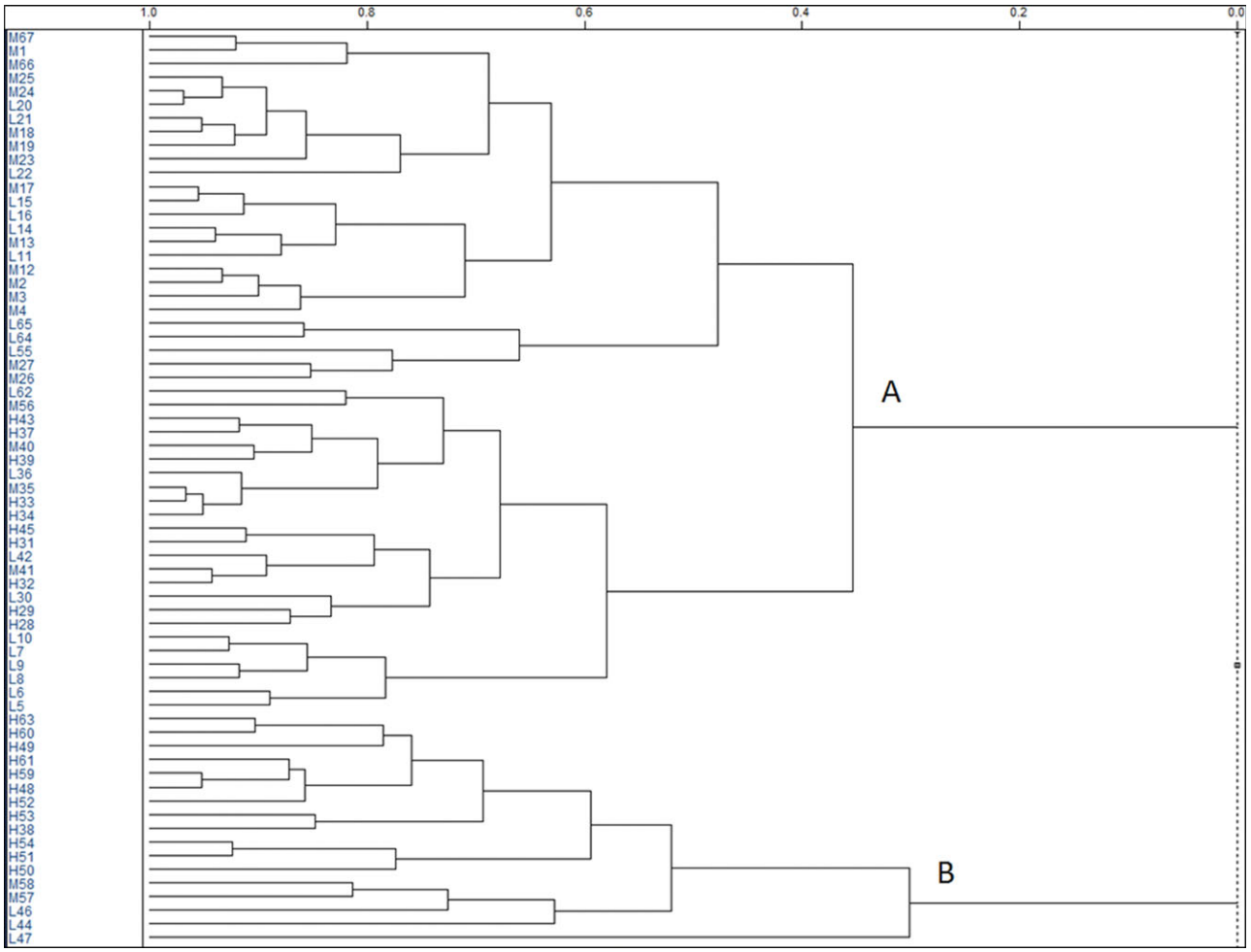


Figure 3. Dendrogram of samples obtained for matrix **X1** (MMP-2 data set). The samples or compounds were classified in two main groups (**A** and **B**). Compounds are indicated by the capital letters H, M, and L, which stand for highly, moderately, and less active inhibitors, respectively.

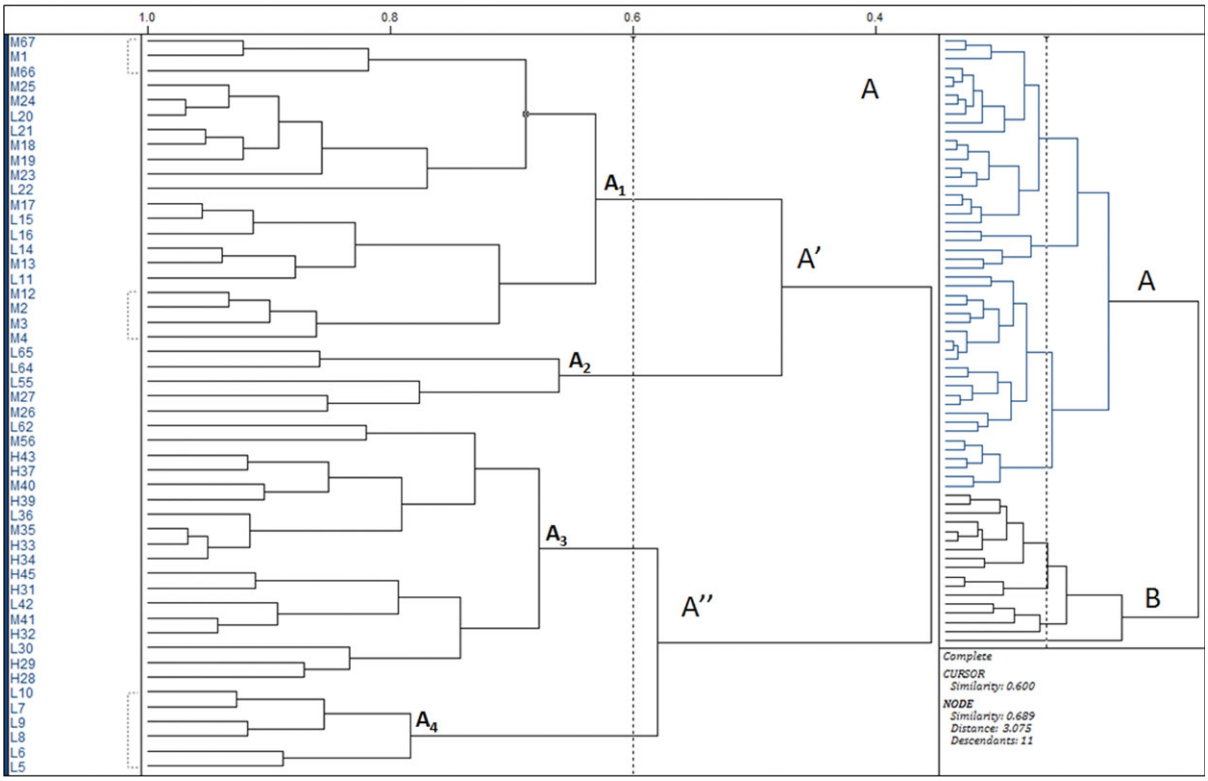


Figure 4. Cluster A from dendrogram of matrix **X1**: two main subgroups can be visualized (A′ and A″). There is mostly a mix of compounds presenting low and moderately activity. Considering a similarity cursor of 0.6 (dashed line), cluster A can be divided in four subclusters: A_1, A_2, A_3, and A_4. In subcluster A_1, two groups only constituted of moderately active compounds were detected (gray dashed line). A_4 is composed by samples with low inhibitory activity. Samples are indicated by the capital letters H, M, and L, which stand for highly, moderately and less active compounds, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

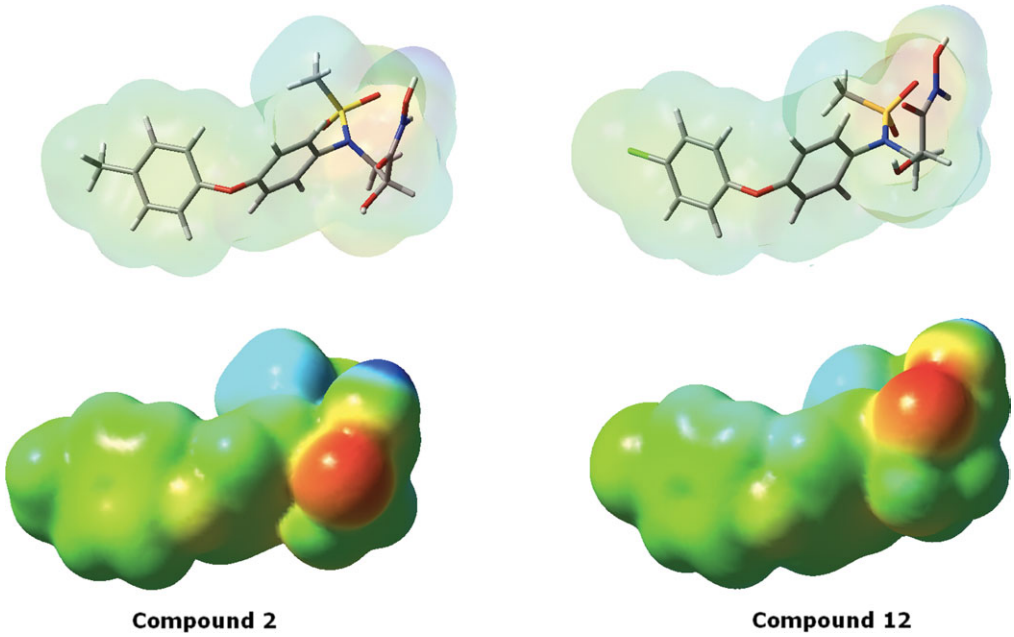


Figure 5. Maps of electrostatic potential (MEPs) for compounds 2 (R_{1c} = 4-CH3; yellow to green color on aromatic ring, regarding transparent surface) and 12 (R_{1c} = 4-Cl; yellow to green color more extended on the R_{1c} group, observing transparent surface). Red color (−0.9) corresponds to regions of higher electronic density distribution whereas intense blue color (0.9) indicates a lower electronic density distribution. The molecules are presented as stick model where the carbon atoms are in gray, oxygen in red, nitrogen in blue, sulphur in yellow, and hydrogen atoms in white. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

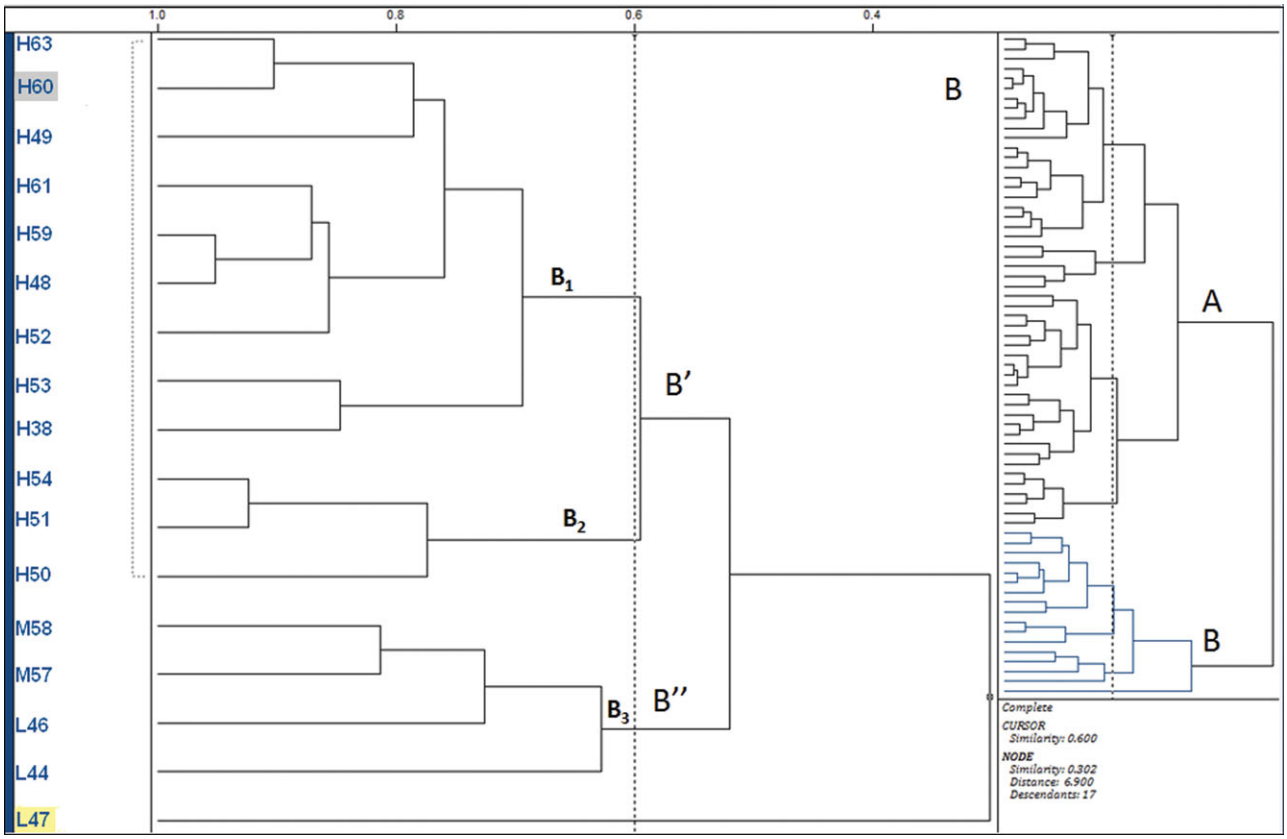


Figure 6. Cluster B from matrix **X1** dendrogram: two main subgroups can be visualized (B′ and B″), and cluster B′ presents the highly active compounds (gray dashed line). Considering a similarity index of 0.6 (cursor, dashed line), cluster B can be divided in three subclusters: B_1, B_2, B_3. The most active inhibitor (H**60**) is in subcluster B_1. Samples are indicated by the capital letters H, M, and L, which stand for highly, moderately and less active compounds, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

polarizablity (α) did not present an explicit linear tendency (but they do present a quite good dispersion; see chart 1 and 2 at Supporting Information section), they were maintained in the pool of selected variables because they showed an important role in inhibitor classification during this preliminary investigation.

The final matrices, **X1** (MMP-2 data set) and **X2** (MMP-9 data set), presented 67 rows, which correspond to the number of inhibitors. Matrix **X1** was composed by nine columns (eight descriptors selected plus biological activity), and the descriptors selected regarding the criteria already mentioned were: E_{Hb} (x_3), E_{solv} (x_4), μ_y(x_{13}), V_{vdW} (x_{14}), V (x_{15}), MR (x_{16}), α (x_{17})and a_3 (x_{26}) (see table II). Matrix **X2** contained 12 columns (11 descriptors plus biological activity), and the selected molecular properties were: E_{vdW} (x_3), E_{solv} (x_5), $ChelpG_{13}$ (x_8), μ_y (x_{10}), V_{vdW} (x_{11}), V (x_{12}), MR (x_{13}), α (x_{14}), MSA (x_{15}), ClogD_1.5 (x_{16}), and a_3(x_{23}) (see Table 2). The exploratory analysis was applied to both final matrices, **X1** and **X2**, and the findings are discussed as follows.

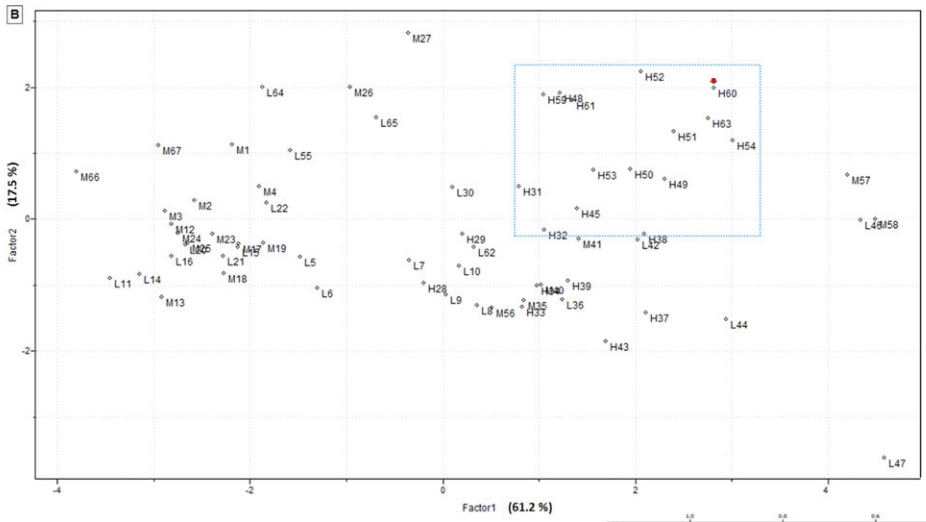
HCA and PCA for matrix X1 (MMP-2 data set)

HCA findings are displayed as a dendrogram (see Figs. 3, 4, 6), constructed from the Euclidean distances that were transformed into similarity indices. The interpretation is quite simple: larger similarity index values are related to smaller distances between any pairs of samples or variables. The larger the similarity index the smaller the distance between any pair of samples or variables. The β-*N*-biaryl ether sulfonamide hydroxamates (samples) were grouped in two main clusters (A and B), as shown in Figure 3, according to the molecular properties or descriptors selected for the MMP-2 data set. Cluster A and B present a similarity index of 0.35 and 0.30, respectively.

