BIBLIOTECA
INSTITUTO DE QUÍMICA
Universidade de São Paulo

21.403

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

**RELAÇÕES QUANTITATIVAS ENTRE ESTRUTURA QUÍMICA E
ATIVIDADE BIOLÓGICA (QSAR/QSAR-3D) DE COMPOSTOS COM
POTENCIAL ATIVIDADE ANTITUBERCULOSE**

Hamilton Mitsugu Ishiki

Tese de Doutorado

ORIENTADORA: Prof. Dra. Antonia Tavares do Amaral

SÃO PAULO
Vinte de maio de dois mil e cinco

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

I79r	<p>Ishiki, Hamilton Mitsugu</p> <p>Relações quantitativas entre estrutura química e atividade biológica (QSAR/QSAR-3D) de compostos com potencial atividade antituberculose / Hamilton Mitsugu Ishiki -- São Paulo, 2005. 253p.</p> <p>Tese (doutorado) – Instituto de Química da Universidade de São Paulo. Departamento de Química Fundamental. Orientador: Amaral, Antonia Tavares do</p> <p>1. Composto orgânico : Propriedades : Química 2. Relações quantitativas estrutura química e atividade biológica (QSAR) : Bioquímica 3. Relações quantitativas estrutura química e atividade biológica (QSAR): Química farmacêutica 4. Química médica I. T. II Amaral, Antonia Tavares do, orientador.</p> <p style="text-align: right;">547 CDD</p>
------	---

“Relações quantitativas entre estrutura química e atividade biológica (QSAR/QSAR-3D) de compostos com potencial atividade antituberculose”

HAMILTON MITSUGU ISHIKI

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Química – Área: Química Orgânica.

Aprovado(a) por:

Profa. Dra. ANTONIA TAVARES DO AMARAL
IQ – USP
(Orientadora e Presidente)

Prof. Dr. FRANK HERBERT QUINA
IQ – USP

Prof. Dr. JOÃO PEDRO SIMON FARAH
IQ – USP

Prof. Dr. YUJI TAKAHATA
IQ – UNICAMP

Prof. Dr. RICARDO BICCA DE ALENCASTRO
UFRJ

SÃO PAULO
25 DE JULHO 2005

O tempo, a ausência e a distância jamais poderão tirar do meu coração, a imagem daqueles que souberam cativar a minha amizade. Cada pessoa que passa em nossa vida, passa sozinha, porque cada pessoa é única e nenhuma substitui a outra. Cada pessoa que passa em nossa vida, passa sozinha, mas não vai sozinha e não nos deixará só, porque levam um pouco de nós e deixa um pouco de si. Há os que levam pouco e deixam muito, e há os que deixam pouco e levam muito. Essa é a mais bela responsabilidade da vida e a prova de que não nos encontramos por acaso.

Chaplin

DEDICATÓRIA

*ESTA TESE É DEDICADA A UMA PESSOA MUITO ESPECIAL
QUE SEMPRE ME AUXILIOU E ME APOIOU EM TODOS OS MOMENTOS
SENDO ELES ALEGRES OU TRISTES.....*

À MINHA QUERIDA MÃE, ELZA

AGRADECIMENTOS

Primeiramente agradeço a Deus por todas as oportunidades fornecidas em minha vida, em especial a de desenvolver este trabalho e pela saúde dada para que eu o completasse.

À Prof. Antonia Tavares do Amaral pela orientação neste estudo e pela paciência e compreensão nestes anos em que trabalhamos juntos.

Ao Prof. Hugo Kubinyi, da Universidade de Heidelberg, Alemanha, pelas sugestões dadas e pelas discussões realizadas.

Ao Prof. Gerhard Klebe, da Universidade Philipps, Alemanha, que nos auxiliou na fase inicial da modelagem da proteína e no início dos estudos empregando as abordagens CoMSIA.

Ao Prof. Célio Lopes Silva, da Faculdade de Medicina de Ribeirão Preto – USP, e a Dra. Ana Olívia de Souza, atualmente no Instituto Butantã, que realizaram os testes biológicos para os compostos preparados.

Ao Prof. Laerte Oliveira, do Departamento de Biofísica - UNIFESP, que auxiliou no início dos estudos da modelagem por homologia da proteína e ao Prof. Gert Vried, da Universidade de Nijmegen, Holanda, que gentilmente nos cedeu o programa WHATIF.

Ao Laboratório de Computação Científica e Avançada da USP, *LCCA*, pelos recursos computacionais disponibilizados, em especial à Rosa e ao Francisco que sempre nos auxiliou prontamente.

Ao CNPq pelo fomento fornecido no período de setembro 2001 a fevereiro de 2002.

Às minhas tias Luisa e Marina que durante os estudos me acolheram em seu lar.

Aos técnicos do laboratório Enio e Inocência pelo carinho com que me receberam e pelos auxílios prestados.

Aos amigos Alberto, Leandro e Marcus, pelas discussões científicas e as ajudas prestadas em vários momentos e pelo companheirismo.

Aos amigos que conheci no laboratório, pelos momentos agradáveis de convivência e descontração: Annelies, Carla, Carlos, Cristiano, Elba, Fabiana, Leonardo, Maurício e Weber.

Às amigas Bianca e Priscilla, por nossa velha amizade.

Aos funcionários da Secretaria de Pós-Graduação do IQ.

SUMÁRIO

ÍNDICE DE FIGURAS.....	i
ÍNDICE DE TABELAS.....	viii
RESUMO.....	x
ABSTRACT.....	xiv
GLOSSÁRIO.....	xvi
OBJETIVOS.....	xvii
1. INTRODUÇÃO.....	1
1.1 Tuberculose.....	1
1.2 Ribonucleotídeo redutase.....	6
1.3 Relações Quantitativas Estrutura-Atividade.....	12
1.3.1 <i>QSAR</i> clássico.....	12
1.3.2 Métodos de modelagem sem o conhecimento do alvo biológico.....	19
1.3.2.1 Análise da Forma Molecular – <i>MSA</i>	19
1.3.2.2 Análise Comparativa dos Campos Moleculares – <i>CoMFA</i>	21
1.3.2.3 Análise Comparativa dos Índices de Similaridade Molecular – <i>CoMSIA</i>	26
1.3.2.4 Grupos de análise, conformação bioativa e alinhamento dos compostos.....	28
1.3.2.5 Análise dos resultados: visualização dos mapas de contorno.....	33
1.3.2.6 Metodologias para validação dos modelos de <i>QSAR</i> clássico.....	34
1.3.2.6.1 A regra <i>QUIK</i>	36
1.3.2.6.2 A regra do Q^2 assintótico.....	36
1.3.2.6.3 A regra da função <i>R</i>	37
1.3.3 Métodos de modelagem com o conhecimento do alvo biológico.....	40
1.3.3.1 Dinâmica Molecular.....	42
1.3.3.2 Método de Monte Carlo.....	43
1.3.3.3 Algoritmos Genéticos.....	45
1.3.3.4 Métodos baseados em fragmentos.....	47
1.3.3.5 Métodos de ponto complementar.....	48
1.3.3.6 Métodos da distância de geometrias.....	49
1.3.3.7 Buscas sistemáticas.....	50
1.3.3.8 Modelagem do alvo: proteínas.....	50
1.3.3.8.1 Identificação e seleção de proteínas molde.....	52
1.3.3.8.2 Alinhamento das seqüências de resíduos de aminoácidos.....	53
1.3.3.8.3 Construção do modelo.....	54
1.3.3.8.3.1 Modelagem das regiões estruturalmente conservadas.....	54
1.3.3.8.3.2 Modelagem das regiões de alças.....	54
1.3.3.8.3.3 Modelagem das cadeias laterais.....	55
1.3.3.8.3.4 Otimização do modelo gerado.....	57
1.3.3.8.3.5 Validação do modelo.....	57
1.3.3.8.4 Limitações do método e possíveis erros.....	59
2. MATERIAIS E METODOLOGIA.....	61
2.1 Reagentes e solventes.....	61
2.2 Equipamentos.....	61
2.3 Séries de compostos estudados e parâmetros biológicos.....	62
2.4 Séries de treinamento e de teste.....	65
2.5 Métodos e recursos computacionais.....	66
2.6 Obtenção dos descritores moleculares por cálculo e seleção das variáveis.....	67
2.7 Conformação bioativa e alinhamentos dos compostos.....	68
2.8 Modelagem molecular.....	69

2.8.1 Abordagem extratermodinâmica (<i>QSAR</i> clássico).....	70
2.8.2 Procedimento pela abordagem <i>CoMFA</i>	70
2.8.3 Procedimento pela abordagem <i>CoMSIA</i>	71
2.8.4 Visualização dos mapas de contorno.....	72
2.8.5 Modelagem da proteína.....	72
2.9 Métodos de preparação.....	73
2.9.1 Preparação do cloridrato de tiossemicarbazida.....	73
2.9.2 Preparação do benzenocarboxaldeidotiossemicarbazona.....	74
2.9.3 Preparação da 2-piridino carboxaldeído tiossemicarbazona.....	74
2.10 Ensaio biológicos.....	75
3. RESULTADOS E DISCUSSÃO.....	76
3.1 Introdução.....	76
3.2 Escolha da série de compostos.....	78
3.3 Otimização das geometrias.....	80
3.4 Obtenção dos descritores moleculares por cálculo e seleção dos descritores.....	81
3.5 Alinhamento dos compostos.....	94
3.6 Estudos de <i>QSAR</i> clássico (abordagem extratermodinâmica).....	95
3.7 Estudos de <i>CoMFA</i>	146
3.8 Estudos de <i>CoMSIA</i>	167
3.9 Proposição de novos compostos.....	188
3.10 Estudos de modelagem por homologia da proteína.....	191
3.11 Preparação dos compostos e ensaios biológicos.....	207
3.11.1 Preparação do cloridrato de tiossemicarbazida.....	207
3.11.2 Preparação do benzenocarboxaldeidotiossemicarbazona.....	207
3.11.3 Preparação da 2-piridino carboxaldeído tiossemicarbazona.....	208
3.11.4 Ensaio biológicos.....	208
4. CONCLUSÕES.....	210
5. REFERÊNCIAS BIBLIOGRÁFICAS.....	213
6. APÊNDICE.....	224

ÍNDICE DE FIGURAS

Figura 1. Estruturas dos principais fármacos de primeira linha utilizados no tratamento da tuberculose.	4
Figura 2. Estruturas dos principais fármacos de segunda linha utilizados no tratamento da tuberculose.	4
Figura 3. Reação da ribonucleotídeo redutase. A redução do ribonucleotídeo 5'-difosfato para o correspondente 2'-desoxi-ribonucleotídeo 5'-difosfato requer a presença do ribonucleotídeo específico, o 5'-trifosfato, como um efetor alostérico positivo e por um sistema de redução auxiliar composto por trioxina, trioxina redutase e NADPH ou por glutarredoxina, glutatona, glutatona redutase e NADPH.	7
Figura 4. Exemplos típicos de subestruturas utilizadas para a determinação dos índices topológicos.	17
Figura 5. Estruturas do <i>p</i> -nitrotolueno (a) e limoneno (b) e o respectivo gráfico molecular (c).	18
Figura 6. Graus de vértice (d^v) corrigidos para as estruturas do <i>p</i> -nitrotolueno (a) e limoneno (b).	18
Figura 7. Etapas do processo da análise comparativa dos campos moleculares (<i>CoMFA</i>).	21
Figura 8. Processo de autovalidação <i>PLS</i> .	24
Figura 9. Perfil da aproximação Gaussiana utilizada na metodologia <i>CoMSIA</i> e perfis dos potenciais de Lennard-Jones e Coulomb, utilizados na metodologia <i>CoMFA</i> , indicando o valor limite do potencial definido pelos valores de corte.	26
Figura 10. Alinhamentos do metotrexato e do dihidrofolato realizados pelo modo "rígido" (átomo a átomo) e através da similaridade de campos.	30
Figura 11. Mapas de contorno estérico relacionado à afinidade dos receptores D2 e D4 da dopamina. Nas áreas dos poliedros verdes a presença de grupos volumosos deverá aumentar a afinidade com o sítio de ligação ao passo que nas áreas dos poliedros vermelhos a presença de grupos volumosos deverá diminuir a afinidade com o sítio de ligação.	34
Figura 12. Representação dos ângulos torsionais da cadeia principal (ϕ e ψ) e das cadeias laterais (χ_n).	56
Figura 13. Estrutura molecular fundamental dos derivados da α -(N)-piridino carboxaldeído tiossemicarbazona substituídos, estudados neste trabalho e retirados de literatura selecionada (French & Blanz-Jr. 1974).	63
Figura 14. Gráficos dos valores de r^2 , Q_{cv}^2 e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste I, obtidos para a série de 40 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	83
Figura 15. Gráficos dos valores de r^2 , Q_{cv}^2 e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste II, obtidos para a série de 40 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	83
Figura 16. Gráficos dos valores de r^2 , Q_{cv}^2 e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste III, obtidos para a série de 40 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	84
Figura 17. Gráficos dos valores de r^2 , Q_{cv}^2 e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste I, obtidos para a série de 39 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	84
Figura 18. Gráficos dos valores de r^2 , Q_{cv}^2 e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste II, obtidos para a série de 39 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	85

Figura 19. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste III, obtidos para a série de 39 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.....	85
Figura 20. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste I, obtidos para a série de 30 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	86
Figura 21. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste II, obtidos para a série de 30 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	86
Figura 22. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste III, obtidos para a série de 30 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.....	87
Figura 23. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste I, obtidos para a série de 23 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	87
Figura 24. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste II, obtidos para a série de 23 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	88
Figura 25. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste III, obtidos para a série de 23 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.....	88
Figura 26. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste I, obtidos para a série de 22 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	89
Figura 27. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste II, obtidos para a série de 22 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.	89
Figura 28. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises <i>PLS</i> , para os grupos de treinamento e de teste III, obtidos para a série de 22 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.....	90
Figura 29. Resultados dos alinhamentos dos 40 derivados da α -(N)-piridino carboxaldeído tiossemicarbazona realizados por dois procedimentos, utilizando o composto 1 como molde: (a) átomo-a-átomo e b) similaridade estérica e eletrostática utilizando-se o programa SEAL.	95
Figura 30. Ilustração da estrutura do composto mais ativo da série E (composto 14) e o valor da distância ($\sim 5,0\text{\AA}$) entre os átomos de nitrogênio do fragmento da tiossemicarbazona e do carbono 5 do anel piridínico.	118
Figura 31. Gráfico da variação dos valores do descritor Jhetv, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 22 compostos (Série E), selecionados da literatura (French & Blanz-Jr. 1974).	119
Figura 32. Gráfico da variação dos valores do descritor Mor07p, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 22 compostos (Série E), selecionados da literatura (French & Blanz-Jr. 1974).	120

Figura 33. Gráfico da variação dos valores do descritor CIC2, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 22 compostos (Série E), selecionados da literatura (French & Blanz-Jr. 1974).	120
Figura 34. Ilustração da estrutura do composto mais ativo da série E (composto 14) e o valor da distância ($\sim 6,5\text{\AA}$) entre os átomos de carbono da tiocarbonila e o carbono 6 do anel piridínico.	121
Figura 35. Gráfico da variação dos valores do descritor SP13, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 22 compostos (Série E), selecionados da literatura (French & Blanz-Jr. 1974).	121
Figura 36. Gráfico dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 22 compostos (Série E) inibidores da RNR, obtido utilizando-se o modelo expresso pela equação 237.	122
Figura 37. Ilustração da estrutura do composto mais ativo da série D (composto 14) e o valor da distância ($\sim 6,0\text{\AA}$) entre os átomos de nitrogênio 4', do fragmento da tiossemicarbazona, e do carbono 2 do anel piridínico.	123
Figura 38. Gráfico da variação dos valores do descritor MAXDP, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 23 compostos (série D), selecionados da literatura (French & Blanz-Jr. 1974).	123
Figura 39. Gráfico da variação dos valores do descritor H2e, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 23 compostos (série D), selecionados da literatura (French & Blanz-Jr. 1974).	124
Figura 40. Gráfico da variação dos valores do descritor SEige, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 23 compostos (série D), retirados da literatura (French & Blanz-Jr. 1974).	124
Figura 41. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 23 compostos (Série D) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 192 (a), 193 (b), 201 (c) e 194 (d), respectivamente.	125
Figura 42. Ilustração da estrutura do composto mais ativo da série D (composto 14) e o valor da distância ($\sim 11,5\text{\AA}$) entre os átomos de hidrogênio do nitrogênio 4', do fragmento da tiossemicarbazona, e do carbono 2 do grupo substituinte etila na posição 5 do anel piridínico.	126
Figura 43. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 23 compostos (Série D) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 211 (a), 207 (b), 208 (c) e 210 (d), respectivamente.	126
Figura 44. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 23 compostos (Série D) inibidores da RNR, obtidos utilizando-se os modelos das equações 219 (a), 220 (b), 215 (c) e 217 (d), respectivamente.	127
Figura 45. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 23 compostos (Série D) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 224 (a), 227 (b) e 229 (c), respectivamente.	128
Figura 46. Gráfico da variação dos valores do descritor GATS4e, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 30 compostos (Série C), selecionados da literatura (French & Blanz-Jr. 1974).	129
Figura 47. Gráfico da variação dos valores do descritor BEHv6, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 30 compostos (Série C), selecionados da literatura (French & Blanz-Jr. 1974).	130
Figura 48. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 30 compostos (Série C) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 150 (a), 145 (b), 147 (c) e 148 (d), respectivamente.	131
Figura 49. Gráfico da variação dos valores do descritor GATS3e, obtidos com o emprego do programa DRAGON em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 30 compostos (Série C), selecionados da literatura (French & Blanz-Jr. 1974).	131

Figura 50. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 30 compostos (Série C) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 161 (a), 158 (b), 156 (c) e 159 (d), respectivamente.....	132
Figura 51. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 30 compostos (Série C) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 170 (a), 169 (b), 166 (c) e 168 (d), respectivamente.....	133
Figura 52. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 30 compostos (Série C) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 179 (a), 174 (b), 178 (c) e 180 (d), respectivamente.....	134
Figura 53. Gráfico da variação dos valores do descritor HATSp, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 39 compostos (Série B), selecionados da literatura (French & Blanz-Jr. 1974).....	135
Figura 54. Gráfico da variação dos valores do descritor MATS8m, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 39 compostos (Série B), selecionados da literatura (French & Blanz-Jr. 1974).....	135
Figura 55. Gráfico dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 39 compostos (Série B) inibidores da RNR, obtido utilizando-se o modelo expresso pela equação 95.	136
Figura 56. Gráficos da variação dos valores dos descritores, respectivamente, GATS3e (a) e Mor08m (b), obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 39 compostos (série B), retirados da literatura (French & Blanz-Jr. 1974).	136
Figura 57. Gráfico dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 39 compostos (Série B) inibidores da RNR, obtido utilizando-se o modelo da equação 102.	137
Figura 58. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 39 compostos (Série B) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 113 (a), 112 (b), 118 (c) e 116 (d), respectivamente.....	138
Figura 59. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 39 compostos (Série E) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 127 (a), 124 (b), 125 (c) e 123 (d), respectivamente.....	139
Figura 60. Gráfico da variação dos valores do descritor R1u, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 40 compostos (série A), retirados da literatura (French & Blanz-Jr. 1974).....	140
Figura 61. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 40 compostos (Série A) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 51 (a), 53 (b), 50 (c) e 46 (d), respectivamente.	141
Figura 62. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 40 compostos (Série A) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 57 (a), 62 (b) e 60 (c), respectivamente.	141
Figura 63. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 40 compostos (Série A) inibidores da RNR, obtidos utilizando-se os modelos das equações 72 (a), 70 (b), 68 (c) e 75 (d), respectivamente.....	142
Figura 64. Gráfico da variação dos valores do descritor IC2, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 40 compostos (Série A), selecionados da literatura (French & Blanz-Jr. 1974).....	143
Figura 65. Gráficos dos valores de pIC_{50} observados <i>versus</i> calculados para a série de 40 compostos (Série A) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 80 (a), 85 (b), 78 (c) e 79 (d), respectivamente.	144
Figura 66. Gráficos dos valores de pIC_{50} calculados <i>versus</i> observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método <i>CoMFA</i> , utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IA), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIA) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIA)	

incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 40 compostos (série A).	147
Figura 67. Gráficos dos valores de pIC_{50} calculados <i>versus</i> observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método <i>CoMFA</i> , utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IB), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIB) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIB) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 39 compostos (série B).	149
Figura 68. Gráficos dos valores de pIC_{50} calculados <i>versus</i> observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método <i>CoMFA</i> , utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IC), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIC) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIC) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 30 compostos (série C).	151
Figura 69. Gráficos dos valores de pIC_{50} calculados <i>versus</i> observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método <i>CoMFA</i> , utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento ID), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IID) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIID) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 23 compostos (série D).	153
Figura 70. Gráficos dos valores de pIC_{50} calculados <i>versus</i> observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método <i>CoMFA</i> , utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IE), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIE) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIE) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 22 compostos (série E).	155
Figura 71. Mapas de contorno gerados por <i>CoMFA</i> (modelo 2, grupo de treinamento IIB, com 39 compostos (série B) alinhados pelo programa SEAL – Tabela 21) para os derivados da α -(N)-piridino carboxaldeído tiossemicarbazona.	165
(a) Mapas de contorno eletrostático: as regiões dos poliedros verdes indicam que a presença de grupos substituintes com densidade negativa aumenta os valores da atividade biológica ao passo que as regiões dos poliedros vermelhos indicam que a presença de grupos com densidade positiva são favoráveis à atividade.	
(b) Mapas de contorno estérico: as regiões dos poliedros verdes indicam que a presença de grupos volumosos favorece a atividade biológica e as regiões dos poliedros vermelhos indicam que a presença de grupos volumosos desfavorece a atividade biológica. Um dos compostos mais ativos, composto 1, está situado no centro.	
Figura 72. Gráficos dos valores de pIC_{50} calculados <i>versus</i> observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método <i>CoMSIA</i> , utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IA), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIA) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIA) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 40 compostos (série A).	168
Figura 73. Gráficos dos valores de pIC_{50} calculados <i>versus</i> observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método <i>CoMSIA</i> , utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IB), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIB) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIB) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 39 compostos (série B).	170
Figura 74. Gráficos dos valores de pIC_{50} calculados <i>versus</i> observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método <i>CoMSIA</i> , utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IC), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIC) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIC) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 30 compostos (série C).	172

- Figura 75. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMSIA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento ID), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IID) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIID) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 23 compostos (série D). 174
- Figura 76. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMSIA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IE), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIE) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIE) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 22 compostos (série E). 176
- Figura 77. Mapas de contorno gerados por *CoMSIA* (modelo 1, grupo de treinamento 1B, com 39 compostos (série B) alinhados pelo N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona (R_B) – Tabela 26) para os derivados da α -(N)-piridino carboxaldeído tiossemicarbazona. 186
- (a) Mapas de contorno eletrostático: as regiões dos poliedros verdes indicam que a presença de grupos com densidade negativa aumentam o valor da atividade biológica enquanto que as regiões dos poliedros vermelhos indicam que a presença de grupos com densidade positiva são favoráveis à atividade.
- (b) Mapas de contorno estérico: as regiões dos poliedros verdes indicam que a presença de grupos relativamente volumosos favorece a atividade biológica e as regiões dos poliedros vermelhos indicam que a presença de grupos relativamente volumosos desfavorece a atividade biológica.
- (c) Mapas de contorno hidrofóbico: as regiões dos poliedros verdes indicam que a presença de grupos hidrofóbicos aumenta os valores da atividade biológica e as regiões dos poliedros vermelhos indicam que grupos hidrofílicos aumentam os valores da atividade.
- (d) Mapas de contorno de doadores de ligação de hidrogênio: as regiões dos poliedros verdes indicam que a presença de grupos doadores de ligação de hidrogênio aumenta os valores da atividade dos compostos e nas regiões dos poliedros vermelhos estes grupos diminuem os valores da atividade biológica.
- (e) Mapas de contorno de aceptores de ligação de hidrogênio: as regiões dos poliedros verdes indicam que a presença de grupos aceptores de ligação de hidrogênio aumentam os valores da atividade dos compostos e nas regiões dos poliedros vermelhos estes grupos diminuem os valores da atividade dos compostos. Um dos compostos mais ativos, composto 1 ($pIC_{50}=6,55$) está situado no centro.
- Figura 78 – Gráfico dos valores de pIC_{50} preditos pelo modelo *QSAR* expresso pela equação 244 em função dos correspondentes preditos pelo modelo *CoMSIA* (modelo 1, para a série de 22 compostos, grupo de treinamento IE, com o alinhamento realizado através dos campos estéricos e eletrostáticos (SEAL), para a série da 2-piridino carboxaldeído tiossemicarbazona. 190
- Figura 79. Gráficos de Ramachandran obtidos para as cadeias A e B da ribonucleotídeo redutase da (a) *S. typhimurium* (1R2F) e do (b) *C. ammoniagenes* (1KGN). Os sinais de X representam as glicinas, os □ representam as prolínas e os sinais de + representam os outros resíduos. 199
- Figura 80. Gráficos de Ramachandran para as proteínas modeladas utilizando-se as estruturas cristalográficas das RNRs da (a) *S. typhimurium* e do (b) *C. ammoniagenes* como moldes. As linhas no gráfico indicam as áreas preferidas. Os sinais de X representam as glicinas, os □ representam as prolínas e os sinais de + representam os outros resíduos. 201
- Figura 81. Alinhamentos resultantes entre os modelos propostos para a RNR do *M. tuberculosis*, através da modelagem por homologia, e os seus respectivos moldes. As estruturas em cor azul indicam os modelos propostos e as estruturas em vermelho indicam os moldes utilizados: (a) estrutura 3D da *S. typhimurium* (código PDB 1R2F) e (b) estrutura 3D do *C. ammoniagenes* (código PDB 1KGN). 202
- Figura 82. Alinhamento resultante entre as estruturas da RNR do modelo proposto através da modelagem por homologia (estrutura em cor azul), utilizando-se como molde os dados cristalográficos da RNR da *S. typhimurium* (código PDB 1R2F), e a RNR do *M. tuberculosis* elucidada recentemente (código PDB 1UZR), (estrutura em cor vermelha). 203
- Figura 83. Alinhamento resultante entre as estruturas da RNR do modelo proposto através da modelagem por homologia (estrutura em cor azul), utilizando-se como molde os dados cristalográficos

da RNR do <i>C. ammoniagenes</i> (código PDB 1KGN), e a RNR do <i>M. tuberculosis</i> elucidada recentemente (código PDB 1UZR), (estrutura em cor vermelha).	203
Figura 84. Esquema do sítio de ferro reduzido na R2F-2 para a RNR do <i>M. tuberculosis</i> , respectivamente, (a) modelada por homologia, utilizando-se a estrutura cristalográfica da <i>S. typhimurium</i> (código PDB 1R2F) e (b) obtida experimentalmente através da cristalografia de raios-X (Uppsten et al. 2004).	204
Figura 85. Representação esquemática da distribuição das características estéricas no sítio dos átomos de ferro para o modelo proposto por homologia da RNR do <i>M. tuberculosis</i> , empregando-se os dados cristalográficos da <i>S. typhimurium</i> , como molde. Os átomos de ferro estão apresentados na cor amarela.	206

ÍNDICE DE TABELAS

Tabela 1. Características das classes de ribonucleotídeo redutases (Jordan & Reichard 1998). ^a	8
Tabela 2. Valores observados da atividade inibitória (pIC_{50}) frente a ribonucleotídeo difosfato redutase derivada de células H.Ep.-2 (carcinoma humano da epiderme) dos derivados da α -(N)-piridino carboxaldeído tiossemicarbazona, selecionados da literatura (French & Blanz-Jr. 1974).	64
Tabela 3. Números dos compostos selecionados para os grupos de treinamento e os correspondentes grupos de teste para as séries A, B, C, D, E, respectivamente.	66
Tabela 4. Valores das variações máximas de atividade biológica (ΔpIC_{50}^a) nos grupos de treinamento e de teste selecionados para as cinco séries, sendo que a identificação dos compostos de cada série estão apresentados na tabela 3.	80
Tabela 5. Números de descritores DRAGON selecionados pelas análises dos parâmetros estatísticos nos modelos <i>PLS</i> apresentados nas figuras 14 a 28, respectivamente, para as séries A, B, C, D e E, que melhor descrevem a atividade inibitória da série de derivados da 2-piridino carboxaldeído tiossemicarbazonas frente a RNR de células H.Ep.-2.	91
Tabela 6. Descritores selecionados pelos modelos <i>PLS</i> obtidos para a série A (40 compostos).	92
Tabela 7. Descritores selecionados pelos modelos <i>PLS</i> obtidos para a série B (39 compostos).	92
Tabela 8. Descritores selecionados pelos modelos <i>PLS</i> obtidos para a série C (30 compostos).	93
Tabela 9. Descritores selecionados pelos modelos <i>PLS</i> obtidos para a série D (23 compostos).	93
Tabela 10. Descritores selecionados pelos modelos <i>PLS</i> obtidos para a série E (22 compostos).	94
Tabela 11. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de <i>QSAR</i> para a série A (n=40 compostos).	111
Tabela 12. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de <i>QSAR</i> para a série B (n=39 compostos).	112
Tabela 13. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de <i>QSAR</i> para a série C (n=30 compostos).	112
Tabela 14. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de <i>QSAR</i> para a série D (n=23 compostos).	113
Tabela 15. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de <i>QSAR</i> para a série E (n=22 compostos).	113
Tabela 16. Matriz de intercorrelação (r) dos descritores selecionados através da análise <i>PLS</i> e, incluídos nos modelos <i>QSAR</i> , para a série contendo 40 compostos (série A).	114
Tabela 17. Matriz de intercorrelação (r) dos descritores selecionados através da análise <i>PLS</i> e, incluídos nos modelos <i>QSAR</i> , para a série contendo 39 compostos (série B).	115
Tabela 18. Matriz de intercorrelação (r) dos descritores selecionados através da análise <i>PLS</i> e, incluídos nos modelos <i>QSAR</i> , para a série contendo 30 compostos (série C).	116
Tabela 19. Matriz de intercorrelação (r) dos descritores selecionados através da análise <i>PLS</i> e, incluídos nos modelos <i>QSAR</i> , para a série contendo 23 compostos (série D).	117
Tabela 20. Matriz de intercorrelação (r) dos descritores selecionados através da análise <i>PLS</i> e, incluídos nos modelos <i>QSAR</i> , para a série contendo 22 compostos (série E).	117
Tabela 21. Resultados obtidos através do emprego da metodologia <i>CoMFA</i> , para o grupo de treinamento, empregando-se a série de 40 derivados (série A) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	146
Tabela 22. Resultados obtidos através do emprego da metodologia <i>CoMFA</i> , para o grupo de treinamento, empregando-se a série de 39 derivados (série B) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974) utilizando-se diferentes	

procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	148
Tabela 23. Resultados obtidos através do emprego da metodologia <i>CoMFA</i> , para o grupo de treinamento, empregando-se a série de 30 derivados (série C) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	150
Tabela 24. Resultados obtidos através do emprego da metodologia <i>CoMFA</i> , para o grupo de treinamento, empregando-se a série de 23 derivados (série D) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	152
Tabela 25. Resultados obtidos através do emprego da metodologia <i>CoMFA</i> , para o grupo de treinamento, empregando-se a série de 22 derivados (série E) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	154
Tabela 26. Resultados obtidos através do emprego da metodologia <i>CoMSIA</i> , para o grupo de treinamento, empregando-se a série de 40 derivados (série A) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	167
Tabela 27. Resultados obtidos através do emprego da metodologia <i>CoMSIA</i> , para o grupo de treinamento, empregando-se a série de 39 derivados (série B) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	169
Tabela 28. Resultados obtidos através do emprego da metodologia <i>CoMSIA</i> , para o grupo de treinamento, empregando-se a série de 30 derivados (série C) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	171
Tabela 29. Resultados obtidos através do emprego da metodologia <i>CoMSIA</i> , para o grupo de treinamento, empregando-se a série de 23 derivados (série D) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	173
Tabela 30. Resultados obtidos através do emprego da metodologia <i>CoMSIA</i> , para o grupo de treinamento, empregando-se a série de 22 derivados (série E) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French et al. 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B) ⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL ⁽²⁾	175
Tabela 31. Série selecionada de grupos substituintes sugeridos por esta tese e pelos modelos gerados por <i>QSAR</i> e por <i>CoMSIA</i> , respectivamente, propostos nesta tese e os respectivos valores de pIC_{50} preditos através do modelo <i>QSAR</i> expresso pela equação 244 e pelo modelo 1E (<i>CoMSIA</i>), alinhamento SC, para a série da 2-piridino carboxaldeído tiossemicarbazona.....	189
Tabela 32. Valores dos descritores DRAGON incluídos no modelo de <i>QSAR</i> expresso pela equação 244 para a série de substituintes selecionados nesta tese, visando a proposição de novos derivados para a série da 2-piridino carboxaldeído tiossemicarbazona.	189

RESUMO

A tuberculose (TB) é uma doença causada pelo *Mycobacterium tuberculosis*. De acordo com estimativas da Organização Mundial da Saúde, a tuberculose é responsável pela morte de ~2 a 3 milhões de pessoas/ano no mundo e nos próximos 15 anos cerca de 1 bilhão de pessoas deverão ser infectadas, e destas, aproximadamente 35 milhões deverão morrer. Apesar de existirem vários medicamentos sendo utilizados no tratamento da doença, constata-se o crescimento no número de casos devido, principalmente, às variedades resistentes do *M. tuberculosis*. Considerando-se o aparecimento de cepas resistentes em TB, recomenda-se que novos medicamentos e/ou alvos biológicos alternativos devam ser intensivamente pesquisados.

A ribonucleotídeo redutase (RNR), por exemplo, é uma proteína de interesse, pois catalisa uma etapa importante e única na síntese *de novo* dos dNTPs, reduzindo o ribonucleosídeo 5'-difosfato ao seu correspondente desoxirribonucleosídeo 5'-difosfato. A RNR é importante na síntese do DNA, e portanto, na divisão das células. Esta enzima importante, que possui 16% de homologia com a RNR de mamíferos, é um alvo potencial para o desenvolvimento de novos fármacos, com provável aplicação no tratamento do câncer, da malária e do tripanossoma. Sabe-se que diferentes classes de compostos, através de diferentes mecanismos de ação, inibem a RNR, incluindo as α -(N)-heterocíclicas carboxaldeído tiosemicarbazonas, um dos inibidores mais potentes da RNR. Sabe-se que alguns derivados da tiosemicarbazona, inibidoras da RNR de células tumorais, apresentam atividade frente o *M. tuberculosis* atuando provavelmente através do mesmo mecanismo, envolvendo a inibição da correspondente RNR.

Neste contexto, nesta tese de doutorado, foram aplicadas diferentes abordagens de *QSAR/QSAR-3D* no estudo de 40 derivados da 2-piridino-carboxaldeído tiosemicarbazona, inibidores da RNR de células H.Ep.-2, retirados de literatura selecionada (French & Blanz-Jr. 1974). Estes compostos foram divididos em cinco séries, a saber: séries A, B, C, D, e E contendo, respectivamente, 40, 39, 30, 23 e 22 compostos, na tentativa de tornar estas séries estruturalmente mais homogêneas. Para cada série, foram criados três grupos de treinamento e os respectivos grupos de teste (I, II e III), visando-se avaliar o poder de predição dos modelos gerados através das análises de *QSAR/QSAR-3D*.

Para as análises de *QSAR* clássico, foram utilizados como variáveis independentes, os descritores mais relevantes gerados através do programa DRAGON e, pré-selecionados por *PLS*. Considerando-se a ausência de informações sobre a estrutura cristalográfica da enzima

RNR do *M. tuberculosis*, os estudos de *QSAR-3D* foram iniciados empregando-se metodologias propostas em *CoMFA* e, em *CoMSIA*, implementadas no programa SYBYL. Além destas, foi realizada a modelagem por homologia da RNR do *M. tuberculosis*, utilizando-se o programa WHATIF.

Para as abordagens *CoMFA* e *CoMSIA* as geometrias otimizadas através do método semi-empírico AM1 foram alinhadas átomo-a-átomo e, através da similaridade dos respectivos campos estéricos e eletrostáticos, utilizando-se o programa SEAL. Nos dois procedimentos a geometria do composto não substituído, um dos mais ativos na série, foi utilizada como molde considerando-se a ausência de informações sobre a conformação bioativa.

A modelagem da RNR por homologia foi realizada utilizando-se como molde as estruturas cristalográficas, respectivamente, do *C. ammoniagenes* (código PDB 1KGN) e da *S. typhimurium* (código PDB 1R2F), sambas apresentando valores de identidade superior a 65%. Mais recentemente foram publicados os dados cristalográficos para a cadeia beta (subunidade menor) da RNR do *M. tuberculosis* (código PDB 1UZR).

Os modelos *CoMFA* e *CoMSIA* gerados apresentaram valores aceitáveis para os coeficientes de correlação de predição, com altos valores para os coeficientes de correlação ajustados e baixos valores para os erros padrões. Os melhores modelos *CoMFA* e *CoMSIA* foram obtidos considerando o grupo com substituintes apenas na posição 5 do anel piridínico. Razoáveis coeficientes de correlação de predição para os modelos *CoMSIA* com altos coeficientes de correlação de ajuste e baixos valores para os erros padrões foram obtidos. Os mapas de contorno gerados em *CoMFA* e *CoMSIA* sugerem que grupos aceptores de ligações de hidrogênio próximos ao nitrogênio do anel piridínico deverá aumentar o valor da atividade inibitória. Esta observação está em boa concordância com os dados da literatura, na qual a formação de um complexo entre a TSC e o íon Ferro foi sugerido para a inibição da RNR.

Estes estudos deverão permitir um melhor entendimento sobre as características estruturais desta classe de TSC inibidoras da RNR, como agentes antitumorais, em termos dos campos estéricos, eletrostáticos, hidrofóbico, doador e aceptor de ligações de hidrogênio, bem como a contribuição para o desenvolvimento racional de novos inibidores para esta importante enzima. Adicionalmente, dois compostos preparados em nosso laboratório, demonstraram atividade frente o *M. tuberculosis*, em testes realizados *in vivo*.

ABSTRACT

Tuberculosis is an illness caused by *Mycobacterium tuberculosis*. Data from World Health Organization (WHO) estimates, that about 2-3 millions of human population died by *Mycobacterium tuberculosis* infection and that during the next 15 years about 1 billion will be infected and 35 million will certainly die. Although, in the clinic it was found several antiTB-drugs, these numbers will increase due several reasons including *M. tuberculosis* resistant strains. It has been stressed the importance of novel medicines and/or alternative biological targets research projects.

It is known that Ribonucleotide reductase (RNR), is an enzyme that catalyses the rate limiting step in the *de novo* synthesis of dNTPs, reducing the ribonucleoside 5'-diphosphates to the corresponding deoxyribonucleoside 5'-diphosphates. RNR has a critical role in the DNA synthesis and, hence, cell division. This key enzyme, that shows 16% homology when compared with mammals RNR, is a potential target for drug design of cell growth inhibitors, with potential application in cancer therapy, antimalaria and trypanosome chemotherapy. It is known that different types of compounds or species by means of different mechanism pathways can show RNR inhibition, including α -(N)-heterocyclic carboxaldehydes thiosemicarbazones that are one of the most potent classes of RNR inhibitors. More than that, some of them, that shows activity against *M. tuberculosis* seems to follow the same mechanism pathways proposed to the thiosemicarbazones tumor cells activity that means, that they probably are RNR inhibitors.

In this study, a series of 40 α -(N)-2-formyl-pyridine thiosemicarbazone derivatives tested against RNR of H.ep.-2-cells (human epidermoid carcinoma), taken from selected literature (French & Blanz-Jr. 1974), has quantitatively analyzed by means of several QSAR/3D-QSAR approaches. These compounds were divided into 5 individual subsets, namely A, B, C, D, and E, having 40, 39, 30, 23 e 22 compounds, respectively. This procedure has been done in order to achieve more structurally homogeneous subsets. For each set, three individual training and test sets (I,II and III) have been created in order to evaluate the predictivity power of the generated QSAR/3D-QSAR models.

QSAR analysis have been done using descriptors generated by DRAGON program that have been further pre-selected by PLS procedures. Considering that crystallographic data of RNR *M. tuberculosis* are not available in the literature, 3D-QSAR studies have been done these applying, initially, CoMFA and CoMSIA approaches, implemented in SYBYL. Homology model studies have been performed with WHATIF program

CoMFA e *CoMSIA* approaches used optimized geometry obtained by semi-empirical AM1 methods that have been aligned by two different methods. Rigid alignment, in which the compounds were fitted atom-by-atom onto a template, based on the root mean square fit. The N(1) and C(2) atoms of the pyridine moiety and the heavy atoms of thiosemicarbazone backbone of TSC were used as template structure. (2) Field based, in which the steric and electrostatic fields, generated by the SEAL program were considered in the alignment. In both procedures the unsubstituted 2-formylpyridine thiosemicarbazone in its *syn* conformation, has been taken as template.

Homology RNR models were done using as template crystallographic data of *ammoniagenes* (1KGN) and *S. typhimurium* (1R2F) as template, respectively, with identity larger than 65%. More recently new crystallographic data have been published for the beta chain (smaller subunity) of RNR do *M. tuberculosis* (1UZR).

CoMFA and CoMSIA generated models showed acceptable predictive correlation coefficients with high fitted correlation coefficients and low standard errors. Better CoMFA and CoMSIA models have been derived considering a homogeneous subset of TSC substituted only at 5-position in pyridine ring. Reasonable predictive correlation coefficients for CoMSIA models with high fitted correlation coefficients and very low standard errors were obtained. The derived CoMFA and CoMSIA countour maps suggested that a hydrogen bond acceptor near the nitrogen pyridine ring could enhance inhibitory activity value. This observation is in good agreement with literature, in which a complex formation between TSC and iron ion has been suggested, to RNR inhibition.

These studies are expected to enhance the understanding of the structural features of this class of TSC-RNR inhibitors as antitumor agents in terms of steric, electrostatic, hydrophobic and hydrogen donor and acceptor fields as well as to contribute to rational design of inhibitors of this key enzyme. Additionally, two compounds that have been prepared by us showed activity against *M. tuberculosis* using *in vivo* test system.

GLOSSÁRIO

5-HP - 5-hidroxi-2-formilpiridino tiossemicarbazona
AG – Algoritmo Genético
AIDS – Síndrome da Imunodeficiência Adquirida
ATP – Adenosina Trifosfato
BCG – *Bacillus Calmette e Guérin*
CDP – Citidina Difosfato
CIM – Concentração Inibitória Mínima
CoMFA - Análise Comparativa dos Campos Moleculares
CoMSIA - Análise Comparativa dos Índices de Similaridade Molecular
dATP – Desoxiadenosina Trifosfato
DM – Dinâmica Molecular
DNA – Ácido Desoxirribonucleico
dNDPs - Desoxirribonucleotídeos Difosfatos
DOT – Terapia Diretamente Observada
ESS - Soma Explicada dos Quadrados
GSH - Glutathione
HIV – Vírus da Imunodeficiência Humana
HOMO - Orbital Molecular Ocupado de Maior Energia
INH - Isoniazida
KatG – Catalase-peroxidase
LOO – Leave One Out
LUMO - Orbital Molecular Desocupado de Menor Energia
MC – Monte Carlo
MSA – Análise da Forma Molecular
NADPH – Nicotinamida Adenina Dinucleotídeo Fosfato
NDPs - Ribonucleotídeo Difosfatos
OMS – Organização Mundial da Saúde
PDB – Protein Data Bank
PLS – Mínimos Quadrados Parciais
QSAR – Relações Quantitativas Estrutura-Atividade
QSAR-3D - Relações Quantitativas Estrutura-Atividade em três dimensões
QSPR – Relações Quantitativas Estrutura-Propriedade
RDR – Ribonucleotídeo Difosfato Redutase
RMN – Ressonância Magnética Nuclear
RMP - Rifampicina
RMS – Desvio Quadrático Médio
RNA – Ácido Ribonucleico
RNR – Ribonucleotídeo Redutase
RSS – Soma Residual dos Quadrados
SDEP - Desvio Padrão na Predição dos Erros
SM - Estreptomicina
TSS – Soma Total dos Quadrados

OBJETIVOS

A tuberculose é uma doença que contamina milhões de pessoas todos os anos, em todo o mundo. Apesar da existência de vários medicamentos no mercado e da vacina BCG, utilizada de forma profilática, muitas pessoas ainda morrem de tuberculose. A desinformação, o surgimento de cepas resistentes aos medicamentos existentes e o surgimento da AIDS agravam a situação. Desta forma, existe uma preocupação imediata na tentativa de se descobrir novos medicamentos ou alvos biológicos alternativos.

Considerando estes fatos, neste trabalho, a partir de dados selecionados da literatura, serão estudados 40 derivados da α -(N)-formilpiridino tiossemicarbazona, inibidores da enzima ribonucleotídeo difosfato redutase de células H.Ep.-2 (carcinoma humano da epiderme). Estudos realizados demonstraram que estes compostos são capazes de inibir o crescimento de células tumorais bem como o crescimento do *Mycobacterium lufu*, pertencente ao mesmo gênero do *Mycobacterium tuberculosis*, através do mesmo mecanismo, isto é, inibindo a enzima ribonucleotídeo redutase. A enzima ribonucleotídeo redutase é um alvo alternativo dos alvos conhecidos para os fármacos atualmente existentes no mercado.

Os objetivos deste trabalho são de verificar as características estruturais que estes derivados da α -(N)-formilpiridino tiossemicarbazona devem possuir para inibirem a enzima ribonucleotídeo redutase de células tumorais H.Ep.-2, contribuindo desta forma para um maior entendimento sobre as interações ligante-receptor, para este sistema. Como ferramentas para a realização do trabalho foram realizados estudos de relação quantitativa estrutura-atividade (QSAR), estudos de relação quantitativa estrutura-atividade em três dimensões (QSAR-3D), utilizando as metodologias da análise comparativa dos campos moleculares (CoMFA) e da análise comparativa dos índices de similaridade moleculares (CoMSIA) e estudos de modelagem de proteínas através de homologia.

1. INTRODUÇÃO

1.1 Tuberculose

A tuberculose (TB) é uma doença causada pelo *Mycobacterium tuberculosis*. Esta doença, presente na população humana desde a Antiguidade, pode afetar diversos órgãos do corpo humano como, por exemplo, o cérebro, os rins e os ossos mas comumente afeta os pulmões (tuberculose pulmonar). A contaminação do indivíduo pode ocorrer de duas formas: (i) através da via respiratória, quando este respira o mesmo ar no qual uma pessoa infectada espirrou ou tossiu ou, (ii) através da via gastrointestinal, quando este ingere leite infectado com o bacilo. No indivíduo já infectado, o bacilo de Koch pode evoluir de vários modos diferentes: (i) o indivíduo, através de suas defesas, elimina o bacilo; (ii) a bactéria se desenvolve, mas não causa a doença; (iii) a tuberculose se desenvolve, causando a doença – chamada de tuberculose primária ou, (iv) ocorre a ativação da doença após vários anos (por reativação endógena) – chamada de tuberculose pós-primária. Dentre todas as pessoas infectadas, apenas uma pequena parcela, aproximadamente de 10% a 20%, possuem a possibilidade de que a infecção progrida realmente para a doença propriamente dita; e este fato pode acontecer a qualquer momento, para o resto de suas vidas, sendo que com o passar dos anos a probabilidade de evolução da doença vai diminuindo gradativamente (OMS). O *M. tuberculosis* pode ficar em um estado dormente, de não reprodução, durante anos no hospedeiro humano e pode ser reativado, causando a doença, sob certas condições imunológicas e/ou ambientais como, por exemplo, alcoolismo, idade avançada, má nutrição, AIDS, diabetes, câncer, insuficiência crônica dos rins, silicose (doença crônica pulmonar) ou em tratamentos com imunossupressores (Yang *et al.* 1997; Helden, P. D. 2002; Shen *et al.* 2004; Chen 2004; Ojcius 2004; Jawahar 2004).

De acordo com estimativas da Organização Mundial da Saúde (OMS), a tuberculose é responsável pela morte anual de aproximadamente 2 a 3 milhões de pessoas em todo o mundo. Nos últimos anos, o número de pessoas infectadas com a tuberculose está novamente aumentando. Estima-se que entre os anos de 2000 a 2020 cerca de 1 bilhão de pessoas deverão ser infectadas com o bacilo, sendo que deste total, 200 milhões poderão evoluir a doença e aproximadamente 35 milhões de pessoas deverão morrer (OMS). Este crescimento no número de casos se deve principalmente as variedades de *M. tuberculosis* que, com o passar do tempo, acabam se tornando resistentes aos medicamentos atualmente existentes no mercado. Este fato, não acarreta apenas em um maior poder de infecção do *M. tuberculosis*,

mas também faz com que a doença se torne difícil de ser tratada. Um fator que acaba contribuindo para o surgimento de cepas resistentes é a falha no controle da doença. O tratamento com os fármacos atualmente existentes deve prosseguir por um período mínimo de seis meses. O que geralmente ocorre em alguns casos é o abandono da medicação pelo paciente após a constatação de uma breve melhora em seu quadro clínico. Esta melhora está condicionada ao fato de que a maioria das bactérias é combatida, mas algumas persistem ao tratamento originando assim as cepas mutantes (Rook & Pando 1996). Uma das soluções propostas é a terapia diretamente observada (*directly observed therapy*, DOT) na qual o paciente é supervisionado enquanto estiver sob medicação. Foi constatado que as pessoas infectadas com o HIV são mais susceptíveis à infecção, aumentando assim a fração da população contaminada com a tuberculose. Segundo a OMS aproximadamente 10 milhões de pessoas estão infectadas com o HIV, sendo que destes, 4 milhões também estão infectadas com a tuberculose (Rook & Pando 1996; Rozwarski *et al.* 1998; Silva *et al.* 1999; WHO).

Atualmente, como forma profilática, existe apenas a vacina BCG – *Bacillus Calmette e Guérin* – isolada a partir do *Mycobacterium bovis*, responsável pela tuberculose bovina. A BCG, surgida em 1930, é uma vacina viva e atenuada durante um prolongado cultivo do *Mycobacterium* em laboratório e possui, de uma maneira geral, um impacto duvidoso. A eficiência desta vacina varia de 0 a 80%, dependendo da população em que é administrada. Em muitos países a BCG não é utilizada, pois além de sua eficiência duvidosa esta vacina pode, posteriormente, impedir os testes para a detecção da doença (Behr & Small 1997; Silva *et al.* 1999; Lee 2002; Hizek *et al.* 2004; Vordermeier *et al.* 2004).

Algumas estratégias com base na combinação de vacinas de DNA, proteínas, BCG, ou vírus atenuados, foram desenvolvidas na tentativa de se melhorar a eficácia contra a tuberculose (Tanghe *et al.* 2001; Skinner *et al.* 2003). A vacina de DNA é uma alternativa para se gerar os antígenos endógenos e também para se gerar uma proteção antituberculose (Silva *et al.* 1999). Uma alternativa que está sendo testada é a aplicação de vírus recombinantes atenuados, como por exemplo, a vacina da cepa modificada Ankara, expressando o antígeno protetor Ag85A (Rv3804c), (MVA85A), em conjunto com a vacina de DNA ou aplicada isoladamente, que mostrou resultados promissores para combater a tuberculose humana em modelos expressos em ratos (Baldwin *et al.* 1998; McShane *et al.* 2001; McShane *et al.* 2004). A fase clínica I, realizada em seres humanos, da vacinação BCG seguida por MVA85A está sendo conduzida em Oxford, Reino Unido (Ginsberg 2002; McShane *et al.* 2004).

Nos casos onde a infecção é constatada utilizam-se dois tipos de medicamentos, classificados como fármacos de primeira linha, nos casos onde não existem indicações de cepas do *M. tuberculosis* resistentes aos medicamentos e, como fármacos de segunda linha, nos casos onde há indícios do aparecimento de cepas resistentes. Fármacos antituberculose de primeira linha são medicamentos como a *isoniazida* que atuam na síntese do ácido micólico da parede celular; a *rifampicina* que inibe a RNA polimerase através da formação de um complexo estável fármaco-enzima; a *estreptomicina* que atua na síntese do polipeptídeo; a *pirazinamida* que atua na biossíntese dos ácidos graxos da parede celular; o *etambutol* que atua na síntese da arabinogalactana da parede celular, dentre outros (vide figura 1). Por outro lado, os fármacos de segunda linha mais utilizados são: o *ácido p-aminossalicílico* um antagonista competitivo do ácido p-aminobenzóico que impede a utilização deste ácido na síntese do ácido fólico; a *cicloserina* que inibe a reação na qual a D-alanina é envolvida na síntese da parede celular da bactéria e a *ciproflaxacina* e a *ofloxacina* que inibem a enzima DNA girase, necessária para a replicação do DNA, (vide figura 2).

A estreptomicina (SM), o primeiro fármaco utilizado em larga escala para o tratamento da tuberculose, foi introduzida em meados da década de 40. Foi observado que após 3 meses de tratamento com a estreptomicina, quando administrada isoladamente, 80% dos pacientes desenvolviam resistência ao fármaco. A supressão de mutantes resistentes somente foi possível quando o ácido p-aminossalicílico (figura 2) e a isoniazida (INH) (figura 1) foram introduzidos em 1946 e 1952, respectivamente. O emprego concomitante destes três fármacos fez com que o tratamento durasse de 18 a 24 meses. Em 1967, com o aparecimento de rifampicina (RMP) (figura 1), o uso da combinação RMP+INH+SM reduziu o tempo do tratamento pela metade. Poucos anos depois a pirazinamida (figura 1), um potente fármaco antibactericida, começou a ser utilizada em combinação com a RMP e a INH e o tempo de tratamento diminuiu para 6 meses. Atualmente, esta combinação é utilizada rotineiramente em vários países (Brandli *et al.* 1989; Mandel & Sande 1990; Gordin *et al.* 2000; Jindani *et al.* 2004; Santha *et al.* 2004).