As shown in Figure 4, cluster A has two subclusters (A′ and A″), which share a similarity index of 0.35. Subcluster A′ (similarity index = 0.48) comprises exclusively less and moderately active compounds whereas A″ (similarity index = 0.58) presents a collection of highly (**25**, **28**, **29**, **31**, **32**, **33**, **34**, **37**, **39**, **43**, **45**), moderately and less active inhibitors.

Considering the dotted vertical line which slices the dendrogram at a similarity value of 0.6 (similarity cursor = 0.6) (see Fig. 4), cluster A can be divided in four clusters: A_1 (similarity index = 0.63), A_2 (similarity index = 0.66), A_3 (similarity index = 0.68), and A_4 (similarity index = 0.78).

Subcluster A_1 presented two groupings that only consist of moderately active compounds (gray dashed line in Fig. 4). The first moderately active grouping is composed of compounds **1**, **66** and **67**. Compound **1** is not substituted at R_3 position (R_3 = H), and because of that the carbon at 7-position is not a chiral center. Compounds **66** and **67** do not have a carbon atom at 7-position and, consequently, they do not have a R_3



A

	Variance	Percent	Cumulative
Factor 1	323.2	61.2	61.2
Factor 2	92.6	17.5	78.8
Factor 3	50.6	9.6	88.3
Factor 4	34.5	6.5	94.9
Factor 5	25.0	4.7	99.6
Factor 6	1.7	0.3	99.9
Factor 7	0.3	0.1	100.0

C

Loadings

Descriptors	Factor 1	Factor 2
X3	0.19	0.66
X4	0.32	0.13
X13	0.17	0.62
X14	0.43	-0.15
X15	0.44	-0.17
X16	0.43	-0.22
X17	0.43	-0.20
X26	0.28	0.15

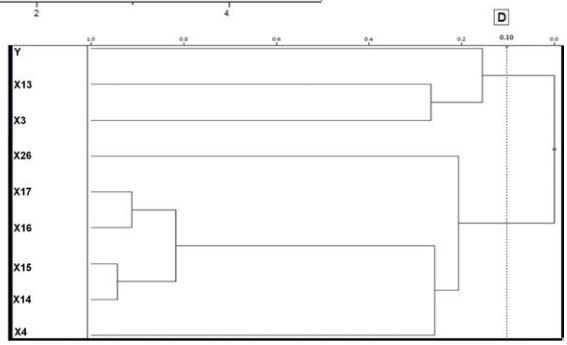


Figure 7. PCA findings for the MMP-2 data set (matrix **X1**). (A) Scores plot of PC1 (factor 1) versus PC2 (factor 2). The highly active compounds found in HCA (cluster B′) are highlighted (light blue rectangle; a red arrow indicates the most active inhibitor). In general, highly (H) active inhibitors are most commonly found on the right side of PC1 (positive values) whereas the less (L) and moderately (M) active compounds on the left side (negative values). (B) Variance, percent, and cumulative found for the seven factors or PCs. (C) Loadings for the original variables or descriptors [E_{Hb} (x_3), E_{solv} (x_4), μ_y (x_{13}), V_{vdW} (x_{14}), V (x_{15}), MR (x_{16}), α (x_{17}), a_3 (x_{26})]in PC1 and PC2. (D) HCA of variables for matrix **X1**. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

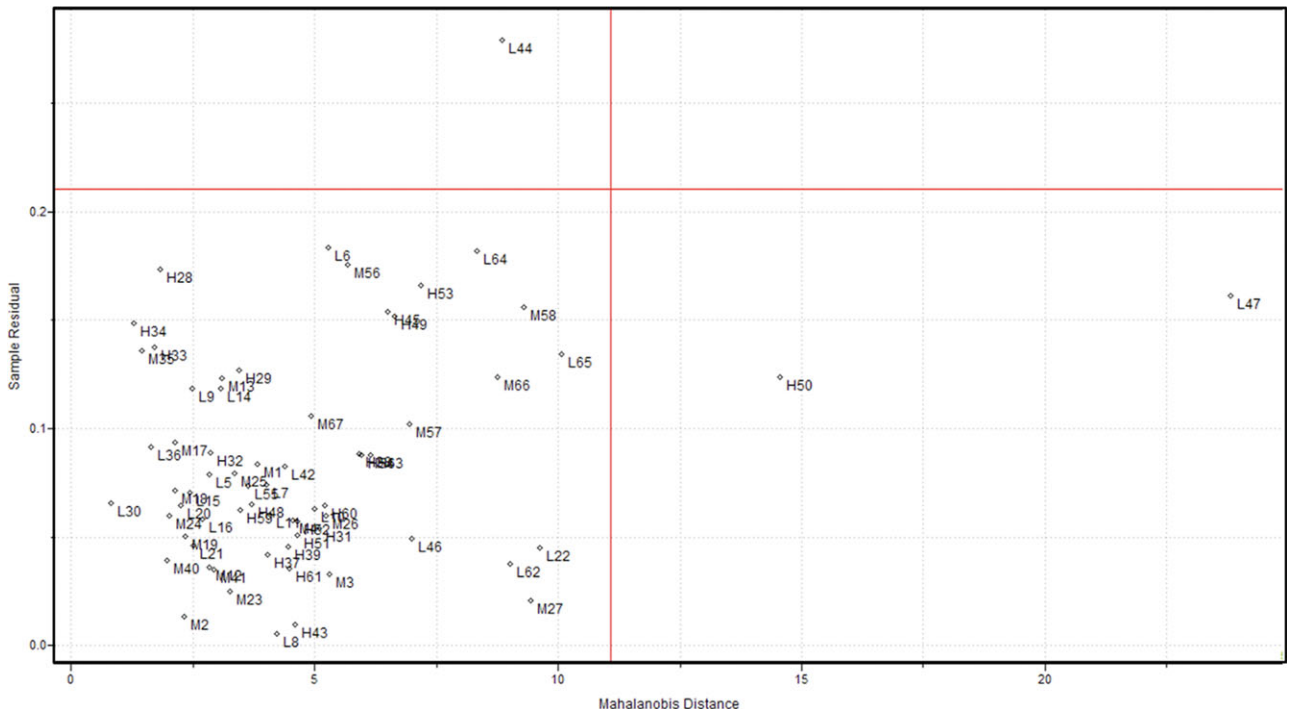


Figure 8. Plot of sample residual versus Mahalanobis distance obtained for matrix **X1**. The sample residual threshold (red line) is based upon a 95% probability limit (set internally in Pirouette). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

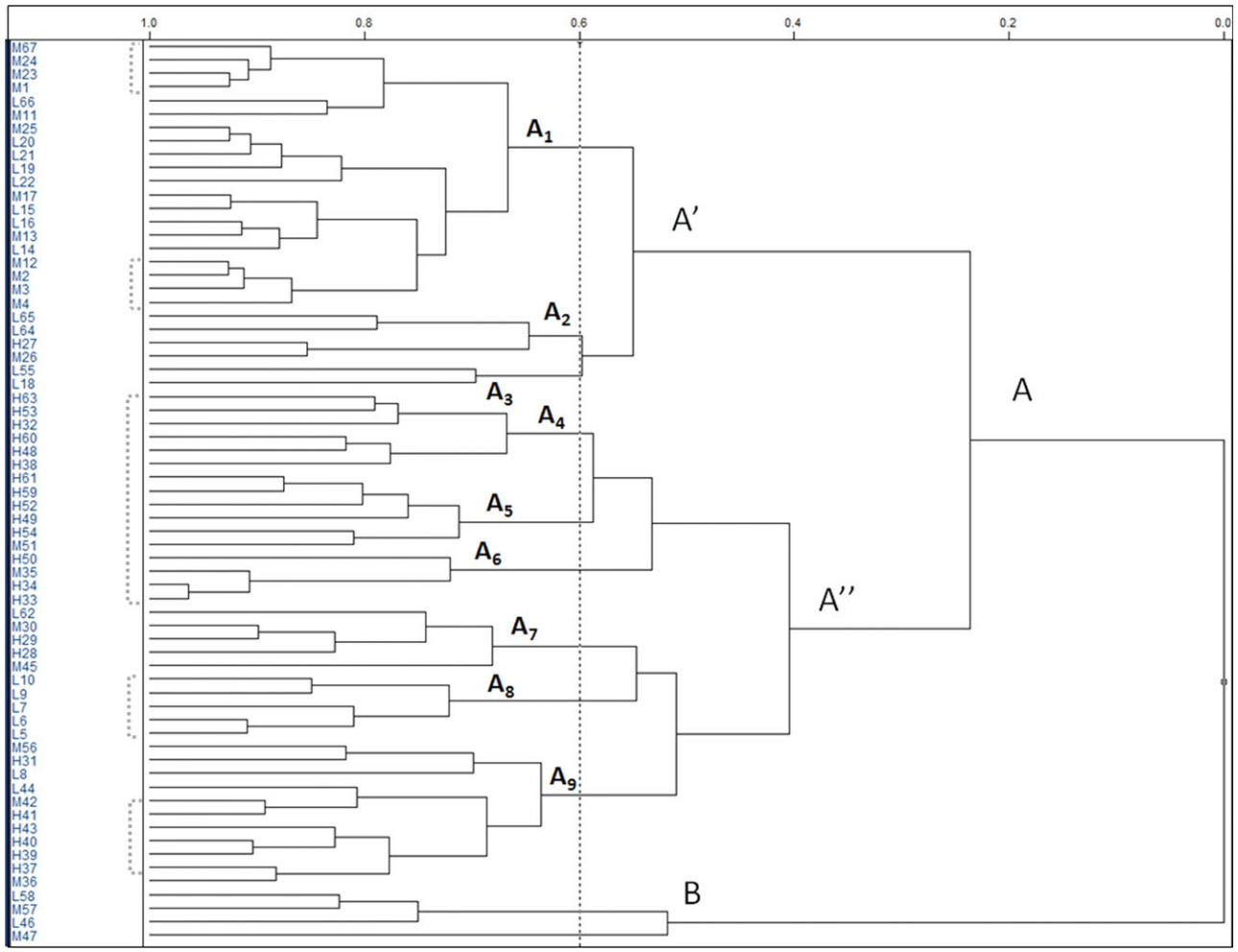


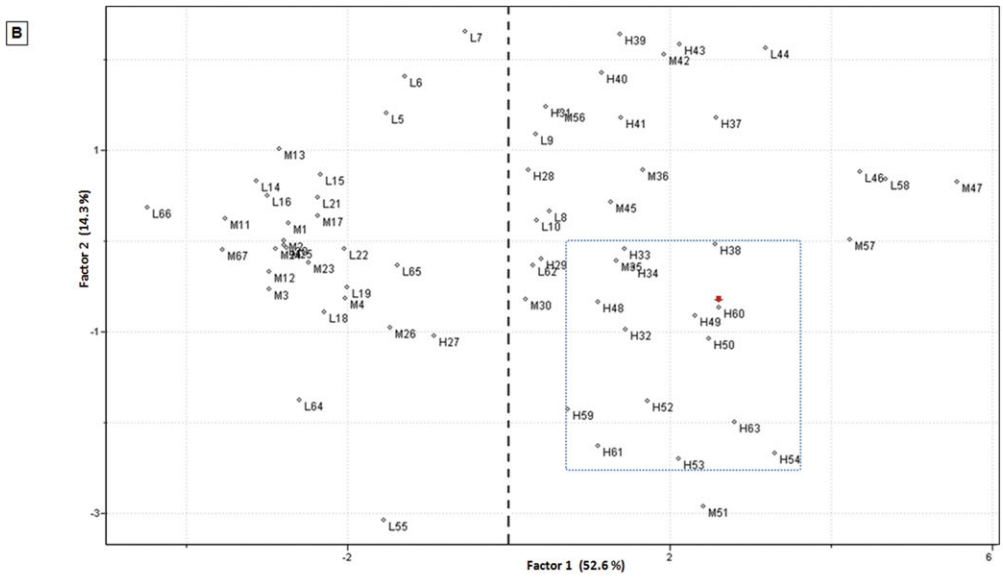
Figure 9. Dendrogram of matrix **X2**: two main clusters can be visualized (A and B). Considering a similarity cursor of 0.6 (dashed line), cluster A can be divided in nine subclusters, $A_{1\text{-}9}$. Subclusters A_4 toA_6 present the major number of highly (H) active compounds. Another grouping is found in subcluster A_9 (gray dashed line). Two groupings only constituted of moderately (M) active compounds were detected in A_1, and A_8 is composed by samples with low (L) inhibitory activity. Cluster B groups less/moderately active compounds. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

substituent group. Compounds **1** and **67** share a similarity index of 0.92 whereas sample **66** presents 82% similarity with both. According to Yang et al.[3,4] derivatives presenting small α-substituent groups in R_3 moiety possibly do not provide a proper interaction with the S1 enzyme subsite. Compounds **1**, **66**, and **67** possibly do not establish any interaction with the S1 subsite since their R_3 group is either too small (H) or does not exist.

Inhibitors **2**, **3**, **4**, and **12** are in the second grouping of moderately active inhibitors. Compounds **2** and **3** are enantiomers, and compound **4** differs from them at the R_2 position; an ethyl instead of methyl group. Compound **12** differs from compounds **2** and **3** in the R_{1c} substituent group; a chlorine instead of methyl. The 4-Cl and 4-CH_3 groups confer a lipophilic character to the molecule since their π Hansch constant value is positive ($\pi_p = 0.71$ and $\pi_p = 0.52$, respectively). Moreover, these substituent groups did not drastically change the molecule's volume (V or V_{vdW}) (see table III of the Supporting Information section). Otherwise, they do present opposite electronic character: 4-Cl is an electron withdrawing whereas 4-CH_3 is an electron-donating group (see Fig. 5). Compounds **2** and **12** share a similarity index of 0.93; compound **3** is 90% similar with both while compound **4** shares a similarity index of 0.86 with all of them (see Fig. 4).

Subcluster A_4 is composed only of compounds with low inhibitory activity (compounds **5** to **10**), as presented in Figure 4. Compounds **5** to **10** differ from each other primarily in the R_2 substituent group. Compounds **7** to **10** have aromatic substituent at R_2 position and share a similarity index of 0.93 and 0.92, respectively; between those pairs the similarity is 86%. Compounds **5** (n-propyl) and **6** (isopropyl) have a similarity index of 0.89.

Cluster B also has two subclusters (B′ and B″), as shown in Figure 6, that share a similarity index of 0.52. Subcluster B′ (similarity index = 0.59) is composed of highly active inhibitors, which present bulky R_3 substituent groups that most likely have the molecular properties to interact with the enzyme S1 subsite.[3,4] Otherwise, subcluster B″ (similarity index = 0.63) is comprised of less (**44** and **46**) and moderately (**57** and **58**) active compounds whose R_3 moieties, that are even bulkier, seem to



A

	Variance	Percent	Cumulative
Factor 1	318.8	52.6	52.6
Factor 2	104.0	14.3	66.9
Factor 3	64.2	8.8	75.8
Factor 4	60.5	8.3	84.1
Factor 5	44.0	6.1	90.1
Factor 6	42.7	5.8	95.9
Factor 7	26.7	3.7	99.6

C

Loadings

Descriptors	Factor1	Factor2
X3	-0.08	-0.41
X5	0.28	0.02
X8	-0.06	0.51
X10	0.14	-0.51
X11	0.41	0.02
X12	0.41	0.04
X13	0.40	0.10
X14	0.41	0.09
X15	0.40	0.01
X16	-0.10	0.53
X23	0.25	0.04

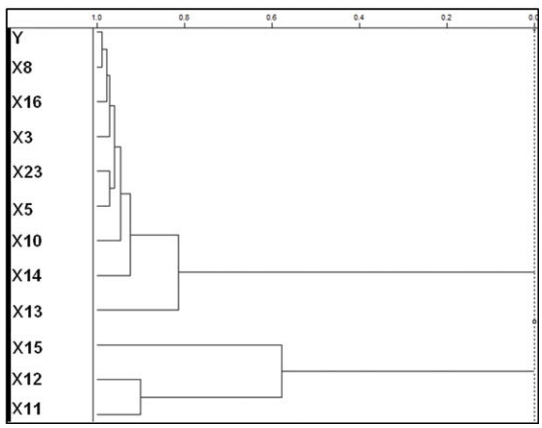


Figure 10. PCA findings for the MMP-9 data set (matrix **X2**). (A) Scores plot of PC1 (factor 1) versus PC2 (factor 2). The highly active compounds found in HCA (subclusters A_4 to A_6) are highlighted (light blue rectangle; a red arrow indicates the most active inhibitor). In general, highly (H) active inhibitors are most commonly found on the right side of PC1 (positive values) whereas the less (L) and moderately (M) active compounds on the left side (negative values) (dashed line). (B) Variance, percent, and cumulative found for the seven factors or PCs. (C) Loadings for the original variables or descriptors [E_{vdW} (x_3), E_{solv} (x_5), ChelpG$_{13}$ (x_8), μ_y (x_{10}), V_{vdW}(x_{11}), V (x_{12}), MR (x_{13}), α (x_{14}), MSA (x_{15}), ClogD_1.5 (x_{16}) and a_3 (x_{23})] in PC1 and PC2. (D) HCA of variables for matrix

contribute to the decreasing of inhibitory activity. Compound **44** has a naphthalene-2-amine ring and **46** a benzyl(propan-2-yl)amine ring as R_3 moiety. Compounds **57** and **58** have a 4-phenylpiperidine ring, unsubstituted and substituted, as R_3 group. Bulky substituents in these moieties seem to decreased potency possibly due to their steric influence.[4]

Considering a similarity cursor of 0.6 (dashed or dotted vertical line through the dendrogram in Figure 6 at a similarity value of 0.6), cluster B can be split into three subclusters: B_1 (similarity index = 0.69), B_2 (similarity index = 0.77), and B_3 (similarity index = 0.63).

The most active inhibitor (compound **60**) is in subcluster B_1 and shares a similarity index of 0.90 with compound **63** whereas compound **49** has 79% similarity with both. Compounds **49**, **60**, and **63** differ from one another in their R_{1c} substituent group, 4-CF3, 4-Cl, and 4-CH3, respectively. The steric, hydrophobic, and electronic character of these substituent groups has been previously discussed and R_3 group is a *cis*-2,6-dimethylmorpholine ring. Compounds **48**, **52**, **59**, and **61** are another grouping of subcluster B_1. Compounds **48** and **59** share a similarity index of 0.95 and compound **61** has 87% similarity with both and these compounds have the same R_3 substituent group (morpholine ring). Compounds **48**, **59**, and **61** differ from one another in their R_{1c} substituent group, 4-CF3, 4-Cl, and 4-CH3, respectively. Compound **52** has a similarity index of 0.86 with the other three compounds and presents 4-CF3 as R_{1c} substituent group and a piperidine ring as R_3 group. Compounds **48**, **49**, **52**, **59**, **60**, **61**, and **63** have a 76% similarity. Compounds **38** and **53** are a distinct group in B_1 (see Fig. 6) and have a similarity index of 0.85. They differ from one another in the substituent groups at R_{1c} (4-CF3 and 4-CH3, respectively) and R_3 moiety (piperidin-4-ol and 1-phenylethan-1-amine ring).

Subcluster B_2 is composed of samples **50**, **51**, and **54**, as presented in Figure 6, and the only difference among them is the R_3 substituent group. Compounds **51** to **54** have a similarity index

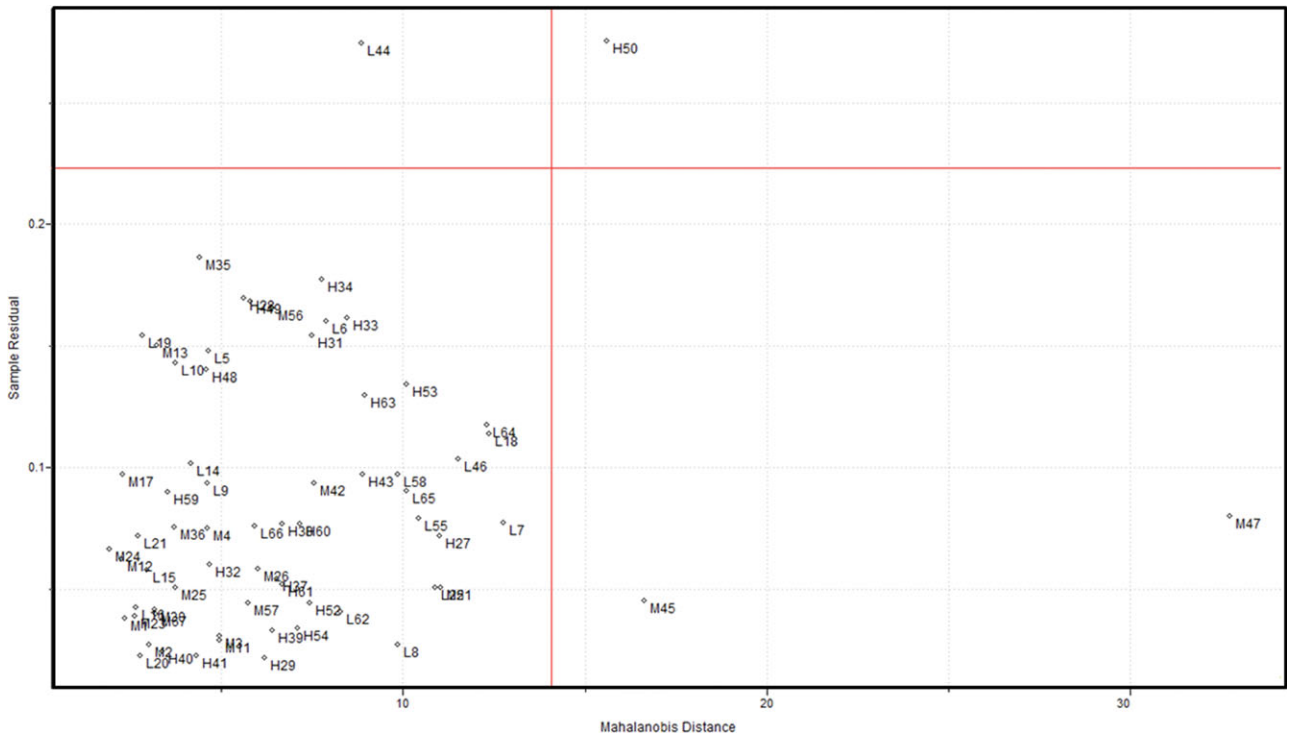


Figure 11. Plot of sample residual versus Mahalanobis distance obtained for matrix **X2**. The sample residual threshold (red line) is based upon a 95% probability limit (set internally in Pirouette). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

of 0.92 whereas compound **50** has 77% similarity with both molecules. The R_3 substituent ring is 1-methanesulfonylpiperazine, 1-(piperazin-1-yl)ethan-1-one, and *N*-methylpiperidine-4-carboxamide for compound **50**, **51**, and **54**, respectively. Subcluster B_2 and B_1 share a similarity index of 0.59.

Compound **47** was not included in any subcluster (see Fig. 6) probably due to the R_3 substituent group [benzyl *N*-[2-propan-2-ylamino)ethyl]carbamate], that is apparently the bulkiest R_3 group of the entire set, reflecting on the total volume (V or V_{vdW}) of molecule, for instance, and it would impair any kind of the interaction with the enzyme's S1 subsite.

PCA and HCA were used as complementary techniques for efficiently exploring the data set. PCA included up to seven factors or PCs, and the findings are indicated in Figure 7A. PC1, or factor 1, contains 61.2% of total variance from the original data whereas PC2, or factor 2, describes 17.5% of total variance (Fig. 7B). Both PCs can be considered as representative for analyzing the MMP-2 inhibitors because together they contain 79% of total variance from the original descriptors or molecular properties.

The scores plot for PC1 versus PC2 is depicted in Figure 7A. PC1 and PC2 discriminated the highly active inhibitors from compounds with low/moderately activity. On the left side of PC1 (negative values; see Fig. 7A) are mostly the samples with smaller substituent groups in R_3 moiety, that were classified as low and moderately active inhibitors whereas on the right side (positive values of PC1) are the compounds with larger R_3 substituent groups. The highly active compounds found in HCA (cluster B') are highlighted as a light blue rectangle (upper on the right side) and a red arrow is indicating the most active inhibitor. There is also some mix of activity (low/moderately) on the right side of scores plot (see Fig. 7A).

The loadings of the original descriptors in PC1 and PC2 are responsible for the discrimination of highly active inhibitors from compounds with low and moderately inhibitory activity as presented in Figure 7C. For PC1, the molecular properties that have a more positively contribution (0.43 or 0.44) were V_{vdW} (x_{14}), V (x_{15}), MR (x_{16}) and α (x_{17}). These descriptors expressed a steric, steric/hydrophobic or steric/electronic nature of the substituent groups in the investigated inhibitors (see table II). For PC2, thermodynamic [E_{Hb} (x_3)] and electronic [μ_y (x_{13})] descriptors have significant contribution for discriminating the MMP-2 inhibitors and their loading values were 0.66 and 0.62, respectively (see Fig. 7C). These PCA findings also corroborate HCA data, since the dendrogram obtained for variables indicated E_{Hb} (x_3) and μ_y (x_{13}) as the descriptors most closely related to the biological data (Y) (see the first cluster in Fig.7D).

As in HCA procedure, compound **47** was again distant from the entire data set (see Fig. 7A). The plot of the sample residual *versus* Mahalanobis distance was evaluated to diagnose possible outliers (see Fig. 8). Samples falling outside of one or both thresholds (red lines in Fig. 8) are potential outliers as compounds **44**, **47**, and **50**. The sample residual threshold is based upon a 95% probability limit (set internally in Pirouette software) and because of that 5% of *normal* samples could fall outside that cutoff. For that reason, in a large set, samples exceeding only one threshold might be normal (as compounds **44** and **50**, for instance).

However, samples lying either significantly beyond one threshold or beyond both are more likely to be outliers, as is the case with compound **47**. Compound **47** has the bulkiest R_3 group of the entire set, as reflected by the total volume (V or V_{vdW}) of molecule, impairing any kind of the interaction with the enzyme S1 subsite.

HCA and PCA for data matrix X2 (MMP-9 data set)

The dendrogram constructed for the MMP-9 data set is showed in Figure 9, and the β-*N*-biaryl ether sulfonamide hydroxamates (samples) were also grouped into two main clusters (A and B). Cluster A and B present a similarity index of 0.24 and 0.52, respectively. Considering a similarity cursor of 0.6, cluster A can be split in nine subclusters (A_1 to A_9) presenting the following similarity indices: 0.67, 0.65, 0.69, 0.67, 0.71, 0.72, 0.68, 0.72, and 0.64, respectively. Moreover, compound **47** is in cluster B, but it is not grouped with the remaining compounds (**46**, **57**, and **58**), which share a similarity index of 0.75.

Subclusters A_4 to A_6 present the major number of highly active compounds. Subcluster A_4 is composed of only highly active inhibitors (**32**, **38, 48, 53**, **60**, **63**), including the most active compound of the investigated set (**60**). Compounds **53** and **63** share a similarity index of 0.79 and compound **32** has 77% similarity with both molecules. Inhibitor **60** shares 82% similarity with **48** and compound **38** is 78% similar to both inhibitors regarding the molecular properties calculated and selected for this analysis. Compounds **35** and **51**, which are moderately inhibitors of MMP-9, are part of the subclusters A_5 (**49**, **51, 52**, **54, 59, 61**) and A_6 (**33, 34, 35**, **50**), respectively. The structural difference of these highly active inhibitors lies on R_{1c} and/or R_3 substituent groups, as already discussed in MMP-2 data set. A smaller grouping of highly active inhibitors (**36, 37**, **39**, **40**, **43**), except to compound **36** (moderately active), can also be found in subcluster A_9 (gray dashed line in Fig. 9).

Two groupings consisting only of moderately active compounds were detected in subcluster A_1 (gray dashed line in Fig. 9). One group is formed by compounds **1**, **23**, **24**, **67**, and the other by compounds **2**, **3**, **4**, **12**. Subcluster A_8 is composed of samples with low inhibitory activity, such as compounds **5**, **6**, **7**, **9**, **10**. Cluster B contains compounds with low/moderately activity. So far, the findings were quite similar to those found for matrix **X1**.

Regarding PCA results presented in Figure 10B, PC1 contains 52.6% of total variance from the original data whereas PC2 describes 14.3% of total variance. Both PCs can also be considered as representative for analyzing the MMP-9 inhibitors because together they discriminate 67% of total variance from the original descriptors or molecular properties.

The scores plot displayed in Figure 10A indicates that PC1 and PC2 discriminated the highly active inhibitors from compounds with low/moderately activity. On the left side of PC1 (negative values; see Fig. 10A) are mostly the samples with smaller substituent groups in R_3 moiety, that were classified as low and moderately active inhibitors, whereas on the right side (positive values of PC1) are the compounds with larger R_3 substituent groups. The highly active compounds found in HCA in subclusters A_4 to A_6 are highlighted as a light blue rectangle (see Fig. 10A, on the right side). There is also some mix of activity (low/moderately) on the right side of the scores plot.

The loadings of original descriptors in PC1 and PC2 are responsible for discriminating highly active inhibitors from compounds with low and moderately inhibitory activity as showed in Figure 10C. For PC1, the molecular properties that have a more positively contribution (0.40) were V_{vdW} (x_{11}), V (x_{12}), MR (x_{13}), α (x_{14}), and MSA (x_{15}). These descriptors expressed a steric, steric/hydrophobic or steric/electronic nature of the substituent groups in the investigated inhibitors (see table II). Thermodynamic, electronic, and hydrophobic descriptors [E_{Hb} (x_3), $ChelpG_{13}$ (x_8), μ_y (x_{10}), and ClogD_1.5 (x_{16})] have significant contribution for discriminating the MMP-9 inhibitors in PC2 (see Fig. 10C). The dendrogram of variables showed a cluster (81% similarity) composed by the biological data and the following descriptors: $ChelpG_{13}$ (x_8), ClogD_1.5 (x_{16}), E_{Hb}(x_3), a_3 (x_{23}), E_{solv} (x_5), μ_y (x_{10}), α (x_{14}), and MR (x_{13}) (see Fig. 10D). Thus, once again, the PCA findings corroborated the previous HCA data.

To diagnose possible outliers, the plot of sample residual versus Mahalanobis distance was evaluated for matrix **X2** (Fig. 11). Compound **47**, with the bulkiest R_3 substituent group, is lying significantly beyond one threshold as shown in Figure 11 and it can be considered an outlier for this data set too.

Conclusion and Perspective

HCA and PCA provide complementary findings for both data set, MMP-2 and MMP-9. Thermodynamic, electronic, and steric were the most significant molecular properties to discriminate highly active from moderately and less active inhibitors in MMP-2 data set whereas apparent partition coefficient at pH 1.5 (hydrophobic and electronic nature; electronic due to the pKa/pH relationship) was also considered as significant descriptor for the classification procedure to MMP-9 data set.

The presence of a bulky substituent group in R_3 is crucial to this set of inhibitors regarding the establishment of molecular interactions with the S1 subsite of both enzymes, MMP-2 and MMP-9. But, there is a limit as demonstrated by compound **47**. It was considered as outlier in all analysis carried out.

The information obtained from PCA, HCA and CEPs generated in MD simulations are going to be considered to perform an independent-receptor (*IR*) 4D-QSAR formalism and, then, establish a 3D-mapping pharmacophore model and also a MMPs binding mode hypothesis, exploring distinct alignments (orientation at the active site). The findings will be helpful to the rational design of novel leads as potential inhibitors of MMPs.

Acknowledgments

The authors thank the São Paulo Research Foundation, FAPESP/Brazil, for the Ph.D. scholarship of MSc. Kely M. Turra (grant 2010/

19142-1) and the Chem21 Group for the academic license of MOLSIM 3.2 program.

Keywords: metalloproteinases · molecular dynamics · melanoma · exploratory data analysis · molecular modeling

How to cite this article: K. M. Turra, K. F. M. Pasqualoto, S. B. M. Barros, *Int. J. Quantum Chem.* **2012,** *112*, 3374–3389. DOI: 10.1002/qua.24222

Additional Supporting Information may be found in the online version of this article.