Figura 1. Estruturas dos principais fármacos de primeira linha utilizados no tratamento da tuberculose.

Figura 2. Estruturas dos principais fármacos de segunda linha utilizados no tratamento da tuberculose.

Dentre os fármacos mais utilizados no tratamento da tuberculose destaca-se a hidrazida do ácido isonicotínico, ou isoniazida (INH), como é mais conhecido. Este fármaco apesar de possuir uma estrutura química bastante simples, como pode ser observado na figura 1, possui o modo de ação bastante complexo. A INH é uma pró-droga que requer a sua ativação através da enzima catalase-peroxidase (KatG) para a sua forma ativa sendo que, na literatura consultada, o mecanismo de ação proposto é ainda incerto. Devido às mudanças físicas e bioquímicas ocorridas coincidentemente com a toxicidade da INH, foi proposto que o efeito pode estar na rota biossintética para os ácidos micólicos (Bernadou *et al.* 2001; Nguyen *et al.* 2002). Outros estudos indicam que a ação pode ocorrer através da inibição do sistema catalase-peroxidase causando a biotransformação da INH (Bodiguel *et al.* 2001; Pierattelli *et al.* 2004) ou na reação da INH com resíduos da tirosina presentes na proteína micobacteriana (Klopman *et al.* 1996; Mdluli *et al.* 1998; Bardou *et al.* 1998). A compreensão do mecanismo de ação das drogas antituberculozes é de grande interesse devido à existência de casos incuráveis oriundos de cepas mutantes resistentes aos medicamentos existentes (Rozwarski *et al.* 1998).

Uma alternativa existente para se tentar contornar o aparecimento de resistência aos medicamentos é a descoberta de novos medicamentos ou até mesmo de novos alvos. A necessidade de novos medicamentos se deve a alguns fatores como, por exemplo: (i) a existência de cepas do *M. tuberculosis* resistentes ao tratamento das multi-drogas; (ii) a propagação de infecções oportunistas devido a micobactéria não tubercular como o *M. avium*, em particular; (iii) a pouca eficácia dos tratamentos atuais realizados em pacientes imunodeprimidos como, por exemplo, os portadores do HIV e (iv) a complexidade dos tratamentos atuais (Bottari *et al.* 2000; Maccari *et al.* 2004).

Como proposto em alguns estudos de literatura, um dos alvos possíveis seria a enzima ribonucleotídeo redutase (RNR). Em 1986, Schaper e colaboradores (Schaper *et al.* 1986) testaram três tiossemicarbazonas α , β e γ -heterocíclicas contra o *Mycobacterium lufu* determinando os valores da concentração inibitória mínima (CIM). Foi observado que o composto α -heterocíclico apresentava a maior atividade. Adicionalmente, French e Blanz (French & Blanz 1966) verificaram um aumento similar na potência antitumoral ao se testar tiossemicarbazonas α -não heterocíclicas e análogos α -heterocíclicos. Vários grupos de pesquisas mostraram que a atividade antitumoral destes compostos α -heterocíclicos ocorre pela inibição da síntese do DNA. Seu sítio de ação é a enzima ribonucleotídeo difosfato redutase (RDR) contendo ferro. Alguns estudos demonstraram que o átomo de ferro (não do heme), no sítio ativo da redutase, parece interagir com as tiossemicarbazonas α -heterocíclicas, conhecidas como quelantes tridentados (French & Blanz 1966; Sartorelli *et al.* 1976). Schaper e colaboradores (Schaper *et al.* 1986) propuseram que tanto a potência antitumoral bem como a atividade contra o *M. lufu* apresentadas por estas tiossemicarbazonas ocorriam pelo mesmo mecanismo de ação, ou seja, através da inibição da RNR. Dois compostos que apresentaram atividade frente o *M. lufu* também apresentaram atividade frente o *M. smegmatis* (ATCC607), o *M. tuberculosis* (H37Rv), o *M. avium* (SN304) e o *M. marinum* (SN1254) (Schaper *et al.* 1986). Posteriormente, este mesmo grupo verificou que alguns compostos com atividade frente o *M. lufu* também inibiam a atividade da ribonucleotídeo redutase *in situ*, na inibição da síntese do DNA, e o crescimento em cultura de células de leucemia L1210 em ratos (Cory *et al.* 1994). Esta hipótese é ainda sustentada pela correlação observada entre a potência antimicobacteriana (*M. lufu*) e a atividade antitumoral apresentada por estes compostos (Schaper *et al.* 1986).

Apesar da possível similaridade entre a RNR de micobactérias e de mamíferos, um potente inibidor da RNR parece ainda ser uma estratégia muito útil na terapia contra o câncer

e também, devido a possibilidade de inibição da RNR da micobactéria, na terapia contra a tuberculose, pois os níveis de desoxi-ribonucleotídeo trifosfato em células de mamíferos são extremamente baixos (Liu *et al.* 1996).

A 5-hidróxi-2-formilpiridino tiossemicarbazona (5-HP) foi o primeiro composto α -(N)-heterocíclico carboxaldeído tiossemicarbazona a ser estudado em seres humanos como parte da fase 1 de estudos clínicos (DeConti *et al.* 1972). A escolha deste composto foi realizada com base em sua atividade frente a tumores transplantados, linfomas espontâneos em cachorros e sua administração na forma de sal de sódio (Creasey *et al.* 1972; DeConti *et al.* 1972). Entretanto, a administração da 5-HP em doses relativamente altas (acima de 4mg/kg) desenvolveu toxicidade gastrointestinal, anemia, hipertensão e hipotensão. Devido a baixa meia vida biológica da 5-HP foi observado uma fraca atividade antileucêmica na fase 1. Atualmente, a 3-aminopiridino-2-carboxaldeído tiossemicarbazona (triapina), está sendo testada na fase I de estudos clínicos juntamente com os estudos fármacos-cinéticos (Wadler *et al.* 2004). Este composto também é um potente inibidor da ribonucleotídeo redutase de células tumorais.

1.2 Ribonucleotídeo redutase

A ribonucleotídeo redutase (RNR) é uma enzima citosólica fundamental para a síntese de DNA que atua na biossíntese dos desoxi-ribonucleotídeos. Esta reação ocorre com o auxílio de radicais livres orgânicos. Os ribonucleotídeos são sintetizados a partir de blocos como, por exemplo, os aminoácidos (Stryer 1995). Por outro lado, não existe uma rota sintética para os desoxi-ribonucleotídeos e toda a síntese *de novo* é catalisada pela enzima ribonucleotídeo redutase (Cory 1988; Stubbe 1998).

A ribonucleotídeo redutase reduz todos os principais ribonucleotídeos aos seus correspondentes desoxi-ribonucleotídeos. Esta reação relativamente simples é acompanhada, entretanto, por um processo complexo. Esta reação está ilustrada na figura 3. A redução da porção D-ribose de um nucleotídeo difosfato para o seu correspondente 2-desoxi-D-ribose requer um par de átomos de hidrogênio, que são doados pelo NADPH via uma proteína que carrega os átomos de hidrogênios, a tiorredoxina. Esta proteína possui pares de grupos -SH que carregam átomos de hidrogênio do NADPH para o ribonucleotídeo difosfato. A forma oxidada ou dissulfeto da tiorredoxina é reduzida pelo NADPH em uma reação catalisada pela enzima tiorredoxina redutase. A tiorredoxina reduzida é então utilizada pela ribonucleotídeo redutase para reduzir os ribonucleotídeos difosfatos (NDPs) para os correspondentes desoxi-

ribonucleotídeos difosfatos (dNDPs). Uma segunda fonte de redutores equivalentes para as ribonucleotídeos redutases é a glutathione (GSH), que serve como redutor para uma proteína semelhante, a tioredoxina, a denominada glutarredoxina. A glutarredoxina reduzida transfere então o poder redutor da glutathione para a ribonucleotídeo redutase (Cory 1988).

Figura 3. Reação da ribonucleotídeo redutase. A redução do ribonucleotídeo 5'-difosfato para o correspondente 2'-desoxi-ribonucleotídeo 5'-difosfato requer a presença do ribonucleotídeo específico, o 5'-trifosfato, como um efetor alostérico positivo e por um sistema de redução auxiliar composto por trioxina, trioxina redutase e NADPH ou por glutarredoxina, glutathione, glutathione redutase e NADPH.

A descoberta da primeira RNR, em *Escherichia coli*, foi vista com considerável descrédito, pois em Química Orgânica, não se conhecia nenhuma reação na qual uma ligação carbono-grupo hidroxila (C-OH) poderia ser substituída diretamente por um átomo de hidrogênio. Poucos anos depois da descoberta desta enzima na *E. coli* uma segunda redutase foi descoberta no *Lactobacillus leishmannii*. Esta enzima necessita da adenosilcobalamina, o que levou os pesquisadores a acreditarem que as RNRs fossem enzimas B₁₂ e o seu mau funcionamento seria responsável pelos defeitos na síntese do DNA, na anemia perniciosa, uma doença que pode ser tratada com a simples administração da vitamina B₁₂.

A primeira evidência de que outras classes de RNR poderiam existir foi a constatação de que a redutase da *E. coli* não necessitava da vitamina B₁₂. A redutase da *E. coli* se tornou o protótipo da classe de enzimas I, na qual pertencem todas as redutases de eucariontes e a enzima da *L. leishmannii* se tornou o protótipo de um grande grupo de redutases microbianas da classe II, que são dependentes da adenosilcobalamina. As RNRs podem ser agrupadas em três classes maiores, com base nos mecanismos utilizados para a geração do radical e em suas diferenças estruturais (Tabela 1). A classe I pode ser subdividida em duas subclasses, denominadas de Ia e Ib, baseadas principalmente nas diferenças existentes nas regulações alostéricas e também no envolvimento de proteínas auxiliares (Jordan & Reichard 1998).

As enzimas da classe I produzem um radical tirosil estável em um tipo de subunidade através da ação do centro dinuclear de ferro. As enzimas da classe II utilizam um radical no cofator cobalamina e as enzimas da classe III formam um radical glicil estável com a ajuda de uma proteína ferro-enzofre e uma S-adenosil metionina (Eklund *et al.* 2001).

Tabela 1. Características das classes de ribonucleotídeo redutases (Jordan & Reichard 1998).^a

Características	Ribonucleotídeo redutases			
	Classe Ia	Classe Ib	Classe II	Classe III
Dependência de oxigênio	aeróbica	aeróbica	aeróbica/ anaeróbica	anaeróbica
Estrutura	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha_2(\alpha_2)$	$\alpha_2\beta_2$
Genes	nrdAB	nrdEF	---- ^b	nrdDG
Sítio metálico	Fe-O-Fe	Fe-O-Fe Mn-O-Mn	Co	Fe-S
Substrato	NDP	NDP	NDP/NTP	NTP
Redutor	Tiorredoxina Glutarredoxina	nrdH-redoxina Glutarredoxina	Tiorredoxina	Formato
Sítios alostéricos/ cadeia polipeptídica	2	1	1	2
Inibição de dATP	sim	não	não	sim
Ocorrência	Eucariotes Eubactérias Bacteriófagos Viroses	Eubactéria	Arcaebactéria Eubactéria Bacteriófagos	Arcaebactéria Eubactéria Bacteriófagos
Protótipo	<i>E. coli</i> Ratos	<i>S. typhimurium</i> <i>C. ammoniagenes</i> <i>M. tuberculosis</i>	<i>L. leishmannii</i>	<i>E. coli</i>

^a Classe I está subdividida em duas subclasses (Ia e Ib) por apresentarem algumas características diferenciais importantes.

^b O gene do protótipo da classe II da *L. leishmannii* nunca foi nomeado.

As RNRs de mamíferos, de *E. coli* e do DNA de vírus como, por exemplo, do vírus da herpes, pertencem à classe I de redutases, que utilizam uma proteína contendo ferro para a geração do radical livre essencial que atuará como catalisador. Nestas redutases, a organização quaternária das haloenzimas é $\alpha_2\beta_2$. O dímero α_2 , denominado de proteína R1, contém os sítios ativos e os sítios de ligações para os efetores alostéricos, ao passo que o dímero β_2 , denominado de proteína R2, contém um centro dinuclear de ferro e um radical tirosil estável, por monômero, ambos essenciais para a atividade enzimática. Muito dos estudos de estrutura e função da RNR pertencentes à classe I foram efetuados na enzima de *E. coli* (Fontecave *et al.* 1992; Nordlund & Eklund 1993) e em RNR de ratos (Thelander & Gräslund 1993). O radical orgânico livre da R2 está localizado na tirosina 122. Estes resultados foram obtidos através da mutagênese direta do sítio (Larsson & Sjöberg 1986). O radical é gerado através do oxigênio molecular em cooperação com o centro dinuclear de ferro (Sahlin *et al.* 1990). A redução do substrato é realizada por um par de cisteínas no sítio

ativo da R1 que são oxidadas durante a reação. Estes resíduos de cisteínas oxidados precisam ser reduzidos antes que a enzima passe para um outro ciclo de redução do substrato. Os sistemas de redução são tiorredoxina e glutaredoxina que na volta são reduzidas pela tiorredoxina redutase e glutathione, respectivamente (Holmgren 1989). O último redutor em ambos os sistemas é o NADPH.

Um subgrupo da classe de enzimas I é a classe de enzimas Ib, que possui diferenças na sequência, regulação e redução (Reichard 1993). A classe Ib da RNR não possui uma atividade regulatória completa (Eliasson *et al.* 1996), mas como são homólogas à classe de enzimas I e também contém o sítio diférrico com um radical tirosil livre, elas pertencem a um subgrupo especial da classe I de enzimas. Os genes da classe Ia foram encontrados apenas em *Enterobacteriaceae*, muito parecidos com o *Hemophilus influenza* e os eucariontes, enquanto a classe Ib é encontrada em organismos procariontes. Os genes da classe Ib foram encontrados em *Lactococcus lactis*, *Mycobacterium tuberculosis*, *Bacillus subtilis* e *Mycoplasma genitalium*.

Ambas as classes Ia e Ib de RNR foram encontradas em *Enterobacteriaceae* (Jordan *et al.* 1995). Nos casos em que as duas classes estão presentes a segunda forma dos genes da RNR (*nrdEF*) é transcrita fracamente (Jordan *et al.* 1996). Em contraste com o gene *nrdAB*, o gene *nrdEF* não é essencial às células tanto durante o crescimento aeróbico quanto durante o crescimento anaeróbico (Jordan *et al.* 1996). A razão para os organismos terem as duas classes funcionais I ainda é desconhecida.

A classe II de RNR contém apenas um tipo de subunidade e são proteínas α ou α_2 , onde o radical necessário é produzido pela adenosilcobalamina (Stubbe 1990). O radical é formado através da clivagem da ligação adenosil-Co. Este tipo de RNR é encontrado em alguns microorganismos e o exemplo mais estudado é a do *Lactobacillus leichmanni*. Recentemente, várias enzimas da classe II foram caracterizadas cujas seqüências mostram que os domínios catalíticos das enzimas dependentes de B12 não são relacionados com as classes I e III da ribonucleotídeo redutase (Riera *et al.* 1997; Tauer & Benner 1997).

As redutases anaeróbicas da classe III utilizam o radical glicil gerado com o auxílio de uma S-adenosil metionina e um grupo de ferro-enxofre (Eliasson *et al.* 1990). A forma ativa é um dímero α_2 (Padovani *et al.* 2001) que contém o sítio ativo do radical glicil e sítios de ligações para os efetores alostéricos. Isto corresponde em vários aspectos a subunidade R1 de ligação do substrato das redutases aeróbicas. A subunidade β_2 é uma pequena proteína contendo ferro-enxofre essencial para a produção do radical glicil (Sun *et al.* 1995).

Infelizmente, na literatura, encontram-se poucos dados sobre a RNR do *M. tuberculosis*. Sabe-se que a enzima pertence à classe Ib das RNRs. A classe Ib se diferencia da classe Ia, pois não são inibidas pelo dATP. Entretanto, o dATP estimula a redução do substrato. Nos casos conhecidos, a classe Ib de RNRs não é normalmente expressas *in vivo* sobre condições de crescimento em laboratório, com exceção da RNR do *M. tuberculosis*. A enzima nativa consiste de dois componentes diferentes, *i.e.*, proteínas R1E e R2F, ambas necessárias para a atividade. A subunidade maior R1E da RNR do *M. tuberculosis* possui um peso molecular predito de 2 x 82.244 kDa. O genoma possui dois genes, *nrdF1* e *nrdF2*, que codificam os pequenos componentes da RNR denotados respectivamente por R2-1 e R2-2. Postula-se que ambos possuem um sítio radical-ferro como outras classes de proteínas I R2. O recombinante R2-1 possui 5 cisteínas que provavelmente atuam na reação de redução mas não estão relacionadas com as subunidades R2-1 de *E. coli* e de mamíferos. O recombinante R2-2 é enzimaticamente ativo, quando testado com a proteína recombinante R1E, na presença de ATP/dATP e substrato CDP. A subunidade R2-2 possui de 21 a 23% de identidade na sequência com o produto genético *nrdB* da *E. coli* e 16% de identidade com a proteína R2 humana. Este dímero tem peso molecular calculado de 2 x 36.957 kDa (Yang *et al.* 1997; Liu *et al.* 1998). Algumas análises de alinhamento estabeleceram que todos os aminoácidos essenciais conservados correspondentes ao resíduo hospedeiro do resíduo radical 122 da *E. Coli*, os resíduos de ligação do ferro, e os resíduos que formam o sítio hidrofóbico ao redor do radical estável são conservados na proteína R2-2 do *M. tuberculosis*. A mutagênese direta do sítio da Y110 para fenilalanina demonstrou que o radical livre está localizado no resíduo tirosil 110 (Liu *et al.* 1998).

Recentemente, Uppsten e colaboradores determinaram a estrutura tridimensional da subunidade menor, a subunidade R2F-2, do *M. tuberculosis* através de estudos de cristalografia de raios-X (Uppsten *et al.* 2004). A subunidade R2F-2 é uma proteína totalmente helicoidal e possui uma dobra similar às primeiras estruturas R2 conhecidas (Nordlund *et al.* 1990). Como esperado esta subunidade é muito similar à estrutura R2F da *S. typhimurium* (RMS 1,04Å para 283 Cα) e a estrutura R2F do *C. ammoniagenes* (RMS 0,86Å para 285 Cα). Esta classe Ib de RNRs possui uma identidade na sequência de 77% e 66%, respectivamente, com a R2F-2 do *M. tuberculosis*.

O monômero R2F-2 é constituído por 11 hélices, das quais 8 (α A-H) formam um pacote. Três hélices curtas, α_1 , α_2 e α_4 estão localizadas periféricamente na estrutura. Existem três monômeros na unidade assimétrica, um dímero e um monômero criando uma interação

dimérica. A forma do dímero R2 da *E. coli* foi comparada com o coração, onde cada monômero forma um lóbulo e duas fitas- β formam a ponta do coração. Estas duas fitas- β estão ausentes em todas as outras estruturas R2 resolvidas até o presente momento, incluindo a R2F-2 do *M. tuberculosis*.

O carbono terminal não estruturado parece possuir uma característica comum em todas as estruturas R2 e deve facilitar a interação com a subunidade maior. Estudos empregando métodos de RMN demonstraram que o peptídeo terminal da subunidade R2 torna-se estruturado sobre a ligação com a subunidade R1 (Lycksell *et al.* 1994) e vários experimentos enzimáticos mostraram que o C-terminal do R2 é um inibidor efetivo para formação do complexo entre R1 e R2 (Yang *et al.* 1997).

Os dois átomos de ferro do R2F-2 estão ligados através de duas histidinas (His 106, His 200), um aspartato (Asp 72), três glutamatos (Glu 103, Glu 163, Glu 197) e uma molécula de água (Uppsten *et al.* 2004). O glutamato 103 e o glutamato 197 estão ligados aos dois átomos de ferro e o glutamato 163 liga-se de forma bidentada ao Fe2. A molécula de água liga-se apenas ao Fe2 tornando-o hexa-coordenado com a forma piramidal tetragonal ao contrário do Fe1 que possui quatro ligantes criando uma coordenação tetraédrica. Esta coordenação é idêntica à estrutura reduzida da *S. typhimurium*, enquanto que na estrutura reduzida do *C. ammoniagenes*, a molécula de água ligada ao Fe2 está ausente. Na forma reduzida da *E. coli* não existem moléculas de água coordenando com nenhum ferro e o glutamato correspondente ao Glu 163 liga-se monodentadamente ao Fe2, não bidentado como a classe Ib de enzimas.

A distância entre os dois átomos de ferro na R2F-2 do *M. tuberculosis* está entre 3,5Å a 3,7Å, possuindo pequenas variações entre as diferentes moléculas, na unidade assimétrica. As distâncias correspondentes nas estruturas R2F reduzidas do *C. ammoniagenes* e da *S. typhimurium* são respectivamente de 3,9Å a 3,7Å. Na estrutura R2 reduzida da *E. coli* a distância ferro-ferro é de 3,9Å (Uppsten *et al.* 2004).

O radical formado pela tirosina na R2F-2, Tyr 110, se situa a uma distância de 6,8Å do átomo de ferro. As distâncias correspondentes no *C. ammoniagenes* e na *S. typhimurium* são 7,0Å e 6,5-7,0Å, respectivamente. Na R2 da *E. coli*, a posição da tirosina responsável pelo armazenamento do radical difere das posições na classe Ib de proteínas. A Tyr 122 na R2 da *E. coli* está situada a 5,0Å do átomo de ferro, ao passo que nas estruturas da classe Ib esta distância é maior e ligada por uma molécula de água.

A RNR é uma proteína atrativa porque catalisa uma etapa importante e única na síntese de DNA. No caso da tuberculose, também pode ser atraente e estratégica por se tratar

de um alvo diferente dos conhecidos para os fármacos de primeira e de segunda linha utilizados no momento contra a tuberculose o que, em princípio, poderá viabilizar o uso dos compostos propostos neste trabalho em conjunto com estes fármacos.

1.3 Relações Quantitativas Estrutura-Atividade

1.3.1 QSAR clássico

O primeiro estudo de que se tem notícia foi realizado por Crum-Brown e Frase, em 1868 (Crum-Brown & Fraser 1868-69), onde as propriedades fisiológicas de várias bases orgânicas, Φ , foram expressas em função das suas estruturas químicas C , equação 1.

$$\Phi = f(C) \quad \text{Equação 1}$$

Richardson, em 1869, observou que os pesos moleculares de alguns álcoois eram proporcionais ao seu efeito narcótico (Richardson 1869). Alguns anos mais tarde, Richet observou que a toxicidade de alguns compostos orgânicos era inversamente proporcional a sua solubilidade em água (Richet 1893).

Meyer (Meyer 1899) e Overton (Overton 1901, 1991), independentemente, encontraram uma relação linear entre a lipofilicidade, expressa como solubilidade ou coeficiente de partição entre óleo-água, e os efeitos biológicos como, por exemplo, a toxicidade e a atividade narcótica. Fühner observou que para uma série homóloga a atividade narcótica aumentava de acordo com uma progressão geométrica, *i.e.*, 1: 3: 3²: 3³: etc (Fühner & Neubauer 1907). Muitos outros estudos confirmaram estas relações utilizando-se de diferentes parâmetros lipofílicos para descrever várias atividades não específicas (Kubinyi 1994a).

Em 1964, dois artigos foram publicados independentemente, por Hansch e Fujita (Hansch & Fujita 1964) e um segundo por Free e Wilson (Free & Wilson 1964). Estas duas contribuições iniciaram o desenvolvimento das abordagens envolvidas nos estudos de relações quantitativas estrutura química-atividade biológica, posteriormente denominada de análise de Hansch (aproximação de relações lineares de energia livre ou extratermodinâmica) e análise de Free-Wilson, respectivamente.

A análise de Hansch é a investigação da relação quantitativa entre a atividade biológica de uma série de compostos congêneres e os seus parâmetros físico-químicos relativos aos substituintes ou globais representados pelos efeitos hidrofóbico, eletrônico, estérico e outros. Assume-se que para uma série de compostos congêneres, possuindo uma

estrutura comum, os substituintes fazem uma contribuição aditiva e constante para a atividade biológica independente de qualquer outra mudança estrutural. A equação a seguir ilustra um exemplo geral do modelo obtido através destas análises:

$$\log 1/C = a \log P + b \sigma + c E_s + \dots + \text{constante} \quad \text{Equação 2}$$

onde $\log 1/C$ é o logaritmo do inverso da dose molar que produz uma certa resposta biológica; $\log P$ é o logaritmo do coeficiente de partição água/octanol; σ é a constante eletrônica do substituinte de Hammett (Hammett 1937; Hammett 1970); E_s é o parâmetro estérico de Taft (Taft 1971); e a , b e c são os valores dos coeficientes de regressão.

A análise de Free-Wilson é uma abordagem não paramétrica e utiliza variáveis indicadoras (I) da presença ou ausência de determinada sub-estrutura ou substituintes.

Como exemplo, a equação 3 ilustra um modelo de equação obtido por Hansch e Lien (Hansch & Lien 1968) no estudo de uma série de α -bromofenetilaminas com atividade antiadrenérgica.

$$\log 1/C = 1,22 \pi - 1,59 \sigma + 7,89$$

Equação 3

(n=22; r=0,918; s=0,238)

A aplicação do método de Free-Wilson à mesma série de α -bromofenetilaminas resultou na seguinte equação:

$$\begin{aligned} \log 1/C = & -0,301 (\pm 0,50) [m-F] + 0,207 (\pm 0,29) [m-Cl] + 0,434 (\pm 0,27) [m-Br] + \\ & + 0,579 (\pm 0,50) [m-I] + 0,454 (\pm 0,27) [m-Me] + 0,340 (\pm 0,30) [p-F] + \\ & + 0,768 (\pm 0,30) [p-Cl] + 1,020 (\pm 0,30) [p-Br] + 1,429 (\pm 0,50) [p-I] + \\ & + 1,256 (\pm 0,33) [p-Me] + 7,821 (\pm 0,27) \end{aligned}$$

Equação 4

(n=22; r=0,969; s=0,194; F=16,99)

Diferentes coeficientes de regressão são obtidos dependendo do composto escolhido como sendo o composto de referência ou se o modelo clássico de Free-Wilson é aplicado. Entretanto, estes valores são apenas trocas lineares dos valores da equação 4, onde todos os parâmetros estatísticos são idênticos, com exceção apenas dos intervalos de confiança de 95%.

O modelo de Free-Wilson foi modificado por Fujita e Ban (Fujita & Ban 1971), resultando na equação 5, onde a_{ij} é a contribuição do grupo do substituinte X_i na posição j e μ é o valor da atividade biológica (teórico) de um composto de referência na série.

$$\log 1/C = \sum a_{ij} + \mu \quad \text{Equação 5}$$

Este modelo foi aperfeiçoado através da combinação da equação de Hansch com os indicadores de variáveis (Hansch & Yoshimoto 1974) e pode ser considerado como um modelo misto (Kubinyi 1976).

Posteriormente, Hansch e colaboradores introduziram o modelo parabólico nas análises de *QSAR* após a constatação de que a atividade biológica de alguns compostos com hidrofobicidade crescente diminui após alcançar um valor máximo (Hansch *et al.* 1968) expresso pela equação 6:

$$\log 1/C = a \log P + b (\log P)^2 + c \quad \text{Equação 6}$$

onde a e b são os coeficientes de regressão relativos aos termos $\log P$ e $(\log P)^2$, respectivamente, e c é um termo constante.

Kubinyi fez uma importante contribuição na metodologia de *QSAR*, em 1977, quando propôs o modelo bilinear (Kubinyi, H. 1977; Kubinyi, H. 1979a). Em um grande número de casos, foi constatado que a atividade biológica aumenta linearmente com a hidrofobicidade até um determinado valor e após este ponto o valor irá decrescer. As diferenças entre os valores observados e calculados são altas se for utilizado o modelo parabólico. O modelo bilinear para descrever esta dependência não linear da atividade biológica de compostos na hidrofobicidade foi expressa como:

$$\log 1/C = a \log P - b \log(\beta P + 1) + c \quad \text{Equação 7}$$

onde a e b são os coeficientes de regressão relativos aos termos $\log P$, β é uma constante de substituinte e c é um termo constante.

Os estudos das relações quantitativas estrutura-atividade - abordagem clássica - correlacionam parâmetros físico-químicos de compostos com a sua respectiva atividade biológica, através de equações matemáticas, e permitem um entendimento das propriedades responsáveis pelo transporte do fármaco, sua distribuição e pela interação intermolecular ligante-receptor.

A qualidade do ajuste do modelo aos valores observados da atividade biológica pode ser avaliada através dos seguintes valores: (i) coeficiente de correlação (r); (ii) desvio padrão na predição dos erros (s) e (iii) (F) do teste de Fischer. A validação do modelo pode ser ampliada utilizando-se o teste F sequencial, a validação dos coeficientes da equação através do cálculo de seus limites de confiabilidade de 95%, a análise dos resíduos e a validação cruzada. O método para a obtenção destes valores para a validação estatística está descrito de forma detalhada no item 1.3.2.2.

Modelos de *QSAR* podem ser gerados através da utilização de um único parâmetro ou através da combinação de vários parâmetros. Os parâmetros podem ser obtidos através de dados tabelados (*e.g.* π , σ), por experimentos (*e.g.* $\log P$) ou através de cálculos computacionais (*e.g.* valor da energia do orbital molecular ocupado de maior energia

(HOMO), valor da energia do orbital molecular desocupado de menor energia (LUMO), valores de carga atômica parcial, etc...).

Os parâmetros físico-químicos mais utilizados nas análises de *QSAR* clássico são basicamente de três naturezas e refletem as interações mais importantes observadas entre o ligante e o seu receptor, sendo eles:

- i) *Parâmetros lipofílicos* expressam a lipofilicidade dos compostos como, por exemplo, os valores do logaritmo do coeficiente de partição n-octanol/água (Kubinyi 1993a), determinado experimentalmente, os valores da constante hidrofóbica de substituintes π de Hansch-Fujita (Kubinyi 1993b) e os valores de ClogP, determinados através de cálculos com base em um banco de dados de fragmentos moleculares (Leo 1993).
- ii) *Parâmetros eletrônicos* como os valores tabelados da constante eletrônica do substituinte (σ), definida por Hammett (Hammett 1937), valores tabelados da constante eletrônica dos efeitos indutivo/campo ($\mathfrak{I}$) e de ressonância ($\mathfrak{R}$), de Swain e Lupton (Swain & Lupton 1968) e valores calculados através da mecânica quântica como, por exemplo, os valores de carga atômica parcial, de densidade de carga, de potencial eletrostático, de polarizabilidade e de energias dos orbitais HOMO e LUMO, dentre outros.
- iii) *Parâmetros estéricos* relacionados com a estrutura dos compostos, como os valores tabelados da constante estérica do grupo substituinte (E_s), definida por Taft (Taft 1971), valores de constantes de Verloop (Verloop 1976) ou valores do volume molecular calculados através de programas computacionais como, por exemplo, o Sybyl (Sybyl Tripos Associates Inc.).

Além dos parâmetros físico-químicos citados anteriormente, encontram-se na literatura os denominados índices topológicos, que são descritores obtidos a partir da topologia das moléculas e determinados pela teoria dos grafos (Trinajstić 1992).

Embora em estudos de *QSAR* a utilização de índices topológicos tem sido criticada por alguns autores (Kubinyi 1979b; Lopez de Compadre *et al.* 1983; Van de Waterbeemd & Testa 1987), na literatura existem vários trabalhos empregando descritores topológicos para a obtenção de correlação com diversas atividades biológicas como, por exemplo, analgésicas (Domenech *et al.* 1996); anestésicas (Kier *et al.* 1975; Di Paolo 1978a, Di Paolo 1978b); anticonvulsivantes (Dang & Madam 1994); inibições enzimáticas (Richard & Kier 1981), etc...

Alguns autores consideram que apesar das correlações serem promissoras, em alguns estudos, as características físico-químicas dos parâmetros topológicos não são evidentes o que

contraria as recomendações gerais para seleção de parâmetros em *QSAR*, ou seja, escolher somente parâmetros que tenham significado biológico. Pode-se ainda observar correlações por acaso pela combinação aleatória de um grande número de termos intercorrelacionados, incluindo os termos de conectividade normal e de valência. Em vários casos, constata-se que os valores do desvio padrão dos coeficientes de regressão são fornecidos no lugar do intervalo de confiança, sugerindo que estes termos são significantes, mas na verdade não o são. Deste modo, recomenda-se que o usuário deve selecionar e analisar de forma correta os resultados gerados nas análises de *QSAR*, considerando os possíveis problemas citados.

A análise da estrutura molecular do ponto de vista da conectividade começa com a adoção de uma representação apropriada para esta estrutura. Esta representação se baseia no esqueleto molecular, o qual contém a rede de ligações químicas incluindo os átomos e as conexões entre estes átomos. Tal representação é denominada de gráfico molecular. Este, por sua vez, é constituído de vértices, representados pelos átomos, e lados, representados pelas ligações. A série de átomos e conexões, no gráfico molecular, contém informações estruturais que devem ser transformadas em um índice numérico que pode ser usado para representar a estrutura molecular. Os índices devem contemplar principalmente os elementos que são variáveis com a estrutura.

Basicamente, existem três tipos de índices topológicos baseados na conectividade molecular: índices qui ($^m c$) de conectividade molecular, que caracterizam atributos estruturais da molécula; índices kappa ($^m k$) da forma molecular; e os valores de equivalência topológica T , que caracterizam átomos e grupos no esqueleto molecular e que são utilizados essencialmente para determinar átomos quimicamente equivalentes dentro de uma molécula.

O menor nível de informação estrutural em um gráfico molecular é simplesmente o número de vértices. Obviamente, este é um índice com baixíssimo conteúdo de informação. Outros índices com maior grau de informação podem ser abstraídos a partir do gráfico molecular. Índices que representam simplesmente o número de lados (índice de Wiener (Wiener 1947; Wiener 1948)) ou o número de pares de lados adjacentes (índice de Platt (Platt 1952)) em um gráfico molecular são exemplos de índices simples, mas que podem fornecer interessantes correlações com uma série de propriedades. Na tentativa de se desenvolver índices que incluam a maior quantidade de informações estruturais possíveis, Randic introduziu o conceito de grau de vértice, δ , o qual, para um dado vértice, é definido como sendo o número de vértices vizinhos (Randic 1975; Randic & Wilkins 1979). Por exemplo, o grupo metila possui apenas um vizinho, portanto $\delta = 1$; um grupo metileno tem dois vizinhos,

$\delta = 2$; e daí por diante. Randic também propôs um peso para o lado entre os vértices i e j , definindo-o por $(\delta_i \times \delta_j)^{-1/2}$. A partir deste ponto, pode-se definir um índice de ramificação como a soma dos pesos dos lados, para todos os lados em um gráfico molecular, $(\sum(\delta_i \times \delta_j)^{-1/2})$. Expressando matematicamente, cada lado entre os vértices i e j é caracterizado por um número definido como:

$$C_{ij} = (\delta_i \cdot \delta_j)^{-1/2} \quad \text{Equação 8}$$

onde C_{ij} é o peso entre os vértices i e j . O índice de ramificação ou de conectividade molecular, c é obtido fazendo-se a soma de todos os C_{ij} .

$$\chi = \sum C_{ij} = \sum (\delta_i \cdot \delta_j)^{-1/2} \quad \text{Equação 9}$$

A generalização do índice de Randic foi feita por Kier e Hall (Kier & Hall 1976). Estes definiram o índice c em várias ordens, $^m c$, em função do número de lados considerados em cada caso, como pode ser mostrado pelas equações 10, 11 e 12.

$$^0 \chi = \sum (\delta_i)^{-1/2} \quad \text{Equação 10}$$

$$^1 \chi = \sum (\delta_i \cdot \delta_j)^{-1/2} \quad \text{Equação 11}$$

$$^2 \chi = \sum (\delta_i \cdot \delta_j \cdot \delta_k)^{-1/2} \quad \text{Equação 12}$$

Segundo Kier e Hall o índice c , em cada ordem, representa uma característica distinta da estrutura molecular. Para ordens maiores que 2 foram definidos novos índices de conectividade que levam em consideração os tipos de subestruturas consideradas. Desta forma, pode-se ter seqüências do tipo linha (c_L), ramo (c_R), linha/ramo (c_{LR}) ou anel (c_A), dependendo da molécula apresentar ramificações ou ciclos, como exemplificadas na figura 4.

Figura 4. Exemplos típicos de subestruturas utilizadas para a determinação dos índices topológicos.

Os índices de conectividades, como definidos acima, não conseguem distinguir entre diferentes átomos nem entre diferentes hibridizações. Por exemplo, o p-nitrotolueno (a) e o limoneno (b) apresentam o mesmo gráfico molecular (c) e, portanto, índices de conectividades iguais (figura 5).

Figura 5. Estruturas do p-nitrotolueno (a) e limoneno (b) e o respectivo gráfico molecular (c).

Para corrigir estas deficiências Kier e Hall sugeriram a definição do grau de vértice (d^v) não mais em função do número de lados vizinhos ao vértice, mas em função da valência do átomo, representado pelo vértice (Kier & Hall 1976). Desta forma, o grau de vértice é definido por:

$$\delta_i^v = \frac{Z_i^v - h_i}{Z_i - Z_i^v - 1} \quad \text{Equação 13}$$

onde, Z_i^v corresponde ao número de elétrons da valência; do átomo i ; h_i corresponde ao número de átomos de hidrogênios ligados ao átomo i ; e Z_i é número atômico do átomo i .

Esta relação acarreta a uma definição equivalente à anterior para carbonos saturados, mas distingue claramente a presença de instaurações ou de heteroátomos (figura 6).

Figura 6. Graus de vértice (d^v) corrigidos para as estruturas do p-nitrotolueno (a) e limoneno (b).

Com base no grau de vértice d^v , calcula-se o índice de conectividade de valência de várias ordens, ${}^m c^v$, de forma similar à definida anteriormente, conforme mostrados nas equações 10, 11 e 12. As relações apresentadas expressam o cálculo dos índices de conectividade conforme desenvolvido por Kier e Hall (Kier *et al.* 1975; Kier *et al.* 1976).

Na literatura são encontrados vários programas utilizados para se determinar os valores dos índices topológicos, dentre os quais pode-se citar os programas TOP (Neves *et al.* 1998), TOPIX (Svozil & Lohninger), GROUND (Katritzky & Gordeeva 1993) e DRAGON (<http://www.disat.unimib.it/chm/Dragon.html>). O programa DRAGON, por exemplo, determina um total de 1497 descritores, distribuídos em 18 classes, a saber: descritores

constitucionais (47), descritores topológicos (266), *molecular walk counts* (21), descritores BCUT (64), índice Galvez de carga topológica (21), auto correlação 2D (96), descritores de carga (14), índice de aromaticidade (4), perfil molecular de Randic (41), descritores geométricos (70), descritores RDF (150), descritores 3D-MorSE (160), descritores WHIM (99), descritores GETAWAY (197), grupos funcionais (121), descritores de átomo centrado (120), descritores empíricos (3) e propriedades (3).

Como exemplo da aplicação destes índices topológicos descreve-se (Fulerand *et al.* 1978) a análise de uma série de 24 inibidores da monoamina oxidase encontrando-se correlações significativas entre os valores de pI_{50} e alguns descritores topológicos relacionados a parâmetros eletrônicos e estéricos, equação 14.

$$pI_{50} = -5,2 - 29 E_{1/2} - 0,82 {}^2X + 1,8 {}^3X^V_p \quad \text{Equação 14}$$

($r = 0,941$; $s = 0,201$; $F = 52$; $n = 24$)

onde, $E_{1/2}$ representa o potencial de meia onda, 2X fornece o peso para as características estruturais de ligação dupla na molécula e ${}^3X^V_p$ representa as ligações triplas e também enfatiza a importâncias dos heteroátomos nos substituintes.

Considerando que as interações ligante-receptor são inerentemente propriedades tridimensionais, vários esforços estão sendo realizados na tentativa de se desenvolver métodos de *QSAR* capazes de explorar as propriedades 3D das moléculas (Kubinyi 1993b). Dentre as diversas metodologias existentes atualmente, pode-se de uma maneira geral classifica-las em dois grupos. Os estudos podem ser realizados sem o conhecimento do possível alvo biológico, como, por exemplo, através da utilização da Análise da Forma Molecular (MSA), da Análise Comparativa dos Campos Moleculares (*CoMFA*) (Cramer III *et al.* 1988) e da Análise Comparativa dos Índices de Similaridade Molecular (*CoMSIA*) (Klebe *et al.* 1994) ou quando existem informações sobre o possível alvo-biológico, como, por exemplo, estudos de ancoragem “*docking*”.

1.3.2 Métodos de modelagem sem o conhecimento do alvo biológico

1.3.2.1 Análise da Forma Molecular - MSA

A análise da forma molecular – *Molecular Shape Analysis* - (MSA) é um formalismo desenvolvido para auxiliar a descoberta de novos ligantes com o auxílio de métodos computacionais (Burke & Hopfinger 1993) sendo representado pela equação 15.

$$\log(1/C_u) = f_0(M(u, v)) + \sum_k H_k(p_k(u)) + p(n_v - n_u) - \beta \Delta E_u + \Psi \quad \text{Equação 15}$$

onde u é qualquer um dos membros da série de compostos em análise e v é o composto de referência na qual todos os membros da série são comparados. A variável dependente, $\log(1/C_u)$, é a atividade biológica medida e $\Psi = [\log(1/C_v) + \beta \Delta E_v + \sum_k H_k(p_k(v))]$ é uma constante característica de v .

Nesta equação, $\sum_k H_k(p_k(u))$ representa as propriedades físico-químicas utilizadas nas análises de *QSAR* clássico e $[f_0(M(u, v)) + p(n_v - n_u) - \beta \Delta E_u]$ é um termo relacionado a estrutura molecular 3D envolvendo a forma molecular e as possíveis conformações termodinâmicas. E, mais especificamente, na equação $f_0(M(u, v))$ incorpora a medida da similaridade relativa entre as formas de u e v ; $p(n_v - n_u)$ é a diferença na entropia conformacional intramolecular (flexibilidade) entre u e v , e $\beta \Delta E_u$ é a medida da estabilidade relativa da conformação bioativa de u comparada com a sua energia global mínima. Deve-se ter em mente de que $p_k(u)$ também pode incluir propriedades obtidas de cálculos 3D e não estão limitadas a relações lineares de energia livre e/ou a descritores topológicos de ligação característicos das análises de *QSAR* clássico.

O formalismo de MSA envolve sete etapas, a saber:

1. Análise da conformação dos compostos;
2. Proposição de hipótese para a conformação bioativa;
3. Seleção da forma para o composto de referência;
4. Sobreposição dos compostos sobre o composto de referência;
5. Medida da forma molecular;
6. Determinação das outras características moleculares;
7. Construção de um modelo de *QSAR*.

Burke e colaboradores (Burke *et al.* 1994) estudaram uma série de análogos da piridobenzodiazepinona, inibidoras de receptores muscarínicos 2 (M2) e 3 (M3), através da análise de relação quantitativa estrutura-propriedade em três dimensões, utilizando o formalismo baseado na análise da forma molecular, MSA. O *PLS* foi utilizado para a construção dos modelos de MSA-3D-QSPR. As análises de QSPR para M2 e M3 sugeriram que a especificidade dos subtipos de receptores pode ser compreendida através do encaixe em um sítio receptor impedido estericamente.

Rhyu e colaboradores (Rhyu *et al.* 1995) através da análise de MSA-3D-*QSAR* efetuada para um grupo de análogos da triazina verificaram que a atividade biológica era governada pela forma estérica dos compostos e pelo valor do momento de dipolo de cada

composto. Foi observado que as correlações obtidas para esta classe de compostos possuíam um poder de predição maior do que as análises anteriores realizadas através de estudos de *QSAR*.

1.3.2.2 Análise Comparativa dos Campos Moleculares - *CoMFA*

Na metodologia *CoMFA*, inicialmente cada um dos compostos é colocado em um cubo com grades regularmente espaçadas. Um átomo de prova, previamente selecionado, irá percorrer todas as interseções deste cubo calculando as energias de interação.

A figura 7 ilustra as etapas do protocolo *CoMFA*. Os valores da atividade biológica são utilizados como sendo a variável dependente e estes valores serão preditos quando as análises forem realizadas. As colunas restantes na tabela devem conter os parâmetros estruturais, isto é, as variáveis independentes como, por exemplo, o campo *CoMFA* ou qualquer parâmetro físico-químico calculado ou obtido experimentalmente.

Figura 7. Etapas do processo da análise comparativa dos campos moleculares (*CoMFA*).

As distâncias entre as grades do cubo irão afetar o número de interações a serem calculadas, sendo que quanto menor o espaçamento maior será o número de interações calculadas. Espaçamento entre as grades com valores pequenos como, por exemplo, 1,0Å, fará com que o custo computacional aumente muito, sendo que a maioria dos trabalhos utilizando a metodologia *CoMFA* utiliza o valor de 2,0Å. O tamanho do cubo deve ser suficientemente grande para acomodar as moléculas e é determinado automaticamente. Em

seguida, o seu tamanho é automaticamente expandido em mais 4Å nas três direções X, Y e Z, respectivamente.

A escolha dos átomos de prova depende das interações a serem calculadas. Assim, por exemplo, a utilização de um átomo de carbono com hibridização sp^3 irá determinar as interações estéricas e o valor da carga deste átomo, *e.g.* +1, irá determinar as interações eletrostáticas. Estas energias serão utilizadas como sendo as variáveis independentes para gerar os modelos de predição.

Para os cálculos das energias estéricas em *CoMFA*, utiliza-se a função potencial de Lennard-Jones 6-12 e para os cálculos das energias eletrostáticas utiliza-se o potencial coulômbico (com o termo da constante dielétrica dependente da distância, $1/r$). Os valores de A e B na equação de Lennard-Jones são obtidos através dos valores padrão do campo de força da Tripos (Vinter *et al.* 1987). Na metodologia *CoMFA* uma energia de corte (*cutoff*) arbitrária, por exemplo, 30 kcal/mol é adotada para as contribuições estéricas e as eletrostáticas. Se a repulsão do átomo de prova, em um determinado ponto do cubo, for maior do que a energia de corte estipulada, esta interação será ajustada para o valor da energia de corte e o método irá calcular a interação pela média das energias das outras moléculas neste ponto do cubo.

Em seguida, as análises de regressão das energias dos campos *CoMFA* são efetuadas através do algoritmo dos mínimos quadrados parciais - *PLS* (*Partial Least Squares*) (Lindberg *et al.* 1983) inserido no módulo *QSAR* do programa Sybyl. *PLS* é uma técnica, dentre as várias existentes, que gera uma equação para descrever ou prever diferenças nas variáveis dependentes (ou propriedades de entrada) através das diferenças dos descritores, variáveis independentes. Esta técnica é uma extensão da técnica de regressão múltipla. Tanto as análises de *PLS* bem como as análises de regressão múltipla fornecem os mesmos resultados numéricos (coeficientes, r^2 e s) quando aplicadas da mesma forma e para o mesmo grupo de treinamento. Quando mais de uma variável está envolvida, uma equação separada é calculada para cada propriedade de entrada, mas os coeficientes são inter-relacionados e geralmente diferem daqueles que são obtidos analisando-se estas propriedades de entrada separadamente.

Na análise *PLS*, o primeiro passo é a utilização do procedimento da validação cruzada (*cross-validation*). A validação cruzada é uma aproximação para selecionar qual modelo, dentre vários, com diferentes níveis de complexidade, deve possuir um bom valor de predição. Este método elimina um, *Leave One Out* (LOO), ou vários grupos de entrada do grupo de treinamento, originando um modelo quantitativo dos objetos restantes e predizendo,

e não reproduzindo, o valor da atividade para um ou vários objetos que não estão incluídos na derivação do modelo (Livingstone 1995; Carrupt *et al.* 1996).

A suposição de uma distribuição normal dos erros permite impor limites de confiança no ajuste da linha para os dados. Isto pode ser realizado através de uma análise da tabela de variância na qual os valores das somas dos quadrados são coletados. A soma total dos quadrados, *total sum of squares* (*TSS*), em outras palavras, a variação total em y , é dado através da somatória da diferença entre os valores de y observados e os seus valores médios.

$$TSS = \sum_i^N (y_{\text{observado}} - y_{\text{médio}})^2 \quad \text{Equação 16}$$

Esta soma dos quadrados é realizada a partir de dois componentes: a variância em y que é explicada através da equação de regressão, conhecida como soma explicada dos quadrados, *explained sum of squares*, *ESS*, e a soma residual dos quadrados, *residual sum of squares*, *RSS*.

Os valores de *ESS* são obtidos através da comparação dos valores de y calculados com os seus valores médios.

$$ESS = \sum_i^N (y_{\text{calculado}} - y_{\text{médio}})^2 \quad \text{Equação 17}$$

Os valores de *RSS* são obtidos através da comparação dos valores observados com os valores calculados.

$$RSS = \sum_i^N (y_{\text{observado}} - y_{\text{calculado}})^2 \quad \text{Equação 18}$$

A soma total dos quadrados é igual a esta equação:

$$TSS = ESS + RSS \quad \text{Equação 19}$$

O valor do quadrado do coeficiente da correlação da validação cruzada, Q_{cv}^2 , é fornecido pela divisão da soma explicada dos quadrados pela soma total dos quadrados:

$$Q_{cv}^2 = \frac{ESS}{TSS} \quad \text{Equação 20}$$

O valor de Q_{cv}^2 pode variar entre zero, quando a regressão não consegue explicar a variância nos dados, e um, quando a regressão explica toda a variância no grupo. A multiplicação de Q_{cv}^2 por 100 fornece a porcentagem da variância no grupo de dados explicada através da equação de regressão.

Os valores do desvio padrão na predição dos erros, *standard deviation in error prediction* (*SDEP*), são obtidos através da raiz quadrada da diferença da somatória do

quadrado dos valores observados e dos valores calculados, *PRESS*, dividido pelo número de compostos, *N*.

$$SDEP = \sqrt{\frac{PRESS}{N}}$$

Equação 21

Na equação anterior, *PRESS*, é a soma dos quadrados dos erros calculados.

$$PRESS = \sum (y_{calculado} - y_{observado})^2$$

Equação 22

Para a obtenção dos valores do desvio padrão da autovalidação, *S_{press}*, utiliza-se a equação apresentada a seguir:

$$S_{press} = \sqrt{\frac{PRESS}{N - C - 1}}$$

Equação 23

onde, *N* representa o número de compostos e *C* representa o número de variáveis latentes utilizadas na regressão.

Geralmente, dez componentes ortogonais são extraídos pelo método *PLS* usando a técnica da validação cruzada. A técnica da validação cruzada exclui um componente por vez, que servirá como teste (*LOO*). Em seguida, as análises *PLS* fornecem uma equação que relaciona as variáveis independentes com a variável dependente, neste caso, a atividade biológica. O componente extraído é testado nesta equação. O processo é repetido até que todos os componentes tenham sido eliminados uma vez, e apenas uma única vez. A melhor equação obtida será utilizada para prever a atividade de outros compostos. O processo *PLS* está ilustrado na figura 8.

Figura 8. Processo de autovalidação *PLS*.

Os campos estéricos e eletrostáticos, respectivamente, são escalonados usando o padrão *CoMFA* para igualar seus pesos no modelo aplicado. O valor mínimo de $\sigma = 2,0$ kcal/mol geralmente é empregado como filtro de coluna para reduzir os ruídos nas análises *PLS*. O melhor número de componentes será obtido através da validação cruzada tendo o menor valor do erro padrão de predição (s_{press}). O melhor número de componentes é utilizado em seguida para derivar a melhor reta do modelo de correlação usando a validação não cruzada (*non-cross-validated*).

Em *CoMFA*, as descrições da interação do complexo ligante-receptor consideram apenas as interações estéricas e eletrostáticas. Entretanto, sabe-se que apenas estes dois campos moleculares não são suficientes para descrever corretamente e completamente todas as interações intermoleculares que ocorrem entre o receptor e o seu respectivo ligante. Uma outra limitação a ser considerada na abordagem *CoMFA* é o fato destas duas interações moleculares serem puramente de natureza entálpica e assim não descrevem o componente entrópico da energia livre de interação. Na tentativa de se contornar estas limitações podem ser consideradas várias aproximações, onde parâmetros físico-químicos adicionais, utilizados em *QSAR* clássico, podem ser incluídos como, por exemplo, parâmetros que representem a interação hidrofóbica (Klebe *et al.* 1994). Um dos problemas adicionais encontrados em *CoMFA* se deve aos campos utilizados. Por exemplo, o potencial de Lennard-Jones, utilizado na descrição das interações estéricas é muito semelhante à superfície de van der Waals. Como consequência, a energia potencial expressa nos pontos de grade perto da superfície da molécula mudam drasticamente. Uma pequena variação conformacional das moléculas pode resultar em uma grande variação nestes descritores. Espera-se que esta região apresente descritores significativos para as análises de *QSAR*. Além disso, os potenciais de Lennard-Jones e coulômbico demonstram singularidades nas posições atômicas. Para permitir valores inaceitavelmente altos, o cálculo dos potenciais é normalmente restrito a regiões fora da molécula e alguns valores de corte são determinados arbitrariamente.

Um exemplo recente da aplicação da metodologia *CoMFA* foi publicado por Zhou e Madura (Zhou & Madura 2004) no estudo de *QSAR*-3D de inibidores não nucleosídeos da transcriptase reversa do vírus HIV-1. No estudo de 50 derivados da TIBO, os autores utilizaram um alinhamento flexível, realizado com o emprego do programa Autodock3, para posteriormente aplicar as metodologias *CoMFA* e *CoMSIA*. Os modelos de *QSAR*-3D obtidos demonstraram uma boa habilidade para prever a atividade dos compostos em estudo ($r^2=0,972$ e $Q^2_{cv}=0,704$). Os resultados de *CoMFA* demonstraram a importância dos campos estéricos e eletrostáticos para a estrutura de ligação do complexo.

Um outro exemplo envolvendo estudos de *QSAR-3D* através das metodologias *CoMFA* e *CoMSIA* foi realizado por Raichurkar e Kulkarni (Raichurkar & Kulkarni 2003) no estudo de análogos da hidroxisemicarbazida com potencial ação de inibição de células tumorais. Os compostos tiveram a sua atividade biológica expressa frente a ribonucleotídeo redutase de células L1210. A inclusão do descritor adicional ClogP melhorou de forma significativa os resultados estatísticos do modelo ($r^2=0,993$ e $Q^2_{cv}=0,708$). O modelo sugere a necessidade da lipofilicidade para a atividade antitumoral.

1.3.2.3 Análise Comparativa dos Índices de Similaridade Molecular – *CoMSIA*

Na tentativa de se contornar os problemas existentes na abordagem *CoMFA* introduziu-se (Klebe *et al.* 1994) a análise comparativa dos índices de similaridade molecular (*CoMSIA*). Com base na função de alinhamento utilizada no programa SEAL (Kearsley & Smith 1990) que revela resultados convincentes para a comparação espacial de moléculas, os índices de similaridade são determinados no espaço. Utilizando-se o mesmo átomo de prova, estes índices de similaridade são enumerados para cada uma das moléculas alinhadas no banco de dados, em pontos da grade regularmente espaçadas. Na determinação desta similaridade a distância entre o átomo de prova e os átomos das moléculas do banco de dados é considerada. Uma função do tipo Gaussiana foi selecionada para descrever esta dependência da distância, sendo que a definição de valores de corte arbitrária não é necessária e os índices podem ser calculados em todos os pontos da grade. Em princípio, qualquer propriedade físico-química relevante pode ser considerada nesta aproximação para o cálculo dos campos dos índices de similaridade. A figura 9 ilustra a variação dos diferentes potenciais nas proximidades das moléculas.

Figura 9. Perfil da aproximação Gaussiana utilizada na metodologia *CoMSIA* e perfis dos potenciais de Lennard-Jones e Coulomb, utilizados na metodologia *CoMFA*, indicando o valor limite do potencial definido pelos valores de corte.

As metodologias *CoMSIA* e *CoMFA* são similares. Uma tabela de dados é construída através dos índices de similaridade calculados através de um átomo de prova comum que é colocado em todas as intersecções de uma grade regularmente espaçada. Os índices de similaridade, $A_{F,k}$, entre os compostos de interesse e o átomo de prova, sistematicamente colocado nas intersecções da grade, são calculados pela equação (e.g., no ponto da grade q para a molécula j do grupo de dados):

$$A_{F,k}^q(j) = -\sum_{i=1}^n w_{prova,k} w_{ik} e^{-\alpha r_{iq}^2} \quad \text{Equação 24}$$

onde, i representa o índice de soma sobre todos os átomos da molécula j sobre investigação; w_{ik} representa o valor atual da propriedade físico-química k do átomo i ; $w_{prova,k}$ representa o átomo de prova com carga +1, raio 1Å, e hidrofobicidade +1; α representa o fator de atenuação; e r_{iq} representa a distância mutua entre o átomo de prova no ponto da grade q e o átomo i da molécula de teste. Valores grandes para α irão resultar em uma grande atenuação, fazendo com que as partes remotas não sintam o átomo de prova e valores pequenos de α farão com que partes remotas de cada molécula sintam o átomo de prova, fazendo com que as características moleculares globais sejam mais importantes. O artigo original utiliza o valor de $\alpha=0,3$ (Klebe *et al.* 1994; Böhm *et al.* 1999).

Em seguida, os índices obtidos são analisados através da *PLS* de modo análogo a abordagem *CoMFA*.

Os exemplos utilizando a metodologia *CoMFA*, citados anteriormente (1.3.2.2), utilizaram também a metodologia *CoMSIA*. Na literatura (Zhou & Madura 2004) em estudos de *QSAR-3D* realizados para 50 derivados da TIBO foram obtidos resultados estatísticos semelhantes para *CoMFA* ($r^2=0,972$ e $Q^2_{cv}=0,704$) e *CoMSIA* ($r^2=0,944$ e $Q^2_{cv}=0,776$). Os modelos *CoMFA* mostraram uma contribuição equivalente para o fator estérico e para o fator eletrostático. Por outro lado, os modelos *CoMSIA* mostraram uma maior contribuição do fator hidrofóbico. Estes resultados são suportados por estudos prévios de *docking* que indicaram a importância das interações hidrofóbicas e hidrofílicas na ligação dos inibidores não nucleosídeos da transcriptase reversa no seu respectivo sítio de ligação.

No estudo dos análogos da hidroxissemicarbazida (Raichurkar & Kulkarni 2003), com potencial ação contra células tumorais, os resultados estatísticos de *CoMFA* ($r^2=0,993$ e $Q^2_{cv}=0,708$), de uma maneira geral, também foram semelhantes aos resultados obtidos através do emprego da metodologia *CoMSIA* (valores de r^2 variando de 0,880 a 0,993 e valores de Q^2_{cv} variando de 0,591 a 0,683). Em *CoMFA* foram introduzidos valores de ClogP gerando-se,

desta forma, modelos com resultados estatísticos mais significativos. Na metodologia *CoMSIA*, o melhor modelo foi obtido ao se utilizar os campos estéricos, eletrostáticos, hidrofóbicos e o relativo à propriedade doadora de ligações de hidrogênio. Os mapas de contorno obtidos por ambas as metodologias foram semelhantes.

1.3.2.4 Grupos de análise, conformação bioativa e alinhamento dos compostos

Na série de compostos, a ser utilizada para os estudos de *QSAR/QSAR-3D*, as propriedades eletrostáticas devem variar de forma contínua e independentemente das propriedades estéricas. Desta forma, os modelos estatísticos obtidos deverão possuir alto poder de predição.

Os compostos a serem analisados devem ser divididos em dois grupos distintos: o grupo de treinamento e o grupo de teste, para averiguar a validade do modelo obtido, denominada de validação externa (Martin 1998). Debnath ao estudar derivados cíclicos da uréia dividiu os compostos em três subgrupos. Em cada subgrupo, com os mesmos 93 compostos cada, foram selecionados aproximadamente 25% dos compostos (25 compostos) para formarem o grupo de teste e os compostos restantes foram utilizados no grupo de treinamento. Os compostos selecionados para compor os grupos de teste eram diferentes em cada um dos três subgrupos e, as variações no espectro da atividade foram também consideradas. Esta validação externa é utilizada para verificar o poder de predição dos modelos gerados (Debnath 1999).

Após a escolha da série é necessária a otimização das geometrias dos compostos em estudo. Existem várias formas para se tentar otimizar as geometrias dos compostos. O ideal seria estudar os compostos em sua conformação bioativa, isto é, a conformação em que o composto se encontra no sítio de ligação, o que infelizmente ainda não é possível, em muitos casos, por se tratar de uma metodologia que geralmente envolve um grande número de compostos. Na existência de informações de cristalografia de raios-X sobre o complexo receptor-ligante, pode-se otimizar os compostos no sítio ativo, para isto é necessário que todos os compostos tenham o mesmo mecanismo de ação e o mesmo alvo. Esta alternativa fornecerá um ponto de partida para a escolha da possível conformação bioativa. Uma alternativa utilizada nos estudos de *QSAR-3D* é a otimização das geometrias dos compostos através de cálculos computacionais, geralmente empregando os cálculos semi-empíricos. Nesta aproximação, as geometrias de mínima energia são utilizadas nas análises.

A etapa de alinhamento é efetuada para que todos os compostos se orientem da mesma forma no espaço em que as análises serão efetuadas simulando, desta forma, a interação com o receptor, de preferência em sua conformação bioativa. A etapa de alinhamento é de fundamental importância nos estudos de *QSAR-3D*, pois o poder de predição na abordagem *CoMFA*, convencionalmente caracterizado através dos valores dos coeficientes de correlação da autovalidação (cross-validated) R^2 (Q^2_{cv}), é altamente dependente da orientação dos compostos alinhados para a formação do banco de dados. Desta forma, o alinhamento correto dos compostos tem uma grande influência nos resultados (Cho e Tropsha 1995). Waller e colaboradores utilizando alinhamentos baseados em dados experimentais, para o estudo da inibição da protease do HIV-1, concluíram que o alinhamento baseado nas minimizações através dos campos estéricos e eletrostáticos fornecia tanto resultados estatísticos melhores como modelos de *QSAR-3D* com maior poder de predição, em relação ao alinhamento baseado em dados cristalográficos dos complexos ligante-receptor (Waller *et al.* 1993). Waller e Marshall investigaram a utilização do alinhamento baseado em dados do sítio ativo do receptor, utilizando a técnica do *campo complementar do receptor* com resultados promissores (Waller e Marshall 1993). Oprea e colaboradores utilizaram um procedimento semi-automatizado denominado de *NewPred*, para analisar o poder de predição para 30 inibidores da protease do HIV-1 através de um modelo contendo 59 inibidores (Oprea *et al.* 1994). Este método utiliza um número limitado de modos alternativos de interação, na qual várias conformações, para cada um dos compostos, são otimizadas no sítio ativo. O poder de predição para o mesmo grupo de teste utilizado previamente por Waller e Marshall não mudou significativamente (Norinder 1998).

Geralmente, em estudos de *QSAR-3D* são empregados dois métodos distintos na etapa de alinhamento. Pode-se utilizar o alinhamento “rígido” ou “átomo a átomo”, onde o próprio operador seleciona intuitivamente os átomos a serem alinhados, ou o alinhamento através da similaridade de campos onde os campos como, por exemplo, estéricos e eletrostáticos, são calculados e sobrepostos através de suas respectivas similaridades. A figura a seguir ilustra os alinhamentos entre o di-hidrofolato e o metotrexato, ambos inibidores da di-hidrofolato redutase (Böhm *et al.* 1996).

Figura 10. Alinhamentos do metotrexato e do di-hidrofolato realizados pelo modo “rígido” (átomo a átomo) e através da similaridade de campos.

Pelo modo “rígido” ou “átomo a átomo”, figura 10 (a), seria esperado que o alinhamento entre o metotrexato e o di-hidrofolato ocorresse através da sobreposição dos heteroátomos, como indicado na área central da figura. O resultado deste alinhamento indica uma grande semelhança estrutural entre os dois compostos. Entretanto, uma análise detalhada dos campos doadores de ligação de hidrogênio, representado pelas flechas que apontam para as moléculas, e dos campos aceptores de ligação de hidrogênio, representado pelas flechas que apontam para fora das moléculas, indicam uma grande dissimilaridade, figura 10 (b). A translação das moléculas, como indicado na figura 10 (c) mostra que desta forma haverá uma maior similaridade entre os campos doadores e aceptores de ligação de hidrogênio. O modo de ligação do metotrexato foi inicialmente predito através da estrutura 3D da di-hidrofolato redutase complexada com o di-hidrofolato e posteriormente confirmado através de estudos de cristalografia de raios-X (Bystroff *et al.* 1990).

No alinhamento realizado de forma rígida, emprega-se a técnica dos mínimos quadrados para a sobreposição das posições atômicas correspondentes. Este método é o mais empregado em estudos *CoMFA* e sobrepõe duas moléculas através da minimização do valor de F (equação 25). O valor de F é calculado para os pares de átomos (X_i e Y_i) que são

determinados manualmente pelo operador, no processo. O átomo X_i , da primeira molécula, e o átomo Y_i , da segunda molécula, devem ser selecionados para que ocorra a maior sobreposição entre as duas moléculas. Uma delas é mantida fixa enquanto a outra é rotacionada e transladada até que o menor valor de F seja alcançado. O valor do rms (desvio quadrático médio) representa uma medida da qualidade dos ajustes.

$$F = \sum_i |X_i - Y_i|^2 \quad \text{Equação 25}$$

Este procedimento é muito eficaz na “descoberta” de diferenças entre moléculas que aparentemente são similares. O ponto fraco deste método é a necessidade de se decidir quais os pares atômicos devem ser selecionados. O alinhamento rígido não pode ser aplicado a sistemas moleculares para os quais as correspondências átomo a átomo não podem ser visualizadas. Entretanto, uma similaridade rigorosa da estrutura atômica não é um pré-requisito para a interação de diferentes moléculas com o mesmo receptor (Itai et al. 1993; Höltje & Folkers 1995). Entretanto, o método da sobreposição convencional dos quadrados médios é considerado inadequado para se tentar ajustar as moléculas dissimilares devido as dificuldades de se encontrar os átomos que são correspondentes entre as duas estruturas, como foi observado na figura 10. Desta forma, existe a necessidade de se empregar um outro método para realizar os alinhamentos.

O programa SEAL, por exemplo, realiza a etapa de alinhamento baseado na sobreposição das moléculas através da similaridade dos campos estéricos, eletrostáticos e hidrofóbicos (Kearsley e Smith 1990). Este método combina os fatores estéricos, eletrostáticos e hidrofóbicos com o método de Monte Carlo, que fornece uma busca automática de todos os possíveis alinhamentos. As vantagens da utilização deste método são:

- i) Não existe um pré-alinhamento átomo a átomo especificado pelo usuário, desta forma, até mesmo os alinhamentos pouco intuitivos não serão descartados;
- ii) O método é automático e exaustivo, permitindo que todos os possíveis alinhamentos sejam considerados e
- iii) As orientações são pontuadas de acordo com a média de similaridade, permitindo que os alinhamentos sejam classificados objetivamente.

O alinhamento através da similaridade dos campos eletrostáticos pode ser calculado pela inversão do sinal do termo coulômbico. Isto fará com que as cargas semelhantes se sobreponham. A função de alinhamento é naturalmente atenuada pelo inverso da distância. Contudo, devido ao fato de que um perfeito recobrimento atômico irá implicar em uma distância igual a zero e a função de alinhamento irá se tornar indefinida, um parâmetro

adicional deve ser adicionado. Por exemplo, o termo coulômbico pode ser ajustado para se prevenir uma distância igual a zero pela introdução de um parâmetro k na lei de Coulomb (Equação 26). O valor apropriado de k deve ser determinado para não afetar os alinhamentos entre as duas estruturas.

$$\frac{-q_i q_j}{\sqrt{(r^2 + k)}} \quad \text{Equação 26}$$

O programa SEAL utiliza a similaridade entre as cargas parciais atômicas para realizar o alinhamento através dos campos eletrostáticos. Existem várias maneiras diferentes de se tratar as propriedades eletrostáticas. O programa considera a magnitude do campo potencial interagindo com uma carga pontual; o gradiente do campo eletrostático interagindo com um dipolo e campos altamente atenuados interagindo com uma enzima através de um meio com constante dielétrica da água. Diferentemente da similaridade estérica que possui um curto alcance, as propriedades eletrostáticas são de alcance muito maior.

O alinhamento através da similaridade dos campos estéricos é realizado pela utilização de uma forma funcional abstrata que calcula o recobrimento dos volumes das moléculas. As regiões ocupadas pelos átomos podem ser representadas através do recobrimento dos volumes das esferas atômicas. Quando duas esferas de mesmo raio são recobertas, elas irão possuir um mesmo volume de $4\pi R^3/3$. A expressão algébrica para a intersecção do volume de duas esferas é ajustada para ser cúbica em r (Equação 27).

$$V(r) = \frac{2\pi\left(\frac{r^3}{8} - \frac{3rR^2}{2} + 2R^3\right)}{3} \quad \text{Equação 27}$$

Na expressão anterior, R é o raio da esfera e r é a distância de separação entre as esferas. Quando r for igual a zero, esta será o volume de uma esfera de raio R e quando r for igual a $2R$ o volume de recobrimento será zero. Infelizmente, fora destes limites, a expressão de volume possui pouco significado físico. O requisito para uma aproximação para $V(r)$ deve ser que esta seja isotropicamente simétrica e ter um máximo quando r é zero e diminuir rapidamente quando r for muito maior em relação a R .

A função de alinhamento para as moléculas é dada pela equação 28, que compreende uma dupla somatória sobre todos os possíveis pares de átomos entre as duas estruturas. O valor de A_F é o escore de similaridade entre um alinhamento particular.

$$A_F = -\sum_{i=1}^m \sum_{j=1}^n W_{ij} \exp\{-\alpha r_{ij}^2\} \quad \text{Equação 28}$$

Na equação 28, o subscrito i percorre sobre os m átomos da primeira estrutura e o subscrito j percorre sobre os n átomos da segunda estrutura. Na exponencial, o termo r é a distância entre o átomo i da primeira estrutura e o átomo j da segunda estrutura. A forma exponencial foi escolhida para atenuar a influência de cada átomo na estrutura a ser alinhada, com respeito ao respectivo átomo da outra estrutura. O parâmetro α serve para ajustar esta influência. O fator pré-exponencial W_{ij} serve para atribuir pesos de cada distância de interação para um par de estruturas. A sua forma é dada pela equação 29 como uma função das cargas parciais, q_i , e raios de van der Waals, v_i .

$$W_{ij} = W_E q_i q_j + W_S v_i v_j + \dots \quad \text{Equação 29}$$

Nesta equação W_E é o peso das contribuições eletrostáticas e W_S é o peso das contribuições estéricas. Os alinhamentos podem ser realizados considerando-se a importância destes dois campos. Desta forma, pode-se atribuir um peso maior para qualquer um dos campos ou se atribuir pesos iguais.

Um estudo sistematizado para se determinar os valores ideais dos pesos estérico e eletrostático foi realizado para uma série de dipiridobenzodiazepinonas inibidoras da transcriptase reversa do vírus HIV-1 (Ishiki *et al.* 2001). Neste trabalho foi constatado a dependência do alinhamento em função dos valores dos pesos atribuídos, fator este que influencia diretamente os modelos *CoMFA* gerados.

Outras propriedades físico-químicas podem ser adicionadas na função do alinhamento, como por exemplo, o campo hidrofóbico. Todos os termos da função de alinhamento são negativos, significando que os recobrimentos dos pares atômicos de sinais semelhantes serão favoráveis e os de sinais opostos serão desfavoráveis ao alinhamento. Os coeficientes dos volumes estéricos utilizados são obtidos através dos raios de van der Waals. Existem quatro parâmetros de alinhamento que devem ser ajustados: α , W_E , W_H e W_S .

A equação 29 pode ser rearranjada para $S(Wq_i q_j + v_i v_j)$, onde S é um fator de escalonamento geral que não deveria possuir efeito nos resultados dos alinhamentos. O único parâmetro de peso, W , decide qual das duas interações, estéricas ou eletrostáticas, será a mais importante.

1.3.2.5 Análise dos Resultados: visualização dos mapas de contorno

Os resultados em *CoMFA* e em *CoMSIA* podem ser visualizados como mapas de contorno tridimensionais. Estes resultados são representados na forma de poliedros coloridos que indicam quais regiões do espaço devem conter, ou não, átomos ou substituintes para

favorecer a atividade dos compostos bem como as características destes átomos/substituintes. Em *CoMFA*, as características dos substituintes se limitam as propriedades estéricas e eletrostáticas. Por outro lado, a metodologia *CoMSIA* permite que adicionalmente sejam estudadas as propriedades hidrofóbicas, doadoras de ligação de hidrogênio e aceptoras de ligação de hidrogênio. Desta forma, a análise dos parâmetros estatísticos juntamente com a análise dos mapas de contorno permite prever com mais precisão as características estruturais que os novos ligantes devem possuir para se obter uma maior atividade biológica.

Descreve-se na literatura (Bostrom *et al.* 2003) o estudo de um grupo de receptores antagonistas da dopamina com o emprego das metodologias *CoMFA* e *CoMSIA*. Estas duas análises forneceram modelos com bom poder de predição e mapas de contorno que auxiliam a interpretação dos resultados como, por exemplo, os mapas de contorno estérico, ilustrado na figura 11.

Figura 11. Mapas de contorno estérico relacionado à afinidade dos receptores D₂ e D₄ da dopamina. Nas áreas dos poliedros verdes a presença de grupos volumosos deverá aumentar a afinidade com o sítio de ligação ao passo que nas áreas dos poliedros vermelhos a presença de grupos volumosos deverá diminuir a afinidade com o sítio de ligação.

Através destes mapas de contorno pode-se notar que a introdução de substituintes relativamente volumosos próxima ao elemento farmacóforo G, figura 11, deve diminuir a afinidade do receptor D₂, mas aumentará a afinidade do receptor D₄.

1.3.2.6 Metodologias para validação dos modelos de *QSAR* clássico

A qualidade dos modelos de regressões multivariadas geralmente é determinada através dos parâmetros estatísticos gerados pelo modelo, como por exemplo: r^2 , r^2 e Q^2 ajustado, critério de informação Akaike, utilizado para verificar se os dados estão sobreestimados, dentre outros parâmetros (Todeschini *et al.* 2004). Vários métodos de *QSAR* implementam o procedimento de autovalidação *leave-one-out* ou *leave-some-out*. O resultado

obtido através do processo de autovalidação é o r^2 (Q^2_{cv}) que geralmente é utilizado como critério de robustez e habilidade de predição do modelo. Vários autores consideram valores altos de Q^2_{cv} ($Q^2_{cv} > 0,5$) como um indicador de que o modelo possui um bom poder de predição (Golbraikh & Tropsha 2002). Embora valores baixos de Q^2_{cv} para o grupo de treinamento possam ser utilizados para indicar um baixo poder de predição do modelo, o oposto não é necessariamente verdadeiro. Realmente, valores altos de Q^2_{cv} não implicam automaticamente que o modelo gerado irá possuir um bom poder de predição.

Para que um modelo seja aceito deve-se levar em consideração a razão do número de objetos sobre o número de variáveis selecionadas. Como proposto por vários estatísticos, o valor mínimo deste valor, para um modelo aceitável, deve ser igual a 3, sendo que um valor igual a 5 seja mais seguro. Desta forma, deve-se utilizar uma variável independente (parâmetro físico-químico) para cada 5 compostos, tornando as análises estatísticas válidas.

De uma forma geral, bons modelos de regressão devem possuir as seguintes características:

- i) Alta capacidade de predição com um número mínimo de variáveis;
- ii) Baixa correlação entre as variáveis do modelo;
- iii) Interpretação aceitável dos modelos gerados.

Por outro lado, modelos de regressão que possuem as seguintes características devem ser evitados:

- i) Modelos deficientes de uma ou mais variáveis relevantes, *e.g.*, com capacidade insuficiente de ajuste;
- ii) Modelos sem poder de predição absoluto;
- iii) Modelos com uma acentuada diferença entre o poder de predição e o poder de ajuste;
- iv) Modelos com baixo poder de predição calculados em um grupo de validação externo;
- v) Modelos com variáveis ruidos, *e.g.*, variáveis com chance de correlação.

Na tentativa de se validar os modelos gerados através das análises de *QSAR* pode-se utilizar diversas metodologias como, por exemplo, a regra QUIK, a regra do Q^2 assintótico, a regra com base na função R, dentre outras (Todeschini *et al.* 2004), que serão discutidas a seguir.

1.3.2.6.1 A regra QUIK

A regra QUIK foi proposta em 1999 (Todeschini *et al.* 1999) e é um critério que permite a rejeição de modelos com alta colinearidade de predição, o que acarreta em uma correlação ao acaso. A regra QUIK se baseia no índice de correlação multivariada K (Todeschini 1997) que mede a correlação total existente entre um grupo de variáveis e, é definida como:

$$K = \frac{\sum_j |(\lambda_j / \sum_j \lambda_j) - (1/p)|}{2(p-1)/p} \quad \text{Equação 30}$$

$j = 1, \dots, p$ e $0 \leq K \leq 1$

onde, λ são os autovalores obtidos através da matriz de correlação do grupo de dados $X(n,p)$, n sendo o número de objetos e p o número de variáveis.

Esta regra é derivada a partir da suposição de que a correlação total em um modelo, de um dado grupo de predição X , somado a resposta Y (K_{XY}) às vezes pode ser maior do que a medida no grupo de predição (K_X). Em outras palavras, a determinação de K_{XY} é realizada considerando a resposta Y como uma variável X e calculando a matriz de correlação correspondente.

Por esta razão, na regra QUIK, apenas modelos com correlação K_{XY} entre as variáveis $[X+Y]$ maiores que a correlação K_X entre as variáveis $[X]$ podem ser aceitas, ou: se $K_{XY} - K_X < \delta K \rightarrow$ rejeitar o modelo, onde, δK é um limiar definido pelo usuário (e.g. 0,01-0,05). O limiar δK também pode ser escolhido como sendo igual a zero. De qualquer forma, valores negativos de limiar não são permitidos, sendo modelos teoricamente inaceitáveis com diferenças negativas $K_{XY} - K_X$.

A regra QUIK mostrou ser muito eficaz em evitar modelos multi-colineares, sem poder de predição. Por outro lado, esta regra não é suficientemente eficaz para rejeitar modelos com mais de uma variável ruído, desde que as variáveis ruídos não estejam usualmente correlacionadas e, então, diminuem a correlação total K_X .

1.3.2.6.2 A regra do Q^2 assintótico

É amplamente aceito que bons modelos devem possuir uma pequena diferença entre a habilidade de ajuste e a de predição. De fato, diferenças marcantes entre os valores de r^2 e Q^2_{cv} podem ser devidas a um sobreajuste (altos valores de r^2) ou para alguns exemplos não preditos (baixos valores de Q^2_{cv}). Mattioni e Jurs (Mattioni & Jurs 2002) relataram

explicitamente a propriedade para este modelo, propondo o seguinte como função objetiva (*custo*) na busca do modelo.

$$Custo = rms_T + 0,4 |rms_T - rms_{cv}| \quad \text{Equação 31}$$

onde, rms_T e rms_{cv} são os erros das raízes dos quadrados médios do grupo de treinamento e teste, respectivamente, e 0,4 é um parâmetro empírico utilizado para ponderar a diferença entre a habilidade de ajuste e predição.

Para permitir esta aproximação empírica, um critério é proposto como regra do sim ou não com base no comportamento assintótico de Q^2 (Q^2_{ass}). Foi demonstrado que Q^2_{LOO} é assintoticamente relacionado com r^2 e, desta forma, um valor assintótico de Q^2 pode ser calculado pela seguinte expressão:

$$Q^2_{ass} = 1 - (1 - R^2)(n/n-p')^2 \quad \text{Equação 32}$$

onde, n é o número de objetos e p' o número de parâmetros utilizados no modelo.

A regra assintótica de Q , proposta para estimar o modelo, é baseada na comparação dos valores do Q^2 assintótico e Q^2 real: se $Q^2_{LOO} - Q^2_{ass} < \delta Q \rightarrow$ rejeitar o modelo.

Para derivar este modelo, assume-se que o modelo com valor atual de Q^2_{LOO} inferior ao valor assintótico não garante uma habilidade de predição. O valor limiar simples é $\delta Q=0$, sendo que um valor mais conservador de $-0,005$ e um valor menos conservador de $0,005$ podem ser utilizados.

1.3.2.6.3 A regra da função R

Esta regra detecta modelos que possuem um excesso de “bom poder preditivo” bem como modelos com um excesso de “mau poder preditivo”.

Dado um modelo de regressão com p variáveis, sendo $F_{j,y}$ o valor absoluto do coeficiente de correlação entre o j ésimo preditor e a resposta Y . Na tentativa de se analisar o papel de cada variável no modelo, a seguinte quantidade é definida como:

$$M_j = \frac{R_{jy}}{R} - \frac{1}{P} \quad -\frac{1}{P} \subseteq M_j \subseteq \frac{p-1}{p} \quad \text{Equação 33}$$

Neste caso, cada variável contribui com $1/p$ para a correlação múltipla R . Cada contribuição R_{jy}/R do modelo atual é comparado com o valor $1/p$ que se refere ao modelo de referência para calcular o papel de uma variável simples no modelo. As diferenças entre estes dois contribuintes positivo e negativo têm sido coletadas, de acordo com seu sinal, em termos de R^P e R^N , definido como o seguinte:

$$R^P = \prod_{j=1}^{p^+} (1 - M_j (\frac{P}{p-1})) \quad \text{Equação 34}$$

$$M_j > 0 \text{ e } 0 \leq R^P \leq 1$$

$$R^N = \sum_{j=1}^{p^-} M_j \quad \text{Equação 35}$$

$$M_j < 0 \text{ e } -1 < R^N \leq 0$$

onde o produto em R^P percorre sobre as variáveis; p^+ fornece uma diferença positiva M_j , a soma em R^N sobre as variáveis p^- fornece uma diferença negativa, $p^+ + p^- = p$.

Cada termo do produto R^P representa o complemento para 1 de cada diferença positiva ($m_j > 0$) escalonado ao seu valor máximo $(p-1)/p$. Esta é uma maneira de se penalizar pesos para os modelos de predição. O valor é baixo se a predição possuir alta correlação com a resposta, assumindo valor zero quando R_{jy} igual a R e $p > 1$. A função R^P é o produto destas funções de penalidades.

O produto dos pesos de penalidade, para as variáveis modeladas, é escolhido com o propósito de possuírem baixos valores para a função R^P , quando apenas uma variável no modelo multivariado ($p > 1$) possui um valor, R_{jy} muito próximo ao valor do coeficiente de correlação múltipla. Neste caso, o produto contribui para zero considerando-se o valor dos outros termos, significando que outras variáveis no modelo são utilizáveis desde que eles não contribuam para o aumento da correlação múltipla, por isso, o modelo sozinho é considerado complexo em relação a sua qualidade. Ao contrário, se cada preditor presente no modelo explicar a mesma fração $1/p$ da correlação multivariada total R , a função R^P é igual a 1.

A regra R^P é finalmente estabelecida como: se $R^P < t^P \rightarrow$ rejeitar o modelo; onde, t^P é um limite definido pelo usuário variando entre 0,01 e 0,1 dependendo dos dados. Sugere-se o valor para $t^P = 0,05$.

Suponha que o coeficiente de correlação r e a simples resposta de correlação predita são: $r = 0,6$; $r_{1y} = 0,6$; $r_{2y} = 0,6$; $r_{3y} = 0,6$; então R^P torna-se:

$$\prod_{p^+} \left[1 - \left(\frac{0,6}{0,6} - \frac{1}{3} \right) \frac{3}{2} \right] = \prod_{p^+} \left[1 - \frac{2}{3} \frac{3}{2} \right] = 0 \quad \text{Equação 36}$$

indicando uma não aceitação do modelo.

A quantidade R^N (equação 35) é a soma das diferenças negativas M_j dadas por todos os preditores cuja razão do R_{jy} sobre R é inferior a $1/p$, por isso ele contém a informação relacionada a todos os preditores que devido as suas baixas correlações com a resposta pode provavelmente ser variáveis inúteis.

Esta quantidade R^N considera as informações diferentes com respeito a R^P . Ela considera para um excesso de variáveis não modeladas ou inúteis e pode ser pensado como sendo uma medida de sobreajuste devido as variáveis com ruídos. Ela adquire um valor máximo igual a zero quando não existem variáveis com ruído no modelo.

Agora, assume-se que todas as variáveis não modeladas possuem a mesma baixa correlação com resposta igual a ε , então, cada uma das variáveis fornece uma contribuição a R^N como:

$$\frac{\varepsilon}{R} - \frac{1}{p} = \frac{p\varepsilon - R}{pR} \quad \varepsilon \ll R \quad \text{Equação 37}$$

O valor de ε pode ser ajustado pelo usuário, dependendo do conhecimento da resposta Y . Além disto, pode ser assumido que não mais de uma variável ruído é permitida no modelo, então um valor limiar para R^N pode ser estimado como:

$$t^N(\varepsilon) = \frac{p\varepsilon - R}{pR} \quad \text{Equação 38}$$

A opção de aceitar eventualmente pelo menos uma variável, com baixa correlação com a resposta, infere sobre a impossibilidade de um conhecimento imediato da variável ser apenas ruído ou útil na modelagem. Além disto, parece ser razoável que apenas uma variável é permitida para a modelagem residual.

Finalmente, dado um valor limiar, a regra R^N pode ser definida como: se $R^N < t^N(\varepsilon) \rightarrow$ rejeitar o modelo.

Diferentemente de R^N que, por definição, pode ser apenas negativa, o limiar t^N pode ser positivo. Obviamente, neste caso, o modelo será rejeitado de acordo com a regra R^N . Contudo, surge uma interessante consideração: um valor positivo do limiar significa que independentemente da correlação do par X - Y o modelo deve ser rejeitado devido ao pequeno valor do coeficiente multivariado R com respeito ao nível de ruído escolhido ε . Neste caso, a correlação do modelo com a resposta é dada apenas pelo fator ao acaso.

Aumento no valor de ε acarreta o aumento no valor do limiar (valores menos negativos) então modelos com variáveis ruídos irão ser rejeitados mais facilmente. Por exemplo, dado um modelo com 3 variáveis e correlação multivariada R de 0,6 valores de ε igual a 0,01; 0,05 e 0,1 originando os seguintes valores de limiar: -0,317; -0,250 e -0,167. Se não for realizada nenhuma suposição no parâmetro de ruído ε (e.g. $\varepsilon = 0$), o valor de limiar torna-se independente para R e igual a $1/p$.

1.3.3 Métodos de modelagem com o conhecimento do alvo biológico

As interações que ocorrem entre um ligante e uma proteína são essenciais para vários processos biológicos. Métodos teóricos e computacionais são muito utilizados para analisar, investigar e prever as interações ligante-macromolécula. Os métodos computacionais podem contribuir para o planejamento e desenvolvimento racional de novos compostos-líder, já apresentando características de fármacos, para uma variedade de doenças, diminuindo desta forma a necessidade de muitos esforços experimentais prolongados e caros (Klebe 2000; Halperin *et al.* 2002; Taylor *et al.* 2002). Como consequência, a pesquisa utilizando métodos *in silico* para a realização de estudos de *docking* está aumentando gradativamente (Bajorath 2002).

O número de algoritmos disponíveis que permitem avaliar e racionalizar as interações que ocorrem entre ligante-proteína é grande e está crescendo. Muitos algoritmos compartilham metodologias comuns embora com novas abrangências. Devido a diversidade na complexidade dos algoritmos observa-se um excesso de técnicas e abordagens tentando resolver os problemas modernos no planejamento de novos fármacos, baseado na estrutura.

Assumindo-se que a estrutura do receptor é conhecida o desafio inicial na descoberta e na otimização de compostos-líder é a predição tanto da orientação do ligante bem como de sua afinidade. Os algoritmos utilizados para solucionar este problema têm recebido uma atenção especial (Blaney & Dixon 1993) indicando, desta forma, a importância do *docking* para projetos de desenvolvimento de novos fármacos. E, devido ao aumento do poder de processamento dos computadores agora é possível realizar o estudo de *docking* de milhares de ligantes em pouco tempo.

As principais técnicas atualmente disponíveis para a realização dos estudos de *docking* envolvem os seguintes métodos: (i) dinâmica molecular; (ii) método de Monte Carlo; (iii) algoritmos genéticos; (iv) métodos baseados em fragmentos; (v) métodos de pontos complementares; (vi) métodos da distância de geometria e (vii) métodos de buscas sistemáticas.

Geralmente utiliza-se *docking* molecular em métodos de busca virtual (*virtual screening*) (Walters *et al.* 1998) nos quais bibliotecas com grande número de compostos virtuais são reduzidas a um subconjunto menor em tamanho. Para que um algoritmo de *docking* seja utilizado como uma ferramenta de busca virtual tanto a velocidade bem como a exatidão deste algoritmo são importantes.

Os procedimentos de *docking* podem ser descritos, de uma maneira geral, como uma combinação de duas componentes: uma estratégia de busca e uma “*scoring function*” (função de pontuação). O algoritmo de busca deve gerar um determinado número de configurações, incluindo o modo de ligação determinado experimentalmente. Um algoritmo rigoroso de busca deve verificar e explorar exaustivamente todos os possíveis modos de ligações entre ligante-receptor. Todos os seis graus de liberdade translacional e rotacional do ligante devem ser explorados juntamente com os graus de liberdade conformacional interno do ligante e da proteína. Entretanto, esta busca é impraticável devido ao tamanho do espaço de busca. Para um sistema simples (Welch *et al.* 1996) que compreende um ligante com quatro ligações flexíveis e seis parâmetros para alinhamento rígido, o espaço de busca foi estimado da seguinte forma: os parâmetros de alinhamento são utilizados para posicionar o ligante relativamente à proteína em um sítio ativo cúbico que mede $10^3 \text{ \AA}^3$. Se os ângulos forem submetidos à incrementos de 10 graus e os parâmetros translacionais à grades de $0,5 \text{ \AA}$ deverá haver uma correspondente busca de aproximadamente 6×10^{14} configurações.

A aplicação prática de tal “busca exaustiva” envolve a amostragem de muitos estados de energia desfavoráveis que podem limitar o sucesso de um algoritmo de otimização. Desta forma, na prática, os custos computacionais são limitados aplicando-se restrições, impedimentos e aproximações para reduzir a dimensão do problema na tentativa de se encontrar eficientemente o mínimo global. Uma aproximação comum nos algoritmos de *docking* foi considerar o ligante e o seu sítio como corpos rígidos e, explorar somente os seis graus de liberdade translacional e rotacional. Um dos primeiros exemplos deste algoritmo foi implementado no programa DOCK (Kuntz *et al.* 1982). Embora este método tenha sido bem sucedido em determinados casos (Shoichet & Kuntz 1996), há uma limitação no *docking* “de corpos rígidos”, sendo que a conformação do ligante deve ser próxima a da conformação observada experimentalmente, quando ligado ao seu sítio. Além disso, vários exemplos de mudança conformacional no alvo, com o ligante no sítio, por exemplo, a ligação da ciclosporina A com a ciclofilina (Wüthrich *et al.* 1991) exigiu incorporar a flexibilidade conformacional no algoritmo de busca. Uma aproximação comum para se modelar a flexibilidade molecular é considerar somente o espaço conformacional do ligante, supondo-se um receptor rígido durante todo o processo e protocolo de *docking*. Entretanto, o algoritmo de busca é apenas uma parte do problema do *docking*; o outro fator a ser incorporado é a “*scoring-function*”.

Gerar vários modos de ligação é ineficaz sem um modelo para classificar as “*scoring-function*” de cada conformação de forma precisa e eficiente. A “*scoring-function*” deve

distinguir os modos de ligação experimentais de todos os modos restantes, através do algoritmo de procura. Uma “*scoring-function*” rigorosa terá alto custo computacional e, desta forma, seu grau de complexidade será reduzido levando a uma perda de exatidão.

Métodos de “*scoring*” podem variar desde campos de força de mecânica molecular tais como AMBER (Cornell *et al.* 1995), OPLS (Jorgensen & Tirado-Rives 1988) ou CHARMM (Brooks *et al.* 1983), até “*scoring-function*” relacionada à energia livre (Eldridge *et al.* 1997) ou “*knowledge-based function*” (funções baseadas no conhecimento) (Muegge & Martin 1999).

Os métodos atualmente disponíveis de *docking* utilizam uma “*scoring-function*” dentre dois modos possíveis. A primeira aproximação utiliza a “*scoring-function*” total para classificar uma conformação do ligante-proteína. O método alternativo é usar uma “*scoring-function*” de dois estágios. Estes métodos fazem suposições sobre a hiper-superfície de energia, omitindo frequentemente os termos com grande custo computacional, tais como os termos eletrostáticos e, considera somente alguns tipos de interação, tais como as ligações de hidrogênio. Assim, tais algoritmos são orientados às áreas importantes como aquelas determinadas por uma “*scoring-function*” reduzida. Os programas GOLD (Jones *et al.* 1997) e DOCK (Ewing & Kuntz 1997) são exemplos de métodos que utilizam a primeira aproximação.

Em muitas “*scoring-functions*” existentes ocorre uma séria limitação que é a tendência de se negligenciar os efeitos de solvatação ou utilizar modelos de solvente de modo instantâneo. Um método instantâneo envolve estruturas geradas no vácuo, que são subsequenteemente ordenadas usando uma “*scoring-function*” que inclui um modelo de solvente. A função de busca é, portanto, dirigida ao espaço conformacional que favorece as conformações no vácuo. Além disso, o papel da estrutura das moléculas de solvente ligadas e de íons não são frequentemente considerados. Por exemplo, nos sistemas da protease do HIV-1 (Miller *et al.* 1989) foi demonstrado que as moléculas de água explícitas possuem um importante papel na interação do ligante com o seu sítio (Lam *et al.* 1994).

1.3.3.1 Dinâmica Molecular

Existem vários programas capazes de executar as simulações de dinâmica molecular (DM) como, por exemplo, o AMBER (Kollman 1996) e o CHARMM (Brooks *et al.* 1983). A dinâmica molecular envolve cálculos para solucionar as equações de movimento Newtoniano. Frequentemente uma trajetória de DM torna-se aprisionada a um mínimo local e não consegue

ultrapassar as barreiras conformacionais de alta energia. Assim sendo, a qualidade dos resultados de uma simulação padrão de DM é extremamente dependente da conformação inicial do sistema.

O trabalho original de Di Nola (Di Nola *et al.* 1994) foi posteriormente refeito (Mangoni *et al.* 1999) utilizando simulações de DM para realizar o *docking*, em solução, de alguns ligantes flexíveis a receptores também flexíveis. O problema em se obter amostras adequadas se deve principalmente a separação do movimento do centro de massa do substrato dos seus movimentos internos e rotacionais.

Pak e Wang (Pak & Wang 2000) aplicaram um novo método de DM, ao *docking* de ligantes flexíveis, utilizando uma técnica onde uma “*scaling function*” foi aplicada às equações de movimento para facilitar a passagem da barreira energética através da redução efetiva da grandeza das forças. A dinâmica molecular multi-canônica considera, em especial, o problema da amostragem limitada de conformações e foi utilizado por Nakajima e colaboradores (Nakajima *et al.* 1997) como uma técnica de *docking* para ligantes flexíveis.

Estes métodos se aplicam à apenas uma única estrutura. Entretanto, em estudos de *docking* é comum se empregar um subconjunto de proteínas usando dinâmica molecular. Tais técnicas foram aplicadas por Carlson e McCammon (Carlson & McCammon 2000), nas quais múltiplas estruturas da proteína foram utilizadas ao invés de uma única estrutura flexível.

1.3.3.2 Métodos de Monte Carlo

Métodos de Monte Carlo (MC) são entre outras, uma técnica de otimização estocástica amplamente utilizada. A combinação dos modelos de energia potencial atomístico com técnicas de busca estocástica produz alguns dos métodos mais poderosos para predição e otimização das geometrias moleculares. Uma vantagem significativa da técnica de MC comparada aos métodos baseados em gradientes, tais como a DM, é que uma função simples de energia pode ser utilizada sem a necessidade de informações adicionais. Além disso, através da seleção criteriosa do tipo de movimento, as barreiras de energia podem ser facilmente ultrapassadas. Os métodos baseados em gradientes são, em geral, eficientes na otimização local, mas apresentam dificuldades para percorrer uma hiper-superfície irregular.

O método padrão de MC (mais corretamente, MC de Metropolis (Metropolis *et al.* 1953)) envolve a aplicação de movimentos cartesianos aleatórios ao sistema, sendo que a aceitação ou a rejeição dos movimentos é baseada em uma distribuição probabilística de Boltzmann.

As implementações iniciais de AutoDock (Goodsell & Olson 1990; Morris *et al.* 1996) utilizavam a simulação de MC de Metropolis com uma avaliação da energia baseada em grades, utilizando o campo de força do AMBER, para o *docking* de ligantes flexíveis na cavidade de ligação de um receptor rígido. O algoritmo foi testado originalmente em seis complexos e, conseguiu reproduzir os correspondentes modos de ligação experimentais, embora as estruturas de mínima energia não correspondessem sempre as conformações encontradas no cristal.

Prodock (Trosset & Scheraga 1999a) é um programa que utiliza uma técnica de minimização de Monte Carlo para o *docking* de ligantes flexíveis a sítios de ligações flexíveis, utilizando as coordenadas internas para representar as estruturas. Este método se diferencia do procedimento padrão de MC, pois, após cada movimento aleatório um gradiente local baseado na minimização é executado, e a estrutura resultante é aceita baseando-se nos critérios do Metropolis. Uma técnica baseada em grades para avaliar a função de energia foi incorporada ao algoritmo usando o procedimento de Bezier (Trosset & Scheraga 1999b) que produz uma função atenuada que possa ser diferenciada. Esta propriedade é crucial para o gradiente local baseado na minimização. O tamanho de cada movimento aleatório é determinado a partir de uma avaliação da curvatura da hiper-superfície usando a segunda derivada da função de energia. Assim, os grandes movimentos são executados nas áreas de pequena curvatura e os pequenos movimentos são executados nas áreas de grande curvatura. O programa Prodock utiliza os campos de força do AMBER (Cornell *et al.* 1995) e do ECEPP/3 (Némethy *et al.* 1992) juntamente com um modelo de solvatação baseado no volume exposto do solvente.

Mecânica de coordenadas internas (ICM) (Abagyan *et al.* 1994) é um programa para executar o *docking* flexível do ligante-proteína utilizando o método de minimização de MC aplicado às coordenadas internas. Um termo eletrostático de solvatação é adicionado à energia ECEPP/3 que é calculada usando a aproximação do MIMEL (Abagyan & Trotoev 1994). Este programa foi utilizado em *docking* de CASP-2 (Liu & Wang 1999), onde para os 8 complexos testados apenas um resultou em um RMS de 1,8 Å em relação a estrutura cristalográfica. Os outros complexos forneceram um valor de RMS de no mínimo 3,0 Å. Entretanto, em vários outros casos as previsões de *docking* são razoáveis, com aproximadamente 50% de acertos.

O programa MCDOCK (Liu e Wang 1999) (versão 1.0) aplica uma estratégia de múltiplos estágios para o *docking* de ligante flexível a um receptor rígido. O primeiro estágio do *docking* é colocar o ligante no sítio de ligação. Em seguida, movimentos aleatórios são aplicados ao ligante na tentativa de se reduzir as sobreposições dos átomos do ligante e da proteína. O método de Monte Carlo de Metropolis é então aplicado executado usando uma

“*scoring-function*” baseada no campo de força do CHARMM (Brooks *et al.* 1983). Esta etapa é seguida de uma simulação de MC que utiliza uma temperatura que pode ser ajustada. Este método permite um ajuste na temperatura de forma a aumentar a razão de aceitação dos resultados. Esta metodologia foi aplicada à inibidores da protease do HIV-1 (Bouzida *et al.* 1999). Neste estudo dois ligantes flexíveis foram utilizados para realizar o *docking* com a estrutura cristalográfica rígida da protease do HIV-1 juntamente com algumas moléculas de água. Apenas um dos ligantes reproduziu com sucesso os resultados experimentais.

1.3.3.3 Algoritmos Genéticos

Desde o início, os algoritmos genéticos (AG) se tornaram populares como uma ferramenta para otimização. Deve-se notar que os AG necessitam de uma população inicial, sendo que os métodos MC e DM convencionais necessitam de apenas uma única estrutura inicial, em sua execução padrão.

A essência de um AG é a evolução de uma população de possíveis soluções através dos operadores genéticos (mutações, cruzamentos e migrações) à uma população final, otimizando uma função de ajuste pré-definida. Os graus de liberdade são codificados em genes ou seqüências binárias e à esta coleção de genes, ou cromossomos, atribui-se um ajuste baseado em uma “*score-function*”. O operador de mutação altera aleatoriamente o valor de um gene, realiza trocas no cruzamento de um grupo de genes de um cromossomo pai a outro e, produz movimentos de genes individuais de uma sub população à outra.

GOLD (Jones *et al.* 1997) é um programa de *docking* que utiliza uma estratégia de busca de AG e, inclui a flexibilidade rotacional para átomos de hidrogênio do receptor previamente selecionados. A codificação dos genes é utilizada para representar os diedros rotacionais e as ligações de hidrogênio do ligante-receptor. Em seguida, um operador de movimento de um AG é aplicado aos cromossomos pai, que são escolhidos aleatoriamente da população existente. As ligações de hidrogênio do ligante-receptor são combinadas através de um protocolo de ajuste de quadrados médios para maximizar o número de ligações de hidrogênio intermoleculares, para cada movimento do AG. Como consequência, a geração da estrutura do AG é influenciada por estas ligações de hidrogênio. Entretanto, cada estrutura é classificada baseando-se em uma função de ajuste mais complexa. A função de ajuste é a soma dos termos da ligação de hidrogênio, de um potencial de dispersão intermolecular e um potencial intramolecular para a energia interna do ligante.

AutoDock 3.0 (Morris *et al.* 1998) usa um algoritmo genético, como um otimizador global, combinado com uma minimização de energia, como um método de busca local. Nesta implementação do AutoDock o ligante é considerado flexível e o receptor é considerado rígido e representado na forma de grades. O algoritmo genético usa operadores de cruzamento de dois pontos e de mutação. Para cada população nova, o usuário determina a fração a ser submetida a um procedimento de busca local usando um operador aleatório de mutação, onde o passo da etapa deve ser ajustado para fornecer uma razão de aceitação apropriada. A função de ajuste compreende cinco termos: um termo de dispersão/repulsão de Lennard-Jones 6-12; um termo de ligação de hidrogênio direcional 10-12; um potencial coulômbico e eletrostático; um termo proporcional ao número das ligações sp^3 no ligante, para representar a entropia desfavorável da ligação do ligante devido à limitação dos graus de liberdade conformacional e um termo de dessolvatação. Esta “*score-function*” se baseia no campo de força do AMBER na qual foram obtidos os parâmetros da proteína e do ligante.

Este algoritmo foi originalmente testado em sete complexos e todas as estruturas de menor energia possuíam, no máximo, um RMS de 1,14Å com a estrutura cristalográfica.

O programa DARWIN (Taylor & Burnett 2000) combina um AG e uma estratégia de otimização do gradiente local com o campo de força de mecânica molecular do CHARMM-AA, para realizar o *docking* flexível de três complexos. Código binário é usado para descrever a conformação e a posição inicial do ligante; a energia potencial é minimizada localmente usando um método de gradiente e o ajuste do cromossomo é marcado usando a função potencial de energia do CHARMM-AA. As populações são modificadas por mutações padrões e operadores de cruzamento mantendo a proteína rígida. As contribuições dos solventes são avaliadas usando uma versão modificada do programa DelPhi (Nicholls & Honig 1991) para fornecer finitas soluções à equação de Poisson-Boltzmann. Embora o algoritmo de busca possa otimizar a superfície de energia, algumas estruturas foram obtidas com energias menores do que os modos de ligação experimentais. Os falsos positivos produzidos foram atribuídos às limitações na “*score-function*”. A inclusão de moléculas de águas explícitas no sítio de ligação aumentou a precisão do programa.

Um dos primeiros relatos da aplicação de AG ao problema de *docking* foi apresentado por Judson e colaboradores (Judson *et al.* 1994). Um ligante flexível foi usado com interação de sub populações e um gradiente de minimização durante a busca. O método foi testado realizando o *docking* do fragmento Cbz-GlyP-Leu-Leu à termolisina e produziram conformações semelhantes às obtidas experimentalmente. Entretanto, em alguns casos, as energias obtidas foram menores em relação à da conformação cristalográfica.

1.3.3.4 Métodos baseados em fragmentos

O conceito dos métodos de *docking* baseados em fragmentos pode ser descrito através da divisão do ligante em porções separadas ou em fragmentos. A seguir realiza-se o *docking* destes fragmentos com o sítio receptor e posteriormente, realiza-se a união dos fragmentos. Estes métodos necessitam de decisões subjetivas para classificar a importância dos vários grupos funcionais do ligante, o que pode resultar na omissão de possíveis soluções devido as suposições feitas sobre a superfície de energia potencial. Além disso, uma escolha criteriosa do fragmento base é essencial para estes métodos podendo afetar significativamente a qualidade dos resultados.

Um dos programas mais populares para realizar *docking* de fragmentos é o algoritmo FlexX (Rarey *et al.* 1996). A etapa inicial é a seleção de fragmentos base para o ligante da qual possíveis conformações são formadas baseando-se no banco de dados de ângulos torsionais do MIMUMBA (Klebe & Mietzner 1994). Como em todos os métodos baseados em fragmento, a escolha do fragmento base é crucial para o algoritmo, necessitando conter as principais interações com o receptor. As implementações iniciais do FlexX requeriam uma seleção manual deste fragmento base, mas este processo foi automatizado posteriormente (Rarey *et al.* 1997). Depois da seleção do fragmento base realiza-se um procedimento de alinhamento para otimizar o número de interações favoráveis. Estas interações são baseadas primeiramente em confinamentos geométricos da ligação de hidrogênio, mas incluem também as interações hidrofóbicas. Neste estágio, o fragmento base é considerado rígido e três posições do fragmento são mapeados em três posições do receptor. Todas as superfícies acessíveis do receptor são agrupadas e as sobreposições dos ligantes sobre o receptor são realizadas usando o método de Kabsch (Kabsch 1976). As sobreposições são removidas e as energias são calculadas para os fragmentos bases usando a função de Böhm (Böhm 1994). Depois da inserção do fragmento base, o ligante é construído através de incrementos, onde cada fragmento novo é adicionado em todas as posições e conformações possíveis. A melhor solução de cada conjunto é usada na interação seguinte. Este processo é repetido até a completa construção do ligante e as estruturas finais são classificadas usando uma “*score-function*” empírica.

O programa DOCK (versão 4.0) (Ewing & Kuntz 1997) realiza uma busca que permite analisar os diferentes modos de ligação do ligante usando as seguintes etapas: descrição do ligante e da cavidade do receptor como um grupo de esferas, comparação dos grupos de esfera, orientação do ligante e classificação da orientação. Este programa inclui a flexibilidade

do ligante usando uma “*score-function*” modificada que incorpora uma classificação intramolecular para o ligante (Makino & Kuntz 1997). Nesta versão, um fragmento base do ligante é selecionado e colocado no receptor seguido por uma minimização simplex de corpo rígido. As conformações das partes restantes do ligante são procuradas por um método retroativo e, em seguida, minimizadas.