[1] R. P. Verma, C. Hansch, *Bioorg. Med. Chem.* **2007**, *15,* 2223.
[2] A. Tandon, S. Sinha, *Bioinformation* **2011**, *5,* 310.
[3] S. M. Yang, R. H. Scannevin, B. Wang, S. L. Burke, Z. Huang, P. Karnachi, L. J. Wilson, K. J. Rhodes, B. Lagu, W. V. Murray, *Bioorg. Med. Chem. Lett.* **2008**, *18,* 1140.
[4] S. M. Yang, R. H. Scannevin, B. Wang, S. L. Burke, L. J. Wilson, P. Karnachi, K. J. Rhodes, B. Lagu, W. V. Murray, *Bioorg. Med. Chem. Lett.* **2008**, *18,* 1135.
[5] K. R. Beebe, R. J. Pell, M. B. Seasholtz, In Chemometrics: A Practical Guide; Wiley: New York, **1998**.
[6] M. Ferreira, *J. Brazil. Chem. Soc.* **2002**, *13,* 742.
[7] M. M. C. Ferreira, C. A. Montanari, A. C. Gaudio, *Quim. Nova* **2002**, *25,* 439.
[8] Hypercube, Inc., HyperChem Program Release 8 for Windows; Hypercube, Inc.: Gainesville, FL, **2007**. Available at: www.hyper.com. Accessed on July 2011.
[9] Y. Feng, J. J. Likos, L. Zhu, H. Woodward, G. Munie, J. J. McDonald, A. M. Stevens, C. P. Howard, G. A. De Crescenzo, D. Welsch, H. S. Shieh, W. C. Stallings, *Biochim. Biophys. Acta* **2002**, *1598,* 10.
[10] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28,* 235.
[11] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107,* 3903.
[12] D. Doherty, MOLSIM: Molecular Mechanics and Dynamics Simulation Software – User's Guide, version 3.2, The Chem21 Group Inc., Lake Forest, IL, **1997**.
[13] K. H. Forsythe, A. J. Hopfinger, *Macromolecules* **1973**, *6,* 423.
[14] Gaussian Inc., Gaussian 03W for Windows, version 6; Gaussian Inc.: Pittsburgh, PA, **1995–2003**.
[15] Chemaxon, Ltd., Marvin Beans 5.8.0 software – Calculator Plugins 2012; Chemaxon, Ltd.: Hungary. Available at: http://www.chemaxon.com. Accessed on July 2011.
[16] V. Fock, *Zeit. fürPhysik.* **1930**, *61,* 126.
[17] C. M. Breneman, K. B. Wiberg, *J. Comput. Chem.* **1990**, *11,* 361.
[18] J. R. Platt, *J. Chem. Phys.* **1947**, *15,* 419.
[19] J. R. Platt, *J. Phys. Chem.* **1952**, *56,* 328.
[20] M. Randic, *Chem. Phys. Lett.* **1993**, *211,* 478.
[21] A. T. Balaban, *Chem. Phys. Lett.* **1982**, *89,* 399.
[22] D. Plavsic, S. Nikolic, N. Trinajstic, Z. Mihalic, *J. Math. Chem.* **1993**, *12,* 235.
[23] I. Lukovits, W. Linert, *J. Chem. Inf. Comput. Sci.* **1994**, *34,* 899.
[24] P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. Dobrynin, I. Gutman, G. Domotor, *J. Chem. Inf. Comput. Sci.* **1995**, *35,* 547.
[25] H. Wiener, *J. Am. Chem. Soc.* **1947**, *69,* 17.
[26] B. Lucic, I. Lukovits, S. Nikolic, N. Trinajstic, *J. Chem. Inf. Comput. Sci.* **2001**, *41,* 527.
[27] G. Klopman, J. Y. Li, S. M. Wang, M. Dimayuga, *J. Chem. Inf. Comput. Sci.* **1994**, *34,* 752.
[28] A. K. Ghose, G. M. Crippen, *J. Chem. Inf. Comput. Sci.* **1987**, *27,* 21.
[29] P. Ferrara, J. Apostolakis, A. Caflisch, *Proteins* **2002**, *46,* 24.
[30] P. Ertl, B. Rohde, P. Selzer, *J. Med. Chem.* **2000**, *43,* 3714.
[31] W. Hasel, F. T. Hendrickson, W. C. Still, *Tetrahedron Comput. Methodol.* **1988**, *1,* 103.
[32] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112,* 6127.
[33] N. Bodor, Z. Gabanyi, C. K. Wong, *J. Am. Chem. Soc.* **1989**, *111,* 3783.
[34] K. J. Miller, *J. Am. Chem. Soc.* **1990**, *112,* 8533.
[35] Pirouette, version 3.11, Woodinville: Infometrix Inc., 1990–2003. Available at: http://www.infometrix.com/software/softdesc.html
[36] M. C. Ferreira, A. M. Antunes, M. S. Melgo, P. L. O. Volpe, *Quim. Nova* **1999**, *22,* 724.
[37] P. C. Mahalanobis, *J. Asiat. Soc. Bengal.* **1930**, *26,* 541.

Received: 1 February 2012
Revised: 14 May 2012
Accepted: 14 May 2012
Published online on 19 June 2012

Full Paper

www.molinf.com

molecular informatics

DOI: 10.1002/minf.201400073

Predicting Novel Antitumor Agents: 3D-Pharmacophore Mapping of β-*N*-biaryl Ether Sulfonamide-Based Hydroxamates as Potentially MMP-2 Inhibitors

Kely Medeiros Turra,[a] Diogo Pineda Rivelli,[a] Silvia Berlanga de Moraes Barros,[a] and Kerly Fernanda Mesquita Pasqualoto*[b]

Abstract: Matrix metalloproteinases (MMP) are a group of enzymes related to extracelular matrix remodeling. Some types of MMP are overexpressed by malignant tumors, mainly the MMP-2 subtype, and have been associated to cancer invasiveness and metastasis. A receptor-independent (*RI*) 4D-QSAR formalism was applied, herein, to a set of forty β-*N*-biaryl ether sulfonamide hydroxamates, previously reported as potent MMP-2 inhibitors, in order to map 3D-pharmacophore models and predict novel antitumor agents. The best *RI* 4D-QSAR model was statistically significant ($N=30$, $r^2=0.93$, $q^2=0.88$, five occupancy descriptors (GCOD), $LSE=0.04$, $LOF=0.11$, outliers = 0), robust and not obtained by chance. The external predictability was 75% (test set; $N=8$). A different orientation (binding mode) in the MMP-2 catalytic site was suggested regarding the most hydrophobic portion (R_1) of the compounds' structure. Compounds were predicted and their inhibitory activity against MMP-2 was calculated by using the optimum *RI* 4D-QSAR model. The findings have provided interesting information to drive the designing and synthesis of novel potentially MMP-2 inhibitors against melanoma invasion.

Keywords: Metalloproteinase · Receptor-independent 4D-QSAR · Drug design · Antitumor agents · Melanoma

1 Introduction

Drug designing is a development process that involves, in general, an ensemble of innovation, advanced experimental and computational methods. Quantitative structure-activity relationship (QSAR) formalisms are part of the theoretical arsenal for drug discovery, and deal with the prediction of biological activity using mathematical relationships based upon structural, physicochemical, and conformational properties of potential hits, leads or drug candidates.[1,2] The three-dimensional (3D) QSAR approach, for instance, is often used in pharmaceutical applications and is considered one of the major methods in computer-assisted molecular design, CAMD. The most popular and one of the first 3D-QSAR schemes is the comparative molecular field analysis, CoMFA,[3] which is essentially a *receptor-independent* (*RI*) CAMD method (*ligand-based drug design* strategy, LBDD).[4] Of note, in CoMFA, the geometry of receptor, or biomacromolecule, is not considered in any calculation step.[3]

When the geometry of the receptor (usually an enzyme) is available along with the corresponding ligand structure-activity data, the *structure-based drug design* (SBDD) approach is the elected *receptor-dependent* (*RD*) CAMD strategy.[4]

Hopfinger and co-workers[4,5] developed a 4D-QSAR formalism, which allows performing 3D pharmacophore models due to its capability of exploring large degrees of both conformational and alignment freedoms in the search for the active conformation and binding mode to a set of investigated compounds. Furthermore, depending upon the available information regarding the molecular system to be exploited, it can be applied to both approaches, LBDD (*RI*)[6–10] and SBDD (*RD*).[11–13]

Molecular interactions in biological systems are complex phenomena and, mostly, generate noisy data, if simulated in silico. Then, activity modeling and predictions can be a risky process. Molecular recognition uncertainty in traditional *RI* multidimensional QSAR cannot be eliminated unless were included the receptor data.[14] When a *RI* 4D-QSAR approach is performed, for instance, providing multi-

[a] *K. Medeiros Turra, D. Pineda Rivelli, S. Berlanga de Moraes Barros*
Laboratory of Cytopathology, Department of Clinical Chemistry and Toxicology, Faculty of Pharmaceutical Sciences, University of São Paulo
Av Prof Lineu Prestes, 580, São Paulo, SP 05508-000, Brazil

[b] *K. Fernanda Mesquita Pasqualoto*
Biochemistry and Biophysics Laboratory, Butantan Institute
Av Vital Brasil, 1500, São Paulo, SP 05503–900, Brazil
phone/fax: 55 11 2627 9856/55 11 2627 9748
**e-mail: kerly.pasqualoto@butantan.gov.br*
kerlypasqualoto@gmail.com

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/minf201400073.

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ple 3D ligand representations and orientations, the bioactive conformations predicted are docked into the interaction site of the 3D target structure whenever possible, improving the quality and prediction accuracy of the QSAR model.

Despite the types of misunderstanding and misuse, QSAR methodology can be inherently considered as a valuable tool based upon sound statistical principles which retrospectively explain SAR and, most of time, provide synthetic guidance leading to experimentally testable hypotheses. These qualities reinforce QSAR as a viable and important medicinal chemistry tool.[15]

Cancer is the second leading cause of mortality worldwide,[16,17] and studies have revealed evidences supporting that extracellular matrix remodeling proteinases, such as matrix metalloproteinases (MMP), are the main mediators of microenvironment alterations observed during cancer progression.[18,19] MMP enzymes, also designated matrixins, belong to a zinc-dependent family of endopeptidases, which hydrolyze components of extracellular matrix (ECM). They play a central role in many biological processes, including embryogenesis, normal tissue remodeling, wound healing, and angiogenesis, as well as in pathological conditions, such as atheroma, arthritis, cancer, and tissue ulceration.[19–23] Also, they are overexpressed in several types of human cancer such as breast, lung cancer, hepatocarcinoma, and melanoma.[24] The idea of exploring MMP as therapeutic cancer targets was introduced about 25 years ago when the metastatic potential of various cancers was correlated to the ability of cancer cells to degrade the basement membrane.[25] Since then, a growing number of MMP inhibitors (MMPI) have been developed and evaluated in several clinical trials.[26,27]

It has been identified 23 different constitutive types of MMP in humans, so far, and they were classified according to their structural, functional and specificity features. MMP-2 and MMP-9, for instance, have been implicated in a number of pathological events leading to cancer, inflammatory diseases, cardiovascular diseases, and neurological disorders.[20,28] In this regard, inhibiting MMP-2 and/or MMP-9 activity would be highly desirable. Recently, a set of β-*N*-biaryl ether sulfonamide hydroxamates was designed, synthesized, and tested against MMP-2, MMP-9 and MMP-1 activity. The hydroxamate derivatives were potential MMP-2 and MMP-9 inhibitors without inhibiting MMP-1,[29,30] indicating a selectivity improvement. Besides a combined molecular modeling study, using CoMFA/CoMSIA (3D-QSAR),[31] any other multidimensional QSAR approach regarding those structure-activity data was not applied yet in order to aid the prediction of novel and more selective MMPI. A 4D-QSAR formalism[4,5] was employed, herein, to develop 3D pharmacophore models for ligand structure-activity data previously reported by Yang et al.,[29,30] regarding especially MMP-2 inhibition. MMP-2 is responsible for basal membrane components degradation, and is expressed at low levels on physiological conditions. However, it is overexpressed on malignant neoplasias and can trigger tumor invasion and metastasis due to its proteolytic activity.[32–35] The findings from 4D-QSAR approach, then, would be quite helpful for the rational design of novel selective MMPI.

2 Methods

2.1 Biological Data Set

A data set of forty β-*N*-bioaryl ether sulfonamide hydroxamate derivatives was selected from the references.[29,30] The compounds, listed in Table 1, were assayed using the same pharmacological protocol. The inhibitory activity values were determined by using a standard peptide-cleavage assay, based upon the concentration of each compound required for fifty-percent of MMP2 inhibition (IC_{50}, nM). The experimental inhibitory values were expressed in negative logarithmic units, pIC_{50} ($-\log IC_{50}$) (see Table 1), and the range in activity is about 3 (5.88–8.60) pIC_{50} units. The pIC_{50} values comprise the set of dependent variables in the 4D-QSAR approach. Previously, QSAR models were constructed with entire data set ($N=38$) making models statistically more representative. Then, the inhibitors were split into training and test set, by applying hierarchical cluster analysis (HCA),[36] in order to perform the external validation (test set).[37] Briefly, the compounds of training set were used to construct QSAR models whereas the compounds of test set (external validation) were used to verify the prediction power of the QSAR models generated with training set. It is recommended to use 20 to 30% of samples for the external validation of smaller data sets.[37–39] The eight compounds selected for test set (21% of total data set) are indicated with [b] in Table 1. Structural diversity and gradient of biological data (highly, moderately, and less active) were also contemplated in training and test sets.

2.2 4D-QSAR Analysis

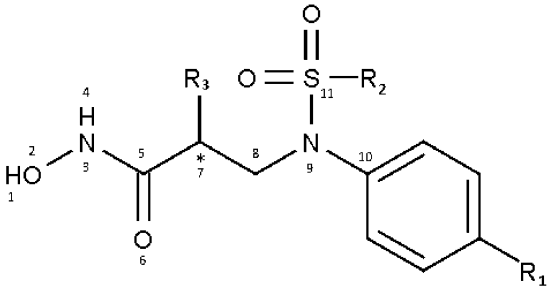
The current methodology of 4D-QSAR analysis[4,5] consists generally of the operational steps listed as follows. The *RI* formalism was applied in this study because of uncertainty in the binding mode of the ligands.

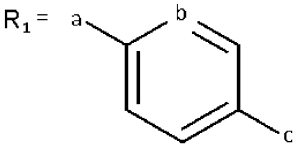
2.2.1 Building Up the Three-Dimensional (3D) Molecular Models (Input Data)

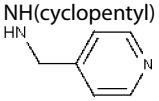
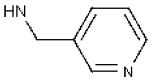
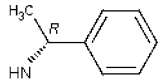
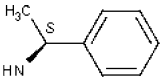
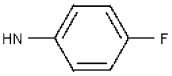
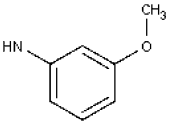
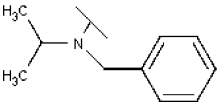
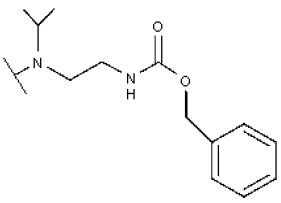
The 3D molecular models of the forty inhibitors, in their neutral forms, were constructed using the HyperChem 8.0 software.[40] The NMR structure of a hydroxamic acid inhibitor in complex with the catalytic domain of MMP-2 was retrieved from Protein Data Bank,[41] PDB, entry code 1HOV,[42] and the inhibitor was used as a geometry reference for constructing all ligands. The chirality of α-carbon center at 7-position (C_7*) was also considered, as shown in Table 1. The 3D molecular model of each ligand was energy-minimized employing MM+ force field (adapted from MM2

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Structural features of the β-*N*-biaryl ether sulfonamide hydroxamate derivatives,[29,30] chemical notation and biological activity of the investigated MMP-2 inhibitors

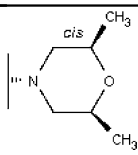
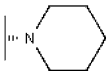
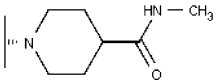
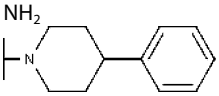
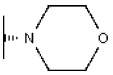
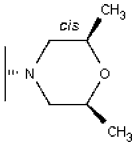
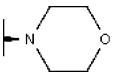