O programa LUDI (Böhm 1992) ajusta os ligantes ao sítio ativo de um receptor através da busca de complementaridade de grupos polares e hidrofóbicos. O programa opera um protocolo de três estágios, onde calcula primeiramente os sítios de interação do ligante e da proteína, que são posições discretas no espaço que podem formar uma ligação de hidrogênio ou preencher espaços hidrofóbicos. Estas posições são obtidas de uma, entre três maneiras possíveis: a partir de distribuições de contatos não ligados baseados em uma busca através do banco de dados estrutural do Cambridge, a partir de um grupo de regras geométricas ou do arquivo de saída do programa GRID (Goodford 1985). O programa GRID calcula as energias de ligação para uma dada “sonda”, por exemplo, uma carbonila, com uma molécula do receptor. A etapa seguinte envolve o ajuste dos fragmentos moleculares aos sítios de interação. A posição do fragmento será aceita se a distância entre os locais de interação no receptor estiver dentro de um valor definido pelo usuário. Uma limitação relativa à flexibilidade interna do ligante pode ser incorporada usando diversos confôrmeros de uma biblioteca de fragmentos. O estágio final do protocolo envolve juntar os fragmentos usando uma base de dados de fragmentos e um algoritmo de ajuste. O programa LUDI foi validado através da predição de modificações para inibidores da protease do HIV e inibidores da diidrofolato resutase (Böhm 1992).

1.3.3.5 Métodos de ponto complementar

Em geral, o *docking* do ligante ao sítio receptor é realizado utilizando-se pontos de complementaridade entre o ligante e a proteína. Assim, de modo discutível, muitos algoritmos de *docking* baseados em fragmentos poderiam ser incluídos nesta categoria, embora uma distinção tenha que ser realizada entre os algoritmos que tratam o ligante inteiro durante todo o *docking* daqueles onde o ligante é dividido em fragmentos.

Jiang e Kim (Jiang & Kim 1991) propuseram o método do “*docking suave*” que permite pequenas alterações conformacionais tanto no ligante bem como no receptor. Neste caso, a superfície molecular é representada por uma série de cubos. A descrição cúbica do ligante é rotacionada e transladada para obter o número máximo de comparações entre o

ligante e os cubos da superfície da proteína, subtraindo o número de cubos do volume sobreposto. Uma comparação da superfície cúbica não deve satisfazer somente uma sobreposição entre os cubos nas duas moléculas, mas também a superfície regular devem estar aproximadamente em sentidos opostos. As soluções do *docking* são agrupadas com base nos vetores de translação e nos ângulos de rotação. O valor médio para cada conjunto é classificado usando uma soma geométrica dos descritores atômicos, que se baseiam nas cargas, nos átomos ou nos grupos doadores e aceptores de ligação de hidrogênio e na hidrofobicidade.

FTDOCK (Gabb *et al.* 1997) é um protocolo de *docking* rígido baseado na complementaridade da forma com um modelo para as interações eletrostáticas. Uma grade é colocada sobre a proteína e sobre o ligante e cada ponto da grade é marcado na superfície ou no interior da molécula. A função de correlação é otimizada pela translação ou rotação de corpo rígido do ligante. Esta função é o produto de um ponto da grade no ligante com o correspondente ponto da grade da proteína, somado sobre todos os pontos. Um “*score*” de correlação alta indica boa superfície de complementaridade entre as moléculas. Esta função é calculada rapidamente usando transformadas de Fourier. Um “*score*” eletrostático também é utilizado baseando-se na carga atômica para a proteína e para o ligante. Este “*score*” adicional é utilizado como filtro binário para remover os falsos positivos com uma elevada contribuição eletrostática.

Jackson *et al.* (Jackson *et al.* 1998) refinaram as interfaces da proteína-proteína para as estruturas geradas pelo FTDOCK, usando o programa MULTIDOCK. Neste trabalho a proteína foi modelada usando uma cópia múltipla das representações das cadeias laterais baseadas em bibliotecas, que foram subsequenteemente refinadas por um método de campo médio seguido pela minimização da energia.

1.3.3.6 Métodos da distância de geometrias

Os métodos da distância de geometrias como os desenvolvidos por Leach e Smellie (Leach & Smellie 1992) determinam os modos de ligação entre o ligante e a proteína considerando apenas as ligações de hidrogênio. Este método averigua os espaços conformacionais identificando os possíveis modos de ligação que serão utilizados para direcionar o algoritmo. Uma aplicação da distância de geometrias para o *docking* de ligantes flexíveis é realizada pelo programa DockIt (Methaphorics LLC). O sítio ativo é representado como um grupo de esferas e as conformações dos ligantes são gerados dentro deste sítio

utilizando a distância de geometrias. As estruturas resultantes são então classificadas usando a “*score-function*” PMF (Muegge & Martin 1999) ou a “*score-function*” PLP (Gehlhaar *et al.* 1995).

1.3.3.7 Buscas sistemáticas

O programa EUDOC (Perola *et al.* 2000) realiza uma busca sistemática de rotações e translações de corpo rígido, de um ligante rígido, dentro de um sítio ativo rígido. Baseia-se no programa SYSDOC (Pang & Kozikowski 1994), anteriormente desenvolvido, que usa uma transformada linear rápida seguida de uma rotação para executar a busca sistemática. A energia intermolecular é calculada pelo campo de força AMBER-AA com o dielétrico dependente da distância.

1.3.3.8 Modelagem do alvo: proteínas

Como apresentado e discutido anteriormente, através dos estudos de *docking* é possível determinar o sítio de ligação de um determinado ligante e desta forma estudar as interações que ocorrem entre ligante-proteína. Entretanto para que estas análises sejam realizadas deve-se conhecer a estrutura tridimensional do receptor, sendo que estes dados podem ser obtidos experimentalmente, por exemplo, através de estudos de cristalografia de raios-X ou de espectroscopia de RMN. Por outro lado, existem várias proteínas que não podem ter as suas estruturas tridimensionais determinadas através destes métodos e os experimentos devem ser complementados através da predição teórica. Embora a estrutura de algumas destas proteínas possam ser elucidadas através da modelagem, freqüentemente existem ambigüidades na disposição das cadeias laterais e incertezas na modelagem das alças que podem ser cruciais nas aplicações de *docking*.

Uma das melhores técnicas computacionais para a predição de estruturas tridimensionais de proteínas é a modelagem baseada na homologia (Johnson *et al.* 1994; Sánchez & Sali 1999; Rost & Sander 2000; D’Alfonso *et al.* 2001; Johnson *et al.* 2003; Santos Filho & de Alencastro 2003). Esta abordagem se baseia em alguns padrões gerais que têm sido observados, em nível molecular, no processo da evolução biológica, a saber:

- i) Homologia entre seqüências de aminoácidos implica em semelhança estrutural e funcional;

- ii) Proteínas homólogas apresentam regiões internas conservadas (hélices- α e fitas- β);
- iii) As principais diferenças estruturais entre proteínas homólogas ocorrem nas regiões externas, constituídas principalmente por alças (*loops*), que liga os elementos das estruturas secundárias.

As proteínas agrupam-se em número limitado de famílias. Estima-se que existam cerca de 5.000 famílias protéicas (Wolf *et al.* 2000). Desta forma, o conhecimento tridimensional de pelo menos um representante de uma família é o suficiente para a modelagem dos demais membros da família através da metodologia de homologia.

A modelagem de uma proteína através da homologia baseia-se no conceito de evolução molecular. Isto é, parte-se do princípio de que a semelhança entre as estruturas primárias desta proteína e das proteínas homólogas de estruturas tridimensionais conhecidas irá implicar em uma similaridade estrutural entre as duas (Chothia & Lesk 1986; Johnson *et al.* 2003; Vogel *et al.* 2004).

Os métodos empregados, atualmente, na modelagem de proteínas por homologia implicam basicamente em quatro passos sucessivos:

- Identificação e seleção de proteínas-molde;
- Alinhamento das seqüências dos resíduos;
- Construção das coordenadas do modelo;
- Validação da proteína modelada.

Existem vários programas computacionais que realizam a modelagem de proteínas através da homologia como, por exemplo, o WHATIF (Vriend 1990), o MODELLER (Sali & Blundell 1993; Sánchez & Sali 1997a; Fiser & Sali 2003), o MULTALIGN (Barton & Sternberg 1987), o MULTAL (Taylor 1988) e o CLUSTALX (Thompson *et al.* 1997).

O programa WHATIF além de realizar a modelagem de proteínas por homologia é eficaz na averiguação da estrutura modelada. Este programa realiza uma série de testes, como a averiguação da simetria, da geometria, das cadeias laterais, etc... A execução da modelagem pode ser realizada via rede, acessando o seguinte endereço: <http://swift.cmbi.kun.nl/WIWWWI/>.

O programa MODELLER utiliza procedimentos de dinâmica molecular para gerar as estruturas. Neste método, o alinhamento é utilizado para produzir restrições geométricas como limites nas distâncias entre pares de átomos de C_α e limites nos ângulos diedros das cadeias laterais, etc... Estas restrições podem ser combinadas para encontrar uma “*scoring function*”

que define a qualidade da estrutura do modelo. A estrutura do modelo que satisfaça estas restrições é gerada através da combinação de métodos de minimização e de dinâmica molecular.

Os programas MULTALIGN, MULTAL e CLUSTALX empregam um algoritmo de alinhamento global para construir um alinhamento sobre toda a seqüência da proteína e diferem principalmente no procedimento empregado para se determinar a ordem do alinhamento destas seqüências. O método mais usual é a ramificação seqüencial para identificar primeiramente duas seqüências semelhantes e, em seguida, alinhar as próximas seqüências semelhantes as que já foram alinhadas.

1.3.3.8.1 Identificação e seleção de proteínas molde

A primeira etapa na modelagem das proteínas é a identificação de pelo menos uma proteína, cuja estrutura cristalográfica seja conhecida, que irá servir posteriormente como molde para a determinação da estrutura 3D da proteína-problema (Peitsch 1997; Peitsch 2002). Geralmente duas situações são possíveis:

- i) Conhece-se a família protéica a que pertence a proteína-problema;
- ii) Não se sabe a que família a proteína problema pertence.

Na primeira situação basta selecionar a proteína a ser utilizada como molde diretamente de um banco de dados como, por exemplo, do *Protein Data Bank*, PDB (Bernstein *et al.* 1977). Esta etapa é realizada através de qualquer algoritmo de procura baseado na similaridade entre seqüências.

No segundo caso, pode-se realizar uma procura sistemática de um ou mais moldes adequados em um banco de dados contendo as estruturas primárias derivadas das proteínas armazenadas no PDB. Adicionalmente, pode-se utilizar ferramentas de busca globais, como o BLAST (Altschul *et al.* 1990), PSI-BLAST (Altschul *et al.* 1997) e o Needleman – Wunsch (Needleman & Wunsch 1970); ferramentas de busca locais como o FastA (Pearson 1990) e o Smith - Waterman (Smith & Waterman 1981) ou bancos de dados de seqüências de pares de bases de DNA ou de aminoácidos, como por exemplo, o GenBank (Benson *et al.* 2000) ou o SWISS-PROT (Appel *et al.* 1994). Nas buscas globais o *score* da similaridade global é calculado entre duas seqüências de uma biblioteca de proteínas e, nas buscas locais a procura pode ser efetuada para seqüências totais ou parciais.

A homologia entre a proteína a ser modelada e o molde é de grande importância nesta abordagem. Se o grau de identidade entre as estruturas primárias da proteína molde e da

proteína problema for igual ou superior a cerca de 25-35%, considerando o número de resíduos superior a 80 aminoácidos, existe uma grande probabilidade de que estas proteínas tenham estruturas tridimensionais semelhantes (Sander & Schneider 1991; Shafferhans & Klebe 2001; Aloy *et al.* 2003, Lewis *et al.* 2004). Neste caso, pode-se realizar o estudo de modelagem por homologia.

Entretanto, haverá casos onde a homologia entre as estruturas primárias da proteína molde e da proteína problema não ultrapassará 25%. Nestes casos deve-se utilizar um procedimento alternativo para a realização da modelagem como, por exemplo, o método de compatibilidade da sequência-estrutura, que se baseia na compatibilidade entre as estruturas terciárias de ambas as proteínas (Bowie *et al.* 1991; Godzik *et al.* 1992; Jones *et al.* 1992; Fischer *et al.* 1996). Neste caso, deve-se escolher uma proteína molde cuja estrutura terciária seja mais relacionada e compatível em relação à sequência da proteína problema. A avaliação da sequência/estrutura é realizada através de um potencial empírico derivado de uma tabela de contatos de resíduos observados em proteínas cujas estruturas são conhecidas (Rost & Sander 1996; Przybylski & Rost 2004). Pode-se também determinar a estrutura 3D da proteína sem nenhum conhecimento prévio da estrutura tridimensional, como por exemplo, através da aplicação de redes neurais (Liu & Rost 2004).

1.3.3.8.2 Alinhamento das sequências de resíduos de aminoácidos

Após a definição do molde, é realizado o alinhamento da sequência problema com a sequência molde. O intuito em se realizar este alinhamento é para alinhar resíduos que são estruturalmente equivalentes considerando-se as características estruturais comuns, como por exemplo, os elementos de estrutura secundária e os resíduos catalíticos. No processo de alinhamento, geralmente vão ocorrer espaços vazios, os denominados “*gaps*”, que são representados nos alinhamentos através de linhas tracejadas e corresponde geralmente às regiões das alças.

Pode ser que na etapa inicial, a procura do molde, mais de uma proteína molde seja encontrada. Neste caso as sequências de todos os moldes devem ser utilizadas no alinhamento com a sequência da proteína problema. Quanto maior o número de proteínas moldes utilizadas melhores serão os resultados, pois este múltiplo alinhamento permite detectar mais facilmente as características estruturas comuns das proteínas homólogas.

Após o alinhamento, pode-se reconhecer as regiões estruturalmente conservadas, que representam as regiões de máxima similaridade, e as regiões variadas, que representam as

regiões sem correspondência estrutural. Os resíduos não utilizados para a construção do modelo como, por exemplo, àqueles situados em regiões de alças não conservadas, devem ser ignorados durante o processo de modelagem (Peitsch 2002).

1.3.3.8.3 Construção do modelo

1.3.3.8.3.1 Modelagem das regiões estruturalmente conservadas

A localização, no espaço tridimensional, do maior número de átomos possível da proteína problema irá depender do alinhamento efetuado (Peitsch 1997; Peitsch 2002). A construção da parte interna da proteína problema baseia-se na hipótese de que a conformação da cadeia principal da estrutura molde pode ser transferida para ela. Existem muitos métodos de modelagem dessas regiões conservadas (Forster 2002; Santos Filho & de Alencastro 2003; Takeda-Shitaka *et al.* 2004), como por exemplo, a modelagem pela união de corpos rígidos (Blundell *et al.* 1987; Browne *et al.* 1969; Greer 1990), a modelagem pela combinação de segmentos (Claessens *et al.* 1989; Levitt 1992) e a modelagem pela satisfação de restrições espaciais (Aszódi & Taylor 1996; Havel & Snow 1991; Srinivasan *et al.* 1993).

1.3.3.8.3.2 Modelagem das regiões de alças

As diferenças funcionais entre membros de uma mesma família protéica são, em geral, consequência de diferenças estruturais na superfície externa das proteínas. Estas diferenças provêm de substituições, eliminações e inserções de resíduos nas cadeias de proteínas homólogas, principalmente nas alças, as regiões mais expostas da proteína. Por serem regiões estruturalmente variáveis, as alças geralmente determinam a especificidade das proteínas, contribuindo para a estrutura dos sítios de ligação. Conseqüentemente, a modelagem das alças é uma etapa essencial na geração de modelos protéicos por homologia (Fiser *et al.* 2000).

A modelagem de alças pode ser vista como um problema de enovelamento das proteínas. A conformação de um segmento da cadeia polipeptídica deve ser obtida a partir de sua seqüência. As alças, entretanto, são geralmente pequenas e isto provoca distorções de enovelamento difíceis de avaliar prontamente (Fiser *et al.* 2000). Além disto, a conformação de um dado segmento é influenciado pela estrutura do resto da proteína. Em outras palavras, as conformações das alças devem ter baixa energia interna e não devem interagir desfavoravelmente com o restante da proteína (Leach 1996). Devido a estas complicações, a

modelagem de alças é, sem dúvida, a etapa mais difícil na predição de modelos de proteínas. Consequentemente, é nestas regiões que a probabilidade de ocorrerem erros de modelagem é maior.

Os métodos de predição de alças são definidos como: *ab initio* (Fine *et al.* 1986; Moult & James 1986; Brucoleri & Karplus 1987) e métodos comparativos, baseados na semelhança com fragmento de estruturas conhecidas (Greer 1980; Jones & Thirup 1986). A predição de alças por métodos *ab initio* baseia-se em uma análise conformacional que depende do ambiente estrutural e é guiada por uma função de energia. Os métodos comparativos utilizam bancos de dados de estruturas de alças de proteínas armazenadas no PDB, nos quais pode-se procurar, para cada alça da proteína problema, estruturas que tenham as conformações energéticas e geometricamente mais favoráveis. Uma vez obtidos os modelos de alça, o sistema é otimizado por mecânica molecular. Esta metodologia é mais eficiente quando o número de aminoácidos das alças é pequeno, geralmente inferior a oito (Peitsch 1997).

Pode-se reduzir, em alguns casos, as limitações inerentes aos métodos comparativos pela aplicação de métodos híbridos, que combinam a abordagem por bancos de dados e os métodos *ab initio* (Chothia *et al.* 1986; Martin *et al.* 1989; Mas *et al.* 1992; van Vlijmen & Karplus 1997; Rohl *et al.* 2004). As alças modeladas são selecionadas de acordo com alguns critérios como a exposição de grupos hidrofóbicos na superfície e a necessidade de baixas energias conformacionais relativas.

Zheng e Kyle (Zheng & Kyle 1996) descreveram uma nova versão para o seu método de modelagem de alças (Zheng *et al.* 1993) onde a conformação aleatória inicial possui todas as ligações escalonadas e a alça se ajusta na região âncora. Em seguida, através da minimização de energia com o campo de força CHARMM (Brooks *et al.* 1983), a alça é relaxada à sua geometria padrão. No trabalho realizado por Zheng e Kyle foram calculadas as conformações para vários resíduos de alças inseridas no ambiente da estrutura cristalográfica da proteína e os valores médios de RMS, para os átomos do esqueleto, diminuíram em média de 1,1Å a 0,7Å, em comparação com o método anterior.

1.3.3.8.3 Modelagem das cadeias laterais

Como no caso das alças, as conformações das cadeias laterais também são modeladas a partir dos dados estruturais e energéticos de estruturas semelhantes, obtidos experimentalmente (Sali 1995; Vásquez 1996; Sánchez & Sali 1997b). A modelagem é realizada com o auxílio de bancos de dados de rotâmeros permitidos de cadeias laterais (χ_n),

0,5Å de RMS, para a cadeia principal, e 0,7Å de RMS para todos os átomos. O algoritmo de busca conformacional demonstrou ser bastante eficiente, sendo que o custo computacional aumenta linearmente com o aumento do tamanho da proteína.

Encontra-se na literatura (Kazmierkiewicz *et al.* 2003) um procedimento automático para a adição automática das cadeias laterais ao esqueleto da proteína. Este procedimento se baseia na otimização de uma função de energia para as cadeias laterais do peptídeo. Este procedimento foi testado em seis estruturas obtidas experimentalmente por raios-X ou por RMN. Os modelos reconstruídos apresentaram um desvio, RMS, entre 0,6-0,9Å para os átomos pesados.

1.3.3.8.3.4 Otimização do modelo gerado

Após a construção do modelo da proteína-problema é necessário otimizar a estrutura. As interações desfavoráveis entre átomos não-ligados ("*nonbonded contacts*"), bem como as energias de ângulos torsionais e de ligação são otimizadas por mecânica molecular (Leach 1996). Como regra geral, esta otimização não pode ser extensiva para não desviar o sistema de seu molde original (Peitsch 1997). Conseqüentemente, as etapas de otimização e validação do modelo devem ser efetuadas simultaneamente.

1.3.3.8.3.5 Validação do modelo

A validação é uma etapa essencial que pode ser executada em diferentes níveis de organização estrutural. Deve-se avaliar a qualidade do empacotamento global da proteína, os possíveis erros estruturais em regiões localizadas e os parâmetros estereoquímicos (Johnson *et al.* 1994).

Como a qualidade do modelo depende muito da estrutura da proteína molde utilizada, é preciso que esta última tenha sido obtida com uma boa resolução e com um fator-R satisfatório, parâmetro para averiguar a qualidade do refinamento cristalográfico. Recomenda-se usar como modelos estruturas cuja resolução seja igual ou superior a 2 Å e fator-R inferior a 20%.

O primeiro requisito que uma proteína adequadamente modelada deve atender é ter uma estrutura terciária satisfatória (Laskowski *et al.* 1998). Sua qualidade depende da proteína escolhida como molde e do alinhamento calculado. É importante verificar se existem grandes

diferenças conformacionais não explicadas entre os elementos da estrutura secundária (regiões conservadas) das estruturas-molde e da estrutura-modelada.

A qualidade estereoquímica do modelo é de fundamental importância. O programa mais utilizado na avaliação dos parâmetros estereoquímicos, o PROCHECK (Laskowski *et al.* 1993), avalia os comprimentos de ligação; os ângulos planos; a planaridade dos anéis das cadeias laterais; a quiralidade; as conformações das cadeias laterais; a planaridade das ligações peptídicas; os ângulos torsionais da cadeia principal e das cadeias laterais; o impedimento estérico entre pares de átomos não-ligados e a qualidade do gráfico de Ramachandran (Ramachandran & Sasisekharan 1968). O gráfico de Ramachandran é particularmente útil porque define os resíduos que se encontram nas regiões energeticamente mais favoráveis e desfavoráveis e orienta a avaliação da qualidade dos modelos teóricos ou experimentais das proteínas. O programa WHATIF também realiza a validação do modelo obtido. Na literatura, pode-se encontrar artigos que utilizaram os programas PROCHECK e/ou WHATIF para a validação dos modelos obtidos por homologia, como o estudo do P450c17 humano, enzima responsável pela catálise das atividades dos esteróides 17 α -hidroxilase e 17,20-liase na biossíntese do andrógeno e estrógeno (Auchus & Miller 1999), o estudo da oligopeptidase B (Gerczei *et al.* 2000) e mais recentemente o estudo da tripsina (Sattar *et al.* 2004).

Além de validar modelos obtidos através da homologia de proteínas, estes programas podem ser utilizados para a validação de dados experimentais de estruturas 3D obtidas tanto por raios-X bem como através de estudos de RMN (Hodsdon *et al.* 1996; Laskowski *et al.* 1996; Doreleijers *et al.* 1998, Laskowski 2001, Laskowski *et al.* 2005).

É necessário avaliar também as interações entre a estrutura modelada e o meio, essencialmente a água. Neste tipo de análise, pode-se usar o programa WHATCHECK (Hooft *et al.* 1996), que dá informações sobre a formação de regiões centrais hidrofóbicas, a acessibilidade de resíduos e átomos a moléculas de solvente (água), a distribuição espacial de grupos iônicos, a distribuição das distâncias atômicas e das ligações de hidrogênio da cadeia principal. No mesmo contexto, o programa PROSAIL (Sippl 1993) avalia o ambiente de cada aminoácido da proteína modelada (validação localizada), tendo como referência o ambiente esperado em proteínas análogas de alta resolução.

Outras metodologias baseadas em mecânica molecular, cálculos de energia livre de solvatação ou métodos estatísticos têm sido testadas (Sauder *et al.* 2000). Estas metodologias são, em princípio, capazes de estimar a qualidade da estrutura terciária dos modelos de proteínas.

1.3.3.8.4 Limitações do método e possíveis erros

Em muitos casos, a probabilidade de ocorrência de erros em um modelo de proteína construído pelo método de homologia é inversamente proporcional ao grau de identidade entre as seqüências de resíduos utilizadas.

No entanto, Chothia e Lesk demonstraram que relações mais complexas entre a qualidade de modelos gerados e o grau de identidade com o molde são possíveis (Chothia & Lesk 1986). Desta forma, a modelagem de uma proteína com baixo grau de identidade com o molde utilizado é geralmente muito mais difícil.

Em geral, os erros não se distribuem igualmente por toda a estrutura (D'Alfonso *et al.* 2001). Como um primeiro critério, os desvios quadráticos médios das distâncias entre os átomos da cadeia principal das estruturas conservadas, dos pares de proteínas, dependem do grau de identidade das seqüências utilizadas (Chothia & Lesk 1986). Pares de proteínas com grau de identidade superior a 50%, aproximadamente, apresentam, em geral, valores do desvio quadrático médio inferiores a cerca de 1,0 Å para a proteína como um todo.

As regiões cuja estrutura é conservada são mais confiáveis em relação as regiões de alças. Quando o grau de identidade tem valor superior a cerca de 30%, os resíduos formados por hélices- α e fitas- β têm, em geral, bom alinhamento com os resíduos correspondentes do molde. Abaixo deste grau de identidade, as fitas aparentemente se alinham melhor que as hélices (D'Alfonso *et al.* 2001).

As etapas mais problemáticas da modelagem de proteínas por homologia são o alinhamento das seqüências e a modelagem das regiões de alças. Muitos avanços têm sido feitos nos métodos de alinhamento de seqüências (Sauder *et al.* 2000). Já a modelagem de alças ainda é um fator limitante, principalmente quando o número de resíduos é grande (D'Alfonso *et al.* 2001). Neste caso, pode-se utilizar as informações contidas em bancos de dados, como já citado anteriormente.

Em resumo, os erros normalmente encontrados em modelos de proteínas gerados por homologia podem ser divididos em cinco categorias (Sánchez & Sali 1997b; Sali *et al.* 1995):

1. *Erros de modelagem das cadeias laterais.* Quando o valor do grau de identidade entre a estrutura-problema e a estrutura-molde é baixo podem surgir problemas de encaixe de algumas cadeias laterais no corpo principal do modelo, formando regiões descontínuas.
2. *Deslizamento em regiões do alinhamento.* Divergências entre as seqüências podem provocar distorções durante a localização dos resíduos da cadeia principal, mesmo que o

enovelamento global do modelo seja satisfatório. A utilização de várias proteínas-molde pode reduzir este tipo de erro.

3. *Erros em regiões sem molde.* Este tipo de erro pode ocorrer em segmentos da estrutura-problema que não têm equivalentes na estrutura-molde, principalmente em regiões de alças grandes. A avaliação crítica do alinhamento proposto e do ambiente em que a alça está localizada pode levar à superação deste tipo de erro.

4. *Erros de alinhamento.* Ocorrem principalmente quando o grau de identidade entre as seqüências envolvidas é inferior ao valor de 30%. Para reduzi-los, recomenda-se alinhamentos múltiplos e, se possível, modificar iterativamente as regiões do alinhamento tidas como problemáticas.

5. *Moldes incorretos.* Este tipo de erro é mais provável quando o grau de identidade entre as duas proteínas é inferior ao valor de 25%. Nestes casos, deve-se buscar parâmetros experimentais.

2. MATERIAIS E METODOLOGIA

2.1 Reagentes e solventes

Ácido Clorídrico (38,0%, Cinética química)

Ácido Sulfúrico (95,0%-98,0%, P.A. – ACS-Synth)

Benzaldeído (99,5%, P.A. Aldrich)

Cloreto de cálcio anidro (CAAL)

Etanol (99,5%, P.A. – ACS-CAAL)

2-piridinocarboxaldeído (99%, Sigma)

Tiossemicarbazida (99%, Aldrich)

Solventes espectroscópicos e/ou deuterados

Dimetilsulfóxido-D6 (DMSO) (99,8%, Merck)

Tetrametilsilano (TMS) (99,9%, Aldrich)

2.2 Equipamentos

Os espectros de ressonância magnética nuclear foram obtidos em um espectrômetro Bruker DPX-300 alocado no laboratório da Central Analítica do Instituto de Química da USP. A precisão nas medidas foi limitada pelo erro sistemático de 0,001% do equipamento, de acordo com as especificações do mesmo.

As análises quantitativas dos elementos: carbono, hidrogênio e nitrogênio dos compostos preparados foram determinadas em um analisador elementar Perkin-Elmer 2400 CHN automatizado alocado no laboratório da Central Analítica do Instituto de Química da USP.

Os valores do ponto de fusão dos produtos obtidos foram determinados em um aparelho tipo Kofler, da marca Thomas/Scientific Aparatus (Micro Hot Stage, modelo 40) e, não foram corrigidos.

Os cálculos utilizando o programa SYBYL (SYBYL Tripos Associates Inc.) e o programa SEAL (Kearsley & Smith 1990) foram realizados em uma estação de trabalho da Silicon Graphics IRIS 2 (R10000) alocada no Laboratório de Computação Científica e Avançada da Universidade de São Paulo.

Os cálculos para a modelagem da proteína através da homologia, utilizando o programa WHATIF (Vriend 1990), a determinação dos descritores, utilizando o programa DRAGON (<http://www.disat.unimib.it/chm/Dragon.html>), e as análises de *QSAR* clássico, utilizando o programa BILIN (Kubinyi 1994b), foram realizados em micro computadores com processadores Pentium III de 300MHz com 64M de memória RAM e Athlon XP de 2.6 MHz com 256M de memória RAM.

2.3 Séries de compostos estudados e parâmetro biológico

Para serem utilizados em análises de *QSAR*, os compostos devem apresentar amplo espectro de variação tanto dos parâmetros físico-químicos/estruturais bem como do parâmetro biológico. Estes requisitos tornam as análises de *QSAR* significativas, tanto do ponto de vista estatístico, bem como, do ponto de vista bioquímico/farmacológico. Adicionalmente, neste trabalho, os compostos estudados foram selecionados da literatura original considerando-se que os experimentos foram realizados por um único grupo de pesquisa seguindo os mesmos protocolos experimentais.

Deste modo, neste trabalho foram previamente selecionados da literatura original (French & Blanz-Jr. 1974) 40 derivados da α -(N)-piridino carboxaldeído tiossemicarbazona substituídos (figura 13) com atividade inibitória frente à RNR de células H.Ep.-2 (compostos 1 a 40, tabela 2), *in vitro*. Os valores de atividade biológica foram expressos por pIC_{50} , ou seja, o logaritmo negativo da concentração molar necessária do composto para inibir 50% da atividade da enzima RNR de células H.Ep.-2 (carcinoma humano da epiderme). A incerteza nos valores de IC_{50} é de aproximadamente $\pm 10\%$.

Este estudo se fundamenta no fato de que alguns compostos, derivados da tiossemicarbazona, testados contra a RNR de células tumorais devem possuir atividade inibitória frente a RNR do *M. tuberculosis* através do mesmo mecanismo (Seydel & Schaper 1995). Adicionalmente, inibidores da RNR do *M. tuberculosis* devem possuir pouco efeito ou nenhum devido a sua baixa homologia (16%) com a RNR de mamíferos (Liu *et al.* 1998).

Figura 13. Estrutura molecular fundamental dos derivados da α -(N)-piridino carboxaldeído tiosemicarbazona substituídos, estudados neste trabalho e retirados de literatura selecionada (French & Blanz-Jr. 1974).

Da série de 61 compostos, original da literatura, apenas 40 destes foram selecionados para efetuar as análises de *QSAR* e *QSAR-3D*, portanto, 21 foram excluídos. Destes compostos, 11 não possuíam os respectivos valores de atividade biológica determinados *in vitro* e os outros 10 possuíam na sua estrutura um grupo hidroxila ligado diretamente no anel piridínico. Admitiu-se, deste modo, que estes 10 compostos em pH fisiológico não se encontram puros na forma fenólica, mas sim como uma mistura das suas formas fenólica e zwitteriônica (French & Blanz-Jr. 1974).

Os valores da atividade inibitória, de cada um dos compostos, determinados sob as mesmas condições e expressos por pIC_{50} estão apresentados na tabela 2. Estes valores foram utilizados como sendo a variável dependente para a obtenção dos modelos de *QSAR* e *QSAR-3D*.

Tabela 2. Valores observados da atividade inibitória (pIC_{50}) frente a ribonucleotídeo difosfato redutase derivada de células H.Ep.-2 (carcinoma humano da epiderme) dos derivados da α -(N)-piridino carboxaldeído tiossemicarbazona, selecionados da literatura (French & Blanz-Jr. 1974).

Número do Composto	R	pIC_{50}
01	2-piridino	6,55
02	3-piridino	2,70
03	4-piridino	3,30
04	1'-CH ₃	5,82
05	2'-CH ₃	2,28
06	4'-CH ₃	5,52
07	4'-n-C ₄ H ₉	5,66
08	4'-C ₆ H ₅	4,96
09	4'-(2-piridila)	5,34
10	4'-4'-(CH ₃) ₂	5,74
11	3-CH ₃	6,59
12	4-CH ₃	6,57
13	5-CH ₃	6,51
14	5-C ₂ H ₅	6,66
15	6-CH ₃	5,11
16	3-OCH ₃	5,89
17	3-OC ₂ H ₅	6,04 ^a
18	3-OOCCH ₃	5,44 ^a
19	3-F	5,42
20	5-F	5,92
21	5-Cl	6,25
22	5-Br	6,30
23	5-I	6,39
24	5-CF ₃	5,62
25	5-OCH ₃	5,92
26	5-OCF ₃	5,60
27	5-OC ₂ H ₅	6,07 ^a
28	5-OC ₂ H ₄ N(CH ₃) ₂	4,62
29	5-O(C ₂ H ₄ O) ₂ C ₂ H ₅	5,69 ^a
30	5-OOCCH ₃	5,44
31	5-OOCC ₂ H ₅	5,28
32	5-n-OOCC ₃ H ₇	5,17
33	5-n-OOCC ₁₅ H ₃₁	3,96
34	5-OOCCH ₂ OCH ₃	5,30
35	5-OOCCH ₂ OC ₂ H ₅	5,25
36	5-OOCCH ₂ N(CH ₃) ₂	5,24
37	5-OOCCH ₂ OC ₆ H ₅	4,89
38	5-NHCOCH ₃	5,92
39	5-N(CH ₃) ₂	6,40
40	3,5-(OC ₂ H ₅) ₂	5,68

^a Valores estimados pelos autores

^b Os números sublinhados em *itálicos* correspondem aos apresentados na literatura original.

2.4 Séries de treinamento e de teste

Para a validação dos modelos de *QSAR/QSAR-3D* os compostos em estudo foram divididos em três grupos de treinamento e em seus três respectivos grupos de teste (I, II e III) (Oprea *et al.* 1994). Os grupos de treinamento e os correspondentes de teste não foram construídos aleatoriamente, mas adotando-se dois critérios: (i) não foram permitidas repetições de compostos nos três grupos de teste e (ii) manteve-se uma variação significativa nos valores da atividade biológica (ver tabela 4).

Inicialmente, ~75% dos 40 compostos pré-selecionados foram divididos em três grupos de treinamento, cada qual com 30 compostos (Grupos de Treinamento IA, IIA e IIIA) e, os correspondentes ~25% restantes, foram utilizados como grupos de teste, cada qual com 10 compostos (Grupos de Teste IA, IIA e IIIA) (tabela 3).

Adicionalmente, quatro novas séries de grupos de treinamento e os correspondentes grupos de teste foram criadas, denominadas séries B, C, D e E, respectivamente. Na segunda série (IB, IIB e IIIB), com 39 compostos, foi excluído das análises o composto 33 (5-n-OOCC₁₅H₃₁); na terceira série (IC, IIC e IIIC), com 30 compostos, foram excluídos das análises os compostos de 2-10 e o composto 33. Por outro lado, a quarta (ID, IID e IIID) e a quinta série (IE, IIE e IIIE), com respectivamente 23 e 22 compostos, possuem somente compostos substituídos na posição 5 do anel piridínico sendo que, para a quinta série, o composto 33 foi excluído.

A exclusão de cada um destes compostos foi realizada considerando-se as suas características e/ou os critérios apresentados a seguir:

- (i) O composto 33 (5-n-OOCC₁₅H₃₁) é muito lipofílico devido à presença de uma longa cadeia carbônica. Admitiu-se, portanto, que o mesmo não contribui significativamente para o modelo, podendo ser considerado um “*outlier*”;
- (ii) Os compostos de 2 a 10 apresentam grupos substituintes no fragmento da tiossemicarbazona. Admitiu-se que as substituições nessa cadeia lateral tornam os compostos estruturalmente muito heterogêneos em relação aos compostos da série original e;
- (iii) Os compostos excluídos da quarta e da quinta séries não apresentam substituintes na posição 5 do anel piridínico sendo considerados inadequados para se realizar uma análise comparativa com modelos de *QSAR* obtidos previamente na literatura, para estas séries de compostos (Gupta *et al.* 1984).

As séries de grupos de treinamento e de teste selecionados estão apresentadas na tabela 3.

Tabela 3. Números dos compostos selecionados para os grupos de treinamento e os correspondentes grupos de teste para as séries A, B, C, D, E, respectivamente.

Série	Grupo	Números dos compostos selecionados	
		Grupo de treinamento	Grupo de teste
A	I	2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 16, 18, 19, 20, 22, 23, 24, 25, 26, 28, 29, 30, 31, 33, 34, 35, 36, 38, 39, 40	1, 4, 7, 10, 13, 17, 21, 27, 32, 37
	II	1, 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, 19, 20, 21, 23, 24, 26, 27, 28, 29, 32, 33, 34, 35, 36, 37, 39	5, 9, 14, 18, 22, 25, 30, 31, 38, 40
	III	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17, 18, 19, 21, 22, 23, 25, 27, 30, 31, 32, 33, 34, 36, 37, 38, 40	6, 12, 15, 20, 24, 26, 28, 29, 35, 39
B	I	2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 16, 18, 19, 20, 22, 23, 24, 25, 26, 28, 29, 30, 31, 34, 35, 36, 38, 39, 40	1, 4, 7, 10, 13, 17, 21, 27, 32, 37
	II	1, 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, 19, 20, 21, 23, 24, 26, 27, 28, 29, 32, 33, 34, 35, 36, 37, 39	5, 9, 14, 18, 22, 25, 30, 31, 38, 40
	III	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17, 18, 19, 21, 22, 23, 25, 27, 30, 31, 32, 33, 34, 36, 37, 38, 40	6, 12, 15, 20, 24, 26, 28, 29, 35, 39
C	I	11, 12, 14, 15, 17, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 32, 34, 35, 36, 38, 39, 40	1, 13, 16, 19, 23, 29, 37
	II	1, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 34, 36, 37, 38	14, 18, 26, 32, 35, 39, 40
	III	1, 11, 12, 13, 14, 16, 18, 19, 20, 21, 24, 25, 26, 27, 29, 30, 32, 35, 36, 37, 38, 39, 40	15, 17, 22, 23, 28, 31, 34
D	I	1, 14, 20, 21, 22, 24, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38	13, 23, 25, 28, 39
	II	1, 13, 20, 21, 22, 23, 25, 26, 28, 29, 30, 32, 33, 35, 36, 37, 38, 39	14, 24, 27, 31, 34
	III	1, 13, 14, 20, 22, 23, 24, 25, 27, 28, 29, 31, 32, 33, 34, 35, 37, 39	21, 26, 30, 36, 38
E	I	1, 14, 20, 21, 22, 24, 26, 27, 29, 30, 31, 32, 34, 35, 36, 37, 38	13, 23, 25, 28, 39
	II	1, 13, 20, 21, 22, 23, 25, 26, 28, 29, 30, 32, 35, 36, 37, 38, 39	14, 24, 27, 31, 34
	III	1, 13, 14, 20, 22, 23, 24, 25, 27, 28, 29, 31, 32, 34, 35, 37, 39	21, 26, 30, 36, 38

2.5 Métodos e recursos computacionais

Os estudos de modelagem molecular, incluindo as abordagens *CoMFA* (Cramer III *et al.* 1988) e as abordagens *CoMSIA* (Klebe *et al.* 1994), foram realizados com o emprego do programa computacional SYBYL versão 6.9 e versões anteriores, da Tripos Inc. (SYBYL Tripos Associates Inc.). Os alinhamentos dos compostos com base na similaridade dos campos estéricos e eletrostáticos foram efetuados com a utilização do programa SEAL (QCPE 634) (Kearsley & Smith 1990). A modelagem por homologia da proteína foi realizada através

do programa WHATIF (Vriend 1990). Todos estes cálculos foram realizados em uma estação de trabalho da Silicon Graphics.

A primeira etapa na modelagem foi a realização da análise conformacional para as geometrias de cada um dos 40 compostos da série, utilizando-se a rotina “*systematic search*” implementada no programa SYBYL (SYBYL Tripos Associates Inc.). Esta análise conformacional foi realizada considerando-se que não existem, na literatura pesquisada até o presente momento, informações cristalográficas sobre esta classe de compostos complexados com a enzima RNR do *M. tuberculosis* ou de proteínas homólogas. Deste modo, a geometria de mínima energia na sua conformação *syn*, para cada um dos compostos, foi considerada como sendo a conformação bioativa.

Em seguida, as geometrias de menor energia, obtidas através da busca sistemática, foram otimizadas com o emprego do método hamiltoniano AM1 (Dewar *et al.* 1985, 1988, 1990) utilizando o programa semiempírico MOPAC 6.0 (Stewart 1990), implementado no programa SYBYL. Neste estudo, o critério da norma do gradiente foi reduzido para 0,0 com o emprego da palavra chave GNORM=0.0 e a otimização do critério de convergência e os cálculos autoconsistentes foram incrementados em 100 vezes com a utilização da palavra chave PRECISE. A palavra chave MMOK foi utilizada para se efetuar as correções de mecânica molecular das “ligações peptídicas” de cada composto. Os compostos foram completamente otimizados sem restrições de geometria com o emprego do algoritmo de Baker (Baker 1986). Neste estudo foram calculadas e empregadas as cargas Gasteiger-Marsili para cada um dos 40 compostos da série (Gasteiger & Marsili 1980; Marsili & Gasteiger 1980).

2.6 Obtenção dos descritores moleculares por cálculo e seleção das variáveis

Após a otimização das geometrias dos compostos, as mesmas foram inseridas no programa DRAGON (<http://www.disat.unimib.it/chm/Dragon.html>). Como já apresentado na Introdução, o programa calcula automaticamente descritores (*i.e.* descritores topológicos, constitucionais, BCUT, RDF, geométricos, de carga, índice de aromaticidade, padrões moleculares de Randic, dentre outros) que foram utilizados como sendo as variáveis independentes nas análises de *QSAR* clássico. Os arquivos de entrada das geometrias dos compostos, gerados pelo programa SYBYL, estavam no formato mol2.

Os arquivos de entrada para as moléculas de cada série foram selecionados na opção “*Calculate Descriptors*”. Na opção “*Descriptor Selection*” selecionou-se todos os descritores

possíveis. Inicialmente todos os 1497 descritores foram calculados. Posteriormente, estes foram submetidos a um pré-tratamento onde foram eliminados os descritores com valores constantes para a série, os descritores que apresentaram valores superiores a 4 desvios padrão (4σ) após o auto-escalamento bem como os que apresentaram intercorrelações entre si com valores superiores a 0,95. Procedendo-se de modo análogo ao utilizado por Marcus T. Scotti (Scotti 2005) em sua dissertação de mestrado, estes critérios foram adotados para que os modelos de *QSAR* gerados não fossem sobreestimados.

Os descritores não eliminados nesta etapa foram submetidos, em seguida, às análises *PLS*, com o emprego do programa SYBYL, cuja rotina foi modificada por Leandro de Rezende, em nosso laboratório. Esta etapa tem como objetivo selecionar apenas os descritores que mais se correlacionam com a atividade biológica dos compostos. Os valores de pIC_{50} dos compostos foram utilizados como sendo a variável dependente e os correspondentes valores dos descritores como sendo as variáveis independentes. Os grupos de treinamento e os respectivos grupos de teste estão apresentados na tabela 3. Nesta etapa foi utilizado o método de *full cross-validation*, com os números de componentes principais variando de 1 a 4. Para a seleção dos melhores modelos obtidos foram considerados os valores dos seguintes parâmetros estatísticos: (i) coeficientes de correlação, r ; (ii) quadrado do coeficiente de correlação da validação cruzada, Q^2_{cv} e (III) valores do desvio padrão da autovalidação, s_{press} .

Para a seleção dos descritores utilizados nas análises de *QSAR* clássico adotou-se os seguintes critérios:

- i) Primeiramente foi escolhido apenas um modelo, para cada variável latente utilizada, com base nos valores dos parâmetros estatísticos r , Q^2_{cv} e s_{press} ;
- ii) Para cada grupo de treinamento e teste foi selecionado um modelo, composto por vários descritores;
- iii) Os descritores selecionados foram aqueles que apareceram em pelo menos dois, dos três, grupos de treinamento e teste.

2.7 Conformação bioativa e alinhamentos dos compostos

Um dos passos de maior importância em *QSAR-3D* é o modo com que os compostos são alinhados. Os campos gerados após o emprego das abordagens *CoMFA* e *CoMSIA* são altamente dependentes dos dados de entrada, ou seja, da disposição das moléculas no espaço. Desta forma, um alinhamento sem considerar a conformação bioativa irá gerar resultados muitas vezes não significativos ou aleatórios.

Neste estudo, foi utilizada a geometria otimizada da *syn*-2-piridino carboxaldeído tiossemicarbazona não substituída, um dos compostos mais ativos ($pIC_{50} = 6,55$) da série, como molde para o alinhamento. Esta geometria foi escolhida considerando-se que: (i) não se dispõe de informações cristalográficas sobre esta classe de compostos complexados com a enzima RNR do *M. tuberculosis* nem de nenhuma enzima homóloga e; (ii) sugere-se na literatura que as tiossemicarbazonas devem estar na conformação *syn* para que ocorra a formação do complexo ligante-ferro (French & Blanz-Jr 1966, 1974).

De acordo com procedimento desenvolvido em nosso mestrado (Ishiki 1999) e apresentado na Introdução desta Tese, os alinhamentos dos compostos foram efetuados por dois métodos distintos, a saber, um alinhamento rígido (átomo a átomo) e um alinhamento através da sobreposição por similaridade dos campos estéricos e eletrostáticos, utilizando-se o programa SEAL (Kearsley 1990). O alinhamento rígido foi realizado com o emprego do programa SYBYL, onde o usuário determina os átomos a serem alinhados. Nesta etapa, todos os átomos “pesados” do anel piridínico e do fragmento da tiossemicarbazona foram utilizados para o alinhamento dos compostos. Para se efetuar o alinhamento átomo a átomo um dos átomos da estrutura molde é selecionado e, em seguida, seleciona-se o correspondente átomo da estrutura a ser alinhada. Este procedimento é realizado, sucessivamente, até que todos os átomos do molde e da molécula a ser alinhada sejam selecionados.

Para o alinhamento realizado através da similaridade dos campos estéricos e eletrostáticos, com a utilização do programa SEAL, há a necessidade de se ajustar previamente três parâmetros, a saber: (i) a contribuição do efeito estérico (W_S); (ii) a contribuição do efeito eletrostático (W_E) e (iii) um fator de atenuamento (α). Neste trabalho foram utilizados os valores propostos na bibliografia original (Kearsley 1990). A seguir, para a realização do alinhamento basta definir a geometria do composto que será utilizado como molde e as geometrias dos compostos que devem ser alinhadas ao molde selecionado. O programa irá calcular automaticamente os campos estéricos e eletrostáticos e os sobreporá através de suas respectivas similaridades.

2.8 Modelagem molecular

Para a série de α -(N)-piridino carboxaldeído tiossemicarbazona substituídas, inibidores da RNR de células H.Ep.-2, a modelagem molecular foi realizada aplicando-se duas metodologias distintas, que já foram descritas na Introdução. No primeiro caso, as metodologias empregadas foram *QSAR* e *QSAR-3D* que não consideram as informações sobre

o alvo, *i.e.*, sobre a enzima RNR, mas somente as relativas aos compostos (ligantes). Na segunda, as informações sobre a enzima RNR foram consideradas e realizadas através da modelagem por homologia da proteína.

2.8.1 Abordagem extratermodinâmica (*QSAR* clássico)

Como apresentado na Introdução, os estudos de *QSAR* clássico (abordagem de Hansch ou abordagem extratermodinâmica) foram realizados para avaliar a natureza e a contribuição relativa de cada um dos diversos descritores calculados para a série de α -(N)-piridino carboxaldeído tiossemicarbazona substituídas, inibidoras da RNR de células H.Ep.-2. Os modelos *QSAR* gerados, expressos pelas equações de correlação, foram obtidos utilizando-se o programa BILIN (Kubinyi 1994b), através do método de regressões lineares múltiplas (MLR). Como variável dependente foram utilizados os valores de pIC_{50} das α -(N)-piridino carboxaldeído tiossemicarbazona substituídas e, como variável independente, foram utilizados os valores dos correspondentes descritores obtidos com o emprego do programa DRAGON (<http://www.disat.unimib.it/chm/Dragon.html>).

As validações estatísticas dos modelos de *QSAR* gerados foram realizadas pela análise dos seguintes parâmetros estatísticos, a saber: (i) valores dos coeficientes de correlação, r ; (ii) valores do desvio padrão na predição dos erros, s ; (iii) valores de (F) do teste de Fischer; (iv) valores do quadrado do coeficiente de correlação da validação cruzada, Q^2_{cv} e (v) valores do desvio padrão da autovalidação, s_{press} .

2.8.2 Procedimento pela abordagem *CoMFA*

De modo geral, procedendo-se como apresentado na Introdução, para cada composto do grupo de treinamento foram calculados os valores das energias de interação estéricas e eletrostáticas, respectivamente, através de um determinado átomo de prova.

Cada um dos compostos foi alocado em um cubo com grades regularmente espaçadas. Neste estudo utilizou-se 2Å como sendo o valor da distância entre as grades. Este valor de distância influencia o número de interações a serem calculadas. Quanto menor a distância entre as grades, maior será o número de interações a serem calculadas. O cubo formado é suficientemente grande para acomodar a molécula e, em seguida, o seu tamanho foi expandido em mais 4Å nas três direções X, Y e Z, respectivamente.

O átomo de prova previamente selecionado percorreu todas as interseções deste cubo calculando os correspondentes valores das energias de interação. Neste estudo, o átomo de prova utilizado foi o átomo de carbono sp^3 , com carga +1. Estes valores de energia são armazenados na forma de uma tabela e utilizados como sendo as variáveis independentes para gerar os modelos *QSAR-3D (CoMFA)* de predição.

Os campos estéricos e eletrostáticos foram escalonados usando o padrão *CoMFA* para igualar os valores de suas contribuições no modelo aplicado. O valor mínimo de $\sigma = 2,0$ kcal/mol foi empregado como filtro de coluna para reduzir os ruídos nas análises *PLS*. Modelos *PLS* com vários números de componentes foram gerados. O modelo *PLS* com o melhor número de componentes foi obtido através da validação cruzada (*cross-validated*) apresentando o menor valor do erro padrão de predição (s_{press}). O melhor número de componentes foi utilizado para derivar a melhor reta do modelo de correlação usando a validação não cruzada (*non-cross-validated*).

De maneira análoga ao estudo de *QSAR* clássico, as validações estatísticas dos modelos de *QSAR-3D* gerados com o emprego da abordagem *CoMFA* foram realizadas pela análise dos seguintes parâmetros estatísticos: (i) valores dos coeficientes de correlação, r ; (ii) valores do desvio padrão na predição dos erros, s ; (iii) valores do teste F, F ; (iv) valores do quadrado do coeficiente de correlação da validação cruzada, Q^2_{cv} e (v) valores do desvio padrão da autovalidação, s_{press} .

2.8.3 Procedimento pela abordagem *CoMSIA*

Como apresentado na Introdução, as abordagens *CoMSIA* e *CoMFA* são similares. De modo análogo, uma tabela de dados foi construída considerando os valores dos índices de similaridade, calculados através de um mesmo átomo de prova que percorreu todas as interseções de uma grade regularmente espaçada.

Neste estudo, novamente foi utilizado um espaçamento de 2Å entre as grades do cubo. Os valores dos índices de similaridade, $A_{F,k}$ (equação 24), entre a molécula de interesse e o átomo de prova, sistematicamente colocado nas interseções da grade, foram calculados. O valor de $\alpha=0,3$ proposto na publicação original foi utilizado neste estudo (Klebe *et al.* 1994; Böhm *et al.* 1999).

Em seguida, analogamente aos procedimentos utilizados na abordagem *CoMFA*, os valores dos índices obtidos por *CoMSIA* foram analisados através da *PLS*. As validações estatísticas foram realizadas de forma análoga às realizadas na abordagem *CoMFA*.

2.8.4 Visualização dos mapas de contorno

Como apresentado na Introdução, os resultados do *QSAR-3D*, obtidos através das abordagens *CoMFA* e *CoMSIA* podem ser visualizados como mapas de contorno tridimensionais. Estes resultados são representados na forma de poliedros coloridos que indicam quais regiões do espaço devem conter, ou não, átomos ou grupos substituintes para favorecer a atividade dos compostos, bem como as características estruturais destes átomos/grupos. No caso das abordagens *CoMFA*, as características dos substituintes se limitam às propriedades estéricas e às eletrostáticas. Adicionalmente, as abordagens *CoMSIA* permitem que sejam incluídas outras propriedades, respectivamente, lipofílicas, doadoras de ligação de hidrogênio e aceptoras de ligação de hidrogênio. Estas propriedades visam mimetizar as possíveis interações significativas existentes entre o ligante e o seu sítio receptor.

Nos mapas de contorno obtidos, os poliedros verdes indicam as regiões favoráveis e/ou permitidas e os poliedros vermelhos indicam as regiões desfavoráveis e/ou proibidas para a presença de grupos substituintes, para cada propriedade (*i.e.* estérica, eletrostática, etc...).

Desta forma, a análise dos parâmetros estatísticos (r , s , F , Q^2_{cv} e s_{press}) conjuntamente com a análise dos mapas de contorno permitiu determinar algumas características físico-químicas que os ligantes devem possuir para se obter maiores valores de atividade biológica.

2.8.5 Modelagem da proteína

Na primeira etapa da modelagem da RNR do *M. tuberculosis* realizou-se uma procura sistemática da estrutura primária no banco de dados do National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>).

A busca foi realizada para se encontrar a sequência primária da proteína. Esta procura foi efetuada ao se ajustar o campo de busca para *proteína* e incluindo as palavras chaves desejadas. Neste caso, as palavras chaves adicionadas na busca foram: *Mycobacterium + tuberculosis + ribonucleotide + reductase*.

Após a determinação das seqüências primárias, estas foram utilizadas na busca por seqüências homólogas à RNR do *M. tuberculosis* através da ferramenta de busca BLAST (Altschul *et al.* 1990). Para este estudo, utilizou-se o banco de dados SWISS-PROT (Appel *et al.* 1994) e apenas as estruturas homólogas com no mínimo 25% de identidade foram

selecionadas. Esta parte do trabalho foi iniciada no laboratório do Prof. Dr. Gerhard Klebe, do Institut für Pharmazeutisch Chemistry, Marburg, Alemanha.

Os estudos de modelagem por homologia foram realizados com o emprego do programa WHATIF (Vriend 1990).

Como estruturas moldes, na modelagem da proteína, foram utilizados os dados cristalográficos obtidos para a RNR da *Salmonella typhimurium* (código PDB 1R2F) (Eriksson *et al.* 1998) e para a RNR do *Corynebacterium ammoniagenes* (código PDB 1KGN) (Hogbom *et al.* 2002). A estrutura cristalográfica da RNR da *Salmonella typhimurium* (código PDB 1R2F) foi obtida com uma resolução de 2,10 Å, sendo que a dimensão da cela utilizada foi de: a = 58,49; b = 71,74 e c = 96,07 e os valores dos ângulos foram: $\alpha = 90,00$; $\beta = 95,01$ e $\gamma = 90,00$. A estrutura cristalográfica da RNR do *Corynebacterium ammoniagenes* (código PDB 1KGN) foi obtida com uma resolução de 1,85 Å, sendo que a dimensão da cela utilizada foi de: a = 50,81; b = 90,71 e c = 136,81 e os valores dos ângulos foram: $\alpha = 90,00$; $\beta = 91,24$ e $\gamma = 90,00$. Comparando-se as seqüências de aminoácidos das RNRs do *M. tuberculosis* e do *C. ammoniagenes*, verifica-se que ambas apresentam 65,3% de identidade para a cadeia R2-1 e 75,9% para a cadeia R2-2. E, comparando-se as seqüências de aminoácidos das RNR do *M. tuberculosis* e da *S. typhimurium* verifica-se que ambas apresentam 64,7% de identidade para a cadeia R2-1 e 71,9% para a cadeia R2-2.

Nos estudos de modelagem por homologia não foram consideradas as moléculas de água existentes.

Adicionalmente, foram utilizados os dados cristalográficos da estrutura da cadeia beta (subunidade menor) da RNR do *M. tuberculosis*, recentemente elucidada (Uppsten *et al.* 2004) para se realizar uma comparação qualitativa dos modelos propostos na modelagem por homologia. A estrutura de raios-X foi extraída do PDB (código 1UZR).

2.9 Métodos de preparação

2.9.1 Preparação do cloridrato de tiosemicarbazida

Procedendo-se de acordo com as indicações da literatura (Ceneviva 1984), em um balão de três bocas de 50 mL foram adicionados 5,99 gramas (75,8 mmol) de tiosemicarbazida e uma solução 50% de etanol-água (v/v) até que se observe a total solubilização do composto. Nesta solução foi borbulhado cloridrato de hidrogênio,

previamente gerado em uma aparelhagem apropriada e montada ao lado, por um período de 30 minutos. Em seguida, mantendo-se a atmosfera de HCl, a mistura de reação foi deixada em repouso por mais 30 minutos, observando-se a formação de cristais brancos.

O gás clorídrico foi gerado através do gotejamento de H₂SO₄ concentrado em HCl concentrado.

O cloridrato de tiossemicarbazida formado cristalizou como um sólido branco e, foi a seguir, isolado. Este foi recristalizado de etanol-água. Os cristais foram filtrados e secos à pressão reduzida por 4 horas.

Por se tratar de um composto intermediário não foram realizadas as análises de RMN e análise elementar. Para este composto foi medido apenas o seu ponto de fusão.

2.9.2 Preparação do benzenocarboxaldeidotiossemicarbazona

Procedendo-se de acordo com as indicações da literatura (Ceneviva 1984), em um béquer de 50 mL, foi preparada uma solução de benzaldeído, a partir de 1,07 g (10,0 mmol) deste composto, previamente purificado por destilação em atmosfera de nitrogênio, dissolvidos em 20,0 mL de etanol tratado e seco, conforme indicações da literatura (Armarego & Perrin 1988). E, em um erlenmeyer de 100 mL foi preparada uma solução de cloridrato de tiossemicarbazida a partir de 1,20 g (10,0 mmol) dissolvidos em 40,0 mL de água destilada.

Em seguida, a solução preparada de benzaldeído foi transferida para um funil de separação de 60 mL e, adicionada lentamente à solução aquosa de cloridrato de tiossemicarbazida, sob agitação, colocada em um erlenmeyer de 40 mL.

Observou-se a formação de um sólido branco cristalino. Este foi recristalizado de uma mistura água/etanol (10%, v/v). A seguir, este foi filtrado e, seco sob pressão reduzida.

O composto foi identificado e caracterizado através dos espectros de RMN ¹H e o seu grau de pureza foi avaliado através da análise elementar (CHN) e ponto de fusão.

2.9.3 Preparação da 2-piridino carboxaldeído tiossemicarbazona

Procedeu-se de acordo com as indicações da literatura (Ceneviva 1984).

Em um béquer de 50 mL, foi preparada uma solução a partir de 1,07g (10,0 mmol) de 2-piridino carboxaldeído, previamente destilado em atmosfera de nitrogênio, e de 20,0 mL de etanol tratado e seco conforme indicações da literatura (Armarego & Perrin 1988). Esta solução foi em seguida transferida para um funil de separação de 60 mL. Em um erlenmeyer

de 25 mL foi preparada uma solução a partir de 1,20g (10,0 mmol) de cloridrato de tiossemicarbazida dissolvida em 40,0 mL de água destilada. A solução de 2-piridino carboxaldeído foi lentamente adicionada, sob agitação constante, à solução aquosa de cloridrato de tiossemicarbazida. Observou-se a formação de um sólido amarelo. Este foi recristalizado de uma mistura água/etanol (10%, v/v). Em seguida, o precipitado de cor amarela cristalino foi filtrado e seco sob pressão reduzida.

O composto foi identificado e caracterizado através do espectro de RMN ^1H e o seu grau de pureza foi avaliado através da análise elementar (CHN) e ponto de fusão.

2.10 Ensaio biológicos

Esta parte do trabalho foi realizada pela Dra. Ana Olímpia de Souza e sob a responsabilidade do Prof. Dr. Célio Lopez Silva no Departamento de Bioquímica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo. E, de acordo com protocolo já estabelecido naquele laboratório. O *M. tuberculosis* H37Rv foi cultivado em meio de cultura na ausência de antibiótico e na presença das amostras em teste. A cultura na ausência de antibiótico foi considerada como 100% de crescimento bacteriano (controle).

Deste modo, foram ensaiadas contra o *M. tuberculosis* H37Rv, os dois compostos por nós preparados, a benzenocarboxaldeidotiossemicarbazona (HBTSC) e a 2-piridino carboxaldeído tiossemicarbazona (H2pyTSC), respectivamente. Os valores correspondentes da Concentração Inibitória Mínima (CIM) de cada composto para a cepa padrão de *M. tuberculosis* H37Rv foram determinados e, os valores de IC_{50} foram obtidos de acordo com a inibição do crescimento do *Mycobacterium tuberculosis* H37Rv. Os dados foram obtidos a partir das curvas dose – resposta.

3. RESULTADOS E DISCUSSÃO

3.1 Introdução

Há mais de quinze anos diversos aspectos dos estudos de *Relações Quantitativas entre Estrutura química e Atividade biológica* (siglas em inglês *QSAR/QSAR-3D*) vem sendo abordados em nosso grupo de pesquisa no Laboratório de *QSAR* e Modelagem Molecular, no Instituto de Química da Universidade de São Paulo, São Paulo. Dentre os estudos mais relevantes realizados neste laboratório estão a determinação de propriedades físico-químicas/estruturais (eletrônicas, lipofílicas, estéricas e àquelas relativas à polarizabilidade) de compostos bioativos, sintetizados no laboratório, visando propor modelos de *QSAR/QSAR-3D*. Neste contexto, um dos trabalhos pioneiros envolveu estudos de *QSAR* de anestésicos locais e de bloqueadores da transmissão do impulso nervoso, estruturalmente análogos da procaína (Tavares 1987; Amaral *et al.* 1991; Amaral *et al.* 1993; Sousa 1997; Amaral *et al.* 1997; Malvezzi *et al.* 2001; Tavares & Amaral 1997; Siqueira 2001; Malvezzi 2003).

Além destes, as seguintes séries de compostos bioativos foram estudados, a saber: antibacterianos (Baroni 1987; Tavares 1993), incluindo-se derivados nitrofurânicos 5-substituídos (Pires 1998; Pires *et al.* 2001; Cheng *et al.* 2002); antitumorais derivados de sais de alquilamônio de N-alkil arilcetonas substituídas (Ramineli 2001) e, mais recentemente de derivados não nucleosídeos inibidores da transcriptase reversa (NNRTI) do vírus HIV-1 (Ishiki 1999; Ishiki, *et al.* 2001), derivados nitrofurânicos (Rando *et al.* 2002) e tiossemicarbazonas (Scotti 2005) antichagásicos e, derivados piridínicos-alfa-substituídos inibidores potências da RNR do *Mycobacterium tuberculosis* tema desta tese de doutorado.

Em todos os projetos de nosso grupo de pesquisa no Laboratório de *QSAR* e Modelagem Molecular, no IQUSP, procuramos, de um modo geral, para séries selecionadas (sintetizados no grupo ou da literatura) desenvolver e/ou aplicar as metodologias utilizadas em *QSAR/QSAR-3D* para descrever, de modo quantitativo, a(s) interação(ões) entre o composto químico e o sistema biológico. Deste modo, procuramos contribuir para: (i) elucidar o(s) mecanismo(s) das interações expressas por cada parâmetro ou, característica estrutural; (ii) esclarecer o mecanismo de ação e (iii) prever derivados mais potentes.

No presente trabalho, destinado à tese de doutorado, foram aplicadas diferentes abordagens de *QSAR* e de *QSAR-3D* no estudo de 40 derivados da 2-piridino carboxaldeído tiossemicarbazona, inibidores da RNR de células H.Ep.-2, selecionados da literatura (French & Blanz-Jr. 1974). Estes estudos foram realizados sabendo-se (Cory *et al.* 1994) que alguns

derivados da tiossemicarbazona são inibidores da RNR de células tumorais e que para estes compostos, a atividade observada frente o *M. tuberculosis* poderia envolver mecanismos semelhantes (Seydel & Schaper 1995). Ainda mais, considerando-se a baixa homologia (16%) observada entre as RNRs de *M. tuberculosis* e de mamíferos. (Liu *et al.* 1998) a RNR de *M. tuberculosis* é reconhecida, atualmente, como um alvo biológico potencialmente relevante para o planejamento destes inibidores.

Considerando-se a ausência de informação sobre a estrutura cristalográfica da enzima RNR do *M. tuberculosis*, nesta tese os estudos de *QSAR-3D* foram iniciados com o emprego das metodologias propostas em *CoMFA* e em *CoMSIA*. Além destas, foi realizada a modelagem por homologia da enzima ribonucleotídeo redutase do *M. tuberculosis*.

No início de nosso trabalho, em 2000, não havia estudo cristalográfico disponível sobre a estrutura 3D da enzima RNR do *M. tuberculosis*. Posteriormente, em 2004, Uppsten e colaboradores (Uppsten *et al.* 2004) publicaram artigo divulgando a estrutura cristalográfica da subunidade menor da RNR do *M. tuberculosis*.

Os 40 derivados da 2-piridino carboxaldeído tiossemicarbazona foram divididos em cinco séries, a saber: séries A, B, C, D, e E contendo 40, 39, 30, 23 e 22 compostos, respectivamente. Esta divisão foi realizada na tentativa de tornar cada série estruturalmente mais homogênea e, por conseqüência obter resultados mais precisos e, estatisticamente mais significativos, sobre as propriedades físico-químicas e suas contribuições relativas para a atividade biológica destes compostos. Para cada série selecionada, foram criados adicionalmente três grupos de treinamento e seus respectivos grupos de teste, (I, II e III). Esta divisão, em grupos de treinamento e de teste, foi realizada com o intuito de se verificar o poder de predição dos modelos gerados através das análises de *QSAR/QSAR-3D*.

Para os estudos de *QSAR* as variáveis independentes utilizadas foram os valores dos descritores gerados através do programa DRAGON (<http://www.disat.unimib.it/chm/Dragon.html>). A partir das geometrias dos compostos, otimizadas através do método semiempírico AM1, o programa DRAGON gera 1497 descritores de várias naturezas. A seguir, estes descritores foram submetidos à uma pré-seleção e analisados através do algoritmo de mínimos quadrados parciais, *PLS (Partial Least Squares)* (Lindberg *et al.* 1983). Deste modo, foi possível selecionar os descritores mais relevantes para a atividade biológica desta série de compostos. Os descritores selecionados foram submetidos à análise *QSAR* clássica, utilizando o programa BILIN (Kubinyi 1994b).

Para os estudos de *QSAR-3D* envolvendo as abordagens *CoMFA* e *CoMSIA*, as geometrias otimizadas dos compostos devem ser primeiramente alinhadas, para que todas elas

estejam na mesma conformação. Idealmente, dever-se-ia realizar o alinhamento das geometrias na conformação bioativa, o que nem sempre é possível, considerando-se a falta dos dados cristalográficos ligante-receptor. Neste estudo, as geometrias de mínima energia foram adotadas como sendo a conformação bioativa.

Estas geometrias foram alinhadas utilizando-se duas abordagens diferentes: o primeiro denominado rígido (átomo à átomo) e, um segundo realizado de forma automática com a sobreposição por similaridade dos campos estéricos e eletrostáticos, do programa SEAL (Kearsley 1990). Em seguida, os compostos foram submetidos às abordagens *CoMFA* (Cramer III *et al.* 1988) e *CoMSIA* (Klebe *et al.* 1994), visando-se determinar as principais características físico-químicas/estruturais que influenciam a atividade biológica dos compostos.

Os estudos envolvendo a modelagem por homologia da enzima ribonucleotídeo redutase do *M. tuberculosis* foram realizadas com o emprego do programa WHATIF (Vriend 1990). Primeiramente, foi realizada uma busca sistemática para se determinar a sequência primária da enzima. E, em seguida, uma nova procura para se encontrar proteínas homólogas à esta, observando-se como resultado duas proteínas homólogas cujas informações cristalográficas estavam disponíveis no banco de dados do Protein Data Bank (PDB). A comparação entre as RNRs do *M. tuberculosis* e do *C. ammoniagenes* (código PDB 1KGN), mostrou que ambas possuem uma identidade de 65,3% para a cadeia R2-1 e, de 75,9% para a cadeia R2-2. E, a comparação entre as RNRs do *M. tuberculosis* e da *S. typhimurium* (código PDB 1R2F), mostrou que ambas possuem uma identidade de 64,7% para a cadeia R2-1 e uma identidade de 71,9% para a cadeia R2-2. As estruturas destas duas proteínas foram utilizadas como molde para a modelagem por homologia. Posteriormente, em literatura recente (Uppsten *et al.* 2004) foram publicados os dados cristalográficos da cadeia beta (subunidade menor) da RNR do *M. tuberculosis*, que foi então comparada ao modelo por nós proposto na modelagem por homologia.

3.2 Escolha da série de compostos

Neste trabalho foram previamente selecionados da literatura original (French & Blanz-Jr. 1974) 40 derivados da α -(N)-piridino carboxaldeído tiossemicarbazona substituídos, com atividade inibitória frente à ribonucleotídeo redutase (RNR) de células H.Ep.-2 (compostos 1 a 40, tabela 2). Na literatura original (French & Blanz-Jr. 1974) foram estudados 61 compostos, dos quais 11 não possuíam os respectivos valores de atividade biológica

determinados *in vitro* e, os 10 restantes possuíam em sua estrutura o grupo hidroxila no anel piridínico. Considerou-se que estes 10 compostos, em pH fisiológico, não se encontram puros, mas sim, como uma mistura das suas formas fenólica e zwitteriônica.

Os valores da atividade biológica, ou seja, a quantidade de cada composto necessária para inibir 50% da atividade da enzima ribonucleotídeo difosfato redutase, foram expressos por pIC_{50} . Estes valores foram utilizados como sendo a variável dependente nas análises de *QSAR/QSAR-3D*. Para esta série, a variação nos valores de pIC_{50} é igual a 4,38.

Os compostos da série original foram divididos em 5 séries, a saber:

- i) Série A, todos os 40 compostos;
- ii) Série B, todos os compostos exceto o composto 33;
- iii) Série C, todos os compostos exceto os compostos 2 a 10 e composto 33;
- iv) Série D, apenas os compostos substituídos na posição 5 do anel piridínico e
- v) Série E, apenas os compostos substituídos na posição 5 do anel piridínico exceto o composto 33.

Este procedimento foi adotado visando-se homogeneizar, estruturalmente, as séries e, verificar a influência dos substituintes apenas na posição 5 do anel piridínico, uma vez que a maioria dos compostos possui substituintes nesta posição.

Cada uma das cinco séries foi novamente subdivida em três grupos de treinamento e, nos respectivos grupos de teste, visando-se validar os modelos *QSAR/QSAR-3D* gerados. Cada grupo de treinamento (I, II e III) engloba ~ 75% dos compostos da série correspondente e, os compostos restantes (~ 25%) formam os grupos de teste. Ainda mais, procurou-se em cada grupo de treinamento e de teste, selecionar compostos que apresentassem faixas de variação, relativamente amplas, de pIC_{50} , visando-se assegurar resultados estatisticamente significantes. Os valores máximo e mínimo observados na variação de pIC_{50} foram, respectivamente, iguais a 4,38 e a 1,77 para o grupo de treinamento, enquanto para o grupo de teste estes foram de 4,38 e de 1,01, respectivamente. Deve-se observar que as séries D e E, respectivamente, com 23 e 22 compostos possuem menor número de compostos do que as séries A e B, respectivamente, com 40 e 39 derivados. A variação nos valores de pIC_{50} em cada grupo está apresentada na tabela a seguir.

Tabela 4. Valores das variações máximas de atividade biológica (ΔpIC_{50}^a) nos grupos de treinamento e de teste selecionados para as cinco séries, sendo que a identificação dos compostos de cada série estão apresentados na tabela 3.

Série	Grupo	ΔpIC_{50} do grupo de treinamento	ΔpIC_{50} do grupo de teste
A	I	4,38	1,66
	II	3,89	4,38
	III	4,48	1,95
B	I	4,38	1,66
	II	3,89	4,38
	III	4,38	1,90
C	I	2,04	1,66
	II	1,97	1,49
	III	1,77	1,77
D	I	2,70	1,89
	II	2,59	1,38
	III	2,70	1,01
E	I	1,77	1,89
	II	1,93	1,38
	III	2,04	1,01

^a ΔpIC_{50} = Valores da diferença entre o maior e o menor valor de pIC_{50} no grupo.

3.3 Otimização das geometrias

Como apresentado anteriormente, no item 2.7, na literatura pesquisada até o presente não existem informações sobre a conformação bioativa desta classe de compostos. Desta forma, para as análises *QSAR* foram utilizadas para cada composto, as geometrias de mínima energia na conformação *syn*, como a conformação bioativa. Esta aproximação foi feita sabendo-se de literatura, que apenas para o confômero *syn* observa-se a formação do complexo entre o ferro da RNR, acarretando sua inativação.

Inicialmente, a geometria de cada composto foi submetida a uma busca sistemática para a determinação das possíveis geometrias de mínima energia através do método de mecânica molecular. Nesta etapa, todas as ligações, exceto às do anel piridínico, foram submetidas à rotação com incrementos de 15°, no intervalo de 0° a 360°. Os grupos substituintes do anel piridínico foram também submetidos à rotação para a determinação da geometria de mínima energia. Como resultados foram obtidas duas conformações de mínima energia, ou seja, os confômeros *syn* e *anti*. Considerando-se que apenas o confômero *syn* é responsável pela atividade inibitória da enzima RNR, este foi selecionado para a segunda otimização.

Em seguida, as conformações de menor energia obtidas foram submetidas novamente a uma minimização através do método semiempírico AM1, com o emprego do programa

MOPAC, implementado no programa SYBYL. Os procedimentos adotados nesta etapa estão apresentados no item 2.5.

3.4 Obtenção dos descritores moleculares por cálculo e seleção dos descritores

Considerando-se que as propriedades físico-químicas de compostos bioativos, são um reflexo de sua estrutura química, estas propriedades podem então ser descritas quantitativamente se forem expressas através de parâmetros físico-químicos ou descritores estruturais adequados.

Neste estudo foram determinados descritores estruturais como, por exemplo, os topológicos, constitucionais, BCUT, RDF, geométricos, de carga, índice de aromaticidade e padrões moleculares de Randic, dentre outros, através do programa DRAGON.

Para a obtenção dos descritores, as geometrias de mínima energia, no formato mol2, foram inseridas no programa DRAGON que calculou automaticamente 1497 descritores de várias naturezas. Todos os 1497 descritores obtidos foram submetidos a um pré-tratamento onde foram eliminados os descritores que apresentaram as seguintes características:

- i) Valores constantes na série;
- ii) Apenas um valor diferente na série;
- iii) Valores maiores do que o limite de 4 desvios padrão ($\pm 4\sigma$) em relação ao valor médio do respectivo descritor na série e
- iv) Valor de intercorrelação entre os descritores superior a 0,95.

Desta forma, o número de descritores foi reduzido e este foi diferente para cada grupo de compostos.

Assim, para as análises realizadas na série com 40 compostos restaram 673 descritores (44,9%), para a série com 39 compostos restaram 641 descritores (42,8%), para a série com 30 compostos restaram 619 descritores (41,3%), para a série com 23 compostos restaram 624 descritores (41,7%) e para a série com 22 compostos restaram 608 descritores (40,6%).

Os descritores que não foram eliminados nesta pré-seleção foram submetidos, em seguida, à análise de mínimos quadrados parciais (PLS), visando-se selecionar apenas os principais descritores que contribuem para pIC_{50} . Os valores de pIC_{50} e dos descritores foram utilizados como sendo as variáveis, respectivamente, dependentes e independentes.

Os compostos foram divididos em três grupos de treinamento e nos respectivos de teste, como apresentado na tabela 3. Para os compostos dos grupos de treinamento e os

respectivos de teste, utilizou-se o método de *full cross-validation*. Foram gerados quatro modelos *PLS*, para cada um dos três grupos de treinamento e para os respectivos de teste. A diferença entre os quatro modelos *PLS* gerados é o número de variáveis latentes utilizado, que, neste estudo variou de um a quatro. Para a realização das análises *PLS* utilizou-se o módulo *QSAR* do programa Sybyl.

Os gráficos apresentados a seguir ilustram os resultados obtidos após a realização das análises *PLS*. Estes gráficos mostram as variações, dos valores dos coeficientes, respectivamente, de correlação (r^2), de correlação da validação cruzada (Q^2_{cv}) e dos correspondentes valores do desvio padrão da autovalidação (s_{press}) obtidos para o grupo de treinamento e de teste em função do número de parâmetros físico-químicos/estruturais gerados pelo programa DRAGON. A seleção dos melhores modelos *PLS* gerados foi realizada considerando-se o maior valor do coeficiente de predição (Q^2_{cv}) gerado pelo método *full cross-validation* para os três grupos de treinamento e de teste.

Os gráficos apresentados nas figura 14 a 28 sugerem o número de descritores DRAGON, ideal e estatisticamente validados pelos modelos *PLS*, que podem descrever a atividade de inibição da enzima RNR de células H.Ep.-2, para a série de compostos estudada nesta tese.

Figura 14. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste I, obtidos para a série de 40 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 15. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste II, obtidos para a série de 40 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 16. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste III, obtidos para a série de 40 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 17. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste I, obtidos para a série de 39 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 18. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste II, obtidos para a série de 39 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 19. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste III, obtidos para a série de 39 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 20. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises PLS, para os grupos de treinamento e de teste I, obtidos para a série de 30 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 21. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises PLS, para os grupos de treinamento e de teste II, obtidos para a série de 30 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 22. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste III, obtidos para a série de 30 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 23. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste I, obtidos para a série de 23 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 24. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste II, obtidos para a série de 23 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 25. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste III, obtidos para a série de 23 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 26. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste I, obtidos para a série de 22 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 27. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste II, obtidos para a série de 22 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Figura 28. Gráficos dos valores de r^2 , Q^2_{cv} e s_{press} versus o número de parâmetros físico-químicos (descritores) obtidos através das análises *PLS*, para os grupos de treinamento e de teste III, obtidos para a série de 22 compostos utilizando-se, respectivamente: a) 1 variável latente, b) 2 variáveis latentes, c) 3 variáveis latentes e d) 4 variáveis latentes.

Analisando-se os gráficos apresentados nas figuras 14 a 28, pôde-se selecionar e determinar o número ideal de descritores DRAGON que descreve, significativamente a atividade de inibição da enzima RNR de células H.Ep.-2, para a série de compostos estudada nesta tese.

A tabela 5 apresenta o número de descritores selecionados para cada série de compostos, para os três grupos de treinamento e de teste, utilizando-se 1 a 4 variáveis latentes.

Tabela 5. Números de descritores DRAGON selecionados pelas análises dos parâmetros estatísticos nos modelos *PLS* apresentados nas figuras 14 a 28, respectivamente, para as séries A, B, C, D e E, que melhor descrevem a atividade inibitória da série de derivados da 2-piridino carboxaldeído tiossemicarbazonas frente a RNR de células H.Ep.-2.

Série	Número de variáveis latentes	Número de descritores selecionados		
		Grupo de Treinamento e de Teste		
		I	II	III
A		Figura 14	Figura 15	Figura 16
	1	5	5	5
	2	25	15	20
	3	40	20	20
	4	20	20	15
B		Figura 17	Figura 18	Figura 19
	1	10	25	25
	2	25	20	15
	3	25	20	10
	4	30	25	25
C		Figura 20	Figura 21	Figura 22
	1	5	5	15
	2	15	10	10
	3	20	40	5
	4	25	30	15
D		Figura 23	Figura 24	Figura 25
	1	5	5	5
	2	10	15	10
	3	15	15	20
	4	15	20	20
E		Figura 26	Figura 27	Figura 28
	1	5	10	5
	2	20	10	25
	3	25	20	20
	4	30	25	25

Para cada uma das séries, respectivamente com 40, 39, 30, 23 e 22 compostos, foram selecionados os melhores descritores obtidos através de análises *PLS* para os grupos de treinamento e de teste (grupos I, II e III). As tabelas de 6 a 10, apresentam os descritores selecionados pelos modelos *PLS* gerados, respectivamente, para as séries A, B, C, D e E.

Na tabela 6 estão apresentados os descritores selecionados através das análises *PLS* realizadas com 4, 4 e 3 variáveis latentes, respectivamente, para os grupos de treinamento IA, IIA e IIIA da série A (40 compostos).

Tabela 6. Descritores selecionados pelos modelos *PLS* obtidos para a série A (40 compostos).

Grupo IA	Grupo IIA	Grupo IIIA	Descritores Selecionados	Freq. Rel. %
MATS8m, GATS3e	Jhetv, IDDE	MAXDP, IC2	IC2	66
GATS4e, GATS4p	IC2, SEige	GATS1p, R4u	BEHv8	66
GATS5p, RDF050m	BEHv3, BEHv8	BEHv8, BELv6	MATS8m	66
RDF055m, RDF060m	BEHp3, MATS8m	RDF085v, H3e	GATS3e	66
RDF070m, RDF075v	GATS1v, GATS3e	GATS4v, HIC	GATS4e	100
RDF020p, RDF070p	GATS4e, RDF030m	GATS4e, R1u	GATS5p	66
RDF075p, Mor03m	RDF050m, RDF055m	GATS5p, R1e	RDF050m	66
Mor08m, Mor07p	RDF070m, RDF075p	HATSe	RDF055m	66
HATSp, R1u	Mor03m, Mor08m		RDF070m	66
nNR2, H-050	HATS0m, HATSp		RDF075p	66
			Mor03m	66
			Mor08m	66
			HATSp	66
			R1u	66

Na tabela 7 estão apresentados os descritores selecionados através das análises *PLS* realizadas com 4 variáveis latentes para os três grupos de treinamento IB, IIB e IIIB da série B (39 compostos).

Tabela 7. Descritores selecionados pelos modelos *PLS* obtidos para a série B (39 compostos).

Grupo IB	Grupo IIB	Grupo IIIB	Descritores Selecionados	Freq. Rel. %
X5sol, RDF105u	Jhetv, IDDE	BEHm7, BELm6	IDDE	66
IDDE, RDF050m	BEHv3, BEHp3	MATS8m, DISPm	BEHv8	66
VRD2, RDF055m	IC2, SEige	BEHv8, BEHp8	MATS8m	66
BEHv8, RDF060m	DISPv, RDF030m	GATS4v, GATS4e	GATS3e	66
BEHp6, RDF020v	MATS8m, GATS3e	RDF115u, HIC	GATS4e	100
GATS3e, RDF035v	GATS4e, GATS4p	RDF055m, H3u	GATS4p	66
GATS4e, RDF070v	RDF070m, RDF115m	RDF085v, Mor07m	GATS5p	66
GATS4p, RDF075v	RDF050m, RDF055m	GATS5p, R1u	RDF050m	66
GATS5p, RDF030e	Mor08m, Mor07v	RDF050e, Mor08m	RDF055m	100
MEcc, nNR2	RDF075p, Mor03m	RDF115e, Mor03p	RDF075p	66
RDF070p, HIC	HATSp, C-001	HATSe, R2m	Mor08m	100
RDF075p, H3u	H-049, H-052	R1e, H-046	Mor07v	66
Mor08m, HATSp	HATS0m	H-052	HIC	66
Mor07v, R1u			H3u	66
Mor07p, R4u			HATSp	66
			R1u	66
			H-052	66

Na tabela 8 estão apresentados os descritores selecionados através das análises *PLS* realizadas com 4 variáveis latentes para os três grupos de treinamento IC, IIC e IIIC da série C (30 compostos).

Tabela 8. Descritores selecionados pelos modelos *PLS* obtidos para a série C (30 compostos).

Grupo IC	Grupo IIC	Grupo IIIC	Descritores Selecionados	Freq. Rel. %
J, JhetZ	J, JhetZ	J	J	100
RDF065m, RDF075v	MAXDP, X4v	JhetZ	JhetZ	100
MAXDP, X3v	CIC2, BEHv6	MAXDP	MAXDP	100
RDF115v, RDF050e	BEHp6, GATS2v	DELS	CIC2	66
X5sol, PJI2	RDF050u, RDF125u	BEHp8	BEHv6	66
C-025, H-046	RDF020v, RDF050v	GATS3e	GATS2v	66
CIC2, BEHv6	RDF140u, RDF065m	GATS4e	GATS3e	66
DISPm, RDF125u	RDF065v, RDF085v	GATS5p	GATS4e	66
GATS2v, GATS4e	GATS4v, GATS7v	J3D	GATS5p	66
RDF065e, H1p	RDF065e, RDF125e	RDF020v	RDF125u	66
RDF125e, R4u	GATS3e, GATS5p	RDF075v	RDF065m	66
Mor07m, nCaR	Mor07m, Mor07v	RDF125v	RDF020v	66
H-052	RDF135e, RDF125p	RDF130e	RDF075v	66
	H-046, H-052	Mor07m	RDF065e	66
	C-024, C-025	C-025	RDF125e	66
			Mor07m	100
			C-025	66
			H-046	66
			H-052	66

Na tabela 9 estão apresentados os descritores selecionados através das análises *PLS* realizadas com 4 variáveis latentes para os três grupos de treinamento ID, IID e IIID da série D (23 compostos).

Tabela 9. Descritores selecionados pelos modelos *PLS* obtidos para a série D (23 compostos).

Grupo ID	Grupo IID	Grupo IIID	Descritores Selecionados	Freq. Rel. %
Jhetv, MAXDP	Jhetv, RDCHI	Jhetv, MAXDP	Jhetv	100
DELS, GATS1e	nRORPh, C-026	SEigv, SEige	MAXDP	66
SP02, SP04	HVcpx, SEige	X2sol, DECC	SEige	66
SP10, SP11	H2e, nROR	GATS5e, GATS7e	SP13	66
SP13, SP14	VEml, BEHp8	SP13, RDF055m	RDF065m	66
SP17, RDF050m	Mor07v, Mor07p	RDF065m, RDF115m	RDF115m	66
Mor06u, nNR2	MATS4m, GATS3e	RDF060v, RDF085v	RDF060v	66
O-060	RDF065v, RDF115v	RDF115v, Mor07v	RDF115v	66
	DP07, RDF065m	H2m, H2e	Mor07v	66
	RDF115m, RDF060v	nNR2, nRORPh	H2e	66
			nRORPh	66

Na tabela 10 estão apresentados os descritores selecionados através das análises *PLS* realizadas com 4 variáveis latentes para os três grupos de treinamento IE, IIE e IIIE da série E (23 compostos).

Tabela 10. Descritores selecionados pelos modelos *PLS* obtidos para a série E (22 compostos).

Grupo IE	Grupo IIE	Grupo IIIE	Descritores Selecionados	Freq. Rel. %
Ss, nDB	Jhetv, RDF065m	nAT, DP05	Jhetv	66
GATS1e, GATS4e	Jhete, RDF115m	nBT, W3D	CIC2	66
ZM2V, Jhetv	CIC2, RDF040v	ISIZ, H3D	SP13	66
piPC09, GGI2	Seige, RDF050v	TIC3, AGDD	RDF050v	66
MAXDP, DELS	BEHp8, RDF060v	TIC4, DDI	RDF065p	66
piPC06, piPC07	GGI1, RDF065v	TIC5, SEig	Mor07p	66
IC2, CIC2	MATS4m, RDF065e	ATS1e, RDF015u		
GATS7e, SP04	GATS3e, RDF135e	ATS2e, RDF020u		
SP13, RDF050m	DP07, RDF065p	ATS4e, RDF050v		
RDF085m, H0e	SP13, RDF115p	ATS5e, RDF015e		
RDF085v, Mor07p	nRORPh, C-025	ATS7e, RDF125e		
RDF085e, Mor06m	Mor07v, Mor07p	RDF065p, Mor01u		
nCOOR, C-002	C-026	Mor01e, Te		
RDF085p, H1p		Mor02e, HIC		
C-040, O-060		HTu, RTu		

A partir dos resultados apresentados nas tabelas de 6 a 11, selecionaram-se os descritores que possuem valores da frequência relativa maiores do que 66% para as análises de *QSAR* clássico.

3.5 Alinhamento dos compostos

Para as análises de *QSAR-3D* aplicando-se as metodologias *CoMFA* e *CoMSIA* as moléculas devem ser alinhadas previamente para que estas estejam todas em suas prováveis conformações bioativas. Esta etapa é crucial, pois somente o alinhamento correto das moléculas permitirá que se determine as características físico-químicas importantes para a atividade biológica. Nesta tese, as geometrias de mínima energia foram consideradas como sendo as conformações bioativas das moléculas. Como molde, para se efetuar os alinhamentos, foi utilizado a geometria do composto não substituído, composto 1, um dos compostos mais ativos desta série ($pIC_{50} = 6,55$) (tabela 2). O alinhamento das estruturas foi realizado por dois métodos distintos, utilizando-se o alinhamento rígido e um outro que considera a similaridade dos campos estéricos e eletrostáticos.

No primeiro método, apenas os átomos “pesados” foram selecionados para o procedimento de alinhamento. Escolheu-se um átomo da estrutura do composto molde e o correspondente átomo da estrutura a ser alinhada, sendo que este procedimento foi realizado sucessivamente para todos os átomos pesados.

No segundo método, o programa SEAL foi utilizado para alinhar as estruturas dos compostos através da sobreposição por similaridade dos campos estéricos e eletrostáticos. Através de rotações e translações nas moléculas, o programa efetua todos os possíveis alinhamentos, até os menos intuitivos e, pontua-se de acordo com os melhores resultados, isto é, quando se observa a máxima sobreposição dos campos. A figura a seguir apresenta os resultados dos alinhamentos obtidos através do método rígido e através do método SEAL.

Figura 29. Resultados dos alinhamentos dos 40 derivados da α -(N)-piridino carboxaldeído tiossemicarbazona realizados por dois procedimentos, utilizando o composto 1 como molde: (a) átomo-a-átomo e b) similaridade estérica e eletrostática utilizando-se o programa SEAL.

Como podem ser observados na figura 29, os dois alinhamentos apresentam resultados muito semelhantes. Uma das principais diferenças está nos átomos pesados presentes no anel piridínico e no fragmento da tiossemicarbazona que, através do alinhamento rígido, ficaram exatamente na mesma posição por terem sido utilizados como moldes na etapa de alinhamento. Pode-se observar que através do alinhamento SEAL os mesmos átomos estão menos sobrepostos. Além disto, os grupos substituintes não estão perfeitamente alinhados. Estes desvios podem ser explicados ao se considerar a flexibilidade dos grupos substituintes e ao se considerar os diferentes tipos de átomos presentes nestes grupos, *i.e.*, oxigênio, cloro, bromo, carbono, etc...

3.6 Estudos de *QSAR* clássico (abordagem extratermodinâmica)

As análises de *QSAR* foram aplicadas aos descritores que melhor se correlacionaram com a atividade biológica e que foram pré-selecionados pelos modelos *PLS* utilizando o procedimento de *full cross-validation*. As equações de regressão linear múltipla foram geradas pelo programa BILIN (Kubinyi 1994b).

Os descritores selecionados para cada série de compostos (séries A até E) estão apresentados, respectivamente, nas tabelas 5 a 9. Estes descritores foram utilizados como

sendo a variável independente e, os correspondentes valores da atividade inibitória frente à ribonucleotídeo redutase (RNR) de células H.Ep.-2, como sendo a variável dependente nas análises de *QSAR*.

A seguir, estão apresentadas as 205 equações mais significativas obtidas para as séries: A (n=40), B (n=39), C (n=30), D (n=23) e E (n=22), considerando a contribuição de um a cinco descritores, respectivamente. Nas análises onde várias equações foram obtidas, apenas as dez mais significativas foram selecionadas, considerando-se os valores de Q^2_{cv} , observados para cada modelo.

Equações de regressão significativas geradas pelas análises de *QSAR* aplicadas para a série de 40 compostos (série A), utilizando-se um a 5 descritores, respectivamente:

Série A (n=40), 1 descritor

$$pIC_{50} = -2,606 (\pm 1,180) R1u + 8,211 (\pm 1,270) \\ (n = 40; r = 0,586; s = 0,805; F = 19,896; Q^2_{cv} = 0,294; s_{press} = 0,835) \quad \text{Equação 39}$$

$$pIC_{50} = -0,209 (\pm 0,130) RDF050m + 6,133 (\pm 0,490) \\ (n = 40; r = 0,474; s = 0,875; F = 10,996; Q^2_{cv} = 0,140; s_{press} = 0,922) \quad \text{Equação 40}$$

$$pIC_{50} = +0,475 (\pm 0,330) HATSp + 3,813 (\pm 1,200) \\ (n = 40; r = 0,425; s = 0,900; F = 8,353; Q^2_{cv} = 0,062; s_{press} = 0,963) \quad \text{Equação 41}$$

$$pIC_{50} = -0,819 (\pm 0,590) MATS8m + 5,803 (\pm 0,370) \\ (n = 40; r = 0,414; s = 0,905; F = 7,873; Q^2_{cv} = 0,065; s_{press} = 0,961) \quad \text{Equação 42}$$

$$pIC_{50} = +0,856 (\pm 0,620) IC2 + 2,331 (\pm 2,290) \\ (n = 40; r = 0,414; s = 0,905; F = 7,847; Q^2_{cv} = 0,047; s_{press} = 0,970) \quad \text{Equação 43}$$

$$pIC_{50} = -0,791 (\pm 0,640) GATS4e + 5,712 (\pm 0,350) \\ (n = 40; r = 0,374; s = 0,922; F = 6,184; Q^2_{cv} = 0,029; s_{press} = 0,980) \quad \text{Equação 44}$$

$$pIC_{50} = -0,141 (\pm 0,120) RDF075p + 5,691 (\pm 0,340) \\ (n = 40; r = 0,372; s = 0,923; F = 6,097; Q^2_{cv} = -0,165; s_{press} = 1,073) \quad \text{Equação 45}$$

Série A (n=40), 2 descritores

$$pIC_{50} = -0,200 (\pm 0,099) RDF050m - 2,532 (\pm 1,000) R1u + 8,763 (\pm 1,100) \\ (n = 40; r = 0,741; s = 0,677; F = 22,475; Q^2_{cv} = 0,492; s_{press} = 0,718) \quad \text{Equação 46}$$

$$pIC_{50} = -0,298 (\pm 0,110) RDF050m - 1,357 (\pm 0,620) Mor08m + 5,228 (\pm 0,570) \\ (n = 40; r = 0,704; s = 0,715; F = 18,202; Q^2_{cv} = 0,346; s_{press} = 0,815) \quad \text{Equação 47}$$

$$pIC_{50} = -0,142 (\pm 0,091) RDF075p - 2,609 (\pm 1,060) R1u + 8,431 (\pm 1,150) \\ (n = 40; r = 0,695; s = 0,724; F = 17,275; Q^2_{cv} = 0,278; s_{press} = 0,856) \quad \text{Equação 48}$$

$$pIC_{50} = +0,395 (\pm 0,270) HATSp - 2,398 (\pm 1,090) R1u + 6,611 (\pm 1,610) \\ (n = 40; r = 0,683; s = 0,736; F = 16,151; Q^2_{cv} = 0,357; s_{press} = 0,808) \quad \text{Equação 49}$$

$$pIC_{50} = -0,674 (\pm 0,530) \text{ GATS4e} - 2,464 (\pm 1,110) \text{ R1u} + 8,264 (\pm 1,180) \\ (n = 40; r = 0,666; s = 0,751; F = 14,782; Q^2_{cv} = 0,369; s_{press} = 0,800) \quad \text{Equação 50}$$

$$pIC_{50} = -0,577 (\pm 0,480) \text{ BEHv8} - 2,681 (\pm 1,110) \text{ R1u} + 8,940 (\pm 1,340) \\ (n = 40; r = 0,660; s = 0,757; F = 14,266; Q^2_{cv} = 0,371; s_{press} = 0,799) \quad \text{Equação 51}$$

$$pIC_{50} = -0,248 (\pm 0,110) \text{ RDF075p} - 1,483 (\pm 0,680) \text{ Mor08m} + 4,555 (\pm 0,600) \\ (n = 40; r = 0,659; s = 0,758; F = 14,173; Q^2_{cv} = 0,115; s_{press} = 0,948) \quad \text{Equação 52}$$

$$pIC_{50} = +0,413 (\pm 0,380) \text{ Mor03m} - 2,921 (\pm 1,160) \text{ R1u} + 9,255 (\pm 1,540) \\ (n = 40; r = 0,648; s = 0,767; F = 13,417; Q^2_{cv} = 0,343; s_{press} = 0,816) \quad \text{Equação 53}$$

$$pIC_{50} = -0,536 (\pm 0,530) \text{ MATS8m} - 2,267 (\pm 1,180) \text{ R1u} + 8,070 (\pm 1,230) \\ (n = 40; r = 0,641; s = 0,773; F = 12,923; Q^2_{cv} = 0,314; s_{press} = 0,834) \quad \text{Equação 54}$$

$$pIC_{50} = -1,008 (\pm 0,530) \text{ MATS8m} - 1,007 (\pm 0,560) \text{ GATS4e} + 6,180 (\pm 0,390) \\ (n = 40; r = 0,624; s = 0,787; F = 11,796; Q^2_{cv} = 0,259; s_{press} = 0,867) \quad \text{Equação 55}$$

Série A (n=40), 3 descritores

$$pIC_{50} = -0,253 (\pm 0,100) \text{ RDF050m} - 0,773 (\pm 0,660) \text{ Mor08m} - 1,801 (\pm 1,130) \text{ R1u} + 7,488 (\pm 1,510) \\ (n = 40; r = 0,781; s = 0,638; F = 18,727; Q^2_{cv} = 0,395; s_{press} = 0,794) \quad \text{Equação 56}$$

$$pIC_{50} = -0,520 (\pm 0,460) \text{ GATS4e} - 0,181 (\pm 0,095) \text{ RDF050m} - 2,429 (\pm 0,950) \text{ R1u} + 8,751 (\pm 1,040) \\ (n = 40; r = 0,779; s = 0,640; F = 18,505; Q^2_{cv} = 0,536; s_{press} = 0,695) \quad \text{Equação 57}$$

$$pIC_{50} = -0,733 (\pm 0,450) \text{ GATS4e} - 0,151 (\pm 0,081) \text{ RDF075p} - 2,455 (\pm 0,950) \text{ R1u} + 8,503 (\pm 1,030) \\ (n = 40; r = 0,776; s = 0,645; F = 18,106; Q^2_{cv} = 0,438; s_{press} = 0,766) \quad \text{Equação 58}$$

$$pIC_{50} = -0,220 (\pm 0,093) \text{ RDF075p} - 1,585 (\pm 0,590) \text{ Mor08m} + 0,464 (\pm 0,250) \text{ HATSp} + 2,800 (\pm 1,070) \\ (n = 40; r = 0,772; s = 0,649; F = 17,704; Q^2_{cv} = 0,452; s_{press} = 0,756) \quad \text{Equação 59}$$

$$pIC_{50} = -0,305 (\pm 0,100) \text{ RDF050m} + 0,117 (\pm 0,087) \text{ RDF070m} - 1,394 (\pm 0,570) \text{ Mor08m} + 4,873 (\pm 0,590) \\ (n = 40; r = 0,763; s = 0,660; F = 16,766; Q^2_{cv} = 0,457; s_{press} = 0,752) \quad \text{Equação 60}$$

$$pIC_{50} = -0,246 (\pm 0,110) \text{ RDF050m} - 1,380 (\pm 0,580) \text{ Mor08m} + 0,328 (\pm 0,270) \text{ HATSp} + 3,894 (\pm 1,240) \\ (n = 40; r = 0,753; s = 0,672; F = 15,685; Q^2_{cv} = 0,381; s_{press} = 0,804) \quad \text{Equação 61}$$

$$pIC_{50} = -0,734 (\pm 0,470) \text{ MATS8m} - 0,855 (\pm 0,490) \text{ GATS4e} - 1,962 (\pm 1,050) \text{ R1u} + 8,085 (\pm 1,070) \\ (n = 40; r = 0,751; s = 0,674; F = 15,514; Q^2_{cv} = 0,450; s_{press} = 0,757) \quad \text{Equação 62}$$

$$pIC_{50} = +0,144 (\pm 0,090) \text{ RDF070m} - 0,272 (\pm 0,097) \text{ RDF075p} - 1,582 (\pm 0,610) \text{ Mor08m} + 4,081 (\pm 0,610) \\ (n = 40; r = 0,750; s = 0,675; F = 15,421; Q^2_{cv} = 0,203; s_{press} = 0,912) \quad \text{Equação 63}$$

$$pIC_{50} = +0,540 (\pm 0,480) \text{ IC2} - 0,277 (\pm 0,110) \text{ RDF050m} - 1,197 (\pm 0,600) \text{ Mor08m} + 3,313 (\pm 1,780) \\ (n = 40; r = 0,748; s = 0,677; F = 15,259; Q^2_{cv} = 0,402; s_{press} = 0,789) \quad \text{Equação 64}$$

$$pIC_{50} = -0,118 (\pm 0,0880) \text{ RDF075p} + 0,317 (\pm 0,260) \text{ HATSp} - 2,442 (\pm 1,010) \text{ R1u} + 7,111 (\pm 1,530) \\ (n = 40; r = 0,747; s = 0,679; F = 15,133; Q^2_{cv} = 0,395; s_{press} = 0,794) \quad \text{Equação 65}$$

Série A (n=40), 4 descritores

$$pIC_{50} = -0,788 (\pm 0,470) \text{ GATS3e} - 0,237 (\pm 0,0820) \text{ RDF075p} - 1,536 (\pm 0,520) \text{ Mor08m} \\ + 0,567 (\pm 0,230) \text{ HATSp} + 3,119 (\pm 0,960) \\ (n = 40; r = 0,835; s = 0,570; F = 20,139; Q^2_{cv} = 0,566; s_{press} = 0,682) \quad \text{Equação 66}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,543 (\pm 0,440) \text{ MATS8m} - 0,857 (\pm 0,440) \text{ GATS4e} - 0,124 (\pm 0,079) \text{ RDF075p} \\ & - 2,085 (\pm 0,940) \text{ R1u} + 8,328 (\pm 0,970) \\ (n = 40; r = 0,814; s = 0,602; F = 17,134; Q^2_{cv} = 0,503; S_{press} = 0,730) \end{aligned} \quad \text{Equação 67}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,594 (\pm 0,490) \text{ GATS3e} - 0,842 (\pm 0,440) \text{ GATS4e} - 0,174 (\pm 0,0780) \text{ RDF075p} \\ & - 2,292 (\pm 0,900) \text{ R1u} + 8,860 (\pm 1,010) \\ (n = 40; r = 0,813; s = 0,604; F = 17,013; Q^2_{cv} = 0,551; S_{press} = 0,694) \end{aligned} \quad \text{Equação 68}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,190 (\pm 0,090) \text{ RDF075p} - 1,083 (\pm 0,680) \text{ Mor08m} + 0,395 (\pm 0,240) \text{ HATSp} \\ & - 1,405 (\pm 1,110) \text{ R1u} + 4,909 (\pm 1,930) \\ (n = 40; r = 0,813; s = 0,603; F = 17,026; Q^2_{cv} = 0,396; S_{press} = 0,805) \end{aligned} \quad \text{Equação 69}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,494 (\pm 0,460) \text{ MATS8m} - 0,673 (\pm 0,460) \text{ GATS4e} - 0,144 (\pm 0,097) \text{ RDF050m} \\ & - 2,099 (\pm 0,950) \text{ R1u} + 8,532 (\pm 1,010) \\ (n = 40; r = 0,809; s = 0,609; F = 16,550; Q^2_{cv} = 0,544; S_{press} = 0,699) \end{aligned} \quad \text{Equação 70}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,216 (\pm 0,110) \text{ RDF050m} - 0,856 (\pm 0,640) \text{ Mor08m} + 0,262 (\pm 0,250) \text{ HATSp} \\ & - 1,599 (\pm 1,100) \text{ R1u} + 6,167 (\pm 1,920) \\ (n = 40; r = 0,808; s = 0,610; F = 16,486; Q^2_{cv} = 0,407; S_{press} = 0,798) \end{aligned} \quad \text{Equação 71}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,504 (\pm 0,490) \text{ GATS3e} - 0,592 (\pm 0,440) \text{ GATS4e} - 0,196 (\pm 0,092) \text{ RDF050m} \\ & - 2,289 (\pm 0,920) \text{ R1u} + 9,064 (\pm 1,040) \\ (n = 40; r = 0,806; s = 0,612; F = 16,275; Q^2_{cv} = 0,553; S_{press} = 0,693) \end{aligned} \quad \text{Equação 72}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,649 (\pm 0,510) \text{ GATS3e} - 0,249 (\pm 0,110) \text{ RDF050m} - 1,306 (\pm 0,540) \text{ Mor08m} \\ & + 0,416 (\pm 0,260) \text{ HATSp} + 4,163 (\pm 1,170) \\ (n = 40; r = 0,798; s = 0,625; F = 15,305; Q^2_{cv} = 0,367; S_{press} = 0,824) \end{aligned} \quad \text{Equação 73}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,571 (\pm 0,450) \text{ GATS4e} + 0,132 (\pm 0,084) \text{ RDF070m} - 0,267 (\pm 0,090) \text{ RDF075p} \\ & - 1,431 (\pm 0,580) \text{ Mor08m} + 4,412 (\pm 0,620) \\ (n = 40; r = 0,795; s = 0,629; F = 14,986; Q^2_{cv} = 0,326; S_{press} = 0,850) \end{aligned} \quad \text{Equação 74}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,567 (\pm 0,520) \text{ GATS3e} - 0,170 (\pm 0,100) \text{ RDF050m} + 0,304 (\pm 0,270) \text{ HATSp} \\ & - 2,249 (\pm 0,950) \text{ R1u} + 7,748 (\pm 1,550) \\ (n = 40; r = 0,795; s = 0,628; F = 15,015; Q^2_{cv} = 0,504; S_{press} = 0,729) \end{aligned} \quad \text{Equação 75}$$