Compound	C_7* (*R*/*S*)	pIC_{50} (MMP-2)	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
5a	*S*	7.21	O	CH	Me	Me	OH
5b	*S*	6.85	O	CH	Me	Et	OH
5c	*S*	6.29	O	CH	Me	*n*-Pr	OH
5f[a]	*S*	6.23	O	CH	Me	Bn	OH
5g	*S*	6.40	O	CH	Me	(pyridin-3-yl)CH_2	OH
5h[a]	*S*	6.46	O	CH	Me	(pyridin-4-yl)CH_2	OH
5j	*S*	6.61	O	CH	H	Me	OH
5l	*S*	6.86	O	CH	CF_3	Me	OH
5o	*S*	6.55	S	CH	F	Me	OH
13a	*S*	7.33	C≡C	CH	Me	Me	OH
13b	*S*	7.25	C≡C	CH	Cl	Me	OH
14b[a]	*S*	6.05	O–CH_2	CH	Cl	Me	OH
5k[a]	*S*	7.45	O	CH	Cl	Me	OH
7a	*S*	7.44	O	CH	Me	Me	OH–CH_3
8a[a]	*S*	8.04	O	CH	CF_3	Me	$NHCH(CH_3)_2$
$8b_1$	*S*	8.40	O	CH	Me	Me	$NHCH(CH_3)_2$
$8b_2$	*R*	6.75	O	CH	Me	Me	$NHCH_2(CH_3)_2$
8c	*S*	7.96	O	CH	Me	Me	$CH_2CH(CH_3)_2$
8d	*S*	8.00	O	CH	Me	Me	NH(cyclopentyl)
8e	*S*	8.02	O	CH	Me	Me	
8f	*S*	7.77	O	CH	Me	Me	
8ir	*S*	7.62	O	CH	Me	Me	
8is	*S*	7.92	O	CH	Me	Me	
8k	*S*	7.19	O	CH	Me	Me	
8n[a]	*S*	7.57	O	CH	Me	Me	
8q	*S*	6.55	O	CH	CF_3	Me	
8r	*S*	6.58	O	CH	CF_3	Me	

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. (Continued)

Compound	C_7* (*R*/*S*)	pIC_{50} (MMP-2)	R_{1a}	R_{1b}	R_{1c}	R_2	R_3
8t[a]	*S*	8.25	O	CH	CF_3	Me	
8x	*S*	8.12	O	CH	CF_3	Me	
8z	*S*	8.26	O	CH	CF_3	Me	
11a	*S*	6.17	O	CH	Me	Me	NH_2
8aa	*S*	7.33	O	CH	CF_3	Me	
8cc	*S*	8.26	O	CH	Cl	Me	
8dd	*S*	8.60	O	CH	Cl	Me	
8ee	*R*	5.88	O	CH	Me	Me	
9a[a][b]	*S*	5.90	O	CH	Cl	Me	OH
9b[b]	*S*	6.54	O	CH	Cl	*i*-Pr	OH
3b	–	7.14	O	CH	Me	Me	–

* The carbon at the 7-position is a chiral center and can confer *S* or *R* configuration to the molecule. Inhibitor **3b** does not present the chiral carbon atom at position 7 (C_7). [a] Compounds of test set. [b] Compounds **9a** and **9b** have an oxygen atom linked to N_9 and R_2 (N_9OR_2), instead of a sulfonyl group ($N_9SO_2R_2$).

force field[43]), without any constrains, and partial atomic charges were calculated using the AM1 semiempirical method,[44] implemented in HyperChem 8.0.[40] The MOLSIM 3.2 program[45] was also employed for the geometry optimization procedure applying the steepest descent and, subsequently, conjugated gradient method. The energy convergence criterion adopted was 0.01 kcal/mol. A dielectric constant of 3.5 was considered because it simulates the biological membrane environment and the inner enzyme binding sites (dielectric constant 2–5).[46]

2.2.2 Selection of the Trial Set of Interaction Pharmacophore Elements (IPE)

Seven types of atomic groups were used herein to define the interaction pharmacophore elements (IPE). The atoms of each inhibitor were partitioned into seven classes: polar atoms of positive charge (*p*+), polar atoms of negative charge (*p*−), nonpolar atoms (*np*), hydrogen bond acceptor (*hba*), hydrogen bond donor (*hbd*), atoms in aromatic system (*ar*), and no differentiation of all-atom occupancy (*any*).

2.2.3 Conformational Ensemble Sampling of Each Compound

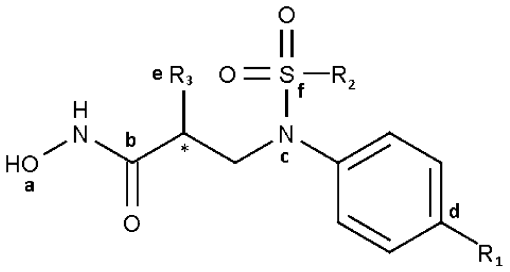
The energy-minimized structure of each inhibitor was used as the initial structure in each molecular dynamics (MD) simulation to generate the conformational ensemble profile (CEP) of each ligand. The MOLSIM 3.2 program[45] was used to perform the MD simulations and to generate the trajectories for, in turn, deriving the CEP. The MD protocol employed 1 000 000 steps (1 ns) for each ligand, the step size was 1 fs, and the simulation temperature was 298 K, the same temperature used in the biological assay.[29,30] Output trajectory files were recorded every 20 simulation steps, resulting 50 000 conformations. Of note, 4D-QSAR analysis does not use a single conformation in constructing a QSAR model, but rather the intrinsic conformational flexibility of each compound is taken into account through its CEP.

2.2.4 Selection of the Trial Alignments

The 4D-QSAR algorithm considers only the unrestricted three ordered-atom match alignment rule. In this study, the chosen atoms cover the common structure of the investigated compounds, and the hydroxamic portion, as well, which interacts with the zinc metal in the MMP-2 active

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Set of trial alignments used for constructing *RI* 4D-QSAR models



Alignment	1st atom	2nd atom	3rd atom
1	a	c	d
2	b	e	c
3	b	f	d

site. The sulfonamide moiety was also considered because it is involved in interactions with the enzyme backbone throughout hydrogen bonds.[29] The corresponding letter sequences for each alignment are listed in Table 2, using the general chemical structure of the β-*N*-biaryl ether sulfonamide hydroxamate derivatives.

2.2.5 Grid Cell Occupancy Descriptors (GCOD) Determination

Each conformation from a 100 ps CEP for each ligand (consisting of 5000 conformations generated by the MD simulation sampling regarding the equilibrium region) was placed in a reference grid cell space according to the trial alignment under consideration. In this study, the grid resolution was set to 1 Å, and the size of the overall grid cell lattice was chosen to enclose all inhibitors of the investigated data set. The atom occupancy of each grid cell is a descriptor in 4D-QSAR analysis, and as described earlier, these grid cell occupancy descriptors (GCOD) were computed for each of seven IPE atom types. Herein, no reference standard inhibitor[4] was used to compute the descriptors (GCOD). Then, the normalized absolute occupancy of each grid cell by each IPE atom type over the CEP for a given alignment comprises a unique set of QSAR descriptors (GCOD). The occupancy descriptors (GCOD) were computed and used as the basis set of trial 4D-QSAR descriptors in this analysis.

2.2.6 Partial Least Square (PLS) Data Reduction of the Entire Set of GCOD

An enormous number of trial QSAR descriptors, GCOD, are generated in a 4D-QSAR analysis due to the large number of grid cells and the seven IPE. Partial least squares (PLS) regression analysis[47] was employed to perform a data reduction fit between the observed dependent variable (pIC_{50} values, Table 1) and the corresponding GCOD values for each of the trial (three) alignments. Furthermore, a variance-filtering constraint was also applied to the entire set of GCOD prior to the PLS analysis. Descriptors (GCOD) having a variance (self-variance) over the set of analogues less than 2.0 (pre-chosen fraction) were eliminated. Before proceeding the PLS analysis, the data were mean-centered. The automated reduction data provides a selection of GCOD having the highest individual weightings to the observed biological activity measures.

Of note, the resulting PLS regression fit is the quantitative QSAR model, and the PLS weightings for GCOD could be used to construct a graphical representation as done, for example, in CoMFA. However, a general finding from the application of *RI* 4D-QSAR analysis is that only a relatively small number of descriptors, GCOD, are significant in the QSAR model. Thus, *RI* 4D-QSAR analysis offers the opportunity of providing a short equation in terms of the number of descriptors, reminiscent of those generated in classic 2D-QSAR approaches.[48] The compact QSAR models are actually constructed in the next step.

2.2.7 Generation of Compact QSAR Models Through a Genetic Algorithm (GA) Optimization

In this study, the 200–700 most highly ranked PLS GCOD were used to form the trial basis set for the genetic algorithm (GA) analysis,[49] and the genetic function approximation (GFA)[50] was employed in QSAR model building and optimization. Of note, the GCOD most highly ranked and only linear terms were used in multiple linear regression (MLR) during the GA optimization (MLR-GA). Mutation probability over the crossover optimization cycle was set at 10%. Fifty thousand to one hundred thousand crossover operations were considered. Smoothing factor values of 1.5 to 3.5 were tested in order to determine the optimal number of descriptors in the 4D-QSAR models[51–53] based upon Friedman's lack-of-fit (*LOF*),[50,54] which is a penalized least-squares measure. The compact QSAR models generated in this step are straightforward to explore and provide substantial insight into the structure-activity information inherent to the training set.

2.2.8 Inspection and Evaluation of the Entire Population of Models

The main purpose of this step is to identify the "best" QSAR model with respect to alignment. Then, the top 10 models for each alignment were selected and their significance measures were evaluated. The diagnostic measures to analyze the resultant QSAR models from GA optimization include descriptor usage as a function of crossover operation, linear cross-correlation among descriptors and/or biological activity measures, and number of significant models. In addition, the evaluation of models considers the following measures of model significance: leave-one-out cross-validation correlation coefficient (q^2), coefficient of multiple determination (r^2), and Friedman's lack-of-fit (*LOF*).[50,53] Herein, compounds of training set whose differences in experimental/observed and predicted/calculated

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

activities (residuals) exceeded 2.0 standard deviations (2 *SD*) (95% confidence level) from the mean of a model were considered as outliers.

The alignment corresponding to the QSAR model with the overall highest r^2 and q^2 measures for all alignments tested was selected as the best alignment. For the best alignment, a cross-correlation matrix of the residuals in error between pairs of the top 10 QSAR models, based upon their q^2, was built in order to determine if the top 10 QSAR models have provided common, or distinct, structure-activity information. It is possible, then, to identify the set of unique best QSAR models. Pairs of models with highly correlated residuals of fit ($R \approx 1$) are judged to be nearly the same model, while pairs of models with poorly correlated residuals ($R < 0.5$) are distinct from one another. Furthermore, the linear cross-correlation matrix of the descriptors, GCOD, for the best QSAR model of the best alignment was calculated to determine if these significant descriptors were, or were not, correlated to one another. Pairs of descriptors presenting *R* values higher than 0.5 (the absolute value) were considered correlated to one another and did not provide reliable information to the model.

2.2.9 External Validation

The eight compounds of test set (see Table 1) were not included in the building of QSAR models, but they were used to validate the best QSAR model constructed from the training set and to evaluate its prediction capacity. The predicted activity value (pIC_{50}) of each compound in the test set was calculated using the equation of the best model/alignment by substitution of the GCOD values found for the Cartesian coordinates indices of the reference grid cell space, GC*i*(*x*,*y*,*z*). Then, GC*i*(*x*,*y*,*z*) represents each GCOD of each compound in the test set and its respective position in the grid space. A confidence level of 99% was considered, herein, in order to verify the prediction power of the QSAR model generated with the training set. In this regard, compounds of the test set whose residual values (observed minus calculated activity) exceeded 1.0 standard deviation (*SD*) from the mean were not accounted as well predicted by the resulting QSAR model.

2.2.10 Generation of the Hypothesized Bioactive Conformation of Each Compound of Training Set

The final step in 4D-QSAR analysis is to hypothesize the *bioactive* conformation of each compound of training set. Firstly, all conformer states sampled for each compound, one at a time, which are within ΔE of the global minimum energy conformation of the CEP are identified. Herein, ΔE was set at 5 kcal/mol. Then, the resulting low-energy conformations are individually evaluated using the correlation equation of the best QSAR model. The single conformation within ΔE that predicts the highest activity is selected as the *bioactive* conformation of the compound by the 4D-QSAR software.[5]

The selected *RI* 4D-QSAR model should sterically fit and provide complementary ligand-receptor interaction sites. As earlier mentioned, the NMR structure of a hydroxamic acid inhibitor in complex with the catalytic domain of MMP-2 (PDB entry code 1HOV)[42] is available, and it was used to explore the structural information within the selected *RI* 4D-QSAR model. The descriptors, GCOD, of the best model and the postulated bioactive conformations of each compound of the training set were docked into the inhibitor catalytic site (HyperChem 8.0)[40] considering the best alignment (three-atoms) as starting point (rigid docking with respect to the ligands). After that, to better accommodate the inhibitors/GCOD into the MMP-2 active site, relaxation MD simulations (50 ps at 298 K) were performed on each docked complex (MOLSIM 3.2).[45] Then, the evaluation of GCOD obtained from the optimum *RI* 4D-QSAR model was carried out using the catalytic site geometry. The preparation of protein (MMP-2) for computational simulation is described in the Supporting Material section.

2.2.11 Leave-N-Out Cross-Validation (LNO) and y-Randomization Procedures

LNO, or leave-many-out, cross-validation[55–57] procedure is highly recommended to test the robustness of a model. Before running LNO, the samples were randomly ordered (in **X** and **y** simultaneously). It is recommended that *N* represents a significant fraction of samples. For smaller data sets, 20 to 30% should leave out.[39] Then, herein, it was considered a fraction of at least 30% of samples. Also, a multiple LNO test was performed (triplicate). The critical *N* is the maximum value of *N* at which q^2_{LNO} is still stable and high. For a good model, q^2_{LNO} should stay close to q^2 from LOO, with small variations at all values for *N* up to the critical *N*. In case of multiple LNO, two standard deviations should not be greater than 0.1 for $N = 2$, 3, etc, including the critical value of *N*.[37]

The aim of the **y**-randomization test[37,55,56,58–60] is to detect and quantify chance correlations between the dependent variable and descriptors, GCOD, in this case. The test consists of several runs for which the original descriptors (block **X**) is kept fixed, an only the vector **y** is randomized (scrambled). The procedure was, then, performed in 10, 30 and 50 randomization runs, in triplicate. The models obtained under such conditions should be of poor quality and without real meaning, and this is reflected on the statistical measures, q^2_{yrand} and r^2_{yrand}. In this study, it was considered the approach of Eriksson and Wold[59] to judge whether the real model was characterized by chance correlation, based on the values of q^2_{yrand} and r^2_{yrand} and their relationship $r^2_{yrand} > q^2_{yrand}$: no chance correlation ($q^2_{yrand} < 0.2$ and $r^2_{yrand} < 0.2$); negligible chance correlation (any q^2_{yrand} and $0.2 < r^2_{yrand} < 0.3$); tolerable chance correlation (any

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

q^2_{yrand} and $0.3 < r^2_{yrand} < 0.4$); and, recognized chance correlation (any q^2_{yrand} and $r^2_{yrand} > 0.4$).

3 Results and Discussion

Before splitting the compounds into training and test set, the 4D-QSAR formalism was applied to the entire data set ($N = 38$) in order to generate more statistically reliable QSAR models, and to optimize the calculation conditions for generating the mathematical models. The statistical measures for the top ten models of each trial alignment using 70.000 crossover operations, mutation probability of 10%, and a smoothing factor value of 2.5 in the GFA-MLR optimization step are shown in Table 1S (Supporting Information).

The three alignments presented models with high values of q^2 (0.83 to 0.93) and r^2 (0.88 to 0.95) and low error values (*LSE* (0.02 to 0.07) and *LOF* (0.17 to 0.22)). Alignment 2, however, generated QSAR models without outliers and the number of grid cell occupancy descriptors (GCOD) was in acceptance with the "rule-of-thumb", which states no more than one descriptor for every five to seven compounds within the training set, since the QSAR models were optimized using GFA-MLR. Thus, the maximum number of descriptors should be between 5 and 8 GCOD.

The statistical measures, number of grid cell occupancy descriptors (GCOD) and outliers for the top ten models of alignment 2 are shown in Table 2S (Supporting Information). Also, the correlation coefficients of the residuals of fit between pairs of the ten top models from alignment 2 were computed, and they were highly correlated to one another (Table 3S, Supporting Information), indicating there is a single unique QSAR model.

The grid cell occupancy descriptors (GCOD or GC) in models 1, 2, 3, 8 and 9 were highly correlated to one another ($R > 0.5$). But, the linear cross-correlation matrix of GCOD for model 4 (see Table 4S in Supporting Information) did not present any descriptors highly correlated to one another, meaning each one provides independent information to the model. Furthermore, model 4 presented high values of q^2 (0.83) and r^2 (0.88), low error values ($LSE = 0.07$ and $LOF = 0.19$), and six occupancy descriptors.