Série A (n=40), 5 descritores

$$\begin{aligned} \text{pIC}_{50} = & -0,852 (\pm 0,440) \text{ GATS3e} - 0,483 (\pm 0,420) \text{ GATS4e} - 0,241 (\pm 0,078) \text{ RDF075p} \\ & - 1,390 (\pm 0,500) \text{ Mor08m} + 0,481 (\pm 0,230) \text{ HATSp} + 3,748 (\pm 1,050) \\ (n = 40; r = 0,860; s = 0,536; F = 19,324; Q^2_{cv} = 0,596; S_{press} = 0,668) \end{aligned} \quad \text{Equação 76}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,700 (\pm 0,450) \text{ GATS3e} - 0,210 (\pm 0,081) \text{ RDF075p} - 1,135 (\pm 0,600) \text{ Mor08m} \\ & + 0,500 (\pm 0,220) \text{ HATSp} - 1,137 (\pm 1,000) \text{ R1u} + 4,789 (\pm 1,730) \\ (n = 40; r = 0,859; s = 0,538; F = 19,165; Q^2_{cv} = 0,518; S_{press} = 0,730) \end{aligned} \quad \text{Equação 77}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,801 (\pm 0,440) \text{ GATS3e} + 0,126 (\pm 0,120) \text{ RDF055m} - 0,250 (\pm 0,079) \text{ RDF075p} \\ & - 1,325 (\pm 0,530) \text{ Mor08m} + 0,562 (\pm 0,210) \text{ HATSp} + 3,015 (\pm 0,910) \\ (n = 40; r = 0,857; s = 0,541; F = 18,883; Q^2_{cv} = 0,616; S_{press} = 0,651) \end{aligned} \quad \text{Equação 78}$$

$$\begin{aligned} \text{pIC}_{50} = & +0,786 (\pm 0,420) \text{ IC2} - 0,806 (\pm 0,470) \text{ GATS3e} + 0,602 (\pm 0,280) \text{ Mor03m} \\ & + 0,518 (\pm 0,220) \text{ HATSp} - 2,042 (\pm 0,890) \text{ R1u} + 4,578 (\pm 2,150) \\ (n = 40; r = 0,853; s = 0,548; F = 18,163; Q^2_{cv} = 0,586; S_{press} = 0,676) \end{aligned} \quad \text{Equação 79}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,744 (\pm 0,470) \text{ GATS3e} - 0,686 (\pm 0,420) \text{ GATS4e} - 0,154 (\pm 0,074) \text{ RDF075p} \\ & + 0,301 (\pm 0,240) \text{ HATSp} - 2,131 (\pm 0,850) \text{ R1u} + 7,678 (\pm 1,310) \\ (n = 40; r = 0,847; s = 0,559; F = 17,203; Q^2_{cv} = 0,603; S_{press} = 0,662) \end{aligned} \quad \text{Equação 80}$$

$$\begin{aligned} \text{pIC}_{50} = & +0,597 (\pm 0,400) \text{ IC2} - 0,747 (\pm 0,460) \text{ GATS3e} - 0,228 (\pm 0,097) \text{ RDF050m} \\ & - 1,116 (\pm 0,510) \text{ Mor08m} + 0,406 (\pm 0,240) \text{ HATSp} + 2,179 (\pm 1,710) \\ (n = 40; r = 0,844; s = 0,564; F = 16,838; Q^2_{cv} = 0,536; S_{press} = 0,716) \end{aligned} \quad \text{Equação 81}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,600 (\pm 0,410) \text{ MATS8m} - 0,721 (\pm 0,420) \text{ GATS4e} + 0,115 (\pm 0,077) \text{ RDF070m} \\ & - 0,223 (\pm 0,087) \text{ RDF075p} - 1,266 (\pm 0,540) \text{ Mor08m} + 4,827 (\pm 0,630) \\ (n = 40; r = 0,842; s = 0,567; F = 16,545; Q^2_{cv} = 0,495; S_{press} = 0,747) \end{aligned} \quad \text{Equação 82}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,560 (\pm 0,470) \text{ GATS3e} - 0,222 (\pm 0,099) \text{ RDF050m} - 0,848 (\pm 0,600) \text{ Mor08m} \\ & + 0,345 (\pm 0,250) \text{ HATSp} - 1,428 (\pm 1,040) \text{ R1u} + 6,156 (\pm 1,810) \\ (n = 40; r = 0,839; s = 0,572; F = 16,181; Q^2_{cv} = 0,403; S_{press} = 0,812) \end{aligned} \quad \text{Equação 83}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,537 (\pm 0,420) \text{ MATS8m} - 0,783 (\pm 0,420) \text{ GATS4e} - 0,171 (\pm 0,088) \text{ RDF075p} \\ & - 0,666 (\pm 0,640) \text{ Mor08m} - 1,492 (\pm 1,060) \text{ R1u} + 7,170 (\pm 1,450) \\ (n = 40; r = 0,838; s = 0,574; F = 15,976; Q^2_{cv} = 0,449; S_{press} = 0,780) \end{aligned} \quad \text{Equação 84}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,451 (\pm 0,430) \text{ MATS8m} - 0,490 (\pm 0,480) \text{ GATS3e} - 0,925 (\pm 0,420) \text{ GATS4e} \\ & - 0,147 (\pm 0,079) \text{ RDF075p} - 2,014 (\pm 0,900) \text{ R1u} + 8,652 (\pm 0,980) \\ (n = 40; r = 0,837; s = 0,575; F = 15,880; Q^2_{cv} = 0,570; S_{press} = 0,689) \end{aligned} \quad \text{Equação 85}$$

Equações de regressão significativas geradas pelas análises de *QSAR* aplicadas para a série de 39 compostos (série B), utilizando-se um a 5 descritores, respectivamente:

Série B (n=39), 1 descritor

$$\begin{aligned} \text{pIC}_{50} = & -1,299 (\pm 0,760) \text{ Mor08m} + 4,415 (\pm 0,700) \\ (n = 39; r = 0,493; s = 0,848; F = 11,909; Q^2_{cv} = 0,009; S_{press} = 0,971) \end{aligned} \quad \text{Equação 86}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,267 (\pm 0,180) \text{ RDF050m} + 6,289 (\pm 0,600) \\ (n = 39; r = 0,436; s = 0,877; F = 8,689; Q^2_{cv} = 0,100; S_{press} = 0,925) \end{aligned} \quad \text{Equação 87}$$

$$\begin{aligned} \text{pIC}_{50} = & +0,453 (\pm 0,330) \text{ HATSp} + 3,923 (\pm 1,190) \\ (n = 39; r = 0,416; s = 0,887; F = 7,759; Q^2_{cv} = 0,050; S_{press} = 0,950) \end{aligned} \quad \text{Equação 88}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,856 (\pm 0,620) \text{ GATS4e} + 5,777 (\pm 0,350) \\ (n = 39; r = 0,416; s = 0,887; F = 7,736; Q^2_{cv} = 0,060; S_{press} = 0,946) \end{aligned} \quad \text{Equação 89}$$

$$\begin{aligned} \text{pIC}_{50} = & +0,771 (\pm 0,590) \text{ IDDE} + 3,400 (\pm 1,640) \\ (n = 39; r = 0,399; s = 0,894; F = 7,009; Q^2_{cv} = -0,0026; S_{press} = 0,976) \end{aligned} \quad \text{Equação 90}$$

$$\begin{aligned} \text{pIC}_{50} = & +1,145 (\pm 0,910) \text{ H3u} + 4,369 (\pm 0,960) \\ (n = 39; r = 0,386; s = 0,899; F = 6,482; Q^2_{cv} = -0,0901; S_{press} = 1,018) \end{aligned} \quad \text{Equação 91}$$

Série B (n=39), 2 descritores

$$\begin{aligned} \text{pIC}_{50} = & -1,507 (\pm 0,640) \text{ Mor08m} + 0,550 (\pm 0,260) \text{ HATSp} + 2,306 (\pm 1,170) \\ (n = 39; r = 0,702; s = 0,704; F = 17,518; Q^2_{cv} = 0,333; S_{press} = 0,808) \end{aligned} \quad \text{Equação 92}$$

$$pIC_{50} = -0,288 (\pm 0,150) \text{ RDF050m} - 1,377 (\pm 0,650) \text{ Mor08m} + 5,182 (\pm 0,720)$$

$$(n = 39; r = 0,680; s = 0,724; F = 15,512; Q^2_{cv} = 0,274; s_{press} = 0,843) \quad \text{Equação 93}$$

$$pIC_{50} = -0,377 (\pm 0,230) \text{ RDF075p} - 1,442 (\pm 0,680) \text{ Mor08m} + 4,729 (\pm 0,650)$$

$$(n = 39; r = 0,650; s = 0,751; F = 13,193; Q^2_{cv} = 0,215; s_{press} = 0,876) \quad \text{Equação 94}$$

$$pIC_{50} = -0,932 (\pm 0,520) \text{ MATS8m} - 1,037 (\pm 0,550) \text{ GATS4e} + 6,192 (\pm 0,380)$$

$$(n = 39; r = 0,627; s = 0,770; F = 11,668; Q^2_{cv} = 0,254; s_{press} = 0,854) \quad \text{Equação 95}$$

$$pIC_{50} = -0,295 (\pm 0,160) \text{ RDF050m} + 1,292 (\pm 0,790) \text{ H3u} + 5,076 (\pm 0,920)$$

$$(n = 39; r = 0,615; s = 0,780; F = 10,944; Q^2_{cv} = 0,177; s_{press} = 0,897) \quad \text{Equação 96}$$

$$pIC_{50} = -1,506 (\pm 0,720) \text{ Mor08m} - 0,853 (\pm 0,650) \text{ HIC} + 7,542 (\pm 2,460)$$

$$(n = 39; r = 0,607; s = 0,785; F = 10,516; Q^2_{cv} = 0,125; s_{press} = 0,925) \quad \text{Equação 97}$$

$$pIC_{50} = -0,660 (\pm 0,530) \text{ MATS8m} - 1,209 (\pm 0,720) \text{ Mor08m} + 4,745 (\pm 0,710)$$

$$(n = 39; r = 0,597; s = 0,793; F = 9,981; Q^2_{cv} = 0,138; s_{press} = 0,918) \quad \text{Equação 98}$$

$$pIC_{50} = -0,697 (\pm 0,570) \text{ GATS4e} - 1,140 (\pm 0,730) \text{ Mor08m} + 4,764 (\pm 0,720)$$

$$(n = 39; r = 0,596; s = 0,794; F = 9,894; Q^2_{cv} = 0,104; s_{press} = 0,936) \quad \text{Equação 99}$$

$$pIC_{50} = -1,504 (\pm 0,740) \text{ Mor08m} - 0,616 (\pm 0,530) \text{ Mor07v} + 5,031 (\pm 0,850)$$

$$(n = 39; r = 0,587; s = 0,800; F = 9,452; Q^2_{cv} = 0,098; s_{press} = 0,939) \quad \text{Equação 100}$$

$$pIC_{50} = -0,287 (\pm 0,170) \text{ RDF050m} + 0,225 (\pm 0,160) \text{ RDF055m} + 5,765 (\pm 0,670)$$

$$(n = 39; r = 0,584; s = 0,803; F = 9,300; Q^2_{cv} = 0,051; s_{press} = 0,963) \quad \text{Equação 101}$$

Série B (n=39), 3 descritores

$$pIC_{50} = -0,964 (\pm 0,500) \text{ GATS3e} - 1,514 (\pm 0,540) \text{ Mor08m} + 0,684 (\pm 0,240) \text{ HATSp} + 2,598 (\pm 1,000)$$

$$(n = 39; r = 0,804; s = 0,596; F = 21,325; Q^2_{cv} = 0,532; s_{press} = 0,686) \quad \text{Equação 102}$$

$$pIC_{50} = -0,263 (\pm 0,210) \text{ RDF075p} - 1,568 (\pm 0,600) \text{ Mor08m} + 0,447 (\pm 0,260) \text{ HATSp} + 2,919 (\pm 1,200)$$

$$(n = 39; r = 0,756; s = 0,656; F = 15,564; Q^2_{cv} = 0,404; s_{press} = 0,774) \quad \text{Equação 103}$$

$$pIC_{50} = -1,663 (\pm 0,620) \text{ Mor08m} - 0,506 (\pm 0,450) \text{ Mor07v} + 0,515 (\pm 0,250) \text{ HATSp} + 2,945 (\pm 1,240)$$

$$(n = 39; r = 0,748; s = 0,665; F = 14,857; Q^2_{cv} = 0,378; s_{press} = 0,791) \quad \text{Equação 104}$$

$$pIC_{50} = -0,178 (\pm 0,160) \text{ RDF050m} - 1,495 (\pm 0,610) \text{ Mor08m} + 0,389 (\pm 0,290) \text{ HATSp} + 3,399 (\pm 1,500)$$

$$(n = 39; r = 0,745; s = 0,669; F = 14,564; Q^2_{cv} = 0,367; s_{press} = 0,798) \quad \text{Equação 105}$$

$$pIC_{50} = -0,538 (\pm 0,500) \text{ GATS5p} - 1,565 (\pm 0,610) \text{ Mor08m} + 0,552 (\pm 0,250) \text{ HATSp} + 2,390 (\pm 1,120)$$

$$(n = 39; r = 0,744; s = 0,670; F = 14,492; Q^2_{cv} = 0,386; s_{press} = 0,785) \quad \text{Equação 106}$$

$$pIC_{50} = -0,225 (\pm 0,150) \text{ RDF050m} - 0,265 (\pm 0,220) \text{ RDF075p} - 1,461 (\pm 0,620) \text{ Mor08m} + 5,236 (\pm 0,680)$$

$$(n = 39; r = 0,736; s = 0,679; F = 13,758; Q^2_{cv} = 0,356; s_{press} = 0,805) \quad \text{Equação 107}$$

$$pIC_{50} = -1,290 (\pm 0,760) \text{ GATS4p} - 0,310 (\pm 0,140) \text{ RDF050m} + 1,623 (\pm 0,720) \text{ H3u} + 4,923 (\pm 0,810)$$

$$(n = 39; r = 0,732; s = 0,683; F = 13,453; Q^2_{cv} = 0,252; s_{press} = 0,867) \quad \text{Equação 108}$$

$$pIC_{50} = -0,831 (\pm 0,470) \text{ MATS8m} - 0,880 (\pm 0,500) \text{ GATS4e} - 0,986 (\pm 0,640) \text{ Mor08m} + 5,271 (\pm 0,690)$$

$$(n = 39; r = 0,726; s = 0,690; F = 12,968; Q^2_{cv} = 0,336; s_{press} = 0,817) \quad \text{Equação 109}$$

$$pIC_{50} = -0,627 (\pm 0,500) \text{ GATS4e} - 0,355 (\pm 0,210) \text{ RDF075p} - 1,291 (\pm 0,650) \text{ Mor08m} + 5,024 (\pm 0,650)$$

$$(n = 39; r = 0,716; s = 0,700; F = 12,253; Q^2_{cv} = 0,302; s_{press} = 0,837) \quad \text{Equação 110}$$

$$pIC_{50} = -0,357 (\pm 0,220) \text{ RDF075p} - 1,618 (\pm 0,660) \text{ Mor08m} - 0,553 (\pm 0,470) \text{ Mor07v} + 5,265 (\pm 0,760) \\ (n = 39; r = 0,710; s = 0,706; F = 11,829; Q^2_{cv} = 0,301; s_{press} = 0,838) \quad \text{Equação 111}$$

Série B (n=39), 4 descritores

$$pIC_{50} = -1,033 (\pm 0,480) \text{ GATS3e} - 0,485 (\pm 0,440) \text{ GATS4e} - 1,369 (\pm 0,530) \text{ Mor08m} \\ + 0,600 (\pm 0,240) \text{ HATSp} + 3,219 (\pm 1,110) \\ (n = 39; r = 0,832; s = 0,564; F = 19,111; Q^2_{cv} = 0,562; s_{press} = 0,674) \quad \text{Equação 112}$$

$$pIC_{50} = -1,091 (\pm 0,500) \text{ GATS3e} - 1,420 (\pm 0,530) \text{ Mor08m} + 0,756 (\pm 0,240) \text{ HATSp} \\ + 0,147 (\pm 0,140) \text{ H-052} + 2,426 (\pm 0,980) \\ (n = 39; r = 0,828; s = 0,570; F = 18,595; Q^2_{cv} = 0,565; s_{press} = 0,671) \quad \text{Equação 113}$$

$$pIC_{50} = -0,814 (\pm 0,500) \text{ GATS3e} - 1,576 (\pm 0,700) \text{ GATS4p} - 0,303 (\pm 0,130) \text{ RDF050m} \\ + 1,690 (\pm 0,640) \text{ H3u} + 5,511 (\pm 0,800) \\ (n = 39; r = 0,805; s = 0,604; F = 15,609; Q^2_{cv} = 0,423; s_{press} = 0,772) \quad \text{Equação 114}$$

$$pIC_{50} = +0,196 (\pm 0,140) \text{ RDF055m} - 1,403 (\pm 0,600) \text{ Mor08m} - 0,734 (\pm 0,440) \text{ Mor07v} \\ + 0,497 (\pm 0,230) \text{ HATSp} + 3,016 (\pm 1,130) \\ (n = 39; r = 0,804; s = 0,605; F = 15,492; Q^2_{cv} = 0,412; s_{press} = 0,780) \quad \text{Equação 115}$$

$$pIC_{50} = -0,252 (\pm 0,200) \text{ RDF075p} - 1,714 (\pm 0,580) \text{ Mor08m} - 0,482 (\pm 0,410) \text{ Mor07v} \\ + 0,418 (\pm 0,250) \text{ HATSp} + 3,503 (\pm 1,230) \\ (n = 39; r = 0,795; s = 0,617; F = 14,604; Q^2_{cv} = 0,458; s_{press} = 0,749) \quad \text{Equação 116}$$

$$pIC_{50} = -0,538 (\pm 0,470) \text{ GATS5p} - 1,720 (\pm 0,590) \text{ Mor08m} - 0,506 (\pm 0,420) \text{ Mor07v} \\ + 0,517 (\pm 0,240) \text{ HATSp} + 3,029 (\pm 1,170) \\ (n = 39; r = 0,788; s = 0,626; F = 13,918; Q^2_{cv} = 0,439; s_{press} = 0,762) \quad \text{Equação 117}$$

$$pIC_{50} = -0,664 (\pm 0,450) \text{ MATS8m} - 0,789 (\pm 0,470) \text{ GATS4e} - 0,275 (\pm 0,200) \text{ RDF075p} \\ - 1,133 (\pm 0,590) \text{ Mor08m} + 5,371 (\pm 0,630) \\ (n = 39; r = 0,784; s = 0,632; F = 13,520; Q^2_{cv} = 0,432; s_{press} = 0,767) \quad \text{Equação 118}$$

$$pIC_{50} = -0,635 (\pm 0,460) \text{ BEHv8} + 0,191 (\pm 0,150) \text{ RDF055m} - 1,248 (\pm 0,620) \text{ Mor08m} \\ + 0,508 (\pm 0,240) \text{ HATSp} + 2,882 (\pm 1,170) \\ (n = 39; r = 0,784; s = 0,631; F = 13,584; Q^2_{cv} = 0,404; s_{press} = 0,785) \quad \text{Equação 119}$$

$$pIC_{50} = +0,219 (\pm 0,150) \text{ RDF055m} - 0,363 (\pm 0,190) \text{ RDF075p} - 1,336 (\pm 0,620) \text{ Mor08m} \\ - 0,802 (\pm 0,450) \text{ Mor07v} + 5,264 (\pm 0,690) \\ (n = 39; r = 0,782; s = 0,635; F = 13,344; Q^2_{cv} = 0,417; s_{press} = 0,777) \quad \text{Equação 120}$$

$$pIC_{50} = -0,495 (\pm 0,480) \text{ GATS5p} - 0,164 (\pm 0,160) \text{ RDF050m} - 1,549 (\pm 0,590) \text{ Mor08m} \\ + 0,402 (\pm 0,280) \text{ HATSp} + 3,391 (\pm 1,430) \\ (n = 39; r = 0,779; s = 0,638; F = 13,088; Q^2_{cv} = 0,419; s_{press} = 0,776) \quad \text{Equação 121}$$

Série B (n=39), 5 descritores

$$pIC_{50} = -0,838 (\pm 0,440) \text{ GATS3e} + 0,187 (\pm 0,120) \text{ RDF055m} - 1,375 (\pm 0,500) \text{ Mor08m} \\ - 0,577 (\pm 0,380) \text{ Mor07v} + 0,624 (\pm 0,210) \text{ HATSp} + 3,081 (\pm 0,96) \\ (n = 39; r = 0,869; s = 0,511; F = 20,396; Q^2_{cv} = 0,611; s_{press} = 0,644) \quad \text{Equação 122}$$

$$pIC_{50} = -0,526 (\pm 0,380) \text{ BEHv8} - 0,898 (\pm 0,450) \text{ GATS3e} + 0,187 (\pm 0,120) \text{ RDF055m} \\ - 1,249 (\pm 0,510) \text{ Mor08m} + 0,639 (\pm 0,210) \text{ HATSp} + 3,022 (\pm 0,97) \\ (n = 39; r = 0,863; s = 0,521; F = 19,282; Q^2_{cv} = 0,614; s_{press} = 0,642) \quad \text{Equação 123}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,968 (\pm 0,470) \text{ GATS3e} - 0,229 (\pm 0,180) \text{ RDF075p} - 1,449 (\pm 0,490) \text{ Mor08m} \\ & + 0,662 (\pm 0,230) \text{ HATSp} + 0,182 (\pm 0,140) \text{ H-052} + 2,873 (\pm 0,970) \\ (n = 39; r = 0,860; s = 0,527; F = 18,740; Q^2_{cv} = 0,623; S_{press} = 0,634) \end{aligned} \quad \text{Equação 124}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,911 (\pm 0,480) \text{ GATS3e} - 0,486 (\pm 0,420) \text{ GATS4e} - 0,181 (\pm 0,180) \text{ RDF075p} \\ & - 1,410 (\pm 0,510) \text{ Mor08m} + 0,512 (\pm 0,240) \text{ HATSp} + 3,606 (\pm 1,130) \\ (n = 39; r = 0,853; s = 0,539; F = 17,575; Q^2_{cv} = 0,586; S_{press} = 0,665) \end{aligned} \quad \text{Equação 125}$$

$$\begin{aligned} \text{pIC}_{50} = & + 0,204 (\pm 0,120) \text{ RDF055m} - 0,264 (\pm 0,170) \text{ RDF075p} - 1,445 (\pm 0,530) \text{ Mor08m} \\ & - 0,717 (\pm 0,390) \text{ Mor07v} + 0,394 (\pm 0,220) \text{ HATSp} + 3,603 (\pm 1,090) \\ (n = 39; r = 0,851; s = 0,542; F = 17,366; Q^2_{cv} = 0,575; S_{press} = 0,673) \end{aligned} \quad \text{Equação 126}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,860 (\pm 0,480) \text{ GATS3e} + 0,126 (\pm 0,120) \text{ RDF055m} - 0,190 (\pm 0,180) \text{ RDF075p} \\ & - 1,344 (\pm 0,540) \text{ Mor08m} + 0,593 (\pm 0,230) \text{ HATSp} + 2,869 (\pm 1,000) \\ (n = 39; r = 0,850; s = 0,544; F = 17,158; Q^2_{cv} = 0,602; S_{press} = 0,651) \end{aligned} \quad \text{Equação 127}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,556 (\pm 0,410) \text{ MATS8m} + 0,243 (\pm 0,130) \text{ RDF055m} - 1,216 (\pm 0,560) \text{ Mor08m} \\ & - 0,748 (\pm 0,400) \text{ Mor07v} + 0,400 (\pm 0,220) \text{ HATSp} + 3,624 (\pm 1,130) \\ (n = 39; r = 0,844; s = 0,553; F = 16,408; Q^2_{cv} = 0,501; S_{press} = 0,730) \end{aligned} \quad \text{Equação 128}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,625 (\pm 0,410) \text{ MATS8m} + 0,267 (\pm 0,130) \text{ RDF055m} - 0,297 (\pm 0,180) \text{ RDF075p} \\ & - 1,145 (\pm 0,570) \text{ Mor08m} - 0,801 (\pm 0,410) \text{ Mor07v} + 5,465 (\pm 0,630) \\ (n = 39; r = 0,837; s = 0,565; F = 15,426; Q^2_{cv} = 0,518; S_{press} = 0,717) \end{aligned} \quad \text{Equação 129}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,288 (\pm 0,190) \text{ RDF075p} - 1,430 (\pm 0,600) \text{ Mor08m} - 0,603 (\pm 0,410) \text{ Mor07v} \\ & + 0,788 (\pm 0,710) \text{ H3u} + 0,349 (\pm 0,240) \text{ HATSp} + 3,396 (\pm 1,170) \\ (n = 39; r = 0,825; s = 0,583; F = 14,117; Q^2_{cv} = 0,447; S_{press} = 0,768) \end{aligned} \quad \text{Equação 130}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,608 (\pm 0,420) \text{ GATS4e} - 0,362 (\pm 0,180) \text{ RDF075p} - 1,075 (\pm 0,610) \text{ Mor08m} \\ & - 0,645 (\pm 0,400) \text{ Mor07v} + 1,111 (\pm 0,690) \text{ H3u} + 4,926 (\pm 0,730) \\ (n = 39; r = 0,825; s = 0,583; F = 14,088; Q^2_{cv} = 0,482; S_{press} = 0,743) \end{aligned} \quad \text{Equação 131}$$

Equações de regressão significativas geradas pelas análises de *QSAR* aplicadas para a série de 30 compostos (série C), utilizando-se um a 5 descritores, respectivamente:

Série C (n=30), 1 descritor

$$\begin{aligned} \text{pIC}_{50} = & -0,274 (\pm 0,120) \text{ MAXDP} + 6,384 (\pm 0,310) \\ (n = 30; r = 0,655; s = 0,431; F = 21,076; Q^2_{cv} = 0,358; S_{press} = 0,457) \end{aligned} \quad \text{Equação 132}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,327 (\pm 0,150) \text{ Mor07m} + 6,488 (\pm 0,360) \\ (n = 30; r = 0,645; s = 0,436; F = 19,943; Q^2_{cv} = 0,322; S_{press} = 0,469) \end{aligned} \quad \text{Equação 133}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,182 (\pm 0,086) \text{ RDF125e} + 6,016 (\pm 0,200) \\ (n = 30; r = 0,633; s = 0,441; F = 18,696; Q^2_{cv} = 0,318; S_{press} = 0,471) \end{aligned} \quad \text{Equação 134}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,201 (\pm 0,096) \text{ RDF125u} + 6,019 (\pm 0,200) \\ (n = 30; r = 0,631; s = 0,442; F = 18,571; Q^2_{cv} = 0,306; S_{press} = 0,475) \end{aligned} \quad \text{Equação 135}$$

$$\begin{aligned} \text{pIC}_{50} = & -0,103 (\pm 0,050) \text{ RDF065e} + 6,459 (\pm 0,360) \\ (n = 30; r = 0,624; s = 0,445; F = 17,845; Q^2_{cv} = 0,179; S_{press} = 0,516) \end{aligned} \quad \text{Equação 136}$$

$$pIC_{50} = -0,180 (\pm 0,110) \text{ RDF065m} + 6,701 (\pm 0,560) \\ (n = 30; r = 0,552; s = 0,475; F = 12,260; Q^2_{cv} = 0,207; S_{press} = 0,508) \quad \text{Equação 137}$$

$$pIC_{50} = +0,719 (\pm 0,500) \text{ C-025} + 5,671 (\pm 0,200) \\ (n = 30; r = 0,487; s = 0,498; F = 8,682; Q^2_{cv} = 0,132; S_{press} = 0,531) \quad \text{Equação 138}$$

$$pIC_{50} = -0,270 (\pm 0,220) \text{ RDF075v} + 5,998 (\pm 0,250) \\ (n = 30; r = 0,432; s = 0,514; F = 6,437; Q^2_{cv} = 0,052; S_{press} = 0,555) \quad \text{Equação 139}$$

$$pIC_{50} = -0,618 (\pm 0,520) \text{ GATS4e} + 5,894 (\pm 0,210) \\ (n = 30; r = 0,418; s = 0,518; F = 5,938; Q^2_{cv} = 0,089; S_{press} = 0,544) \quad \text{Equação 140}$$

$$pIC_{50} = -0,533 (\pm 0,460) \text{ BEHv6} + 6,733 (\pm 0,840) \\ (n = 30; r = 0,410; s = 0,520; F = 5,651; Q^2_{cv} = 0,061; S_{press} = 0,552) \quad \text{Equação 141}$$

Série C (n=30), 2 descritores

$$pIC_{50} = -0,222 (\pm 0,100) \text{ MAXDP} - 0,0806 (\pm 0,040) \text{ RDF065e} + 6,795 (\pm 0,320) \\ (n = 30; r = 0,808; s = 0,342; F = 25,467; Q^2_{cv} = 0,533; S_{press} = 0,397) \quad \text{Equação 142}$$

$$pIC_{50} = -0,208 (\pm 0,110) \text{ MAXDP} - 0,147 (\pm 0,083) \text{ RDF125u} + 6,408 (\pm 0,260) \\ (n = 30; r = 0,786; s = 0,359; F = 21,803; Q^2_{cv} = 0,527; S_{press} = 0,399) \quad \text{Equação 143}$$

$$pIC_{50} = -0,0760 (\pm 0,043) \text{ RDF065e} - 0,250 (\pm 0,130) \text{ Mor07m} + 6,817 (\pm 0,350) \\ (n = 30; r = 0,778; s = 0,365; F = 20,728; Q^2_{cv} = 0,513; S_{press} = 0,405) \quad \text{Equação 144}$$

$$pIC_{50} = -0,149 (\pm 0,082) \text{ RDF065m} - 0,160 (\pm 0,073) \text{ RDF125e} + 6,742 (\pm 0,430) \\ (n = 30; r = 0,777; s = 0,366; F = 20,532; Q^2_{cv} = 0,539; S_{press} = 0,394) \quad \text{Equação 145}$$

$$pIC_{50} = -0,202 (\pm 0,110) \text{ MAXDP} - 0,129 (\pm 0,078) \text{ RDF125e} + 6,389 (\pm 0,260) \\ (n = 30; r = 0,775; s = 0,367; F = 20,318; Q^2_{cv} = 0,510; S_{press} = 0,406) \quad \text{Equação 146}$$

$$pIC_{50} = -0,175 (\pm 0,081) \text{ RDF125u} - 0,147 (\pm 0,083) \text{ RDF065m} + 6,737 (\pm 0,440) \\ (n = 30; r = 0,773; s = 0,368; F = 20,035; Q^2_{cv} = 0,528; S_{press} = 0,399) \quad \text{Equação 147}$$

$$pIC_{50} = -0,263 (\pm 0,110) \text{ MAXDP} - 0,245 (\pm 0,160) \text{ RDF075v} + 6,550 (\pm 0,290) \\ (n = 30; r = 0,764; s = 0,375; F = 18,910; Q^2_{cv} = 0,500; S_{press} = 0,410) \quad \text{Equação 148}$$

$$pIC_{50} = -0,298 (\pm 0,130) \text{ Mor07m} + 0,599 (\pm 0,380) \text{ C-025} + 6,328 (\pm 0,330) \\ (n = 30; r = 0,760; s = 0,377; F = 18,451; Q^2_{cv} = 0,479; S_{press} = 0,419) \quad \text{Equação 149}$$

$$pIC_{50} = -0,244 (\pm 0,110) \text{ MAXDP} + 0,553 (\pm 0,390) \text{ C-025} + 6,227 (\pm 0,300) \\ (n = 30; r = 0,751; s = 0,383; F = 17,495; Q^2_{cv} = 0,496; S_{press} = 0,412) \quad \text{Equação 150}$$

$$pIC_{50} = -0,571 (\pm 0,390) \text{ GATS3e} - 0,348 (\pm 0,130) \text{ Mor07m} + 7,009 (\pm 0,470) \\ (n = 30; r = 0,751; s = 0,383; F = 17,429; Q^2_{cv} = 0,467; S_{press} = 0,424) \quad \text{Equação 151}$$

Série C (n=30), 3 descritores

$$pIC_{50} = -0,490 (\pm 0,300) \text{ JhetZ} - 0,193 (\pm 0,093) \text{ MAXDP} - 0,240 (\pm 0,091) \text{ RDF125u} + 7,739 (\pm 0,850) \\ (n = 30; r = 0,856; s = 0,306; F = 23,775; Q^2_{cv} = 0,648; S_{press} = 0,351) \quad \text{Equação 152}$$

$$pIC_{50} = -0,200 (\pm 0,094) \text{ MAXDP} - 0,433 (\pm 0,280) \text{ GATS5p} - 0,168 (\pm 0,073) \text{ RDF125u} + 6,515 (\pm 0,240) \\ (n = 30; r = 0,851; s = 0,311; F = 22,754; Q^2_{cv} = 0,612; S_{press} = 0,368) \quad \text{Equação 153}$$

$$pIC_{50} = -0,192(\pm 0,098) \text{ MAXDP} - 0,439(\pm 0,290) \text{ GATS5p} - 0,150(\pm 0,069) \text{ RDF125e} + 6,495(\pm 0,240)$$

(n = 30; r = 0,843; s = 0,319; F = 21,204; $Q^2_{cv} = 0,599$; $s_{press} = 0,375$) Equação 154

$$pIC_{50} = -0,210(\pm 0,094) \text{ MAXDP} - 0,0679(\pm 0,039) \text{ RDF065e} + 0,372(\pm 0,340) \text{ C-025} + 6,625(\pm 0,340)$$

(n = 30; r = 0,842; s = 0,319; F = 21,137; $Q^2_{cv} = 0,597$; $s_{press} = 0,375$) Equação 155

$$pIC_{50} = -0,473(\pm 0,320) \text{ JhetZ} - 0,182(\pm 0,099) \text{ MAXDP} - 0,212(\pm 0,088) \text{ RDF125e} + 7,662(\pm 0,880)$$

(n = 30; r = 0,841; s = 0,320; F = 20,961; $Q^2_{cv} = 0,620$; $s_{press} = 0,365$) Equação 156

$$pIC_{50} = +0,705(\pm 0,500) \text{ CIC2} - 0,0812(\pm 0,039) \text{ RDF065e} - 0,279(\pm 0,120) \text{ Mor07m} + 6,866(\pm 0,320)$$

(n = 30; r = 0,838; s = 0,323; F = 20,418; $Q^2_{cv} = 0,610$; $s_{press} = 0,369$) Equação 157

$$pIC_{50} = -0,462(\pm 0,340) \text{ GATS3e} - 0,0654(\pm 0,039) \text{ RDF065e} - 0,278(\pm 0,120) \text{ Mor07m} + 7,193(\pm 0,420)$$

(n = 30; r = 0,836; s = 0,325; F = 20,085; $Q^2_{cv} = 0,614$; $s_{press} = 0,368$) Equação 158

$$pIC_{50} = -0,462(\pm 0,330) \text{ GATS4e} - 0,148(\pm 0,073) \text{ RDF065m} - 0,144(\pm 0,066) \text{ RDF125e} + 6,797(\pm 0,390)$$

(n = 30; r = 0,835; s = 0,325; F = 20,022; $Q^2_{cv} = 0,604$; $s_{press} = 0,372$) Equação 159

$$pIC_{50} = -0,468(\pm 0,330) \text{ GATS4e} - 0,158(\pm 0,073) \text{ RDF125u} - 0,147(\pm 0,074) \text{ RDF065m} + 6,793(\pm 0,390)$$

(n = 30; r = 0,834; s = 0,327; F = 19,729; $Q^2_{cv} = 0,600$; $s_{press} = 0,374$) Equação 160

$$pIC_{50} = -0,152(\pm 0,110) \text{ MAXDP} - 0,144(\pm 0,076) \text{ RDF125u} - 0,101(\pm 0,081) \text{ RDF065m} + 6,794(\pm 0,390)$$

(n = 30; r = 0,833; s = 0,327; F = 19,698; $Q^2_{cv} = 0,613$; $s_{press} = 0,368$) Equação 161

Série C (n=30), 4 descritores

$$pIC_{50} = -0,605(\pm 0,270) \text{ JhetZ} + 0,824(\pm 0,420) \text{ CIC2} - 0,256(\pm 0,087) \text{ RDF125u}$$

$$- 0,237(\pm 0,110) \text{ Mor07m} + 8,084(\pm 0,770)$$

(n = 30; r = 0,893; s = 0,271; F = 24,740; $Q^2_{cv} = 0,699$; $s_{press} = 0,331$) Equação 162

$$pIC_{50} = +0,751(\pm 0,430) \text{ CIC2} - 0,0644(\pm 0,034) \text{ RDF065e} - 0,276(\pm 0,100) \text{ Mor07m}$$

$$+ 0,475(\pm 0,290) \text{ C-025} + 6,668(\pm 0,300)$$

(n = 30; r = 0,891; s = 0,274; F = 24,017; $Q^2_{cv} = 0,720$; $s_{press} = 0,319$) Equação 163

$$pIC_{50} = -0,391(\pm 0,280) \text{ JhetZ} - 0,190(\pm 0,084) \text{ MAXDP} - 0,335(\pm 0,260) \text{ GATS5p}$$

$$- 0,238(\pm 0,082) \text{ RDF125u} + 7,553(\pm 0,780)$$

(n = 30; r = 0,890; s = 0,275; F = 23,810; $Q^2_{cv} = 0,682$; $s_{press} = 0,340$) Equação 164

$$pIC_{50} = -0,582(\pm 0,280) \text{ JhetZ} + 0,751(\pm 0,440) \text{ CIC2} - 0,222(\pm 0,083) \text{ RDF125e}$$

$$- 0,237(\pm 0,120) \text{ Mor07m} + 8,014(\pm 0,810)$$

(n = 30; r = 0,881; s = 0,285; F = 21,667; $Q^2_{cv} = 0,689$; $s_{press} = 0,336$) Equação 165

$$pIC_{50} = -0,421(\pm 0,290) \text{ JhetZ} - 0,152(\pm 0,095) \text{ MAXDP} - 0,225(\pm 0,086) \text{ RDF125u}$$

$$- 0,0766(\pm 0,073) \text{ RDF065m} + 7,844(\pm 0,800)$$

(n = 30; r = 0,880; s = 0,286; F = 21,533; $Q^2_{cv} = 0,691$; $s_{press} = 0,335$) Equação 166

$$pIC_{50} = -0,374(\pm 0,300) \text{ JhetZ} - 0,178(\pm 0,090) \text{ MAXDP} - 0,350(\pm 0,270) \text{ GATS5p}$$

$$- 0,212(\pm 0,079) \text{ RDF125e} + 7,480(\pm 0,810)$$

(n = 30; r = 0,879; s = 0,288; F = 21,167; $Q^2_{cv} = 0,662$; $s_{press} = 0,351$) Equação 167

$$pIC_{50} = -0,140(\pm 0,100) \text{ MAXDP} - 0,405(\pm 0,300) \text{ GATS3e} - 0,0630(\pm 0,035) \text{ RDF065e}$$

$$- 0,178(\pm 0,130) \text{ Mor07m} + 7,221(\pm 0,370)$$

(n = 30; r = 0,878; s = 0,289; F = 21,000; $Q^2_{cv} = 0,681$; $s_{press} = 0,341$) Equação 168

$$pIC_{50} = -0,502(\pm 0,300) \text{ GATS4e} - 0,131(\pm 0,067) \text{ RDF065m} - 0,121(\pm 0,061) \text{ RDF125e}$$

$$+ 0,423(\pm 0,310) \text{ C-025} + 6,617(\pm 0,370)$$

(n = 30; r = 0,877; s = 0,289; F = 20,885; $Q^2_{cv} = 0,702$; $s_{press} = 0,329$) Equação 169

$$pIC_{50} = -0,508 (\pm 0,300) \text{ GATS4e} - 0,133 (\pm 0,068) \text{ RDF125u} - 0,130 (\pm 0,067) \text{ RDF065m} \\ + 0,423 (\pm 0,310) \text{ C-025} + 6,614 (\pm 0,370) \\ (n = 30; r = 0,875; s = 0,291; F = 20,516; Q^2_{cv} = 0,697; s_{press} = 0,332) \quad \text{Equação 170}$$

$$pIC_{50} = +0,721 (\pm 0,450) \text{ CIC2} + 0,478 (\pm 0,390) \text{ BEHv6} - 0,111 (\pm 0,042) \text{ RDF065e} \\ - 0,356 (\pm 0,130) \text{ Mor07m} + 6,379 (\pm 0,490) \\ (n = 30; r = 0,874; s = 0,293; F = 20,179; Q^2_{cv} = 0,620; s_{press} = 0,372) \quad \text{Equação 171}$$

Série C (n=30), 5 descritores

$$pIC_{50} = -0,520 (\pm 0,260) \text{ JhetZ} + 0,752 (\pm 0,400) \text{ CIC2} - 0,266 (\pm 0,240) \text{ GATS5p} \\ - 0,252 (\pm 0,081) \text{ RDF125u} - 0,231 (\pm 0,100) \text{ Mor07m} + 7,916 (\pm 0,730) \\ (n = 30; r = 0,914; s = 0,250; F = 24,207; Q^2_{cv} = 0,715; s_{press} = 0,329) \quad \text{Equação 172}$$

$$pIC_{50} = +0,657 (\pm 0,400) \text{ CIC2} - 0,323 (\pm 0,280) \text{ GATS4e} - 0,0661 (\pm 0,032) \text{ RDF065e} \\ - 0,229 (\pm 0,100) \text{ Mor07m} + 0,494 (\pm 0,270) \text{ C-025} + 6,637 (\pm 0,270) \\ (n = 30; r = 0,913; s = 0,252; F = 23,956; Q^2_{cv} = 0,741; s_{press} = 0,314) \quad \text{Equação 173}$$

$$pIC_{50} = -0,523 (\pm 0,260) \text{ JhetZ} + 0,691 (\pm 0,410) \text{ CIC2} - 0,237 (\pm 0,083) \text{ RDF125u} \\ - 0,0682 (\pm 0,064) \text{ RDF065m} - 0,203 (\pm 0,110) \text{ Mor07m} + 8,133 (\pm 0,720) \\ (n = 30; r = 0,912; s = 0,252; F = 23,843; Q^2_{cv} = 0,759; s_{press} = 0,302) \quad \text{Equação 174}$$

$$pIC_{50} = -0,476 (\pm 0,280) \text{ JhetZ} + 0,827 (\pm 0,400) \text{ CIC2} - 0,211 (\pm 0,093) \text{ RDF125u} \\ - 0,242 (\pm 0,110) \text{ Mor07m} + 0,306 (\pm 0,300) \text{ C-025} + 7,661 (\pm 0,840) \\ (n = 30; r = 0,911; s = 0,254; F = 23,313; Q^2_{cv} = 0,726; s_{press} = 0,322) \quad \text{Equação 175}$$

$$pIC_{50} = -0,476 (\pm 0,280) \text{ JhetZ} + 0,827 (\pm 0,400) \text{ CIC2} - 0,211 (\pm 0,093) \text{ RDF125u} \\ - 0,242 (\pm 0,110) \text{ Mor07m} + 0,306 (\pm 0,300) \text{ C-025} + 7,661 (\pm 0,840) \\ (n = 30; r = 0,911; s = 0,254; F = 23,313; Q^2_{cv} = 0,726; s_{press} = 0,322) \quad \text{Equação 176}$$

$$pIC_{50} = -0,466 (\pm 0,260) \text{ JhetZ} - 0,199 (\pm 0,080) \text{ MAXDP} - 0,376 (\pm 0,270) \text{ BEHv6} \\ - 0,418 (\pm 0,300) \text{ GATS3e} - 0,174 (\pm 0,088) \text{ RDF125u} + 8,629 (\pm 0,900) \\ (n = 30; r = 0,906; s = 0,261; F = 21,945; Q^2_{cv} = 0,714; s_{press} = 0,329) \quad \text{Equação 177}$$

$$pIC_{50} = -0,450 (\pm 0,260) \text{ JhetZ} - 0,428 (\pm 0,290) \text{ GATS3e} - 0,0964 (\pm 0,063) \text{ RDF065m} \\ - 0,149 (\pm 0,081) \text{ RDF125e} - 0,221 (\pm 0,120) \text{ Mor07m} + 8,447 (\pm 0,800) \\ (n = 30; r = 0,906; s = 0,261; F = 21,875; Q^2_{cv} = 0,733; s_{press} = 0,318) \quad \text{Equação 178}$$

$$pIC_{50} = -0,498 (\pm 0,270) \text{ JhetZ} + 0,613 (\pm 0,430) \text{ CIC2} - 0,0755 (\pm 0,066) \text{ RDF065m} \\ - 0,206 (\pm 0,077) \text{ RDF125e} - 0,197 (\pm 0,110) \text{ Mor07m} + 8,085 (\pm 0,750) \\ (n = 30; r = 0,905; s = 0,262; F = 21,646; Q^2_{cv} = 0,750; s_{press} = 0,308) \quad \text{Equação 179}$$

$$pIC_{50} = -0,450 (\pm 0,270) \text{ JhetZ} - 0,429 (\pm 0,300) \text{ GATS3e} - 0,163 (\pm 0,090) \text{ RDF125u} \\ - 0,0945 (\pm 0,064) \text{ RDF065m} - 0,225 (\pm 0,120) \text{ Mor07m} + 8,446 (\pm 0,810) \\ (n = 30; r = 0,904; s = 0,263; F = 21,537; Q^2_{cv} = 0,713; s_{press} = 0,330) \quad \text{Equação 180}$$

$$pIC_{50} = +0,661 (\pm 0,420) \text{ CIC2} - 0,415 (\pm 0,300) \text{ GATS4e} - 0,0780 (\pm 0,032) \text{ RDF065e} \\ - 0,226 (\pm 0,110) \text{ Mor07m} + 0,109 (\pm 0,068) \text{ H-046} + 6,726 (\pm 0,270) \\ (n = 30; r = 0,904; s = 0,263; F = 21,401; Q^2_{cv} = 0,645; s_{press} = 0,367) \quad \text{Equação 181}$$

Equações de regressão significativas geradas pelas análises de *QSAR* aplicadas para a série de 23 compostos (série D), utilizando-se um a 5 descritores, respectivamente:

Série D (n=23), 1 descritor

$$pIC_{50} = -0,199 (\pm 0,051) \text{ SP13} + 7,319 (\pm 0,440) \\ (n = 23; r = 0,871; s = 0,342; F = 66,223; Q^2_{cv} = 0,614; S_{press} = 0,432) \quad \text{Equação 182}$$

$$pIC_{50} = -1,078 (\pm 0,350) \text{ Mor07v} + 7,287 (\pm 0,550) \\ (n = 23; r = 0,810; s = 0,408; F = 40,001; Q^2_{cv} = 0,586; S_{press} = 0,448) \quad \text{Equação 183}$$

$$pIC_{50} = -1,068 (\pm 0,370) \text{ SEige} + 6,297 (\pm 0,280) \\ (n = 23; r = 0,796; s = 0,422; F = 36,268; Q^2_{cv} = 0,566; S_{press} = 0,459) \quad \text{Equação 184}$$

$$pIC_{50} = -0,385 (\pm 0,150) \text{ RDF060v} + 6,564 (\pm 0,400) \\ (n = 23; r = 0,748; s = 0,462; F = 26,660; Q^2_{cv} = 0,481; S_{press} = 0,501) \quad \text{Equação 185}$$

$$pIC_{50} = -0,434 (\pm 0,180) \text{ RDF115v} + 5,920 (\pm 0,220) \\ (n = 23; r = 0,742; s = 0,467; F = 25,673; Q^2_{cv} = 0,446; S_{press} = 0,518) \quad \text{Equação 186}$$

$$pIC_{50} = -0,354 (\pm 0,160) \text{ MAXDP} + 6,555 (\pm 0,440) \\ (n = 23; r = 0,716; s = 0,486; F = 22,056; Q^2_{cv} = 0,420; S_{press} = 0,530) \quad \text{Equação 187}$$

$$pIC_{50} = -0,989 (\pm 0,450) \text{ H2e} + 7,026 (\pm 0,640) \\ (n = 23; r = 0,708; s = 0,492; F = 21,092; Q^2_{cv} = 0,395; S_{press} = 0,542) \quad \text{Equação 188}$$

$$pIC_{50} = -0,482 (\pm 0,240) \text{ RDF115m} + 5,882 (\pm 0,240) \\ (n = 23; r = 0,667; s = 0,519; F = 16,837; Q^2_{cv} = 0,232; S_{press} = 0,610) \quad \text{Equação 189}$$

$$pIC_{50} = -0,276 (\pm 0,140) \text{ RDF065m} + 7,135 (\pm 0,780) \\ (n = 23; r = 0,659; s = 0,524; F = 16,121; Q^2_{cv} = 0,296; S_{press} = 0,584) \quad \text{Equação 190}$$

$$pIC_{50} = +0,863 (\pm 0,500) \text{ Jhetv} + 4,530 (\pm 0,710) \\ (n = 23; r = 0,618; s = 0,547; F = 12,970; Q^2_{cv} = 0,278; S_{press} = 0,591) \quad \text{Equação 191}$$

Série D (n=23), 2 descritores

$$pIC_{50} = -0,554 (\pm 0,300) \text{ SEige} - 0,140 (\pm 0,050) \text{ SP13} + 7,152 (\pm 0,350) \\ (n = 23; r = 0,929; s = 0,264; F = 63,102; Q^2_{cv} = 0,834; S_{press} = 0,291) \quad \text{Equação 192}$$

$$pIC_{50} = -0,136 (\pm 0,054) \text{ SP13} - 0,554 (\pm 0,310) \text{ Mor07v} + 7,626 (\pm 0,390) \\ (n = 23; r = 0,926; s = 0,270; F = 59,916; Q^2_{cv} = 0,824; S_{press} = 0,299) \quad \text{Equação 193}$$

$$pIC_{50} = -0,231 (\pm 0,075) \text{ RDF065m} - 0,339 (\pm 0,092) \text{ RDF060v} + 7,665 (\pm 0,430) \\ (n = 23; r = 0,925; s = 0,271; F = 59,266; Q^2_{cv} = 0,818; S_{press} = 0,304) \quad \text{Equação 194}$$

$$pIC_{50} = -0,173 (\pm 0,110) \text{ MAXDP} - 0,156 (\pm 0,049) \text{ SP13} + 7,391 (\pm 0,360) \\ (n = 23; r = 0,920; s = 0,280; F = 55,126; Q^2_{cv} = 0,777; S_{press} = 0,337) \quad \text{Equação 195}$$

$$pIC_{50} = -0,796 (\pm 0,270) \text{ SEige} - 0,294 (\pm 0,120) \text{ RDF115v} + 6,297 (\pm 0,190) \\ (n = 23; r = 0,919; s = 0,281; F = 54,604; Q^2_{cv} = 0,801; S_{press} = 0,318) \quad \text{Equação 196}$$

$$pIC_{50} = -0,165 (\pm 0,048) \text{ SP13} - 0,137 (\pm 0,087) \text{ RDF065m} + 7,751 (\pm 0,460) \\ (n = 23; r = 0,918; s = 0,283; F = 53,637; Q^2_{cv} = 0,677; S_{press} = 0,406) \quad \text{Equação 197}$$

$$pIC_{50} = -0,273 (\pm 0,099) \text{ MAXDP} - 0,344 (\pm 0,120) \text{ RDF115v} + 6,538 (\pm 0,260) \\ (n = 23; r = 0,912; s = 0,293; F = 49,338; Q^2_{cv} = 0,723; S_{press} = 0,375) \quad \text{Equação 198}$$

$$pIC_{50} = -0,773 (\pm 0,290) \text{ SEige} - 0,252 (\pm 0,110) \text{ RDF060v} + 6,700 (\pm 0,270) \\ (n = 23; r = 0,908; s = 0,299; F = 47,061; Q^2_{cv} = 0,747; s_{press} = 0,359) \quad \text{Equação 199}$$

$$pIC_{50} = +0,394 (\pm 0,310) \text{ Jhetv} - 0,170 (\pm 0,050) \text{ SP13} + 6,550 (\pm 0,710) \\ (n = 23; r = 0,907; s = 0,300; F = 46,438; Q^2_{cv} = 0,730; s_{press} = 0,371) \quad \text{Equação 200}$$

$$pIC_{50} = -0,265 (\pm 0,100) \text{ MAXDP} - 0,300 (\pm 0,110) \text{ RDF060v} + 7,015 (\pm 0,320) \\ (n = 23; r = 0,905; s = 0,304; F = 45,124; Q^2_{cv} = 0,765; s_{press} = 0,346) \quad \text{Equação 201}$$

Série D (n=23), 3 descritores

$$pIC_{50} = -0,179 (\pm 0,056) \text{ RDF065m} - 0,244 (\pm 0,074) \text{ RDF060v} - 0,495 (\pm 0,210) \text{ Mor07v} + 7,910 (\pm 0,310) \\ (n = 23; r = 0,968; s = 0,183; F = 94,558; Q^2_{cv} = 0,916; s_{press} = 0,212) \quad \text{Equação 202}$$

$$pIC_{50} = +0,429 (\pm 0,190) \text{ Jhetv} - 0,220 (\pm 0,053) \text{ RDF065m} - 0,273 (\pm 0,072) \text{ RDF060v} + 6,878 (\pm 0,470) \\ (n = 23; r = 0,965; s = 0,191; F = 86,674; Q^2_{cv} = 0,906; s_{press} = 0,224) \quad \text{Equação 203}$$

$$pIC_{50} = +0,425 (\pm 0,220) \text{ Jhetv} - 0,131 (\pm 0,040) \text{ SP13} - 0,145 (\pm 0,066) \text{ RDF065m} + 6,948 (\pm 0,540) \\ (n = 23; r = 0,957; s = 0,212; F = 69,441; Q^2_{cv} = 0,829; s_{press} = 0,303) \quad \text{Equação 204}$$

$$pIC_{50} = -0,153 (\pm 0,089) \text{ MAXDP} - 0,160 (\pm 0,073) \text{ RDF065m} - 0,304 (\pm 0,076) \text{ RDF060v} + 7,586 (\pm 0,350) \\ (n = 23; r = 0,956; s = 0,215; F = 67,159; Q^2_{cv} = 0,863; s_{press} = 0,271) \quad \text{Equação 205}$$

$$pIC_{50} = -0,116 (\pm 0,044) \text{ SP13} - 0,115 (\pm 0,067) \text{ RDF065m} - 0,481 (\pm 0,250) \text{ Mor07v} + 7,949 (\pm 0,360) \\ (n = 23; r = 0,956; s = 0,214; F = 67,725; Q^2_{cv} = 0,858; s_{press} = 0,276) \quad \text{Equação 206}$$

$$pIC_{50} = -0,198 (\pm 0,066) \text{ RDF065m} - 0,276 (\pm 0,086) \text{ RDF060v} - 0,386 (\pm 0,250) \text{ H2e} + 7,869 (\pm 0,380) \\ (n = 23; r = 0,953; s = 0,223; F = 62,164; Q^2_{cv} = 0,877; s_{press} = 0,257) \quad \text{Equação 207}$$

$$pIC_{50} = -0,434 (\pm 0,280) \text{ SEige} - 0,155 (\pm 0,079) \text{ RDF065m} - 0,279 (\pm 0,085) \text{ RDF060v} + 7,380 (\pm 0,400) \\ (n = 23; r = 0,953; s = 0,223; F = 62,108; Q^2_{cv} = 0,876; s_{press} = 0,258) \quad \text{Equação 208}$$

$$pIC_{50} = -0,305 (\pm 0,078) \text{ MAXDP} - 0,321 (\pm 0,090) \text{ RDF115v} - 0,426 (\pm 0,230) \text{ nRORPh} + 6,715 (\pm 0,220) \\ (n = 23; r = 0,952; s = 0,224; F = 61,098; Q^2_{cv} = 0,861; s_{press} = 0,273) \quad \text{Equação 209}$$

$$pIC_{50} = -0,181 (\pm 0,091) \text{ MAXDP} - 0,271 (\pm 0,099) \text{ RDF115v} - 0,492 (\pm 0,270) \text{ Mor07v} + 7,003 (\pm 0,330) \\ (n = 23; r = 0,951; s = 0,226; F = 60,221; Q^2_{cv} = 0,869; s_{press} = 0,265) \quad \text{Equação 210}$$

$$pIC_{50} = -0,133 (\pm 0,072) \text{ RDF065m} - 0,233 (\pm 0,100) \text{ RDF115v} - 0,647 (\pm 0,240) \text{ Mor07v} + 7,467 (\pm 0,420) \\ (n = 23; r = 0,948; s = 0,233; F = 56,183; Q^2_{cv} = 0,861; s_{press} = 0,273) \quad \text{Equação 211}$$

Série D (n=23), 4 descritores

$$pIC_{50} = -0,0945 (\pm 0,072) \text{ MAXDP} - 0,146 (\pm 0,054) \text{ RDF065m} - 0,243 (\pm 0,064) \text{ RDF060v} - 0,389 \\ (\pm 0,200) \text{ Mor07v} + 7,809 (\pm 0,280) \\ (n = 23; r = 0,978; s = 0,158; F = 97,120; Q^2_{cv} = 0,912; s_{press} = 0,224) \quad \text{Equação 212}$$

$$pIC_{50} = +0,337 (\pm 0,180) \text{ Jhetv} - 0,102 (\pm 0,072) \text{ MAXDP} - 0,174 (\pm 0,055) \text{ RDF065m} - 0,264 \\ (\pm 0,061) \text{ RDF060v} + 6,994 (\pm 0,400) \\ (n = 23; r = 0,977; s = 0,161; F = 94,240; Q^2_{cv} = 0,921; s_{press} = 0,211) \quad \text{Equação 213}$$

$$pIC_{50} = +0,248 (\pm 0,220) \text{ Jhetv} - 0,190 (\pm 0,051) \text{ RDF065m} - 0,239 (\pm 0,067) \text{ RDF060v} - 0,321 \\ (\pm 0,240) \text{ Mor07v} + 7,369 (\pm 0,550) \\ (n = 23; r = 0,976; s = 0,164; F = 90,444; Q^2_{cv} = 0,917; s_{press} = 0,217) \quad \text{Equação 214}$$

$$pIC_{50} = +0,342 (\pm 0,190) \text{ Jhetv} - 0,202 (\pm 0,051) \text{ RDF065m} - 0,248 (\pm 0,068) \text{ RDF060v} - 0,236 (\pm 0,210) \text{ H2e} + 7,162 (\pm 0,490)$$

$$(n = 23; r = 0,974; s = 0,171; F = 82,306; Q^2_{cv} = 0,920; s_{press} = 0,212) \quad \text{Equação 215}$$

$$pIC_{50} = -0,128 (\pm 0,074) \text{ MAXDP} - 0,144 (\pm 0,060) \text{ RDF065m} - 0,258 (\pm 0,068) \text{ RDF060v} - 0,313 (\pm 0,200) \text{ H2e} + 7,764 (\pm 0,300)$$

$$(n = 23; r = 0,973; s = 0,174; F = 79,604; Q^2_{cv} = 0,904; s_{press} = 0,233) \quad \text{Equação 216}$$

$$pIC_{50} = -0,231 (\pm 0,085) \text{ MAXDP} - 0,276 (\pm 0,083) \text{ RDF115v} - 0,346 (\pm 0,250) \text{ Mor07v} - 0,303 (\pm 0,210) \text{ nRORPh} + 6,991 (\pm 0,280)$$

$$(n = 23; r = 0,968; s = 0,190; F = 66,014; Q^2_{cv} = 0,894; s_{press} = 0,245) \quad \text{Equação 217}$$

$$pIC_{50} = +0,424 (\pm 0,260) \text{ Jhetv} - 0,437 (\pm 0,310) \text{ SEige} - 0,130 (\pm 0,087) \text{ RDF065m} - 0,266 (\pm 0,130) \text{ RDF115m} + 6,153 (\pm 0,520)$$

$$(n = 23; r = 0,947; s = 0,241; F = 39,150; Q^2_{cv} = 0,750; s_{press} = 0,376) \quad \text{Equação 218}$$

$$pIC_{50} = -0,266 (\pm 0,076) \text{ MAXDP} - 0,281 (\pm 0,086) \text{ RDF115v} - 0,281 (\pm 0,230) \text{ H2e} - 0,353 (\pm 0,210) \text{ nRORPh} + 6,958 (\pm 0,280)$$

$$(n = 23; r = 0,965; s = 0,197; F = 60,873; Q^2_{cv} = 0,889; s_{press} = 0,251) \quad \text{Equação 219}$$

$$pIC_{50} = -0,127 (\pm 0,090) \text{ MAXDP} - 0,0874 (\pm 0,069) \text{ RDF065m} - 0,247 (\pm 0,088) \text{ RDF115v} - 0,489 (\pm 0,230) \text{ Mor07v} + 7,311 (\pm 0,370)$$

$$(n = 23; r = 0,965; s = 0,196; F = 61,431; Q^2_{cv} = 0,883; s_{press} = 0,257) \quad \text{Equação 220}$$

$$pIC_{50} = -0,256 (\pm 0,085) \text{ MAXD} - 0,0763 (\pm 0,072) \text{ RDF065m} - 0,301 (\pm 0,085) \text{ RDF115v} - 0,406 (\pm 0,210) \text{ nRORPh} + 6,977 (\pm 0,320)$$

$$(n = 23; r = 0,963; s = 0,204; F = 56,672; Q^2_{cv} = 0,858; s_{press} = 0,284) \quad \text{Equação 221}$$

Série D (n=23), 5 descritores

$$pIC_{50} = +0,262 (\pm 0,160) \text{ Jhetv} - 0,0963 (\pm 0,063) \text{ MAXDP} - 0,160 (\pm 0,049) \text{ RDF065m} - 0,241 (\pm 0,056) \text{ RDF060v} - 0,216 (\pm 0,170) \text{ H2e} + 7,248 (\pm 0,400)$$

$$(n = 23; r = 0,984; s = 0,138; F = 102,865; Q^2_{cv} = 0,939; s_{press} = 0,191) \quad \text{Equação 222}$$

$$pIC_{50} = +0,211 (\pm 0,190) \text{ Jhetv} - 0,0831 (\pm 0,066) \text{ MAXDP} - 0,160 (\pm 0,050) \text{ RDF065m} - 0,239 (\pm 0,058) \text{ RDF060v} - 0,254 (\pm 0,220) \text{ Mor07v} + 7,361 (\pm 0,470)$$

$$(n = 23; r = 0,983; s = 0,141; F = 98,302; Q^2_{cv} = 0,926; s_{press} = 0,211) \quad \text{Equação 223}$$

$$pIC_{50} = -0,178 (\pm 0,080) \text{ MAXDP} - 0,0793 (\pm 0,056) \text{ RDF065m} - 0,254 (\pm 0,072) \text{ RDF115v} - 0,355 (\pm 0,210) \text{ Mor07v} - 0,279 (\pm 0,180) \text{ nRORPh} + 7,271 (\pm 0,300)$$

$$(n = 23; r = 0,979; s = 0,158; F = 77,782; Q^2_{cv} = 0,922; s_{press} = 0,216) \quad \text{Equação 224}$$

$$pIC_{50} = +0,346 (\pm 0,180) \text{ Jhetv} - 0,279 (\pm 0,230) \text{ SEige} - 0,0875 (\pm 0,041) \text{ SP13} - 0,110 (\pm 0,061) \text{ RDF065m} - 0,138 (\pm 0,110) \text{ RDF115m} + 6,729 (\pm 0,450)$$

$$(n = 23; r = 0,976; s = 0,168; F = 69,114; Q^2_{cv} = 0,831; s_{press} = 0,318) \quad \text{Equação 225}$$

$$pIC_{50} = -0,194 (\pm 0,077) \text{ MAXDP} - 0,0890 (\pm 0,060) \text{ RDF065m} - 0,298 (\pm 0,088) \text{ RDF115m} - 0,423 (\pm 0,190) \text{ H2e} - 0,298 (\pm 0,190) \text{ nRORPh} + 7,395 (\pm 0,300)$$

$$(n = 23; r = 0,975; s = 0,172; F = 65,164; Q^2_{cv} = 0,899; s_{press} = 0,245) \quad \text{Equação 226}$$

$$pIC_{50} = -0,219 (\pm 0,077) \text{ MAXDP} - 0,0742 (\pm 0,061) \text{ RDF065m} - 0,262 (\pm 0,077) \text{ RDF115v} - 0,274 (\pm 0,200) \text{ H2e} - 0,335 (\pm 0,190) \text{ nRORPh} + 7,208 (\pm 0,320)$$

$$(n = 23; r = 0,975; s = 0,172; F = 65,160; Q^2_{cv} = 0,903; s_{press} = 0,241) \quad \text{Equação 227}$$

$$\begin{aligned} \text{pIC}_{50} = & + 0,308 (\pm 0,230) \text{ Jhetv} - 0,135 (\pm 0,087) \text{ MAXDP} - 0,106 (\pm 0,070) \text{ RDF065m} \\ & - 0,245 (\pm 0,087) \text{ RDF115v} - 0,255 (\pm 0,230) \text{ H2e} + 6,632 (\pm 0,550) \\ (n = 23; r = 0,969; s = 0,192; F = 51,622; Q^2_{cv} = 0,859; S_{press} = 0,291) \end{aligned} \quad \text{Equação 228}$$

$$\begin{aligned} \text{pIC}_{50} = & + 0,270 (\pm 0,230) \text{ Jhetv} - 0,121 (\pm 0,086) \text{ MAXDP} - 0,116 (\pm 0,068) \text{ RDF065m} \\ & - 0,282 (\pm 0,099) \text{ RDF115m} - 0,399 (\pm 0,220) \text{ H2e} + 6,879 (\pm 0,540) \\ (n = 23; r = 0,969; s = 0,190; F = 52,927; Q^2_{cv} = 0,872; S_{press} = 0,277) \end{aligned} \quad \text{Equação 229}$$

$$\begin{aligned} \text{pIC}_{50} = & - 0,295 (\pm 0,290) \text{ SEige} - 0,0798 (\pm 0,054) \text{ SP13} - 0,094 (\pm 0,073) \text{ RDF065m} \\ & - 0,172 (\pm 0,130) \text{ RDF115m} - 0,285 (\pm 0,270) \text{ H2e} + 7,454 (\pm 0,450) \\ (n = 23; r = 0,964; s = 0,204; F = 45,272; Q^2_{cv} = 0,780; S_{press} = 0,363) \end{aligned} \quad \text{Equação 230}$$

$$\begin{aligned} \text{pIC}_{50} = & - 0,166 (\pm 0,100) \text{ MAXDP} - 0,0997 (\pm 0,072) \text{ RDF065m} - 0,268 (\pm 0,110) \text{ RDF115m} - 0,406 \\ & (\pm 0,270) \text{ Mor07v} - 0,271 (\pm 0,230) \text{ nRORPh} + 7,394 (\pm 0,390) \\ (n = 23; r = 0,964; s = 0,207; F = 44,202; Q^2_{cv} = 0,798; S_{press} = 0,348) \end{aligned} \quad \text{Equação 231}$$

Equações de regressão significativas geradas pelas análises de *QSAR* aplicadas para a série de 22 compostos (série E), utilizando-se um a 5 descritores, respectivamente:

Série E (n=22), 1 descritor

$$\begin{aligned} \text{pIC}_{50} = & - 0,264 (\pm 0,076) \text{ SP13} + 7,802 (\pm 0,600) \\ (n = 22; r = 0,850; s = 0,313; F = 51,936; Q^2_{cv} = 0,663; S_{press} = 0,345) \end{aligned} \quad \text{Equação 232}$$

$$\begin{aligned} \text{pIC}_{50} = & - 0,352 (\pm 0,140) \text{ RDF065p} + 7,052 (\pm 0,530) \\ (n = 22; r = 0,763; s = 0,383; F = 27,879; Q^2_{cv} = 0,513; S_{press} = 0,414) \end{aligned} \quad \text{Equação 233}$$

$$\begin{aligned} \text{pIC}_{50} = & - 0,668 (\pm 0,310) \text{ Mor07p} + 6,592 (\pm 0,430) \\ (n = 22; r = 0,706; s = 0,420; F = 19,885; Q^2_{cv} = 0,411; S_{press} = 0,455) \end{aligned} \quad \text{Equação 234}$$

$$\begin{aligned} \text{pIC}_{50} = & + 0,761 (\pm 0,420) \text{ Jhetv} + 4,735 (\pm 0,600) \\ (n = 22; r = 0,647; s = 0,452; F = 14,402; Q^2_{cv} = 0,309; S_{press} = 0,493) \end{aligned} \quad \text{Equação 235}$$

$$\begin{aligned} \text{pIC}_{50} = & - 0,175 (\pm 0,170) \text{ RDF050v} + 6,227 (\pm 0,510) \\ (n = 22; r = 0,425; s = 0,537; F = 4,415; Q^2_{cv} = 0,048; S_{press} = 0,578) \end{aligned} \quad \text{Equação 236}$$

Série E (n=22), 2 descritores

$$\begin{aligned} \text{pIC}_{50} = & - 0,191 (\pm 0,095) \text{ SP13} - 0,158 (\pm 0,140) \text{ RDF065p} + 7,818 (\pm 0,550) \\ (n = 22; r = 0,886; s = 0,282; F = 34,706; Q^2_{cv} = 0,705; S_{press} = 0,331) \end{aligned} \quad \text{Equação 237}$$

$$\begin{aligned} \text{pIC}_{50} = & + 0,483 (\pm 0,320) \text{ Jhetv} - 0,277 (\pm 0,130) \text{ RDF065p} + 6,123 (\pm 0,760) \\ (n = 22; r = 0,852; s = 0,319; F = 25,067; Q^2_{cv} = 0,624; S_{press} = 0,373) \end{aligned} \quad \text{Equação 238}$$

$$\begin{aligned} \text{pIC}_{50} = & + 0,762 (\pm 0,740) \text{ CIC2} - 0,721 (\pm 0,290) \text{ Mor07p} + 6,587 (\pm 0,390) \\ (n = 22; r = 0,773; s = 0,386; F = 14,076; Q^2_{cv} = 0,251; S_{press} = 0,527) \end{aligned} \quad \text{Equação 239}$$

$$\begin{aligned} \text{pIC}_{50} = & + 1,382 (\pm 1,030) \text{ CIC2} - 0,303 (\pm 0,180) \text{ RDF050v} + 6,432 (\pm 0,470) \\ (n = 22; r = 0,648; s = 0,464; F = 6,868; Q^2_{cv} = -0,620; S_{press} = 0,774) \end{aligned} \quad \text{Equação 240}$$

Série E (n=22), 3 descritores

$$\text{pIC}_{50} = -0,235 (\pm 0,089) \text{ SP13} + 0,140 (\pm 0,110) \text{ RDF050v} - 0,210 (\pm 0,130) \text{ RDF065p} + 7,980 (\pm 0,490)$$

$$(n = 22; r = 0,920; s = 0,245; F = 33,056; Q^2_{cv} = 0,764; s_{press} = 0,304) \quad \text{Equação 241}$$

$$\text{pIC}_{50} = +0,434 (\pm 0,400) \text{ CIC2} - 0,190 (\pm 0,086) \text{ SP13} - 0,169 (\pm 0,130) \text{ RDF065p} + 7,809 (\pm 0,500)$$

$$(n = 22; r = 0,913; s = 0,255; F = 30,012; Q^2_{cv} = 0,772; s_{press} = 0,299) \quad \text{Equação 242}$$

$$\text{pIC}_{50} = +0,297 (\pm 0,290) \text{ Jhetv} - 0,151 (\pm 0,095) \text{ SP13} - 0,154 (\pm 0,130) \text{ RDF065p} + 7,085 (\pm 0,870)$$

$$(n = 22; r = 0,911; s = 0,258; F = 29,311; Q^2_{cv} = 0,745; s_{press} = 0,315) \quad \text{Equação 243}$$

Série E (n=22), 4 descritores

$$\text{pIC}_{50} = +0,281 (\pm 0,240) \text{ Jhetv} - 0,195 (\pm 0,086) \text{ SP13} + 0,135 (\pm 0,098) \text{ RDF050v} - 0,204 (\pm 0,110)$$

$$\text{RDF065p} + 7,280 (\pm 0,740)$$

$$(n = 22; r = 0,942; s = 0,217; F = 33,209; Q^2_{cv} = 0,812; s_{press} = 0,279) \quad \text{Equação 244}$$

A partir das análises de *QSAR*, foram obtidos n=205 modelos (Equações 39-244) para a predição da atividade biológica. O programa BILIN calcula os valores das contribuições dos parâmetros estatísticos (r , s , F , Q^2_{cv} e s_{press}) eliminando automaticamente as equações não significativas, ou seja, as equações cujos valores dos desvios são maiores do que os correspondentes valores das contribuições dos parâmetros físico-químicos. Todas as 205 equações apresentadas são significativas. Entretanto, apenas a análise dos parâmetros estatísticos dos modelos gerados como, por exemplo, os valores de r e de Q^2_{cv} , não irá assegurar que o modelo tenha um bom poder de predição (Golbraikh & Tropsha 2002; Todeschini *et al.* 2004).

Desta forma, utilizou-se a regra do Q^2 assintótico. Como descrito anteriormente, na Introdução, através desta regra é possível realizar a validação das equações obtidas com o intuito de se verificar quais as equações realmente possuem poder de predição. As tabelas, a seguir, mostram os valores de Q^2 assintótico calculados, respectivamente, para as séries de 40, 39, 30, 23 e 22 compostos (Séries A – E).

Tabela 11. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de *QSAR* para a série A (n=40 compostos).

Equação Número	Valores de Q^2 assintótico	Equação Número	Valores de Q^2 assintótico
Equação 39	0,060	Equação 63	-0,197
Equação 40	0,044	Equação 64	0,006
Equação 41	0,018	Equação 65	0,001
Equação 42	0,031	Equação 66	0,018
Equação 43	0,013	Equação 67	0,007
Equação 44	0,032	Equação 68	0,057
Equação 45	-0,160	Equação 69	-0,098
Equação 46	0,062	Equação 70	0,060
Equação 47	-0,017	Equação 71	-0,075
Equação 48	-0,069	Equação 72	0,076
Equação 49	0,031	Equação 73	-0,091
Equação 50	0,072	Equação 74	-0,125
Equação 51	0,084	Equação 75	0,053
Equação 52	-0,170	Equação 76	0,020
Equação 53	0,076	Equação 77	-0,055
Equação 54	0,058	Equação 78	0,048
Equação 55	0,031	Equação 79	0,030
Equação 56	-0,070	Equação 80	0,063
Equação 57	0,075	Equação 81	0,004
Equação 58	-0,016	Equação 82	-0,031
Equação 59	0,006	Equação 83	-0,115
Equação 60	0,030	Equação 84	-0,066
Equação 61	-0,025	Equação 85	0,058
Equação 62	0,048		

Tabela 12. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de *QSAR* para a série B (n=39 compostos).

Equação Número	Valores de Q^2 assintótico	Equação Número	Valores de Q^2 assintótico
Equação 86	-0,105	Equação 109	-0,010
Equação 87	0,048	Equação 110	-0,024
Equação 88	0,018	Equação 111	-0,013
Equação 89	0,028	Equação 112	0,027
Equação 90	-0,017	Equação 113	0,040
Equação 91	-0,093	Equação 114	-0,046
Equação 92	-0,022	Equação 115	-0,054
Equação 93	-0,042	Equação 116	0,013
Equação 94	-0,051	Equação 117	0,011
Equação 95	0,026	Equação 118	0,014
Equação 96	-0,032	Equação 119	-0,014
Equação 97	-0,072	Equação 120	0,003
Equação 98	-0,044	Equação 121	0,012
Equação 99	-0,076	Equação 122	0,015
Equação 100	-0,069	Equação 123	0,035
Equação 101	-0,111	Equação 124	0,053
Equação 102	0,021	Equação 125	0,036
Equação 103	-0,003	Equação 126	0,030
Equação 104	-0,013	Equação 127	0,060
Equação 105	-0,017	Equação 128	-0,024
Equação 106	0,004	Equação 129	0,012
Equação 107	-0,010	Equação 130	-0,026
Equação 108	-0,106	Equação 131	0,009

Tabela 13. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de *QSAR* para a série C (n=30 compostos).

Equação Número	Valores de Q^2 assintótico	Equação Número	Valores de Q^2 assintótico
Equação 132	0,060	Equação 157	0,067
Equação 133	0,040	Equação 158	0,077
Equação 134	0,055	Equação 159	0,069
Equação 135	0,046	Equação 160	0,068
Equação 136	-0,070	Equação 161	0,083
Equação 137	0,062	Equação 162	0,049
Equação 138	0,070	Equação 163	0,076
Equação 139	0,052	Equação 164	0,041
Equação 140	0,104	Equação 165	0,076
Equação 141	0,084	Equação 166	0,081
Equação 142	0,009	Equação 167	0,055
Equação 143	0,051	Equação 168	0,077
Equação 144	0,054	Equação 169	0,101
Equação 145	0,083	Equação 170	0,102
Equação 146	0,058	Equação 171	0,028
Equação 147	0,080	Equação 172	0,036
Equação 148	0,071	Equação 173	0,066
Equação 149	0,058	Equação 174	0,088
Equação 150	0,094	Equação 175	0,058
Equação 151	0,065	Equação 176	0,058
Equação 152	0,059	Equação 177	0,064
Equação 153	0,036	Equação 178	0,083
Equação 154	0,043	Equação 179	0,103
Equação 155	0,044	Equação 180	0,070
Equação 156	0,070	Equação 181	0,002

Tabela 14. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de *QSAR* para a série D (n=23 compostos).

Equação Número	Valores de Q^2 assintótico	Equação Número	Valores de Q^2 assintótico
Equação 182	-0,069	Equação 207	0,040
Equação 183	0,038	Equação 208	0,039
Equação 184	0,047	Equação 209	0,027
Equação 185	0,060	Equação 210	0,039
Equação 186	0,036	Equação 211	0,041
Equação 187	0,060	Equação 212	0,003
Equação 188	0,050	Equação 213	0,016
Equação 189	-0,039	Equação 214	0,016
Equação 190	0,039	Equação 215	0,027
Equação 191	0,090	Equação 216	0,015
Equação 192	0,042	Equação 217	0,025
Equação 193	0,041	Equação 218	-0,035
Equação 194	0,038	Equação 219	0,032
Equação 195	0,011	Equação 220	0,026
Equação 196	0,037	Equação 221	0,010
Equação 197	-0,084	Equação 222	0,018
Equação 198	-0,021	Equação 223	0,009
Equação 199	0,014	Equação 224	0,025
Equação 200	0,000	Equação 225	-0,052
Equação 201	0,040	Equação 226	0,021
Equação 202	0,028	Equação 227	0,025
Equação 203	0,028	Equação 228	0,010
Equação 204	-0,022	Equação 229	0,023
Equação 205	0,016	Equação 230	-0,045
Equação 206	0,011	Equação 231	-0,027

Tabela 15. Valores de Q^2 assintótico correspondentes às equações de regressão calculados pelas análises de *QSAR* para a série E (n=22 compostos).

Equação Número	Valores de Q^2 assintótico	Equação Número	Valores de Q^2 assintótico
Equação 232	0,032	Equação 239	-0,124
Equação 233	0,069	Equação 240	-0,719
Equação 234	0,079	Equação 241	0,044
Equação 235	0,083	Equação 242	0,076
Equação 236	0,139	Equação 243	0,056
Equação 237	0,039	Equação 244	0,056
Equação 238	0,050		

Além dos valores de Q^2 assintótico calculados para cada modelo de equação obtido através das análises de *QSAR* clássico, foram observados os valores das correlações existentes entre os descritores selecionados através das análises *PLS* para cada série de compostos. Uma alta correlação entre os descritores utilizados acarreta em uma sobreestimação dos modelos gerados. Neste trabalho, nas equações de regressão (*QSAR*) considerou-se não significativas as intercorrelações de descritores com até o valor máximo de 0,6. Os valores dos coeficientes de correlação entre os descritores incluídos nos modelos *QSAR*, para cada série de compostos, estão apresentados nas tabelas de 15 a 19, respectivamente.

Tabela 18. Matriz de intercorrelação (r) dos descritores selecionados através da análise *PLS* e, incluídos nos modelos *QSAR*, para a série contendo 30 compostos (série C).

	J	JhetZ	MAXDP	CIC2	BEHv6	GATS	2v	GATS	3e	GATS	4e	GATS	5p	RDF	125u	RDF	065m	RDF	075v	RDF	020v	RDF	Mor	C-	H-	H-
	1,00	0,15	0,00	0,13	0,01	0,13	0,02	0,02	0,02	0,03	0,18	0,12	0,01	0,12	0,00	0,18	0,07	0,02	0,03	0,01	0,18	0,07	0,02	0,03	0,01	
JhetZ		1,00	0,02	0,00	0,03	0,07	0,04	0,00	0,11	0,40	0,01	0,08	0,10	0,02	0,40	0,11	0,02	0,08	0,00	0,40	0,11	0,02	0,08	0,00		
MAXDP			1,00	0,06	0,02	0,11	0,00	0,31	0,00	0,12	0,20	0,02	0,00	0,06	0,15	0,35	0,04	0,01	0,07	0,15	0,35	0,04	0,01	0,07		
CIC2				1,00	0,02	0,00	0,06	0,01	0,02	0,05	0,03	0,02	0,14	0,03	0,03	0,05	0,01	0,02	0,02	0,03	0,05	0,01	0,02	0,02		
BEHv6					1,00	0,00	0,06	0,00	0,09	0,13	0,19	0,05	0,15	0,39	0,13	0,33	0,00	0,01	0,07	0,13	0,33	0,00	0,01	0,07		
GATS2v						1,00	0,06	0,09	0,08	0,03	0,01	0,00	0,03	0,01	0,02	0,10	0,06	0,01	0,01	0,02	0,10	0,06	0,01	0,01		
GATS3e							1,00	0,00	0,06	0,09	0,00	0,02	0,04	0,02	0,08	0,01	0,27	0,07	0,04	0,08	0,01	0,27	0,07	0,04		
GATS4e								1,00	0,00	0,03	0,00	0,02	0,00	0,00	0,03	0,13	0,00	0,07	0,04	0,03	0,13	0,00	0,07	0,04		
GATS5p									1,00	0,03	0,07	0,00	0,06	0,15	0,04	0,01	0,06	0,01	0,14	0,04	0,01	0,06	0,01	0,14		
RDF125u										1,00	0,03	0,07	0,00	0,43	0,35	0,29	0,09	0,01	0,06	0,98	0,29	0,09	0,01	0,06		
RDF065m											1,00	0,03	0,05	0,06	0,41	0,03	0,09	0,05	0,10	0,03	0,09	0,05	0,10	0,00		
RDF020v												1,00	1,00	0,00	0,00	0,06	0,00	0,05	0,03	0,06	0,00	0,00	0,05	0,03		
RDF075v													1,00	1,00	0,53	0,39	0,06	0,08	0,03	0,39	0,06	0,08	0,03	0,26		
RDF065e														1,00	1,00	0,33	0,11	0,11	0,02	0,33	0,11	0,11	0,02	0,28		
RDF125e																1,00	0,30	0,08	0,01	1,00	0,30	0,08	0,01	0,04		
Mor07m																	1,00	0,02	0,00	1,00	1,00	0,02	0,00	0,01		
C-025																		1,00	0,37	0,04						
H-046																			1,00	0,05						
H-052																				1,00						

Tabela 19. Matriz de intercorrelação (r) dos descritores selecionados através da análise *PLS* e, incluídos nos modelos *QSAR*, para a série contendo 23 compostos (série D).

	Jhetv	MAXDP	SEige	SP13	RDF065m	RDF115m	RDF060v	RDF115v	Mor07v	H2e	nRORPh
Jhetv	1,00	0,19	0,22	0,20	0,03	0,10	0,19	0,11	0,49	0,28	0,19
MAXDP		1,00	0,55	0,29	0,34	0,06	0,10	0,08	0,36	0,20	0,04
SEige			1,00	0,39	0,38	0,14	0,20	0,16	0,71	0,41	0,02
SP13				1,00	0,21	0,43	0,75	0,70	0,44	0,38	0,03
RDF065m					1,00	0,07	0,03	0,12	0,17	0,13	0,00
RDF115m						1,00	0,57	0,78	0,20	0,08	0,01
RDF060v							1,00	0,63	0,30	0,24	0,07
RDF115v								1,00	0,23	0,22	0,01
Mor07v									1,00	0,58	0,05
H2e										1,00	0,04
nRORPh											1,00

Tabela 20. Matriz de intercorrelação (r) dos descritores selecionados através da análise *PLS* e, incluídos nos modelos *QSAR*, para a série contendo 22 compostos (série E).

	Jhetv	CIC2	SP13	RDF050v	RDF065p	Mor07p
Jhetv	1,00	0,01	0,30	0,11	0,15	0,36
CIC2		1,00	0,00	0,19	0,01	0,02
SP13			1,00	0,41	0,47	0,48
RDF050v				1,00	0,38	0,25
RDF065p					1,00	0,54
Mor07p						1,00

Para avaliar o poder de predição dos modelos obtidos através das análises de *QSAR* clássico foram observados os valores de Q^2 assintótico (Q^2_{ass}), calculados para cada modelo.

Considerando que o número de equações obtidas, com bons valores para o Q^2 assintótico, foi muito grande, utilizou-se o valor limiar simples ($\delta Q = 0,005$) como valor de corte. Desta forma, foram analisadas apenas as equações com valores de Q^2 assintótico maiores que o valor estipulado. Do número total de 205 equações obtidas nas análises de *QSAR* clássico, restaram 145 equações que possuem os valores de Q^2 assintótico iguais ou superiores ao valor limiar de 0,005.

Para facilitar as análises dos modelos, primeiramente foram averiguadas as séries menores. Desta forma, são apresentados os resultados obtidos para as séries E, D, C, B e A, respectivamente.

Análises realizadas para a série de 22 compostos (série E, n=22)

Para a série de 22 compostos (Série E), considerando-se a contribuição de 1 descritor para o modelo *QSAR*, a equação 236 apresentou o maior valor de Q^2 assintótico ($Q^2_{ass}=0,139$) entre todas as equações de regressão obtidas. O modelo *QSAR* expresso por esta equação foi obtido com a contribuição do descritor RDF050v, ponderado pelo volume de van der Waals, explicando 42,5% do valor da variância da atividade inibitória da RNR de células tumorais. Por definição, a função de distribuição radial (RDF) de um grupo de N átomos pode ser interpretada como sendo a distribuição da probabilidade de se encontrar um átomo em um volume esférico de raio r . Para este modelo obtido, a presença do descritor RDF050v indica que a distância de 5,0 Å entre dois átomos não adjacentes na molécula deve ser importante para a atividade biológica. Para a série E, observa-se um aumento no valor da atividade biológica quando o átomo de nitrogênio, do fragmento da tiossemicarbazona, indicado na figura 30, estiver a uma distância de 5,0 Å do carbono 5, do anel piridínico. O descritor RDF050v apresenta correlação direta com os parâmetros estruturais relativos a conformação.

Figura 30. Ilustração da estrutura do composto mais ativo da série E (composto 14) e o valor da distância ($\sim 5,0\text{\AA}$) entre os átomos de nitrogênio do fragmento da tiossemicarbazona e do carbono 5 do anel piridínico.

Considerando-se Q^2_{ass} como critério, a segunda equação que apresentou o maior valor de Q^2_{ass} foi a equação 235 ($Q^2_{ass} = 0,083$). Esta equação explica 64,7% do valor da variância da atividade biológica. Esta equação foi obtida apenas com a contribuição do descritor Jhetv, ponderado pelo volume de van der Waals. O modelo expresso pela equação 235 mostra que o aumento no valor de Jhetv proporciona um aumento no valor da atividade biológica. Por definição, o valor de Jhetv é inversamente proporcional à raiz quadrada dos valores de volume. Conseqüentemente, a introdução de grupos substituintes relativamente volumosos diminui o valor da atividade biológica dos compostos.

O gráfico apresentado na figura 31 ilustra a variação dos valores do descritor Jhetv em função dos valores da atividade biológica (valores de pIC_{50}).

Figura 31. Gráfico da variação dos valores do descritor Jhetv, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 22 compostos (Série E), selecionados da literatura (French & Blanz-Jr. 1974).

Ainda mais, observa-se que a equação 234 apresenta valor de $Q^2_{ass} = 0,079$ e que o modelo é descrito apenas pela contribuição do descritor Mor07p. Este explica 70,6% do valor da variância da atividade inibitória da RNR. O descritor 3D-MoRSE (*Molecule Representation of Structure based on Electron diffraction*) expressa características estruturais importantes como, por exemplo, a massa e a quantidade de ramificações presentes na estrutura. Desta forma, a presença do descritor Mor07p, ponderado pela polarizabilidade, no modelo *QSAR*, indica de modo análogo ao observado para o descritor RDF50v, a importância do fator estérico para a atividade biológica dos compostos da série E. O gráfico apresentado na figura 32 ilustra a variação entre os valores de Mor07p e os valores observados da atividade biológica. A interpretação deste modelo indica que o aumento nos valores de Mor07p acarretará uma diminuição nos valores da atividade biológica dos compostos. Ou seja, a introdução de grupos substituintes relativamente volumosos diminuirá os valores da atividade dos compostos. Esta observação pode ser constatada ao se analisar, por exemplo, os seguintes compostos: 34 (5-OOCCH₃, $pIC_{50}=5,44$), 35 (5-OOCC₂H₅, $pIC_{50}=5,28$) e 36 (5-OOCC₃H₇, $pIC_{50}=5,17$).

Figura 32. Gráfico da variação dos valores do descritor Mor07p, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 22 compostos (Série E), selecionados da literatura (French & Blanz-Jr. 1974).

Observa-se que a equação 242 ($Q^2_{ass} = 0,076$) explica 91,3% da variância da atividade biológica. E, o modelo *QSAR* correspondente é descrito pelas contribuições dos descritores CIC2, SP13 e RDF065p, respectivamente. Por definição, o descritor CIC2 considera os átomos de hidrogênio nas moléculas. Valores iguais a ordem zero representam grupos de átomos isolados em classes equivalentes e, iguais a ordem 1 representam pares de átomos ligados covalentemente, agrupados em classes equivalentes. O valor da contribuição isolada deste descritor para a atividade biológica dos compostos não demonstrou ser significativa, como observado na figura 33.

Figura 33. Gráfico da variação dos valores do descritor CIC2, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 22 compostos (Série E), selecionados da literatura (French & Blanz-Jr. 1974).

Por outro lado, o descritor RDF065p descreve a contribuição estérica para o modelo. Esta função de distribuição radial, ponderada pela polarizabilidade, indica que a distância entre dois átomos no valor de 6,5 Å é importante para a atividade biológica dos compostos. Para esta série, se os átomos de carbono da tiocarbonila, do fragmento da tiossemicarbazona, e do carbono 6, do anel piridínico, de um composto da série estiverem a uma distância de 6,5Å, pode-se verificar um aumento no valor da atividade biológica para este composto, como pode ser observado na figura 34.

Figura 34. Ilustração da estrutura do composto mais ativo da série E (composto 14) e o valor da distância ($\sim 6,5\text{\AA}$) entre os átomos de carbono da tiocarbonila e o carbono 6 do anel piridínico.

Adicionalmente, o descritor SP13 expressa o perfil de similaridade entre as formas das moléculas. Quanto maior a similaridade entre duas moléculas, menor será o valor de S, enquanto que quanto menor a similaridade entre as duas moléculas, maior será o valor de S. Para a série E, observa-se na figura 35 que os valores da contribuição do descritor SP13 para o modelo *QSAR*, expresso pela equação 242, são baixos para os compostos mais ativos e altos para os compostos menos ativos. Estes resultados indicam que os compostos ativos são estruturalmente similares entre si enquanto que os inativos não apresentam similaridade estrutural com os ativos.

Figura 35. Gráfico da variação dos valores do descritor SP13, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 22 compostos (Série E), selecionados da literatura (French & Blanz-Jr. 1974).

Na figura 33, verifica-se que o descritor CIC2, quando empregado isoladamente, não contribui significativamente para o modelo *QSAR*. Adicionalmente, observando-se os valores dos parâmetros estatísticos das equações 242 ($r=0,913$ e $F=30,012$) e 237 ($r=0,886$ e $F=34,706$), expressas respectivamente por três (CIC2, SP13 e RDF065p) e por dois (SP13 e RDF065p) descritores, pode-se considerá-las equivalentes. Desta forma, pode-se substituir o modelo *QSAR* expresso pela equação 242 pelo da equação 237, sem perda de informação.

A figura 36 mostra o gráfico dos valores da atividade biológica observados *versus* os calculados através do modelo *QSAR* expresso pela equação 237.

$$r = 0,886, SD = 0,275, n = 22$$

Figura 36. Gráfico dos valores de pIC_{50} observados *versus* calculados para a série de 22 compostos (Série E) inibidores da RNR, obtido utilizando-se o modelo expresso pela equação 237.

A equação 233 possui o valor de $Q^2_{ass} = 0,069$ e 76,3% da variância da atividade biológica foi explicada. Pode-se observar que apenas o descritor RDF065p contribui para o modelo, indicando novamente a importância da contribuição estérica para a atividade dos compostos.

Em resumo, para esta série contendo 22 compostos, o fator estérico foi o que mais contribuiu para a atividade biológica dos compostos. Sugerindo-se que os compostos devem possuir uma certa distância entre os átomos do anel piridínico e os átomos do fragmento da tiossemicarbazona e que os grupos substituintes no anel não devem ser relativamente volumosos para que estes apresentem altos valores de atividade biológica.

Análises realizadas para a série de 23 compostos (série D, n=23)

Para a série contendo 23 compostos (Série D), considerando-se a contribuição de 1 descritor para o modelo *QSAR*, as equações 191 ($Q^2_{ass} = 0,090$); 185 ($Q^2_{ass} = 0,060$); 187 ($Q^2_{ass} = 0,060$) e 188 ($Q^2_{ass} = 0,050$) foram as que apresentaram os maiores valores para o Q^2_{ass} , entre todas as equações de regressão obtidas.

O modelo *QSAR* expresso pela equação 191, explica 61,8% da variância da atividade biológica na inibição da enzima ribonucleotídeo redutase de células tumorais e apresenta uma contribuição significativa do descritor Jhetv. Este descritor, como discutido anteriormente, indica a importância do parâmetro estérico para a atividade biológica. Por outro lado, o modelo expresso pela equação 185, que explica 74,8% da variância da atividade biológica, indica uma contribuição significativa do descritor RDF060v, ponderado pelo volume de van der Waals. Este descritor mostra novamente a importância da distância entre dois átomos para a atividade biológica, para esta série de compostos. Sugere-se portanto, que o átomo de

nitrogênio 4' do fragmento da tiossemicarbazona, deva estar a uma distância de 6,0 Å do átomo de carbono 2, do anel piridínico, para que os compostos tenham uma alta atividade inibitória frente a RNR, como demonstrado na figura 37.

Figura 37. Ilustração da estrutura do composto mais ativo da série D (composto 14) e o valor da distância ($\sim 6,0\text{\AA}$) entre os átomos de nitrogênio 4', do fragmento da tiossemicarbazona, e do carbono 2 do anel piridínico.

O modelo expresso pela equação 187 explica 71,6% da variância da atividade biológica e observa-se uma contribuição significativa do descritor MAXDP. Por definição, este descritor é um índice de estado eletrotopológico baseado nos gráficos constantes de cada átomo presente na molécula, sendo considerado um descritor molecular local. O descritor MAXDP expressa a eletrofilicidade do composto, e, verifica-se que quanto menor o valor do MAXDP maior será a atividade biológica do composto, como ilustrado na figura 38. O valor deste descritor indica que a presença de grupos substituintes polares na estrutura diminui o valor da atividade biológica do composto como, pode ser observado, por exemplo, para os compostos 34 ($5\text{-OOCCH}_2\text{OCH}_3$, $\Delta\text{pIC}_{50}=5,30$) e 35 ($5\text{-OOCCH}_2\text{OC}_2\text{H}_5$, $\Delta\text{pIC}_{50}=5,25$).

Figura 38. Gráfico da variação dos valores do descritor MAXDP, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 23 compostos (série D), selecionados da literatura (French & Blanz-Jr. 1974).

O modelo expresso pela equação 188 explica 70,8% do valor da variância da atividade biológica, observando-se uma contribuição significativa do descritor H2e. Este descritor é do tipo *GETAWAY* (*Geometry, Topology and Atom Weights Assembly*) e compara a geometria molecular tridimensional obtida através da matriz de influência molecular com a ligação dos átomos através da topologia molecular. Para o descritor H2e a ponderação é realizada através da eletronegatividade dos átomos com incremento 2. Este descritor também

pode ser utilizado para avaliar a similaridade/dissimilaridade entre padrões moleculares. Neste modelo, os compostos mais ativos possuem um valor para H2e, *i.e.* 1, e os compostos menos ativos apresentam outro valor, *i.e.* 2, como pode ser observado na figura 39.

Figura 39. Gráfico da variação dos valores do descritor H2e, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC₅₀), para a série de 23 compostos (série D), selecionados da literatura (French & Blanz-Jr. 1974).

Para os modelos *QSAR* gerados para a série contendo 23 compostos (Série D), considerando-se a contribuição de dois descritores para o modelo *QSAR*, as equações 192 ($Q^2_{ass} = 0,042$), 193 ($Q^2_{ass} = 0,041$), 201 ($Q^2_{ass} = 0,040$) e 194 ($Q^2_{ass} = 0,038$) apresentaram os maiores valores para o Q^2_{ass} entre todas as equações de regressão obtidas.

O modelo *QSAR* expresso pela equação 192 explica 92,9% da variância da inibição da ribonucleotídeo redutase de células tumorais e foi obtido através das contribuições dos descritores SEige e SP13. O descritor SEige é determinado pela soma dos autovalores da eletronegatividade ponderada pela matriz de distância. Este descritor descreve o parâmetro eletrostático associado às ramificações dos compostos. Para a série D, verifica-se que quanto menor o valor do descritor SEige maior será o valor da atividade biológica do composto, indicando que a presença de grupos volumosos diminui a atividade biológica dos compostos, figura 40.

Figura 40. Gráfico da variação dos valores do descritor SEige, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC₅₀), para a série de 23 compostos (série D), retirados da literatura (French & Blanz-Jr. 1974).

Os modelos *QSAR* expressos pelas equações 193, 201 e 194 explicam 92,6%, 90,5% e 92,5% da variância da atividade biológica, respectivamente. Nesta tese, a natureza e o significado dos descritores incluídos nestes modelos (SP13, Mor07v, RDF065m, RDF060v e MAXDP) já foram discutidos, observando-se para a série D uma contribuição significativa dos descritores relativos aos efeitos estérico e eletrostático para os valores de pIC_{50} .

Os gráficos (a), (b), (c) e (d) da figura 41 mostram os valores observados da atividade biológica dos compostos *versus* os calculados através dos modelos *QSAR* expressos pelas equações 192, 193, 201 e 194, respectivamente, observando-se correlações significativas.

Figura 41. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 23 compostos (Série D) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 192 (a), 193 (b), 201 (c) e 194 (d), respectivamente.

Para a série contendo 23 compostos (Série D), considerando-se a contribuição de 3 descritores para o modelo *QSAR*, as equações 211 ($Q^2_{ass} = 0,041$), 207 ($Q^2_{ass} = 0,040$), 208 ($Q^2_{ass} = 0,039$) e 210 ($Q^2_{ass} = 0,039$) foram as que apresentaram os maiores valores para o Q^2_{ass} entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam respectivamente 94,8%, 95,3%, 95,3% e 95,1% da variância da inibição de células tumorais.

Nesta tese, a natureza e o significado dos descritores incluídos nestes modelos (RDF065m, Mor07v, RDF060v, H2e, SEige e MAXDP) já foram discutidos, exceto o

descriptor RDF115v, que possui interpretação análoga ao do RDF065m. Este descritor da função de distribuição radial indica que a distância entre dois átomos no valor de 11,5Å é importante para a atividade biológica, figura 42. Sugere-se, portanto, que os compostos não devem possuir substituintes relativamente volumosos para que valores altos de pIC_{50} sejam observados.

Figura 42. Ilustração da estrutura do composto mais ativo da série D (composto 14) e o valor da distância ($\sim 11,5\text{\AA}$) entre os átomos de hidrogênio do nitrogênio 4', do fragmento da tiossemicarbazona, e do carbono 2 do grupo substituinte etila na posição 5 do anel piridínico

Os gráficos (a), (b), (c) e (d) da figura 43 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* expressos pelas equações 211, 207, 208 e 210, respectivamente, observando-se correlações significativas.

Figura 43. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 23 compostos (Série D) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 211 (a), 207 (b), 208 (c) e 210 (d), respectivamente.

Para a série contendo 23 compostos (Série D), considerando-se a contribuição de 4 descritores para o modelo *QSAR*, as equações 219 ($Q^2_{ass} = 0,032$), 220 ($Q^2_{ass} = 0,026$), 215 ($Q^2_{ass} = 0,027$) e 217 ($Q^2_{ass} = 0,025$) foram as que apresentaram os maiores valores para o Q^2_{ass} entre todas as equações de regressão obtidas. O modelo *QSAR* descrito pela equação 219 explica 96,5% da variância da atividade biológica dos compostos e, inclui os descritores MAXDP, RDF115v, H2e e nRORPh. Nesta tese, a natureza e o significado destes descritores já foram discutidos, exceto o descritor nRORPh que indica a contribuição do número de grupos éteres presentes nas moléculas para o modelo. Neste modelo, a presença deste grupo funcional diminui o valor da atividade biológica do composto.

Os modelos *QSAR* descritos pelas equações 220, 215 e 217 explicam 96,5%, 97,4% e 96,8%, respectivamente, do valor da variância da atividade biológica. A natureza e o significado dos descritores incluídos nestes modelos (MAXDP, RDF065m, RDF115v, Mor07v, Jhetv, RDF060v e H2e) já foram discutidos, observando-se uma contribuição significativa dos descritores relativos ao efeito estérico para a atividade biológica.

Figura 44. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 23 compostos (Série D) inibidores da RNR, obtidos utilizando-se os modelos das equações 219 (a), 220 (b), 215 (c) e 217 (d), respectivamente.

Os gráficos (a), (b), (c) e (d) da figura 44 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* expressos pelas equações 219, 220, 215 e 217, respectivamente.

Para a série D, considerando-se a contribuição de 5 descritores para o modelo *QSAR*, as equações 224 ($Q^2_{ass} = 0,025$), 227 ($Q^2_{ass} = 0,025$) e 229 ($Q^2_{ass} = 0,023$) foram as que apresentaram os maiores valores para o Q^2_{ass} entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam 97,9%, 97,5% e 96,9%, respectivamente, da variância da atividade biológica. Nesta tese, a natureza e o significado dos descritores incluídos nestes modelos (MAXDP, RDF065m, RDF115v, Mor07v, nRORPh, H2e, Jhetv e RDF115m) já foram discutidos. Observando-se os valores dos parâmetros estatísticos das equações 224 ($r=0,979$ e $F=77,782$) e 210 ($r=0,951$ e $F=60,221$) pode-se considerá-las equivalentes. Desta forma, sem perda de informação, pode-se substituir o modelo *QSAR* expresso pela equação 224 pelo da equação 210. Analogamente, a equação 227 ($r=0,975$ e $F=65,160$) pode ser substituída, sem perda de informação, pelo modelo *QSAR* expresso pela equação 219 ($r=0,965$ e $F=60,873$) obtido com a contribuição dos descritores MAXDP, RDF115v, H2e e nRORPh.

Os gráficos (a), (b) e (c) da figura 45 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* expressos pelas equações 224, 227 e 229, respectivamente, observando-se correlações significativas.

Figura 45. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 23 compostos (Série D) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 224 (a), 227 (b) e 229 (c), respectivamente.

Análises realizadas para a série de 30 compostos (série C, $n=30$)

Para a série contendo 30 compostos (Série C), considerando-se a contribuição de apenas um descritor para o modelo *QSAR*, as equações 140 ($Q^2_{ass} = 0,104$), 141 ($Q^2_{ass} = 0,084$), 138 ($Q^2_{ass} = 0,070$) e 137 ($Q^2_{ass} = 0,062$) foram as que apresentaram os maiores valores para o Q^2_{ass} . O modelo *QSAR* expresso pela equação 140, que apresenta contribuição significativa do

descriptor GATS4e, explica 41,8% da variância da atividade biológica. Por definição, o descriptor GATS4e é obtido através de uma autocorrelação espacial e mede a similaridade dos objetos existentes em uma dada região, o grau no qual um evento é correlacionado com ele mesmo no espaço, o nível de interdependência entre as variáveis, a natureza e a grandeza desta interdependência. No modelo *QSAR* obtido, este descriptor classifica os compostos inativos atribuindo um determinado valor, *i.e.* 1, e para os ativos um outro valor, *i.e.* 0, como apresentado na figura 46. Pode-se verificar que GATS4e não classifica satisfatoriamente todos os compostos da série.

Figura 46. Gráfico da variação dos valores do descriptor GATS4e, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 30 compostos (Série C), selecionados da literatura (French & Blanz-Jr. 1974).

O modelo *QSAR* expresso pela equação 141 explica 41,0% da variância da atividade biológica e apresenta uma contribuição significativa do descriptor BEHv6. O descriptor BEHv6 pertence ao grupo de descritores moleculares BCUT, definidos como autovalores da matriz modificada de conectividade, também denominada de matriz B de Burden (Todeschini & Consonni 2000). Estes descritores refletem aspectos relevantes da estrutura molecular e são úteis na avaliação da similaridade entre compostos. Para o modelo de *QSAR* obtido, os compostos inativos possuem um determinado valor, *i.e.* 2, e os compostos ativos possuem um outro valor, *i.e.* 1, figura 47. De modo análogo aos resultados obtidos com a contribuição do descriptor GATS4e, este descriptor não classifica corretamente todos os compostos desta série.

Figura 47. Gráfico da variação dos valores do descritor BEHv6, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 30 compostos (Série C), selecionados da literatura (French & Blanz-Jr. 1974).

O modelo *QSAR* expresso pela equação 138, obtido com a contribuição do descritor C-025, explica 48,7% da variância da atividade biológica. O descritor C-025 descreve a importância de se ter um átomo de carbono centrado em um fragmento do tipo R- -CR'- -R, onde R são átomos de carbono pertencentes a um sistema aromático e R' um grupo alquila. Para esta série, observa-se que grupos substituintes (*i.e.* metila ou etila) no anel piridínico podem contribuir para o aumento do valor de pIC_{50} como, por exemplo, os compostos 11 (3-CH₃; pIC_{50} =6,59), 12 (4-CH₃; pIC_{50} =6,57), 13 (5-CH₃; pIC_{50} =6,51) e 14 (5-CH₂H₅; pIC_{50} =6,66). O modelo *QSAR* expresso pela equação 137 explica 55,2% da variância do valor de pIC_{50} e foi gerado com a contribuição do descritor RDF065m. A natureza e o significado do descritor RDF065m já foi discutido anteriormente neste trabalho.

Para a série C, considerando a contribuição de dois descritores para o modelo *QSAR*, as equações 150 ($Q^2_{ass} = 0,094$), 145 ($Q^2_{ass} = 0,083$), 147 ($Q^2_{ass} = 0,080$) e 148 ($Q^2_{ass} = 0,071$) foram as que apresentaram os maiores valores de Q^2_{ass} entre todas as equações de regressão obtidas. Os modelos expressos por estas equações explicam respectivamente 75,1%, 77,7%, 77,3% e 76,4% da variância da atividade biológica. Neste trabalho, a natureza e o significado dos descritores incluídos nestes modelos (MAXDP, C-025, RDF065m, RDF125e, RDF125u, RDF065m, RDF075v) já foram discutidos, observando-se uma contribuição predominante dos descritores relativos aos efeitos estéricos.

Os gráficos (a), (b), (c) e (d) da figura 48 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR*, para a série C considerando-se a contribuição de 2 descritores, expressos pelas equações 150, 145, 147 e 148, respectivamente.

Para a série C, considerando-se a contribuição de 3 descritores para o modelo *QSAR*, as equações 161 ($Q^2_{ass} = 0,083$), 158 ($Q^2_{ass} = 0,077$), 156 ($Q^2_{ass} = 0,070$) e 159 ($Q^2_{ass} = 0,069$)

foram as que apresentaram os maiores valores de Q^2_{ass} , entre todas as equações de regressão obtidas. O modelo *QSAR* expresso pela equação 161, que explica 83,3% da variância da atividade biológica, foi obtido através das contribuições dos descritores MAXDP, RDF125u e RDF065m. O modelo *QSAR* expresso pela equação 158, que explica 83,6% da variância da atividade biológica, foi obtido através da contribuição dos descritores GATS3e, RDF065e e Mor07m. O descritor GATS3e, de forma análoga ao GATS4e classifica os compostos em ativos e inativos, figura 49.