According to Table 2S (Supporting Information), model 4 from alignment 2 has 1 outlier, which is compound **14b** ($pIC_{50} = 6.05$). The residual value (difference between experimental and calculated activity) for compound **14b** was 0.68, which exceeded 2 *SD* from the mean of a model (2 $SD = 0.53$) (Table 5S, Supporting Information). From the structural point of view, compound **14b** has a CH_2 group linked to the oxygen atom in R_{1a} substitution. This kind of substitution seems to be responsible for reducing the inhibitory activity against MMP-2, since it provides more freedom degree on that region of molecule. Compound **13b**, for instance, which presents C≡C as substituent in R_{1a}, but all further substitutions similar to those of compound **14b**, is more potentially active ($pIC_{50} = 7.25$). It is possible that an outlier can come from the sampled population, and compound **14b** is the only molecule in the entire set with an heteroatom linked to a CH_2 group in R_{1a}. Even though, the correlation coefficient (*R*) between the experimental and calculated values of pIC_{50} considering model 4 ($N = 38$) from alignment 2 was 0.97, and the plot is presented in Figure 1A.

The robustness and chance correlation of model 4 from alignment 2 were, then, evaluated through a multiple LNO test (triplicate) and **y**-randomization procedure, respectively, as reported in Methods section.

Regarding LNO test, *N* varied from 2 to 16 (42% of samples), and the critical *N* for model 4 was set up when a fraction comprising twelve samples was removed. Thus, q^2_{LNO} was close to q^2_{LOO} up to twelve samples (32%) were removed from the data set ($2\ SD < 0.1$), meaning the model 4 is robust. Table 6S (Supporting Information) contains the q^2_{LNO} values for each procedure (triplicate) varying *N* from 2 to 16, the q^2_{LNO} average ($< q^2_{LNO} >$), and the respective two standard deviations (2 *SD*).

The **y**-randomization procedure was performed in 10 (Figure 1C), 30 and 50 runs, in triplicate, and the models obtained under such conditions were of poor quality and without real meaning as reflected on the statistical measures, q^2_{yrand} and r^2_{yrand}. According to the approach of Eriksson and Wold[59] model 4 from alignment 2 was characterized as negligible chance correlation. The q^2_{yrand} values were equal to or lower than zero (negative values), and the r^2_{yrand} values varied from 0.07 to 0.27, regarding 10, 30, and 50 runs.

The plots of LNO and **y**-randomization (10 runs) for model 4 from alignment 2 are displayed in Figure 1B and 1C, respectively. Table 7S (Supporting Information) contains the q^2_{yrand} and r^2_{yrand} values for 10 randomization runs. Figure 1S and 2S (Supporting Information) show the findings (table and plot) for 30 and 50 randomization runs, respectively.

Despite a QSAR model constructed from a larger set of compounds be more statistically reliable, mostly it is recommended to define a training set and a test set. The test set is crucial in order to carry out the external validation and, consequently, verify the prediction power of the QSAR model built up from the training set.

Then, the thirty-eight inhibitors were split into two sets: a training set containing thirty compounds and a test set comprising eight compounds, as earlier described in the Methods Section.

The *RI* 4D-QSAR formalism was applied to the training set ($N = 30$) using the optimized previously conditions (70000 crossover operations and 10% of mutation probability) and a smoothing factor value of 2.4 in the GFA-MLR optimization step. The statistical measures for the top ten models of each trial alignment are shown in Table 3.

The three alignments presented models with higher values of q^2 (0.86 to 0.97) and r^2 (0.91 to 0.98) and low

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

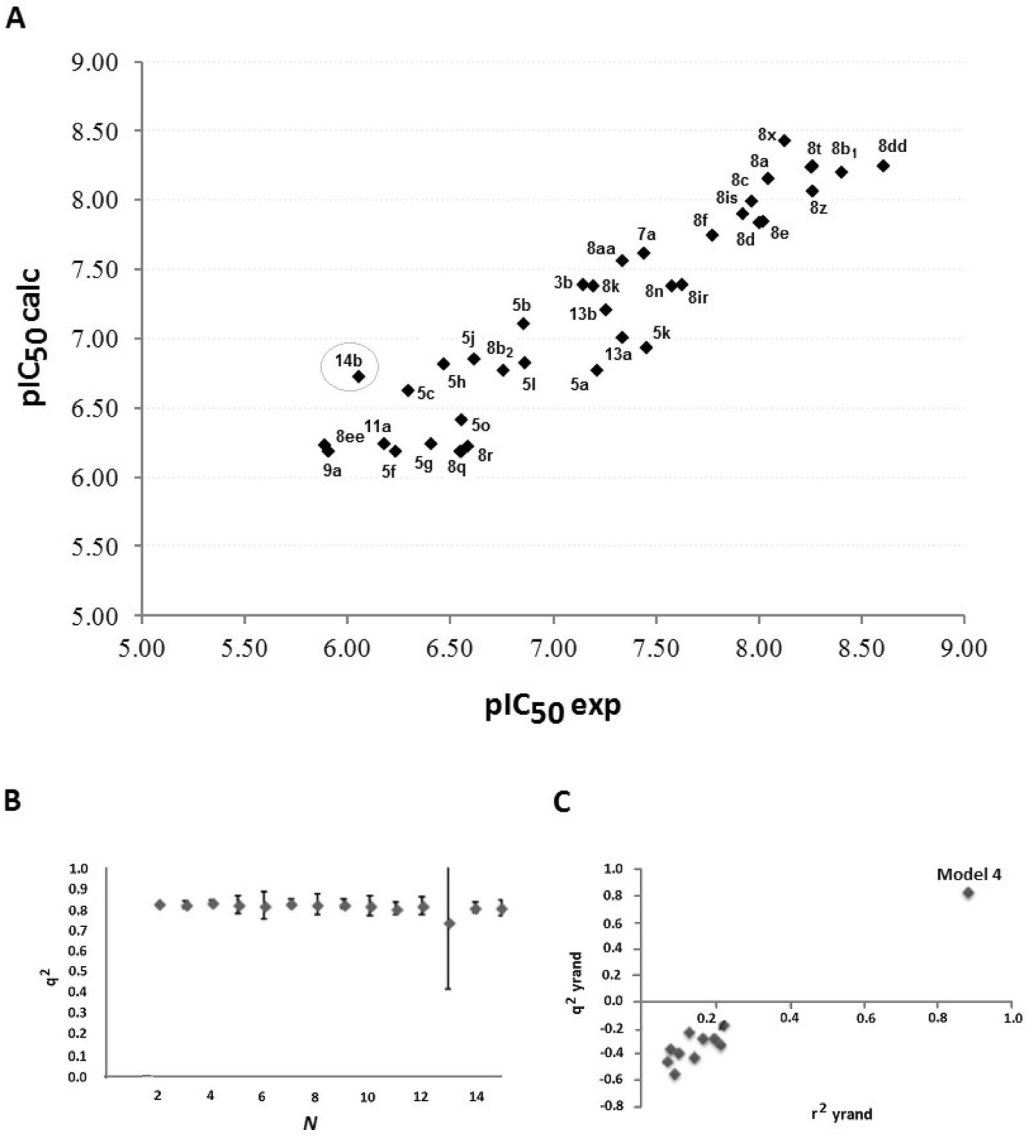


Figure 1. (A) Experimental versus calculated pIC_{50} values for model 4 ($N = 38$) from alignment 2.The correlation coefficient (R) between the predicted/calculated and observed/experimental activities for the data set was 0.97. Compound **14b** (in the light-gray circle) was the only sample whose the residual value (0.68) exceeded 2SD (confidence level of 95%) from the mean of a model (2SD = 0.53). (B) LNO and (C) **y**-randomization (10 runs) plots for model 4 ($N = 38$) from alignment 2. N critical was reached when a fraction of twelve compounds was removed (**B**). Model 4 was characterized as negligible chance correlation ($0.2 < r^2_{yrand} < 0.3$) (C), regarding Eriksson and Wold approach.[59]

Table 3. Statistical measures, number of occupancy descriptors (GCOD) and outliers for the top ten QSAR models ($N = 30$) of each trial alignment (1, 2, 3) using smoothing factor value of 2.4.

Alignment	Number of GCOD	r^2	q^2	*LSE*	*LOF*	Outliers
1	7–8	0.96–0.97	0.95–0.96	0.09–0.10	0.01–0.02	1–2
2	**5–6**	**0.91–0.94**	**0.86–0.89**	**0.03–0.05**	**0.11–0.14**	**0–2**
3	8	0.98	0.97	0.06–0.07	0.08	2

error values (*LSE* (0.03 to 0.10) and LOF (0.01 to 0.14)). Furthermore, only alignment 2 generated QSAR models having no outliers, and the number of descriptors (GCOD) was in acceptance with the "rule-of-thumb" (between 4 to 6 descriptors), regarding the optimization step (GFA-MLR). Thus, alignment 2 provides the best spatial representation, regarding all CEP, for the set of MMP-2 inhibitors investigated. It considers the carbonyl carbon of hydroxamic portion, the R_3 substituent group, and the sulfur of sulfonamide moiety. The statistical measures, number of descriptors

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 4. Test set predictions. pIC_{50exp}: pIC_{50} experimental values; pIC_{50calc}: pIC_{50} calculated values.

Compound	pCI_{50exp}	pCI_{50calc}	Residuals
5f	6.23	6.95	0.72
5h	6.46	6.77	0.31
14b	6.05	7.51	1.46
5k	7.45	7.51	0.06
8a	8.04	7.41	−0.62
8n	7.57	7.10	−0.46
8t	8.25	8.92	0.67
9a	5.90	6.94	1.04

(GCOD) and outliers for the top ten models of alignment 2 are shown in Table 8S (Supporting Information). Model 4 presented high values of q^2 (0.88) and r^2 (0.93), low error values (*LSE* = 0.04 and *LOF* = 0.11), five grid cell occupancy descriptors and no outliers (see Table 9S, Supporting Information). Thus, it was chosen as the best model from alignment 2, and considered for further analysis.

The correlation coefficients of the residuals of fit between pairs of the ten top models from alignment 2 were computed, and they were highly correlated to one another (Table 10S, Supporting Information), indicating there is a unique QSAR model (model 4). Also, a linear cross-correlation matrix of the cell grid occupancy descriptors (GCOD or GC) for model 4 was built (Table 11S, Supporting Information), and none of the descriptors are highly correlated to one another, meaning each one provides independent information to model 4 ($N = 30$). Model 4 is defined by the following equation:

$$pIC_{50} = 8.91\ GC1_{(any)} - 0.77\ GC2_{(any)} - 0.78\ GC3_{(any)} + 2.09\ GC4_{(np)} + 1.11\ GC5_{(any)} + 6.70 \quad (1)$$

$N = 30$; $r^{2=}0.93$; $q^{2=}0.88$; LSE = 0.04; LOF = 0.11;

Outliers = 0

Regarding Equation 1, model 4 is composed of only two classes of IEP: no differentiation of all-atom occupancy (*any*) and nonpolar atoms (*np*). In addition, considering the coefficients' signs, the descriptors GC2 and GC3 negatively favor the inhibitory activity (pIC_{50}) whereas GC1, GC4, and GC5 descriptors contribute positively to the inhibitory activity of MMP-2.

In order to verify the robustness and chance correlation of model 4 a multiple LNO test and **y**-randomization procedure, respectively, were performed. Regarding the LNO test, *N* varied from 2 to 10 (33% of samples), and the critical *N* for model 4 was set up when a fraction of nine samples was removed. Thus, q^2_{LNO} was close to q^2_{LOO} up to nine samples (29.7%) were removed from the data set (2 *SD* < 0.1), meaning the model 4 is indeed robust. Table 12S (Supporting Information) contains the q^2_{LNO} values for each procedure (triplicate) varying *N* from 2 to 10, the q^2_{LNO} average ($< q^2_{LNO} >$), and the respective two standard deviations (*2 SD*). The plots of LNO and **y**-randomization (10 runs) for model 4 ($N = 30$) from alignment 2 are displayed in Figure 2A.

The **y**-randomization procedure was carried out in 10 (Figure 2B), 30 and 50 runs, in triplicate, and the models obtained under such conditions were, again, of poor quality and without real meaning as reflected on the statistical measures, q^2_{yrand} and r^2_{yrand}. Considering the approach of Eriksson and Wold,[59] model 4 from alignment 2 was characterized as tolerable chance correlation. The q^2_{yrand} values varied from lower than zero (negative values) to 0.1 and the r^2_{yrand} values from 0.1 to 0.39, regarding 10, 30, and 50 runs. Table 13S (Supporting Information) contains the q^2_{yrand} and r^2_{yrand} values for 10 randomization runs. Figure 3S and 4S (Supporting Information) show the findings (table and plot) for 30 and 50 randomization runs, respectively.

To ascertain the external predictive power of model 4 based on screening a test set compounds (Table 1), the pIC_{50} value of each of compounds was calculated using Equation 1, as reported in the external validation step in Methods section. In Table 14S (Supporting Information) the test set GCOD indices, GC*i*(*x*,*y*,*z*), for model 4 are presented. GC*i*(*x*,*y*,*z*) represents each descriptor, GCOD, of model 4 (GC*i* = GC1, GC2, GC3, GC4, and GC5), and *x*, *y*, and *z* are the Cartesian coordinates of the reference grid cell space. The test set predictions are given in Table 4. Six of the eight MMP-2 inhibitors in the test set have residuals ($pIC_{50exp} - pIC_{50calc}$) whose absolute values are less than 0.73, which is the SD value, considering 99% of confidence level. This finding indicates that model 4 has a capacity of prediction of about 75%, which is a good predictive power. Also, model 4 has a high correlation coefficient value between the predicted/calculated and observed/experimental activities of compounds from test set ($R = 0.80$), reinforcing its reliability. The plot of experimental versus calculated activities is presented in Figure 2C.

Among the two compounds not well-predicted by model 4 is compound **14b**. The possible reasons in this regard were earlier discussed. Compound **9a** ($pIC_{50} = 5.90$) has exactly the same substituent groups as compound **5k** ($pIC_{50} = 7.45$), but instead of a sulfonyl group linked to N_9 ($N_9SO_2R_2$) it has an oxygen atom (N_9OR_2). Compound **9b** (training set; $pIC_{50} = 6.54$) has the same chemical change in N_9. But, instead of a methyl group (**9a**) in R_2 position it presents an isopropyl group. Thus, when the sulfonyl group is replaced, bulkier substituent groups in R_2 position seem to improve the MMP-2 inhibition.

After the validation procedures, the bioactive conformation of each inhibitor in the training set was hypothesized using model 4 (Equation 1) from alignment 2 by first identifying all conformer states sampled for each compound within ΔE equal to 5 kcal/mol of the global minimum energy conformation of the CEP. Then, the values of the grid cell occupancy descriptors of each resulting set of low-energy conformers were used to calculate the inhibitory ac-

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

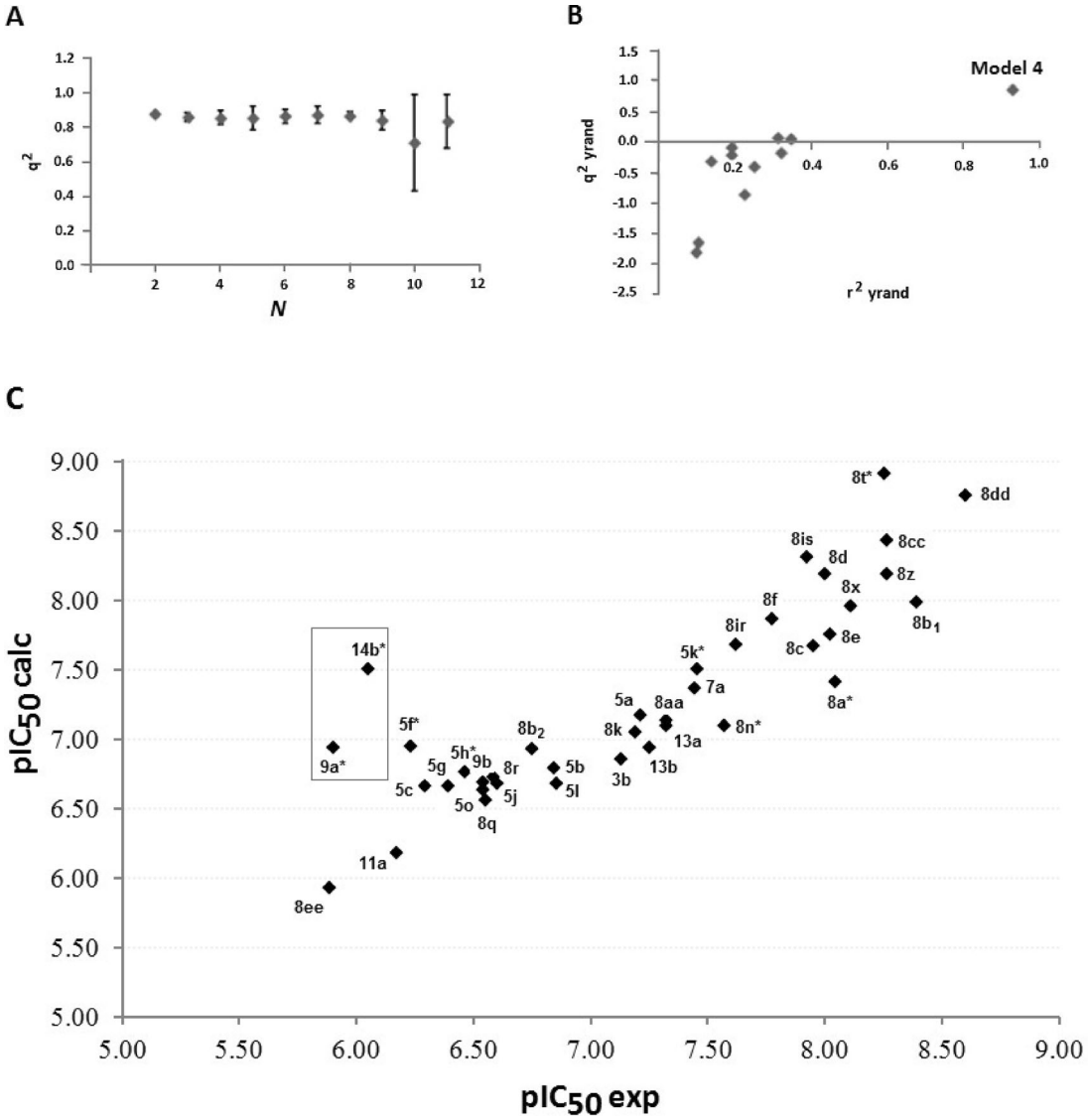


Figure 2. (A) LNO and (B) **y**-randomization (10 runs) plots for model 4 ($N = 30$) from alignment 2. *N* critical was reached when a fraction of nine compounds was removed (A). Model 4 was characterized as tolerable chance correlation ($0.3 < r^2_{yrand} < 0.4$) (B), regarding Eriksson and Wold approach.[59] (C) Experimental versus calculated pIC_{50} values for model 4 ($N = 30$) from alignment 2. Compounds of test set are marked with an asterisk (*). Compound **9a*** and **14b*** (inside the gray square) were not well-predicted by model 4 from alignment 2 considering a confidence level of 99%.

tivities for each compound employing Equation 1 (Model 4/alignment 2), and the conformer with the highest predicted activity was selected as the bioactive conformation of each inhibitor.

The predicted bioactive conformations for two inhibitors are shown in Figure 3. The two inhibitors have been singled out from the training set based upon their activities against MMP-2: compound **8dd** (**A**) being the most active and compound **8ee** (**B**) the least active inhibitor. These conformer states were developed using a grid cell resolution of 1 Å. The grid cell occupancy descriptors (GCOD or GC) that positively contribute (Equation 1) to the inhibitory activity are shown as light-gray spheres (GC1, GC4, and GC5), and those that negatively favor the biological activity are displayed as dark-gray spheres (GC2 and GC3).

Regarding the most active compound, the grid cell occupancy is favored near the R_3 group (GC4, 0.87 Å), the sulfonamide moiety (GC5, 0.39 Å to the oxygen atom), and the R_{1b} region (GC1, 0.82 Å to the carbon atom of CH in the aromatic ring).The IEP class for GC4 is non-polar atom (*np*), and this descriptor is placed closer to the carbon atom $C_{(a)}$ (Figure 3) linked to nitrogen of the dimethylmorpholine group (0.87 Å), meaning the substitution in that part of molecule should be preferentially apolar. The grid cell occupancy descriptors GC2 and GC3 have no differentiation regarding the type of atom (IEP *any*). Then, molecular substi-

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

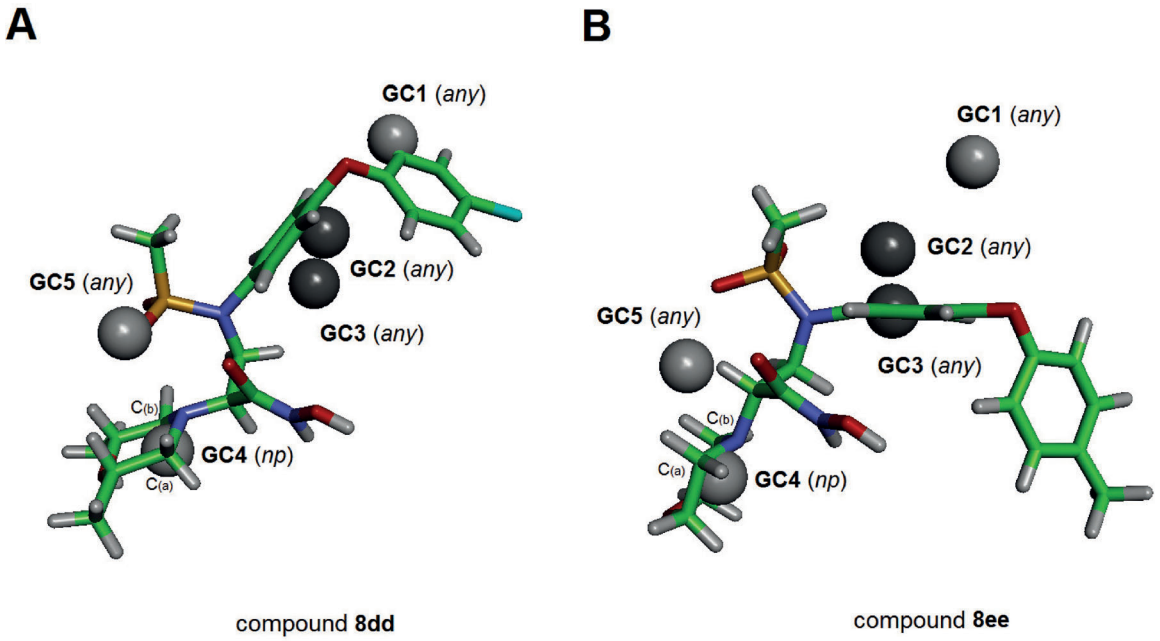


Figure 3. The predicted active conformations for two inhibitors using model 4 from alignment 2: compound **8dd** (the most active; *S* configuration at C_7) (A) and compound **8ee** (the least active, *R* configuration at C_7) (B). The compounds are displayed in stick model. The nitrogen atoms are shown in blue, hydrogens in white, oxygen atoms in red, sulfur in yellow, chlorine in cyan, and carbon atoms are shown in green. The grid cell occupancy descriptors are presented as light-gray (GC1, GC4, and GC5) and dark-gray spheres (GC2 and GC3) (Discovery Studio v. 3.5, Accelrys Software, Inc., 2005–2012), according to their contribution to the biological activity (Eq. 1). GC4 has IPE type atom = non-polar (*np*), and GC1, GC2, GC3, and GC5 have IPE type atom = any. The orientation of molecules was based on the hydroxamic moiety (reference). $C_{(a)}$ and $C_{(b)}$ are the carbon atoms linked to nitrogen of the morpholine group (R_3 group).

tutions of any type should be avoided near the region covered by the aromatic ring attached to the sulfonamide moiety (GC2, 1.21 Å to the carbon atom linked to R_{1a}; and GC3, 2.17 Å to the carbon atom close to the sulfonamide moiety).

Furthermore, compound **8ee** (the least active), except to GC4 (0.94 Å close to $C_{(a)}$), presented different distribution of the grid cell occupancy descriptors, as expected. GC5 is closer to the carbon atom $C_{(b)}$ linked to nitrogen of the morpholine group (2.36 Å; R_3 region) instead of the oxygen atom of the sulfonamide moiety (2.56 Å). GC2 and GC3 are on the region of the aromatic ring attached to the sulfonamide moiety, but 0.10 Å and 1.79 Å, respectively, to the carbon atom next the carbon bound to the nitrogen of the sulfonamide portion. Conversely, the GC1 region is not occupied by the R_{1b} group. GC1 is 6.6 Å distant from the carbon atom (CH, R_{1b}) in the aromatic ring. All these differences were probably responsible for reducing the inhibitory activity against MMP-2.

To determine whether model 4 from alignment 2 sterically fits in the active site and provides complementary inhibitor-enzyme interaction sites, the predicted active conformation of each compound of training set and its respective GCOD were docked into the active site of MMP-2 using the binding alignment 2. This procedure can also aid the differentiation of atom type, particularly, for grid cell occupancy descriptors defined as *any* in the QSAR model selected. The solution structure of a catalytic domain of MMP-2 complexed with a hydroxamic acid inhibitor (SC-74020; Figure 4) (PDB ID 1HOV), determined by three-dimensional heteronuclear NMR spectroscopy,[42] was employed to perform this investigation.

The coordinates of SC-74020 inhibitor were used as reference to dock the predicted active conformations plus grid cell occupancy descriptors (model 4/alignment 2) for the compounds of training set into the MMP-2 catalytic domain. A molecular dynamics simulation of 50 ps sampling time at 298 K, using a time step of 1 fs, was carried out for each complex (MMP-2/predicted active conforma-

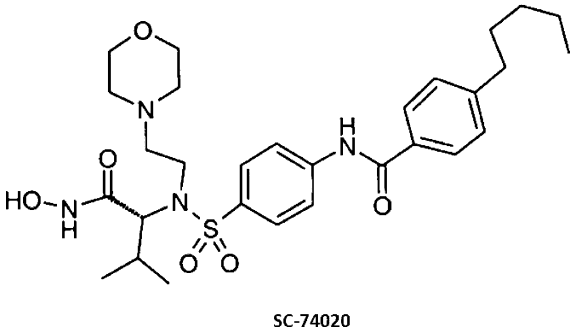


Figure 4. Chemical structure of the MMP-2 inhibitor SC-74020.[42]

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

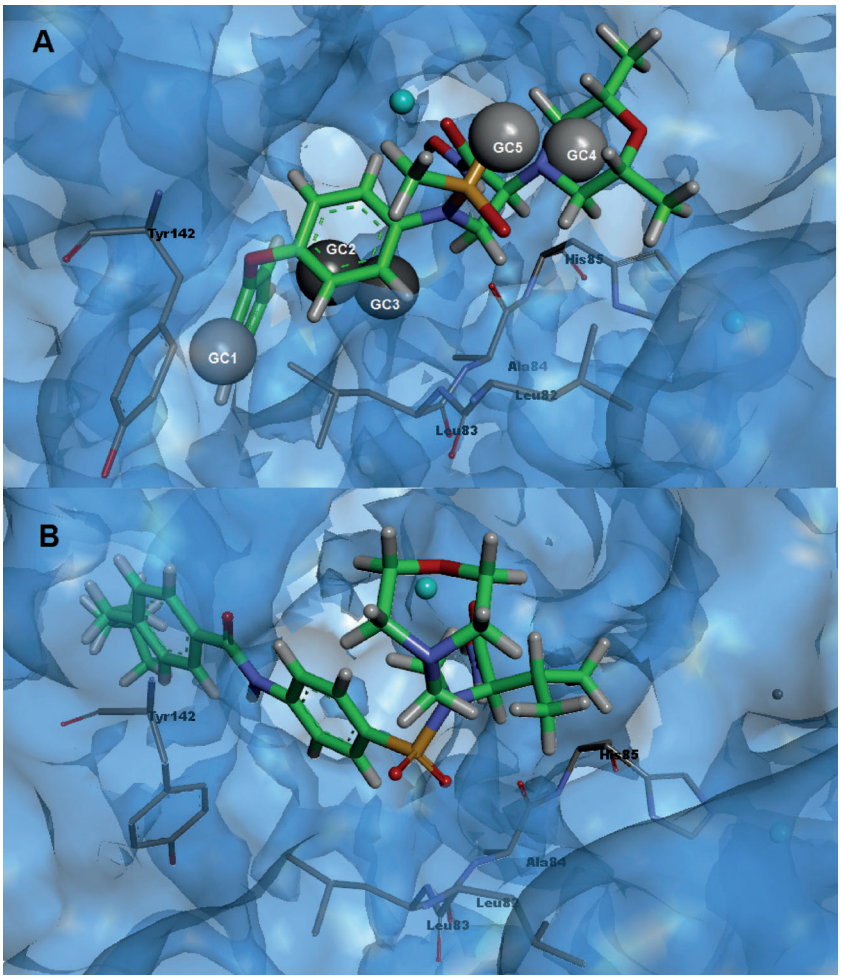


Figure 5. (A) The predicted active conformation for ligand **8dd** (the most active) and its respective descriptors (GCOD, GC1-GC5) and (B) the SC-74020 inhibitor docked in the active site of MMP-2. The amino acid residues Leu82, Leu83, Ala84, His85, and Tyr142 are displayed as stick model and the hydrogen atoms are hidden. The carbon atoms are in gray, oxygen in red, and nitrogen in blue. The soft molecular surface of MMP-2 protein is shown in blue. The ligands are presented as stick model. The carbon atoms are displayed in green, oxygen in red, nitrogen in blue, sulfur in orange, chlorine in cyan, and hydrogen atoms in white. The metals zinc (light blue) and calcium (gray) as well as the descriptors (GCOD, GC; light and dark gray) are shown as CPK (space filling) model.

tion + GCOD), and each energy-minimized complex was considered for further analysis.

The distances between the amino acid residues of active site (Leu82, Leu83, Ala84, His85, Leu116, Val117, His130, Leu137, Met139, Pro140, Ile141, Tyr142, Thr143, Thr145, Lys146, Asn147, Leu150) and the descriptors (GCOD or GC) plus the predicted active conformations were measured. All possible complementary sets of inhibitor-enzyme interactions, based on the selected model/alignment (model 4/alignment 2) of *RI* 4 D-QSAR analysis, can be identified from this process.

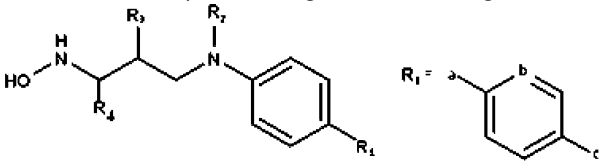
Herein, the distances found for the most active compound of the training set (**8dd**; Figure 5A) were compared to the corresponding distances for the inhibitor SC-74020, in the solution structure of MMP-2 (1HOV; Figure 5B).[42] Despite the ligands **8dd** and SC-74020 belong to related chemical classes, some differences can be pointed out regarding the primary intermolecular interactions established with the amino acid residues in the active site of MMP-2.

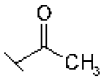
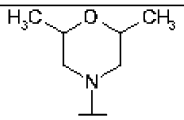
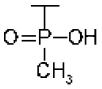
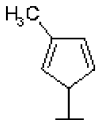
The residues of the bottom of S_1' subsite (Thr145, Lys146, Asn147), for instance, do not establish any kind of interaction with any portion of the most active inhibitor, **8dd**. The **8dd** inhibitor is, actually, shorter than SC-74020, and is aligned in the catalytic site by accommodating the most hydrophobic portion of its structure (R_1) in an adjacent pocket to the S_1' subsite, suggesting a different orientation (binding mode) in MMP-2 for these derivatives. Interestingly, the hydroxamic portion in both inhibitors (**8dd** and SC-74020) interacts with the Zn metal into the active site. The carbonyl and hydroxyl oxygen atoms have showed practically the same distance from Zn (2.1 Å and 2.0 Å).

Biological activity (pIC_{50}) is predicted to increase with ligand atom occupancy of grid cells of GC1, GC4 and GC5 by the appropriate IPE types. The opposite is true for occu-

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 5. Predicted ligands and their calculated inhibitory activities against MMP-2-using model 4/alignment 2.



ligands	R_{1a}	R_{1b}	R_{1c}	R_2	R_3	R_4	$pIC_{50}calc$
L4	O	CH	Cl			H_2	8.12
L6	O	CH	Cl			O	7.33
L7	O	CH	Cl			O	7.03

pancy of GC2 and GC3 (see Equation 1). Except to GC4, which has the IPE atom type as non-polar, the grid cell descriptors do not present any specificity regarding their IPE types.

On the basis of the docking of ligand **8dd** into the active site (Figure 5A), GC4 is placed close to the side chain of His85 residue (β IV strand) (3.9 Å from carbon linked to the imidazole ring), just like the isopropyl group of SC-74020 inhibitor, reinforcing the corresponding hydrophobic interaction in the MMP-2 active site.

The GC1 descriptor is located near the carbon of aromatic ring of the Tyr142 residue side chain (2.4 Å), and also close to the methyl group of Leu82 residue (1.9 Å from carbon, 1.6 Å to 2.3 Å from hydrogen atoms). The IPE atom type of GC1 is *any*, which means all occupancy by any type of atom. The nonspecificity in atom type for GC1 might be explained through the location of GC1 in the active site. GC1 can establish hydrogen bonds with the methyl group of the side chain of Leu82 or have hydrophobic interactions with the side chain of Tyr142.

The GC5 descriptor is near the carbon of the methyl group of the side chain of Leu83 (5.8 Å) as well as close to nitrogen of the imidazole ring of His130 (5.2 Å). Then, GC5 can establish hydrophobic and polar interaction with the side chain of Leu83 and His130, respectively. It is noteworthy that to explore the interactions with the His130 residue could be a novel approach to develop new MMP-2 inhibitors.

Otherwise, substituent groups of any kind in GC2 and GC3 positions would decrease the biological activity. These grid cell occupancy descriptors are related to the aromatic ring attached to the nitrogen of the sulfonamide moiety, which must establish interactions with the Leu82, Leu83 and Ala84 residues into the MMP-2 active site.

4 Conclusions

The best QSAR (model 4/alignment 2) was statistically validated, robust, and did not present chance correlation. Furthermore, the interpretation was ease from the chemical and biological point of view. Thus, the QSAR model is reliable for predicting novel derivatives capable of inhibiting MMP-2. In this regard, a chemical library ($N = 11$) was built up and the inhibitory activity of the ligands was calculated using model 4/alignment 2. The three more promising ligands predicted are shown in Table 5.

Regarding Table 5, the most potent predicted compound, **L4**, pointed out that one zinc-binding group (ZBG) could be enough to maintain the catalytic activity (the Zn coordination). The sulfonamide moiety was replaced by amide in **L4**, but R_1 and R_3 groups were remained the same as the compound **8dd** (the most experimentally active).

The hydrophobic portion (R_1) of the predicted compounds was also accommodated into an adjacent pocket to S_1' subsite, and a different binding mode in the MMP-2 catalytic site might be considered for these derivatives.

Thus, the in silico findings have already provided interesting information to drive the designing and synthesis of novel potentially MMP-2 inhibitors against melanoma invasion.

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Acknowledgements

The authors thank the *São Paulo Research Foundation* (*FAPESP*; Grant# 2010/19142-1), the *Chem21 Group* for the academic license of MOLSIM 3.2, and Professor *Anton J. Hopfinger* (University of New Mexico, Albuquerque, USA) for the collaborative license of a beta-version of FC10-4DQSAR software.

References

[1] J. P. Martins, E. G. Barbosa, K. F. M. Pasqualoto, M. M. C. Ferreira, *J. Chem. Inf. Model.* **2009**, *49*, 1428 – 1436.
[2] C. H. Andrade, K. F. M. Pasqualoto, E. I. Ferreira, A. J. Hopfinger, *Molecules* **2010**, *15*, 3281 – 3294.
[3] R. D. Cramer, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.* **1988**, *110*, 5959 – 5967.
[4] A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, C. Duraiswami, *J. Am. Chem. Soc.* **1997**, *119*, 10509 – 10524.
[5] A. J. Hopfinger, *4D-QSAR Software, Vers. 3.0 – User's Manual*, The Chem21 Group, 1780 Wilson Dr., Lake Forest, IL, 60045, **1999**.
[6] X. Hong, A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 324 – 336.
[7] K. F. M. Pasqualoto, E. I. Ferreira, O. A. Santos-Filho, A. J. Hopfinger, *J. Med. Chem.* **2004**, *47*, 3755 – 3764.
[8] N. C. Romeiro, M. G. Albuquerque, R. B. Alencastro, M. Ravi, A. J. Hopfinger, *J. Comput. Aided Mol. Des.* **2005**, *19*, 385 – 400.
[9] C. H. Andrade, K. F. M. Pasqualoto, E. I. Ferreira, A. J. Hopfinger, *J. Chem. Inf. Model.* **2009**, *49*, 1070 – 1078.
[10] C. H. Andrade, K. F. Pasqualoto, E. I. Ferreira, A. J. Hopfinger, *J. Comput. Aided Mol. Des.* **2010**, *24*, 157 – 172.
[11] D. Pan, J. Liu, C. Senese, A. J. Hopfinger, Y. Tseng, *J. Med. Chem.* **2004**, *47*, 3075 – 3088.
[12] O. A. Santos-Filho, A. J. Hopfinger, *J. Chem. Inf. Model.* **2006**, *46*, 345 – 354.
[13] S. S. Rocha-Pita, M. G. Albuquerque, C. R. Rodrigues, H. C. Castro, A. J. Hopfinger, *Chem. Biol. Drug Des.* **2012**, *79*, 740 – 748.
[14] J. Polanski *Curr. Med. Chem.* **2009**, *16*, 3243 – 3257.
[15] A. M. Doweyko *J. Comput. Aided Mol. Des.* **2008**, *22*, 81 – 89.
[16] A. Jemal, R. C. Tiwari, T. Murray, A. Ghafoor, A. Samueis, E. Ward, E. J. Feuer, M. J. Thun, *CA Cancer J. Clin.* **2004**, *54*, 8 – 29.
[17] WHO – World Health Organization, *Cancer, Fact sheet n°297*, **2014**.
[18] K. Kessenbrock, V. Plaks, Z. Werb, *Cell* **2010**, *141*, 52 – 67.
[19] A. Page-McCaw, A. J. Ewald, Z. Werb, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 221 – 233.
[20] W. C. Parks, C. L. Wilson, Y. S. López-Boado, *Nat. Rev. Immunol.* **2004**, *4*, 617 – 629.
[21] R. Visse, H. Nagase, *Circ. Res.* **2003**, *92*, 827 – 839.
[22] M. Egeblad, Z. Werb, *Nat. Rev. Cancer* **2002**, *2*, 161 – 174.
[23] H. Nagase, R. Visse, G. Murphy, *Cardiovasc. Res.* **2006**, *69*, 562 – 573.
[24] P. Vihinen, V. M. Kähäri, *Int. J. Cancer* **2002**, *99*, 157 – 66.
[25] L. A. Liotta, K. Tryggvason, S. Garbisa, I. Hart, C. M. Foltz, S. Shafie, *Nature* **1980**, *84*, 67 – 68.
[26] C. Gialeli, A. D. Theocharis, N. K. Karamanos, *FEBS J.* **2011**, *278*, 16 – 27.
[27] R. P. Verma, C. Hansch, *Bioorg. Med. Chem.* **2007**, *15*, 2223 – 2268.
[28] B. Fingleton, *Curr. Pharm. Des.* **2007**, *13*, 333 – 346.
[29] S. M. Yang, R. H. Scannevin, B. Wang, S. L. Burke, Z. Huang, P. Karnachi, L. J. Wilson, K. J. Rhodes, B. Lagu, W. V. Murray, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1140 – 1145.
[30] S. M. Yang, R. H. Scannevin, B. Wang, S. L. Burke, L. J. Wilson, P. Karnachi, K. J. Rhodes, B. Lagu, W. V. Murray, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1135 – 1139.
[31] L. Xi, J. Du, S. Li, J. Li, H. Liu, X. Yao, *J. Comput. Chem.* **2010**, *31*, 24 – 42.
[32] C. P. Pereira, E. D. Carmo, V. A. S. Silveira, S. U. Amadei, L. E. B. Rosa, *Rev. Bras. Cancerol.* **2006**, *52*, 257 – 262.
[33] E. Kerkelä, U. Saarialho-Kere, *Exp. Derma.* **2003**, *12*, 109 – 125.
[34] W. P. Li, C. J. Anderson, *Q. J. Nucl. Med.* **2003**, *47*, 201 – 208.
[35] M. Hidalgo, S. G. Eckhardt, *J. Natl. Cancer Inst.* **2001**, *93*, 178 – 193.
[36] K. M. Turra, K. F. M. Pasqualoto, S. B. M. Barros, *Int. J. Quantum Chem.* **2012**, *112*, 3374 – 3389.
[37] R. Kiralj, M. M. C. Ferreira, *J. Braz. Chem. Soc.* **2009**, *20*, 770 – 787.
[38] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **2002**, *20*, 269 – 276.
[39] P. Gramatica, *QSAR Comb. Sci.* **2007**, *26*, 694 – 701.
[40] *HyperChem Program Release 8 for Windows*, Hypercube, Inc., Gainesville, FL, **2007** (evaluation copy).
[41] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235 – 242.
[42] Y. Feng, J. J. Likos, L. Zhu, H. Woodward, G. Munie, J. J. Mcdonald, A. M. Stevens, C. P. Howard, G. A. de Crescenzo, D. Welsch, H. S. Shieh, W. C. Stallings, *Biochim. Biophys. Acta* **2002**, *1598*, 10 – 23.
[43] N. L. Allinger, *J. Am. Chem. Soc.* **1977**, *99*, 8127 – 8134.
[44] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902 – 3909.
[45] D. Doherty, *MOLSIM: Molecular Mechanics and Dynamics Simulation Software, Version 3.2 – User's Guide*, The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA, **2002**.
[46] S. H. Chung, O. S. Andersen, V. V. Krishnamurthy, *Biological Membrane Ion Channels: Dynamics, Structure, and Applications*, Springer, Berlin, **2007**.
[47] W. J. Dunn III, D. R. Scott, W. G. Glen, *Tetrahedron Comp. Methodol.* **1989**, *2*, 349 – 376.
[48] C. Hansch, A. Leo, D. H. Hoekman, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, **1995**.
[49] J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT, Cambridge, **1992**.
[50] D. Rogers, A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854 – 866.
[51] D. Rogers, *WOLF Genetic Function Approximation, Reference Manual, Vers. 5.5*, Molecular Simulation Inc., Burlington, MA, **1994**.
[52] W. J. Dunn III, D. Rogers, in *Genetic Algorithms in Molecular Modeling* (Ed: J. Devillers), Academic Press, London, **1996**, pp. 109 – 130.
[53] D. Rogers, in *Proc. 4th Int. Conf. on Genetic Algorithms* (Eds: R. K. Belew, L. B. Booker) Morgan Kaufmann, San Francisco, **1991**, pp. 38 – 46.
[54] J. H. Friedman, *Ann. Statist.* **1991**, *19*, 1 – 67.
[55] K. Baumann, N. Stiefl, *J. Comput. Aided Mol. Des.* **2004**, *18*, 549 – 562.
[56] R. D. Clark, P. C. Fox, *J. Comput. Aided Mol. Des.* **2004**, *18*, 563 – 576.
[57] K. Baumann, *Trends Anal. Chem.* **2003**, *22*, 395 – 406.

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[58] L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, P. Gramatica, *Environ. Health Perspect.* **2003**, *111*, 1361 – 1375.
[59] S. Wold, L. Eriksson, S. Clementi, in *Chemometric Methods in Molecular Design* (Ed: H. van de Waterbeemd), Wiley, Weinheim, **1995**, pp 309 – 338.
[60] C. Rücker, G. Rücker, M. Meringer, *J. Chem. Inf. Model.* **2007**, *47*, 2345 – 2357.
[61] J. S. Tokarski, A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 779 – 791.
[62] K. H. Forsythe, A. J. Hopfinger, *Macromolecules* **1973**, *6*, 423 – 437.

Received: May 12, 2014
Accepted: June 25, 2014
Published online: August 26, 2014

 © 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Janus - Sistema Administrativo da Pós-Graduação

Universidade de São Paulo
Faculdade de Ciências Farmacêuticas
FICHA DO ALUNO

9138 - 6603535/2 - Kely Medeiros Turra

Email:	kely.turra@usp.br
Data de Nascimento:	04/04/1985
Cédula de Identidade:	RG - 41.490.439-4 - SP
Local de Nascimento:	Estado de São Paulo
Nacionalidade:	Brasileira
Graduação:	Farmacêutico - Universidade de Mogi das Cruzes - São Paulo - Brasil - 2008
Mestrado:	Mestre em Ciências - Área: Insumos Farmacêuticos - Faculdade de Ciências Farmacêuticas - Universidade de São Paulo - São Paulo - Brasil - 2010

Curso:	Doutorado
Programa:	Fármaco e Medicamentos
Área:	Insumos Farmacêuticos
Data de Matrícula:	01/02/2011
Início da Contagem de Prazo:	01/02/2011
Data Limite para o Depósito:	02/02/2015
Orientador:	Prof(a). Dr(a). Silvia Berlanga de Moraes Barros - 01/02/2011 até o presente. E.Mail: smbarros@usp.br
Coorientador:	Prof(a). Dr(a). Kerly Fernanda Mesquita Pasqualoto - 18/04/2012 até o presente. E.Mail: kerlypasqualoto@gmail.com
Proficiência em Línguas:	Inglês, Aprovado em 01/02/2011
Data de Aprovação no Exame de Qualificação:	Aprovado em 22/04/2013
Data do Depósito do Trabalho:	17/12/2014
Título do Trabalho:	".."
Data Máxima para Aprovação da Banca:	19/02/2015
Data de Aprovação da Banca:	
Data Máxima para Defesa:	
Data da Defesa:	
Resultado da Defesa:	

Aluno matriculado no Regimento da Pós-Graduação USP (Resolução nº 5473 em vigor de 18/09/2008 até 19/04/2013)
Última ocorrência: Matrícula de Acompanhamento em 15/07/2014
Impresso em: 15/01/15 14:06:21

Miriam Alves Wrigg
SPG/FCF/USP
Nº USP 3646184

Faculdade de Ciências Farmacêuticas - USP
Secretaria de Pós-Graduação

Universidade de São Paulo
Faculdade de Ciências Farmacêuticas
FICHA DO ALUNO

9138 - 6603535/2 - Kely Medeiros Turra

Sigla	Nome da Disciplina	Início	Término	Carga Horária	Cred.	Freq.	Conc.	Exc.	Situação
FBF5779-1/3	Preparo de Artigos Científicos na Área de Farmácia	08/04/2011	09/06/2011	90	6	100	A	N	Concluída
FBF5780-1/3	Estratégias Farmacêuticas de Prevenção do Dano Cutâneo Induzido por Radiação Solar: Antioxidantes Naturais	24/05/2011	20/06/2011	60	4	100	A	N	Concluída
FBF5794-1/4	Tópicos Gerais de Fármaco e Medicamentos II	16/08/2011	24/10/2011	30	2	100	A	N	Concluída
FBF5734-7/2	Mecanismos de Ação dos Fármacos aos Níveis Molecular e Eletrônico	06/03/2012	18/06/2012	90	6	100	A	N	Concluída
Atividade do Programa	Publicação do trabalho "A novel set of beta-N-biaryl ether sulfonamide hydroxamates as potential MMPs inhibitors: molecular dynamics simulations and molecular properties evaluation", no International Journal of Quantum Chemistry, v. 112, p. 3374-3389, Estados Unidos, 2012 (1)	15/10/2012	15/10/2012	-	2	0	-	-	-

	Créditos mínimos exigidos		Créditos obtidos
	Para exame de qualificação	Para depósito de tese	
Disciplinas:	0	20	20
Estágios:			
Total:	0	20	20

Créditos Atribuídos à Tese: 167

Observações:

1) Créditos atribuídos de acordo com o disposto no Artigo 65 do Regimento de Pós-Graduação e aprovados pela Comissão de Pós-Graduação, em Sessão de 20/03/2013.

Conceito a partir de 02/01/1997:

A - Excelente, com direito a crédito; B - Bom, com direito a crédito; C - Regular, com direito a crédito; R - Reprovado; T - Transferência.

Um(1) crédito equivale a 15 horas de atividade programada.

Última ocorrência: Matrícula de Acompanhamento em 15/07/2014

Impresso em: 15/01/15 14:06:21

Miriam Alves Wrigg
SPG/FCF/USP
Nº USP 3646184

Faculdade de Ciências Farmacêuticas - USP - Secretaria de Pós-Graduação

Janus - Sistema Administrativo da Pós-Graduação

Universidade de São Paulo
Faculdade de Ciências Farmacêuticas
FICHA DO ALUNO

9138 - 6603535/2 - Kely Medeiros Turra

Comissão julgadora da tese de doutorado:

NUSP	Nome	Vínculo	Função
2083397	Silvia Berlanga de Moraes Barros	FCF - USP	Presidente

Última ocorrência: Matrícula de Acompanhamento em 15/07/2014
Impresso em: 15/01/15 14:06:22

Miriam Alves Wrigg
SPG/FCF/USP
Nº USP 3646184

Faculdade de Ciências Farmacêuticas - USP - Secretaria de Pós-Graduação