Figura 48. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 30 compostos (Série C) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 150 (a), 145 (b), 147 (c) e 148 (d), respectivamente.

Figura 49. Gráfico da variação dos valores do descritor GATS3e, obtidos com o emprego do programa DRAGON em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 30 compostos (Série C), seleccionados da literatura (French & Blanz-Jr. 1974).

O modelo *QSAR* expresso pela equação 156, explica 84,1% da variância da atividade biológica e foi obtido com as contribuições dos descritores *Jhetz*, *MAXDP* e *RDF125e*, sendo este último ponderado pela eletronegatividade atômica de Sanderson (Sanderson 1976). O modelo *QSAR* expresso pela equação 159, obtido com a contribuição dos descritores *GATS4e*, *RDF065m* e *RDF125e*, explica 83,5% da variância da atividade biológica.

Os gráficos (a), (b), (c) e (d) da figura 50 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* expressos pelas equações 161, 158, 156 e 159, respectivamente, observando-se correlações significativas.

(a)
 $r = 0,833$, $SD = 0,315$, $n = 30$

(b)
 $r = 0,836$, $SD = 0,313$, $n = 30$

(c)
 $r = 0,841$, $SD = 0,308$, $n = 30$

(d)
 $r = 0,835$, $SD = 0,313$, $n = 30$

Figura 50. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 30 compostos (Série C) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 161 (a), 158 (b), 156 (c) e 159 (d), respectivamente.

Para a série C, considerando-se a contribuição de 4 descritores para o modelo *QSAR*, as equações 170 ($Q^2_{ass} = 0,102$), a 169 ($Q^2_{ass} = 0,101$), 166 ($Q^2_{ass} = 0,081$) e 168 ($Q^2_{ass} = 0,077$) foram as que apresentaram os maiores valores para o Q^2_{ass} entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam respectivamente 87,5%, 87,7%, 88,0% e 87,8% da variância da atividade biológica. Neste trabalho, a natureza e o significado

dos descritores incluídos nestes modelos (GATS4e, RDF125u, RDF065m, C-025, MAXDP, GATS3e, RDF065e, Mor07m e JhetZ) já foram discutidos.

Os gráficos (a), (b), (c) e (d) da figura 51 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* expressos pelas equações 170, 169, 166 e 168, respectivamente, observando-se correlações significativas.

Figura 51. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 30 compostos (Série C) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 170 (a), 169 (b), 166 (c) e 168 (d), respectivamente.

Para a série C, considerando-se a contribuição de cinco descritores para o modelo *QSAR*, as equações 179 ($Q^2_{ass} = 0,103$), 174 ($Q^2_{ass} = 0,088$), 178 ($Q^2_{ass} = 0,083$) e 180 ($Q^2_{ass} = 0,070$) foram as que apresentaram os maiores valores para o Q^2_{ass} , entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam respectivamente 90,5%, 91,2%, 90,6% e 90,4% da variância da atividade biológica. Neste trabalho, a natureza e o significado dos descritores incluídos nestes modelos (JhetZ, CIC2, RDF065m, RDF125e, Mor07m, RDF125u e GATS3e) já foram discutidos.

Os gráficos (a), (b), (c) e (d) da figura 52 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* obtidos com a contribuição de cinco descritores para a série C, observando-se correlações significativas.

Figura 52. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 30 compostos (Série C) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 179 (a), 174 (b), 178 (c) e 180 (d), respectivamente.

Análises realizadas para a série de 39 compostos (série B, n=39)

Para a série de 39 compostos (Série B), considerando-se a contribuição de apenas um descritor, as equações 87 ($Q^2_{ass} = 0,048$), 89 ($Q^2_{ass} = 0,028$) e 88 ($Q^2_{ass} = 0,018$) foram as que apresentaram valores de Q^2_{ass} maiores que o valor limiar. Os modelos *QSAR* explicam respectivamente 43,6%, 41,6% e 41,6% da variância da atividade biológica. Os descritores incluídos nestes modelos foram: o RDF050m, o GATS4e e o HATSp. Neste trabalho, a natureza e o significado dos descritores RDF050m e GATS4e já foram discutidos. Por definição, os descritores HATSp e H2e são do tipo GATEWAY, verificando a similaridade/dissimilaridade entre padrões moleculares. Na figura 53 observa-se que o descritor HATSp discrimina os compostos mais ativos dos menos ativos. Comparando-se estes resultados com os da figura 39, pode-se observar que o descritor HATSp classifica melhor os compostos, em relação ao descritor H2e.

Figura 53. Gráfico da variação dos valores do descritor HATSp, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 39 compostos (Série B), selecionados da literatura (French & Blanz-Jr. 1974).

Para a série B, considerando-se a contribuição de dois descritores, apenas a equação 95 ($Q^2_{ass} = 0,026$) apresentou um valor de Q^2_{ass} maior do que o valor limiar estipulado. Este modelo explica 62,7% da variância da atividade biológica e foi obtido com a contribuição dos descritores MATS8m e GATS4e. Por definição, o descritor MATS8m pertence à classe de descritores de autocorrelação 2D, assim como o GATS4e, medindo a similaridade dos objetos existentes em uma determinada região. Para esta série, o descritor MATS8m também classifica os compostos em ativos e inativos, figura 54. Ao contrário do descritor HATSp que considera vários valores para agrupar os compostos, o descritor MATS8m considera apenas dois valores, *i.e.* 1 e 0, respectivamente, o que torna o resultado da classificação dos compostos inferior em relação ao primeiro.

Figura 54. Gráfico da variação dos valores do descritor MATS8m, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 39 compostos (Série B), selecionados da literatura (French & Blanz-Jr. 1974).

Considerando que os dois descritores que contribuem para o modelo *QSAR*, expresso pela equação 95, apenas classificam os compostos em ativos ou inativos, o gráfico dos valores observados da atividade biológica *versus* os calculados através deste modelo também apresentará um perfil de classificação dos compostos, como pode ser observado na figura 55.

$$r = 0,627, SD = 0,759, n = 39$$

Figura 55. Gráfico dos valores de pIC_{50} observados *versus* calculados para a série de 39 compostos (Série B) inibidores da RNR, obtido utilizando-se o modelo expresso pela equação 95.

Para a série B, considerando-se a contribuição de três descritores para o modelo *QSAR*, apenas a equação 102 ($Q^2_{ass} = 0,021$) apresentou o valor de Q^2_{ass} maior que o valor limiar. Este modelo explica 80,4% da variância da atividade biológica e foi obtido com a contribuição dos descritores GATS3e, Mor08m e HATSp. O significado do descritor GATS3e é semelhante ao do descritor GATS4e enquanto que o significado do descritor Mor08m é semelhante ao do Mor07p, discutidos anteriormente. Os gráficos apresentados na figura 56 mostram a variação dos valores dos descritores GATS3e e Mor08m em função dos valores de pIC_{50} .

Figura 56. Gráficos da variação dos valores dos descritores, respectivamente, GATS3e (a) e Mor08m (b), obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 39 compostos (série B), retirados da literatura (French & Blanz-Jr. 1974).

O gráfico da figura 57 mostra os valores observados da atividade biológica *versus* os calculados através do modelo *QSAR* expresso pela equação 102.

$$r = 0,804, SD = 0,580, n = 39$$

Figura 57. Gráfico dos valores de pIC_{50} observados *versus* calculados para a série de 39 compostos (Série B) inibidores da RNR, obtido utilizando-se o modelo da equação 102.

Para a série B, considerando-se 4 descritores para o modelo *QSAR*, as equações 113 ($Q^2_{ass} = 0,040$), 112 ($Q^2_{ass} = 0,027$) 118 ($Q^2_{ass} = 0,014$) e 116 ($Q^2_{ass} = 0,013$) foram as que apresentaram os maiores valores para o Q^2_{ass} , entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam respectivamente 82,8%, 83,2%, 78,4% e 79,5% da variância da atividade biológica. Para o modelo *QSAR* expresso pela equação 113, verificou-se a contribuição dos descritores GATS3e, Mor08m, HATSp e H-052, sendo que este último é um descritor de fragmentos de átomos centrados. O descritor H-052 está correlacionado com a hidrofobicidade atômica e com a refratividade atômica, indicando a contribuição do efeito lipofílico e estérico para a atividade biológica (Viswanadhan *et al.* 1989). A natureza e o significado dos descritores GATS3e, Mor08m e HATSp já foram discutidos.

Os gráficos (a), (b), (c) e (d) da figura 58 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* expressos pelas equações 113, 112, 118 e 116, respectivamente, observando-se correlações significativas. Observando-se, entretanto, valores do desvio padrão altos ($\sim 0,60$).

Figura 58. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 39 compostos (Série B) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 113 (a), 112 (b), 118 (c) e 116 (d), respectivamente.

Para série B, considerando-se a contribuição de 5 descritores para o modelo *QSAR*, as equações 127 ($Q^2_{ass} = 0,060$), 124 ($Q^2_{ass} = 0,053$), 125 ($Q^2_{ass} = 0,036$) e 123 ($Q^2_{ass} = 0,035$) foram os que apresentaram os maiores valores para o Q^2_{ass} , entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam respectivamente 85,0%, 86,0%, 85,3% e 86,3% da variância da atividade biológica. Neste trabalho, a natureza e o significado dos descritores incluídos nestes modelos (GATS3e, RDF055m, RDF075p, Mor08m, HATSp, H-052 e GATS4e) já foram discutidos, exceto o descritor BEHv8, que contribui com o modelo *QSAR* expresso pela equação 123. Por definição, o descritor BEHv8 pertence à classe de descritores BCUT que calcula os maiores e os menores autovalores da matriz original de Burden e as três variantes introduzidas por Pearlamn (Pearlman & Smith 1998). As três variantes são, respectivamente: carga atômica, polarizabilidade atômica e propriedades doadoras e aceptoras de ligação de hidrogênio. O descritor BEHv8 indica o maior autovalor da matriz de Burden e expressa a importância da contribuição do parâmetro eletrônico para a atividade dos compostos.

Os gráficos (a), (b), (c) e (d) da figura 59 mostram os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* expressos pelas equações 127, 124, 125 e 123, respectivamente, observando-se correlações significativas. Observando-se, entretanto, valores do desvio padrão altos ($\sim 0,5$).

Figura 59. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 39 compostos (Série E) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 127 (a), 124 (b), 125 (c) e 123 (d), respectivamente.

Análises realizadas para a série de 40 compostos (série A, $n=40$)

Para as análises contendo 40 compostos (Série A), considerando-se a contribuição de apenas um descritor para o modelo *QSAR*, as equações 39 ($Q^2_{ass} = 0,060$), 40 ($Q^2_{ass} = 0,044$), 44 ($Q^2_{ass} = 0,032$) e 42 ($Q^2_{ass} = 0,031$) foram as que apresentaram os maiores valores para o Q^2_{ass} entre todas as equações de regressão obtidas. Estas equações explicam respectivamente 58,6%, 47,4%, 37,4% e 41,4% da variância da atividade biológica. Neste trabalho, a natureza e o significado dos descritores incluídos nestes modelos (RDF050m, GATS4e e MATS8m), já foram discutidos, exceto R1u. Por definição, o descritor R1u é um descritor *GETAWAY*, semelhante aos descritores HATSp e H2e e classifica por similaridade os compostos em

ativos e inativos. Assim como o descritor H2e, observa-se que o descritor R1u isoladamente não classifica satisfatoriamente os compostos, figura 60.

Figura 60. Gráfico da variação dos valores do descritor R1u, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC₅₀), para a série de 40 compostos (série A), retirados da literatura (French & Blanz-Jr. 1974).

Para os modelos *QSAR* gerados para a série A, considerando-se a contribuição de dois descritores para o modelo *QSAR*, as equações 51 ($Q^2_{ass} = 0,084$), 53 ($Q^2_{ass} = 0,076$), 50 ($Q^2_{ass} = 0,072$) e 46 ($Q^2_{ass} = 0,062$) apresentaram os maiores valores para o Q^2_{ass} , entre todas as equações de regressão obtidas. Os modelos explicam respectivamente 66,0%, 64,8%, 66,6% e 74,1% da variância da atividade biológica. A natureza e o significado dos descritores incluídos nestes modelos (BEHv8, R1u, Mor03m, GATS4e e RDF050m) já foram discutidos anteriormente. Os descritores R1u, Mor03m e GATS4e classificam os compostos em ativos e inativos de acordo com a similaridade existente entre eles. De forma análoga ao gráfico apresentado na figura 55, observa-se no gráfico da figura 61 um perfil de classificação dos compostos. Observa-se ainda que o composto 5 (2'-CH₃, pIC₅₀ observado=2,28) é um "outlier". Este fato pode ser explicado considerando-se as séries utilizadas para gerar modelos *QSAR* não possuem substituintes para esta posição. Desta forma, não fornecem informações suficientes sobre as características estruturais para a correta predição do valor de pIC₅₀ deste composto.

Para a série A, considerando-se a contribuição de 3 descritores para o modelo *QSAR*, as equações 57 ($Q^2_{ass} = 0,075$), 62 ($Q^2_{ass} = 0,048$) e 60 ($Q^2_{ass} = 0,030$) foram as que apresentaram os maiores valores para Q^2_{ass} , entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam 77,9%, 75,1% e 76,3%, respectivamente, da variância da atividade biológica. Neste trabalho, a natureza e o significado dos descritores incluídos nestes modelos (GATS4e, RDF050m, R1u, MATS8m, RDF070m e Mor08m) já foram discutidos. Os gráficos (a), (b) e (c) da figura 62 apresentam os valores de pIC₅₀ observados *versus* os calculados através dos modelos *QSAR* expressos pelas equações 57, 62 e 60, respectivamente.

Figura 61. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 40 compostos (Série A) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 51 (a), 53 (b), 50 (c) e 46 (d), respectivamente.

Figura 62. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 40 compostos (Série A) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 57 (a), 62 (b) e 60 (c), respectivamente.

Para a série A, considerando-se a contribuição de 4 descritores para o modelo *QSAR*, as equações 72 ($Q^2_{ass} = 0,076$), 70 ($Q^2_{ass} = 0,060$), 68 ($Q^2_{ass} = 0,057$) e 75 ($Q^2_{ass} = 0,053$) foram as que apresentaram os maiores valores para o Q^2_{ass} , entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam respectivamente 80,6%, 80,9%, 81,3% e

79,5% da variância da atividade biológica. Neste trabalho, a natureza e o significados dos descritores incluídos nestes modelos (GATS3e, GATS4e, RDF050m, R1u, MATS8m, RDF075p e HATSp) já foram discutidos. Os gráficos (a), (b), (c) e (d) da figura 63 apresentam os valores observados da atividade biológica *versus* os calculados através dos modelos *QSAR* expressos pelas equações 72, 70, 68 e 75, observando-se valores do desvio padrão altos 0,6.

Figura 63. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 40 compostos (Série A) inibidores da RNR, obtidos utilizando-se os modelos das equações 72 (a), 70 (b), 68 (c) e 75 (d), respectivamente.

Finalmente, para a série A, considerando-se a contribuição de 5 descritores para o modelo *QSAR*, as equações 80 ($Q^2_{ass} = 0,063$), 85 ($Q^2_{ass} = 0,058$), 78 ($Q^2_{ass} = 0,048$) e 79 ($Q^2_{ass} = 0,030$) foram as que apresentaram os maiores valores de Q^2_{ass} , entre todas as equações de regressão obtidas. Os modelos *QSAR* explicam respectivamente 84,7%, 83,7%, 85,7% e 85,3% da variância da atividade biológica. Com exceção do descritor IC2, que contribui para o modelo *QSAR* expresso pela equação 79, a natureza e o significado de todos os outros descritores (GATS3e, GATS4e, RDF075p, HATSp, R1U, MATS8m, RDF055m,

Mor08m e Mor03m) já foram discutidos anteriormente. Por definição, o IC2 é um descritor do *conteúdo de informação topológico*, assim como o descritor CIC2 que considera a vizinhança de incremento 2. Assim como os valores de CIC2, os valores de IC2 isolados não são significativos para os modelos de *QSAR*, pois não agrupa corretamente os compostos, com respeito a atividade, como observado na figura 64.

Figura 64. Gráfico da variação dos valores do descritor IC2, obtidos com o emprego do programa DRAGON, em função dos valores observados para a atividade biológica (pIC_{50}), para a série de 40 compostos (Série A), selecionados da literatura (French & Blanz-Jr. 1974).

Os gráficos (a), (b), (c) e (d) da figura 65 mostram os valores da atividade biológica observados *versus* os valores calculados através dos modelos *QSAR*, para a série A considerando-se a contribuição de 5 descritores, expressos pelas equações 80, 85, 78 e 79, respectivamente.

Figura 65. Gráficos dos valores de pIC_{50} observados *versus* calculados para a série de 40 compostos (Série A) inibidores da RNR, obtidos utilizando-se os modelos expressos pelas equações 80 (a), 85 (b), 78 (c) e 79 (d), respectivamente.

Em resumo, as análises de *QSAR* clássico aplicadas à série de derivados da α -(N)-piridino carboxaldeído tiossemicarbazonas substituídas, com atividade inibitória frente à ribonucleotídeo redutase (RNR) de células H.Ep.-2, resultaram em:

- A contribuição do fator estérico é importante para a atividade biológica na inibição da RNR de células tumorais. Através do descritor RDF, função de distribuição radial, foi verificado que deve haver uma distância considerada “ideal” entre alguns átomos nos compostos. Além disto, os descritores Jhetv, SEige e Mor07p, indicam que a presença de substituintes relativamente volumosos pode diminuir o valor da atividade biológica dos compostos.
- Alguns descritores como, por exemplo, o SP13, GATS4e, H2e e HATSp conseguem discriminar os compostos em ativos e em inativos, através da similaridade estrutural.
- A contribuição do fator eletrostático, representado pela contribuição dos descritores MAXDP e BEHv8, também demonstrou ser importante para a inibição da RNR. O descritor

MAXDP, por exemplo, indicou que a presença de compostos relativamente polares pode diminuir o valor da atividade biológica dos compostos.

- A contribuição do fator hidrofóbico, representado pelo descritor H-052, também demonstrou ser importante para os modelos *QSAR* gerados.

As análises realizadas previamente por Gupta e colaboradores (Gupta *et al.* 1984) para a série de derivados da α -(N)-piridino carboxaldeído tiossemicarbazona substituídos, com atividade inibitória frente a ribonucleotídeo redutase (RNR) de células H.Ep.-2, mostraram que a atividade biológica é descrita apenas pelo volume de van der Waals, determinado para o grupo substituinte sendo que a lipofilicidade, expressa por logP, foram calculados pelo método descrito por Hansch e Leo (Hansch & Leo 1979). Os resultados obtidos por Gupta e colaboradores mostraram que a contribuição do fator lipofílico não foi significativa para os modelos obtidos.

De modo análogo, neste estudo, ao se considerar somente as séries de 23 e 22 compostos, com substituintes apenas na posição 5 do anel piridínico, semelhante aos estudos da literatura (Gupta *et al.* 1984), foi novamente constatada a importância do parâmetro estérico para a atividade dos compostos. O aumento do volume dos substituintes, expressos pelos descritores Mor07p e Jhetv, irá acarretar uma diminuição nos valores da atividade biológica. Adicionalmente, foi constatado que deve haver uma distância “ideal” entre os átomos dos compostos para que a atividade inibitória da RNR seja alta. Neste estudo, além da contribuição estérica foi constatada a importância do fator eletrostático, representados pelos descritores MAXDP e SEige. Observou-se que a presença de grupos relativamente polares desfavorece a atividade biológica dos compostos.

3.7 Estudos de *CoMFA*

Os resultados obtidos através da metodologia *CoMFA* aplicada para as séries A (n=40), B (n=39), C (n=30), D (n=23) e E (n=22) estão apresentados nas tabelas e nos gráficos das figuras apresentados a seguir. Os gráficos foram construídos com os valores de pIC_{50} preditos através dos modelos *CoMFA*, gerados para cada uma das cinco séries (A-E), para cada um dos três grupos de treinamento e teste (I-III) e considerando-se os três alinhamentos utilizados (R_A , R_B e S_C). Os valores de pIC_{50} preditos estão no APÊNDICE, tabelas de I a XV.

Tabela 21. Resultados obtidos através do emprego da metodologia *CoMFA*, para o grupo de treinamento, empregando-se a série de 40 derivados (série A) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

		Modelo 1 (grupo treinamento IA)			Modelo 2 (grupo treinamento IIA)			Modelo 3 (grupo treinamento IIIA)		
		Rígido		SEAL	Rígido		SEAL	Rígido		SEAL
		R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
Número de		5	2	2	2	5	5	4	2	2
Componentes										
Q^2_{cv}		0,480	0,478	0,431	0,335	0,566	0,617	0,357	0,403	0,462
s_{press}		0,844	0,797	0,832	0,781	0,669	0,628	0,938	0,869	0,825
r^2		0,977	0,895	0,877	0,847	0,965	0,979	0,983	0,893	0,884
s		0,176	0,357	0,387	0,375	0,190	0,146	0,152	0,368	0,383
Teste F		208,103	115,383	96,056	74,555	132,099	225,867	361,939	112,642	103,326
Contribuições:										
Estérica		0,635	0,670	0,679	0,636	0,635	0,629	0,646	0,654	0,654
Eletrostática		0,365	0,330	0,321	0,364	0,365	0,371	0,354	0,346	0,346

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 66. Gráficos dos valores de pIC₅₀ calculados *versus* observados, sendo que os pIC₅₀ calculados foram obtidos através dos modelos gerados pelo método *CoMFA*, utilizando-se três diferentes modos de alinhamento (R_A, R_B e S_C): modelo 1 (1R_A, 1R_B e 1S_C) (grupo de treinamento IA), modelo 2 (2R_A, 2R_B e 2S_C) (grupo de treinamento IIA) e, modelo 3 (3R_A, 3R_B e 3S_C) (grupo de treinamento IIIA) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 40 compostos (série A).

Nestes gráficos, as letras R_A, R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Tabela 22. Resultados obtidos através do emprego da metodologia *CoMFA*, para o grupo de treinamento, empregando-se a série de 39 derivados (série B) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

		Modelo 1 (grupo treinamento IB)			Modelo 2 (grupo treinamento IIB)			Modelo 3 (grupo treinamento IIIB)		
		Rígido		SEAL	Rígido		SEAL	Rígido		SEAL
Número de		R_A 4	R_B 3	S_C 3	R_A 5	R_B 4	S_C 5	R_A 4	R_B 2	S_C 3
Componentes										
Q^2_{cv}		0,528	0,537	0,478	0,250	0,543	0,609	0,342	0,360	0,455
S_{press}		0,780	0,756	0,803	0,853	0,651	0,616	0,937	0,887	0,835
r^2		0,965	0,944	0,950	0,976	0,953	0,981	0,986	0,881	0,955
s		0,211	0,263	0,249	0,152	0,209	0,135	0,138	0,383	0,240
Teste F		167,035	140,816	158,022	187,568	121,412	240,448	416,917	96,253	176,840
Contribuições:										
Estérica		0,623	0,660	0,664	0,634	0,634	0,627	0,641	0,632	0,651
Eletrostática		0,377	0,340	0,336	0,366	0,366	0,373	0,359	0,368	0,349

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 67. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMFA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 (1R_A, 1R_B e 1S_C) (grupo de treinamento IB), modelo 2 (2R_A, 2R_B e 2S_C) (grupo de treinamento IIB) e, modelo 3 (3R_A, 3R_B e 3S_C) (grupo de treinamento IIIB) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 39 compostos (série B).

Nestes gráficos, as letras R_A , R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B - alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Tabela 23. Resultados obtidos através do emprego da metodologia *CoMFA*, para o grupo de treinamento, empregando-se a série de 30 derivados (série C) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

		Modelo 1 (grupo treinamento IC)			Modelo 2 (grupo treinamento IIC)			Modelo 3 (grupo treinamento IIIC)		
		Rígido		SEAL	Rígido		SEAL	Rígido		SEAL
		R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
Número de		5	1	3	3	1	1	2	2	5
Componentes										
Q^2_{cv}		0,210	0,117	0,145	0,310	0,280	0,242	0,239	0,247	0,455
s_{press}		0,556	0,528	0,547	0,508	0,494	0,507	0,477	0,475	0,438
r^2		0,980	0,469	0,820	0,902	0,532	0,506	0,818	0,845	0,987
S		0,088	0,410	0,251	0,192	0,398	0,409	0,233	0,215	0,066
Teste F		169,994	18,518	28,890	58,223	23,845	21,491	45,055	54,636	267,113
Contribuições:										
Estérica		0,629	0,656	0,680	0,719	0,706	0,751	0,626	0,622	0,625
Eletrostática		0,371	0,344	0,320	0,281	0,294	0,249	0,374	0,378	0,375

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 68. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMFA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 (1 R_A , 1 R_B e 1 S_C) (grupo de treinamento IC), modelo 2 (2 R_A , 2 R_B e 2 S_C) (grupo de treinamento IIC) e, modelo 3 (3 R_A , 3 R_B e 3 S_C) (grupo de treinamento IIIC) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 30 compostos (série C).

Nestes gráficos, as letras R_A , R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B - alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Tabela 24. Resultados obtidos através do emprego da metodologia *CoMFA*, para o grupo de treinamento, empregando-se a série de 23 derivados (série D) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

	Modelo 1			Modelo 2			Modelo 3		
	(grupo treinamento ID)			(grupo treinamento IID)			(grupo treinamento IIID)		
	Rígido	SEAL		Rígido	SEAL		Rígido	SEAL	
	R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
Número de Componentes	5	2	3	5	2	2	5	4	4
Q^2_{cv}	0,586	0,563	0,595	0,699	0,596	0,570	0,652	0,681	0,666
s_{press}	0,499	0,458	0,457	0,468	0,486	0,501	0,526	0,484	0,495
r^2	0,992	0,936	0,983	0,997	0,945	0,941	0,997	0,994	0,996
s	0,071	0,176	0,093	0,045	0,180	0,186	0,047	0,067	0,057
Teste F	282,381	109,278	272,290	871,440	127,712	119,290	871,555	523,931	741,821
Contribuições:									
Estérica	0,594	0,606	0,676	0,703	0,691	0,732	0,656	0,631	0,682
Eletrostática	0,406	0,394	0,324	0,297	0,309	0,268	0,344	0,369	0,318

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B - alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 69. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMFA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento ID), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IID) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIID) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 23 compostos (série D).

Nestes gráficos, as letras R_A , R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B – alinhamento rígido (alinhamento utilizando $N(1)$ e $C(2)$ do anel da piridino e os átomos $C=NNC=SN$ do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Tabela 25. Resultados obtidos através do emprego da metodologia *CoMFA*, para o grupo de treinamento, empregando-se a série de 22 derivados (série E) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974) utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

	Modelo 1			Modelo 2			Modelo 3		
	(grupo treinamento IE)			(grupo treinamento IIE)			(grupo treinamento IIIE)		
	Rígido	SEAL		Rígido	SEAL		Rígido	SEAL	
	R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
Número de Componentes	6	2	2	6	5	2	6	5	3
Q^2_{cv}	0,578	0,486	0,563	0,677	0,582	0,521	0,652	0,683	0,670
s_{press}	0,426	0,398	0,367	0,428	0,464	0,440	0,470	0,428	0,402
r^2	0,993	0,896	0,917	0,999	0,997	0,934	0,997	0,993	0,994
s	0,055	0,179	0,160	0,029	0,040	0,164	0,040	0,065	0,052
Teste F	239,368	60,158	77,510	1132,419	701,753	98,297	660,875	301,310	768,431
Contribuições:									
Estérica	0,536	0,541	0,651	0,648	0,628	0,710	0,637	0,592	0,642
Eletrostática	0,464	0,459	0,349	0,352	0,372	0,290	0,363	0,408	0,358

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 70. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMFA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IE), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIE) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIE) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 22 compostos (série E).

Nestes gráficos, as letras R_A , R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Análises realizadas para a série de 40 compostos (série A, n=40)

Para o modelo 1, grupo de treinamento IA, Tabela 21, as análises *PLS* com autovalidação forneceram os seguintes valores para os parâmetros estatísticos: $Q^2_{cv} = 0,480$ ($s_{press}=0,844$), para o alinhamento rígido considerando apenas o anel da piridino (R_A), $Q^2_{cv}=0,478$ ($s_{press}=0,797$), para o alinhamento rígido utilizando os átomos de N(1) e C(2) da piridino e os átomos C=NNC=SN, do fragmento da tiossemicarbazona (R_B), e $Q^2_{cv} = 0,431$ ($s_{press}=0,832$), para o alinhamento através do programa SEAL (S_C), utilizando os seguintes valores dos parâmetros: $\alpha=0,30$; $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 5, 2 e 2 componentes principais, respectivamente. Apesar da diferença no número de componentes principais do primeiro cálculo (R_A) se comparado com o segundo (R_B) e o terceiro (S_C), pode-se observar que os valores de Q^2_{cv} , na autovalidação, são próximos, variando de 0,431 a 0,480. Esta observação demonstra a robustez dos modelos *CoMFA* obtidos. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, forneceram: $r^2 = 0,977$ ($s = 0,176$), $r^2 = 0,895$ ($s = 0,357$) e $r^2 = 0,877$ ($s = 0,387$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 98%, 89% e 88%, respectivamente, da variância da atividade inibitória da enzima ribonucleotídeo difosfato redutase de células tumorais H. ep-2 dos compostos utilizados no grupo de treinamento. A contribuição da interação estérica variou entre 64% a 68% e da eletrostática variou entre 32% a 37%.

Os gráficos dos valores de pIC_{50} calculados *versus* observados apresentados na figura 66 (1 R_A , 1 R_B e 1 S_C), para o grupo de treinamento e teste (IA), modelo 1, mostram uma boa correlação (APÊNDICE, Tabela I). Observa-se que um dos compostos não teve o respectivo valor de pIC_{50} predito corretamente. Este foi o composto 4 (1'-CH₃, $pIC_{50} = 5,82$), indicado no lado direito inferior dos gráficos 1 R_A , 1 R_B e 1 S_C da figura 66. Como este composto foi selecionado para o grupo de teste e, não possui nenhum análogo estrutural no grupo de treinamento os modelos não conseguiram realizar a predição correta para o valor de pIC_{50} .

Para o modelo 2, grupo de treinamento IIA, Tabela 21, as análises *PLS* com autovalidação forneceram os seguintes valores para os parâmetros estatísticos: $Q^2_{cv} = 0,335$ ($s_{press}=0,781$), $Q^2_{cv} = 0,566$ ($s_{press}=0,669$) e $Q^2_{cv} = 0,617$ ($s_{press}=0,628$), para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 2, 5 e 5 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, forneceram: $r^2 = 0,847$ ($s=0,375$), $r^2 = 0,965$ ($s = 0,190$) e $r^2 = 0,979$ ($s = 0,146$) para os alinhamentos (R_A), (R_B) e

(S_C), respectivamente. Estes modelos explicam 85%, 97% e 98%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica variou entre 63% a 64% e da eletrostática variou entre 36% a 37%, sendo estes valores semelhantes aos valores obtidos com os grupos de treinamentos do modelo 1.

Observando-se a figura 66 (2R_A, 2R_B e 2S_C), pode-se constatar no lado esquerdo superior a presença de um “outlier”, composto 5 (2'-CH₃, pIC₅₀ = 2,28) (APÊNDICE, Tabela II). Como observado nas análises do modelo 1, este composto foi selecionado para o grupo de teste e não possui nenhum análogo estrutural no grupo de treinamento. Desta forma o modelo novamente não consegue prever o valor de sua atividade biológica. Os outros compostos selecionados para o grupo de teste tiveram os respectivos valores de atividade preditos sem grandes desvios em relação aos valores observados.

Para o modelo 3, grupo de treinamento IIIA, Tabela 21, as análises *PLS* com autovalidação forneceram os seguintes valores dos parâmetros estatísticos: $Q^2_{cv} = 0,357$ ($s_{press}=0,938$), $Q^2_{cv} = 0,403$ ($s_{press}=0,869$) e $Q^2_{cv} = 0,462$ ($s_{press}=0,825$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 4, 2 e 2 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, forneceram os seguintes valores dos parâmetros estatísticos: $r^2 = 0,983$ ($s = 0,152$), $r^2 = 0,893$ ($s = 0,368$) e $r^2 = 0,884$ ($s = 0,383$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 98%, 89% e 88%, respectivamente, da variância da atividade biológica. De forma análoga aos modelos anteriores, a contribuição da interação estérica foi de ~65% e da eletrostática foi de ~35%.

Todos os modelos obtidos possuem uma maior contribuição estérica ~65% e uma menor contribuição eletrostática ~35%. Este fato revela que para estes modelos, o fator estérico é o que mais contribui para os valores da atividade biológica dos compostos.

A figura 66 (3R_A, 3R_B e 3S_C) mostra os valores de pIC₅₀ calculados *versus* preditos pelos grupos de treinamento IIIA, modelo3. Um dos compostos que não teve o valor de sua atividade biológica corretamente predito foi o composto 15 (6-CH₃, pIC₅₀ = 5,11) (APÊNDICE, Tabela III). Este composto foi selecionado para o grupo de teste não possuindo análogo estrutural no grupo de treinamento, o que dificulta a predição do seu valor de pIC₅₀ pelo modelo. O outro composto que aparece como “outlier” é o 28 (5-OC₂H₄N(CH₃)₂). Apesar da série possuir vários compostos com substituintes na posição 5 do anel piridínico, o composto 28 é o único que apresenta um grupo substituinte com estas características, assim como esperado, tornando não significativa a predição do valor de pIC₅₀ pelo modelo.

Análises realizadas para a série de 39 compostos (série B, n=39)

Na tentativa de tornar os grupos de treinamento estruturalmente mais homogêneos, retirou-se da série o composto 33 (5-n-OOCC₁₅H₃₁) que apresenta uma grande cadeia carbônica lateral. Nenhum composto desta série possui um substituinte tão volumoso, fato este que o torna estatisticamente insignificante para o modelo. Além disto, este composto possui uma baixa atividade biológica ($pIC_{50} = 3,96$) ao se comparar com os demais compostos da série, fato que deve estar relacionado diretamente com as características do seu grupo substituinte.

Para o modelo 1, grupo de treinamento IB, Tabela 22, as análises *PLS* com autovalidação forneceram os seguintes resultados estatísticos: $Q^2_{cv} = 0,528$ ($s_{press}=0,780$), $Q^2_{cv} = 0,537$ ($s_{press}=0,756$) e $Q^2_{cv} = 0,478$ ($s_{press}=0,803$), para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 4, 3 e 3 componentes principais, respectivamente. Nestes cálculos, os valores de Q^2_{cv} foram maiores em relação ao observado no estudo considerando-se todos os 40 compostos e os valores de s_{press} menores. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2 = 0,965$ ($s = 0,211$), $r^2 = 0,944$ ($s = 0,263$) e $r^2 = 0,950$ ($s = 0,249$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 97%, 94% e 95%, respectivamente, da variância da atividade biológica. Com exceção do alinhamento R_A , os resultados da predição interna foram melhores ao se comparar os resultados obtidos com a série de 40 compostos. A contribuição da interação estérica variou entre 63% a 66% e da eletrostática variou entre 34% a 38%. Os valores das contribuições estéricas e eletrostáticas são semelhantes ao se considerar a série contendo todos os 40 compostos.

A figura 67 (1 R_A , 1 R_B e 1 S_C) mostra os gráficos dos valores de pIC_{50} calculados *versus* observados para o modelo 1, grupo de treinamento IB. De modo análogo, o composto 4 (1'-CH₃, $pIC_{50} = 5,82$) (APÊNDICE, Tabela IV) não teve a sua atividade predita de forma correta, pois não haviam compostos estruturalmente semelhantes no grupo de treinamento.

Para o modelo 2, grupo de treinamento IIB, Tabela 22, as análises *PLS* com autovalidação forneceram os seguintes valores para os parâmetros estatísticos: $Q^2_{cv} = 0,250$ ($s_{press}=0,853$), $Q^2_{cv} = 0,543$ ($s_{press}=0,651$) e $Q^2_{cv} = 0,609$ ($s_{press}=0,616$) empregando os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 5, 4 e 5 componentes principais, respectivamente. Comparando-se os valores de Q^2_{cv} ao se retirar o composto 33, com os valores obtidos sem a retirada do mesmo,

observa-se que estes foram menores. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2 = 0,976$ ($s = 0,152$), $r^2 = 0,953$ ($s = 0,209$) e $r^2 = 0,981$ ($s = 0,135$) para os alinhamentos (R_A), (R_B) e (S_C). Estes modelos explicam 98%, 95% e 98%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica foi de ~63% e da eletrostática foi de ~37%. Estes valores são semelhantes aos obtidos para o modelo 2, para a série de 40 compostos.

Observando-se a figura 67 (2 R_A , 2 R_B e 2 S_C) pode-se verificar a presença de um “outlier” no lado esquerdo superior dos gráficos. Este “outlier” é o composto 5 (2'-CH₃, pIC₅₀ = 2,28) (APÊNDICE, Tabela V) que, como já mencionado, não possui análogos estruturais no grupo de treinamento.

Para o modelo 3, grupo de treinamento IIIB, Tabela 22, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,342$ ($s_{press}=0,937$), $Q^2_{cv}=0,360$ ($s_{press}=0,887$) e $Q^2_{cv} = 0,455$ ($s_{press}=0,835$) utilizando os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 4, 2 e 3 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2=0,986$ ($s = 0,138$), $r^2 = 0,881$ ($s = 0,383$) e $r^2 = 0,955$ ($s = 0,240$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 99%, 88% e 96%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica variou entre 63% a 65% e da eletrostática variou de 35% a 37%.

Observando-se a figura 67 (3 R_A , 3 R_B e 3 S_C) pode-se verificar a existência de dois “outliers”. Estes compostos que não tiveram os valores de suas respectivas atividades biológicas corretamente preditas são, novamente, os compostos 15 (6-CH₃, pIC₅₀ = 5,11) e 28 (5-C₂H₄N(CH₃)₂, pIC₅₀ = 4,62) (APÊNDICE, Tabela VI).

Análises realizadas para a série de 30 compostos (série C, n=30)

De forma análoga ao realizado para a série contendo 39 compostos, na tentativa de tornar os grupos de treinamento e teste mais homogêneos, foram retirados das análises o composto 33 e os compostos de 2 a 10, cujos substituintes estavam posicionados no fragmento pertencente a tiossemicarbazona. Esta análise visou estudar apenas os compostos substituídos no anel da piridino.

Para o modelo 1, grupo de treinamento IC, Tabela 23, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,210$ ($s_{press}=0,556$),

$Q^2_{cv}=0,117$ ($s_{press}=0,528$) e $Q^2_{cv} = 0,145$ ($s_{press}=0,547$), para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 5, 1 e 3 componentes principais, respectivamente. Nestes cálculos, os valores de Q^2_{cv} foram menores em relação ao estudo considerando todos os 40 compostos, e os valores de s_{press} maiores. O decréscimo nos valores de Q^2_{cv} se devem à diminuição da variação no valor da atividade biológica no grupo de compostos utilizados tanto no treinamento quanto no teste. A variação do valor da atividade biológica ΔpIC_{50} , que na série contendo os 40 compostos era de 4,38 passou a ser de 2,04 para esta série com 30 compostos. Por outro lado, a diminuição nos valores de s_{press} indica que os modelos possuem um melhor poder de predição. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2 = 0,980$ ($s = 0,088$), $r^2 = 0,469$ ($s = 0,410$) e $r^2 = 0,820$ ($s=0,251$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 98%, 47% e 82%, respectivamente, da variância da atividade biológica. Com exceção do alinhamento R_A , os valores da predição interna foram menores em relação à série de 40 compostos. A contribuição da interação estérica variou entre 63% a 68% e da eletrostática variou entre 32% a 37%, sendo que estes valores são semelhantes aos obtidos para a análise realizada com a série de 40 compostos.

A figura 68 (1 R_A , 1 R_B e 1 S_C) mostra os gráficos dos valores de pIC_{50} calculados *versus* observados para o modelo 1, grupo de treinamento IC, e indica a presença de “outliers”. O composto 15 (6-CH₃, $pIC_{50} = 5,11$), possui um substituinte na posição 6 do anel piridínico, não possuindo análogos estruturais no grupo de treinamento. O composto 24 (5-CF₃, $pIC_{50} = 5,62$) é estruturalmente semelhante ao composto 26 (5-OCF₃, $pIC_{50} = 5,60$), mas os dois não possuem uma boa predição. E o composto 28 (5-OC₂H₄N(CH₃)₂) que, como comentado anteriormente, não possui análogos estruturais no grupo de treinamento (APÊNDICE, Tabela VII).

Para o modelo 2, grupo de treinamento IIC, Tabela 23, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,310$ ($s_{press}=0,508$), $Q^2_{cv}=0,280$ ($s_{press}=0,494$) e $Q^2_{cv} = 0,242$ ($s_{press}=0,507$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 3, 1 e 1 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2=0,902$ ($s = 0,192$), $r^2 = 0,532$ ($s = 0,398$) e $r^2 = 0,506$ ($s = 0,409$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 90%, 53% e 51%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica variou de 71% a 75% e

da eletrostática variou de 25% a 29%. Assim como nos modelos obtidos para as análises realizadas com as séries de 40 e 39 compostos, a contribuição do fator estérico é maior em relação ao fator eletrostático.

A figura 68 (2R_A, 2R_B e 2S_C) mostra os valores de pIC₅₀ calculados *versus* preditos pelo grupo de treinamento IIC, modelo 2. Como pode ser observado, existe uma baixa correlação entre os valores de pIC₅₀ calculados e observados.

Para o modelo 3, grupo de treinamento IIIC, Tabela 23, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,239$ ($s_{press}=0,477$), $Q^2_{cv}=0,247$ ($s_{press}=0,475$) e $Q^2_{cv} = 0,455$ ($s_{press}=0,438$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 2, 2 e 5 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2=0,818$ ($s = 0,233$), $r^2 = 0,845$ ($s = 0,215$) e $r^2 = 0,987$ ($s = 0,066$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 82%, 84% e 99%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica variou entre 62% a 63% e da eletrostática variou de 37% a 38%.

A figura 68 (3R_A, 3R_B e 3S_C) mostra os valores de pIC₅₀ calculados *versus* preditos pelo modelo 3, grupos de treinamento IIIC. Dentre os três modelos gerados, este foi o que apresentou os melhores valores tanto para a predição interna, bem como para a predição externa. Como “*outliers*” estão os compostos 15 (6-CH₃, pIC₅₀ = 5,11), 28 (5-OC₂H₄N(CH₃)₂, pIC₅₀ = 4,62) e 31 (5-OOCC₂H₅, pIC₅₀ = 5,28) (APÊNDICE, Tabela IX). Os motivos pelos quais os dois primeiros não tiveram os valores de suas respectivas atividades biológicas corretamente preditos já foram discutidos anteriormente. O composto 31 deveria apresentar o valor de sua atividade biológica predito com um desvio menor em relação ao valor observado, visto que nesta série existem outros dois compostos análogos (30 (5-OOCCH₃) e 32 (5-n-OOCC₃H₇)) incluídos no grupo de treinamento.

Análises realizadas para a série de 23 compostos (série D, n=23)

Para o modelo 1, grupo de treinamento ID, Tabela 24, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,586$ ($s_{press}=0,499$), $Q^2_{cv}=0,563$ ($s_{press}=0,458$) e $Q^2_{cv} = 0,595$ ($s_{press}=0,457$), para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 5, 2 e 3 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo

número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2=0,992$ ($s = 0,071$), $r^2 = 0,936$ ($s = 0,176$) e $r^2 = 0,983$ ($s = 0,093$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 99%, 94% e 98%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica variou entre 60% a 68% e da eletrostática variou entre 32% a 40%.

A figura 69 (1 R_A , 1 R_B e 1 S_C) mostra os valores de pIC_{50} calculados *versus* observados através do modelo 1, grupo de treinamento ID. Este modelo apresenta um alto poder de predição interno e externo. Apenas o composto 28 (5-OC₂H₄N(CH₃)₂, $pIC_{50} = 4,62$) (APÊNDICE, Tabela X) não teve o seu valor de pIC_{50} corretamente predito por motivos já discutidos anteriormente.

Para o modelo 2, grupo de treinamento IID, Tabela 24, as análises *PLS* com autovalidação forneceram os seguintes resultados estatísticos: $Q^2_{cv} = 0,699$ ($s_{press}=0,468$), $Q^2_{cv}=0,596$ ($s_{press}=0,486$) e $Q^2_{cv} = 0,570$ ($s_{press}=0,501$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação, foram obtidos utilizando 5, 2 e 2 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2=0,997$ ($s = 0,045$), $r^2 = 0,945$ ($s = 0,180$) e $r^2 = 0,941$ ($s = 0,186$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 100%, 95% e 94%, respectivamente, da variância na atividade biológica. A contribuição da interação estérica variou entre 69% a 73% e da eletrostática variou entre 27% a 31%.

Através da análise visual dos gráficos da figura 69 (2 R_A , 2 R_B e 2 S_C), pode-se notar uma alta correlação entre os valores de pIC_{50} calculados e preditos para os compostos dos grupos de treinamento e de teste.

Para o modelo 3, grupo de treinamento IIID, Tabela 24, as análises *PLS* com autovalidação forneceram os seguintes resultados estatísticos: $Q^2_{cv} = 0,652$ ($s_{press}=0,526$), $Q^2_{cv}=0,681$ ($s_{press}=0,484$) e $Q^2_{cv} = 0,666$ ($s_{press}=0,495$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 5, 4 e 4 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2=0,997$ ($s = 0,047$), $r^2 = 0,994$ ($s = 0,067$) e $r^2 = 0,996$ ($s = 0,057$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Todos os três modelos explicam 100% da variância da atividade biológica. A contribuição da interação estérica variou entre 63% a 68% e da eletrostática variou entre 32% a 37%.

Os gráficos da figura 69 (3R_A, 3R_B e 3S_C), apresentam alta correlação entre os valores de pIC₅₀ calculados e observados.

Para esta série de 23 compostos, foram obtidos valores significativos do quadrado do coeficiente de correlação da validação cruzada, Q^2_{cv} . Os valores variaram entre 0,563 e 0,699. Este fato pode ser explicado considerando que a série é mais homogênea em relação as que possuem 40, 39 e 30 compostos, possuindo substituições apenas na posição 5 do anel piridínico.

Análises realizadas para a série de 22 compostos (série E, n=22)

O composto 33 foi retirado novamente das análises para tornar o grupo de treinamento e de teste mais homogêneo. Para o modelo 1, grupo de treinamento IE, Tabela 25, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,578$ ($s_{press}=0,426$), $Q^2_{cv} = 0,486$ ($s_{press}=0,398$) e $Q^2_{cv} = 0,563$ ($s_{press}=0,367$), para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 6, 2 e 2 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2 = 0,993$ ($s = 0,055$), $r^2 = 0,896$ ($s = 0,179$) e $r^2 = 0,917$ ($s = 0,160$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 99%, 90% e 92%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica variou entre 54% a 65% e da eletrostática variou entre 35% a 46%. Em relação às análises realizadas com o composto 33, pode-se notar que houve uma pequena diminuição da contribuição estérica e conseqüentemente um aumento da contribuição eletrostática.

Os gráficos apresentados na figura 70 (1R_A, 1R_B e 1S_C), mostram uma alta correlação entre os valores de pIC₅₀ calculados e observados. O valor da atividade biológica do composto 28 (5-OC₂H₄N(CH₃)₂, pIC₅₀ = 4,62) (APÊNDICE, Tabela XIII) novamente não foi predito de forma correta pelo modelo.

Para o modelo 2, grupo de treinamento IIE, Tabela 25, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,677$ ($s_{press}=0,428$), $Q^2_{cv}=0,582$ ($s_{press}=0,464$) e $Q^2_{cv} = 0,521$ ($s_{press}=0,440$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 6, 5 e 2 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2=0,999$ ($s=0,029$), $r^2 = 0,997$ ($s = 0,040$) e $r^2 = 0,934$ ($s = 0,164$) para os alinhamentos (R_A), (R_B) e

(S_C), respectivamente. Os dois primeiros modelos explicam 100% da variância da atividade inibitória da enzima e o terceiro modelo explica 93% da variância da atividade. A contribuição da interação estérica variou entre 63% a 71% e da eletrostática variou entre 29% a 37%.

Os gráficos da figura 70 (2R_A, 2R_B e 2S_C), mostram que os modelos gerados possuem um alto poder de predição interno e externo. Apenas o composto 14 (5-C₂H₅, pIC₅₀ = 6,66) (APÊNDICE, Tabela XIV) não teve o seu valor de atividade biológica corretamente predito por este modelo. Esta observação pode ser explicada ao se considerar que este composto é o mais ativo da série e, desta forma, as informações contidas no grupo de treinamento não são suficientemente significativas para a predição do valor de pIC₅₀ deste composto.

Para o modelo 3, grupo de treinamento IIIE, Tabela 25, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,652$ ($s_{press}=0,470$), $Q^2_{cv}=0,683$ ($s_{press}=0,428$) e $Q^2_{cv} = 0,670$ ($s_{press}=0,402$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação, foram obtidos utilizando 6, 5 e 3 componentes principais, respectivamente. As análises *PLS* sem autovalidação com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2=0,997$ ($s=0,040$), $r^2 = 0,993$ ($s = 0,065$) e $r^2 = 0,994$ ($s = 0,052$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes três modelos explicam 100% da variância da atividade inibitória da enzima. A contribuição da interação estérica variou entre 59% a 64% e da eletrostática variou entre 36% a 41%.

Os gráficos da figura 70 (3R_A, 3R_B e 3S_C), mostram os valores de pIC₅₀ calculados *versus* observados. Observa-se que os valores da atividade para os compostos do grupo de treinamento foram significativamente preditos. Por outro lado, três compostos do grupo de teste (26 (5-OCF₃), pIC₅₀ = 5,60; 30 (5-OOCCH₃), pIC₅₀ = 5,44 e 38 (5-NHCOCH₃), pIC₅₀=5,92) (APÊNDICE, Tabela XV) não tiveram os valores de pIC₅₀ significativamente preditos.

A figura 71 mostra os mapas de contornos do *CoMFA*, para um dos modelos mais significativos, obtidos através das análises *PLS* empregando-se o modelo 2, para a série contendo 39 compostos (série B), grupo de treinamento IIB, com as moléculas alinhadas com o emprego do programa SEAL, tabela 21.

A análise visual da figura 71(a), permite sugerir um aumento no valor da atividade biológica para os compostos com grupos com cargas atômicas parciais negativas situados na região dos poliedros verdes (80% da contribuição) e/ou grupos com cargas atômicas parciais positivas estiverem situados na região dos poliedros vermelhos (20% da contribuição). Um

aumento no valor da atividade inibitória da RNR poderá ocorrer se substituintes eletronegativos estiverem ocupando a posição 5 do anel piridínico como, por exemplo, os compostos 23 (5-I), 21 (5-Cl) e 20 (5-F). Assim como apresentado na figura 71(a), um aumento no valor da atividade poderá ocorrer se grupos menos eletronegativos estiverem próximos ao nitrogênio do anel piridínico e, na posição 6 do anel piridínico, nas regiões dos poliedros vermelhos. Por outro lado, em algumas regiões próximas da posição 5, do anel piridínico, grupos menos eletronegativos devem favorecer a atividade biológica.

Figura 71. Mapas de contorno gerados por *CoMFA* (modelo 2, grupo de treinamento IIB, com 39 compostos (série B) alinhados pelo programa SEAL – Tabela 21) para os derivados da α -(N)-piridino carboxaldeído tiossemicarbazona.

(a) Mapas de contorno eletrostático: as regiões dos poliedros verdes indicam que a presença de grupos substituintes com densidade negativa aumenta os valores da atividade biológica ao passo que as regiões dos poliedros vermelhos indicam que a presença de grupos com densidade positiva são favoráveis à atividade.

(b) Mapas de contorno estérico: as regiões dos poliedros verdes indicam que a presença de grupos volumosos favorece a atividade biológica e as regiões dos poliedros vermelhos indicam que a presença de grupos volumosos desfavorece a atividade biológica. Um dos compostos mais ativos, composto 1, está situado no centro.

Por outro lado, os mapas de contorno eletrostático, figura 71(a), sugerem, não como em parte seria esperado, que grupos menos eletronegativos na região do nitrogênio, da piridino, devem contribuir para a atividade biológica dos compostos. Entretanto, a literatura indica que a presença de um átomo eletronegativo nesta região favorece a formação do complexo com o átomo de ferro (French & Blanz 1966; Sartorelli *et al.* 1976). A observação desta discrepância pode ser atribuída, em parte, pelo fato de que no modelo *CoMFA* apenas os campos estéricos e eletrostáticos são considerados e, sabe-se que para poder descrever corretamente a atividade biológica dos compostos são necessárias informações sobre a estrutura de várias naturezas, além daquelas consideradas no modelo *CoMFA*.

Analisando-se a figura 71(b), pode-se esperar que ocorra um aumento no valor da atividade biológica dos compostos se grupos substituintes volumosos estiverem nas regiões dos poliedros verdes (80% da contribuição) e/ou um decréscimo no valor de pIC_{50} para grupos substituintes volumosos presentes nas regiões dos poliedros vermelhos (20% da

contribuição). Nas posições 4 e 5 do anel piridínico, substituintes relativamente pouco volumosos são favoráveis para a atividade inibitória da RNR. Por exemplo, os compostos 12 (4-CH₃) e 13 (5-CH₃) possuem altos valores de pIC₅₀, iguais a 6,57 e 6,51, respectivamente. Por outro lado, os compostos 28 (5-OC₂H₄N(CH₃)₂, pIC₅₀ = 4,62) e 37 (5-OOCCH₂OC₆H₅, pIC₅₀ = 4,89), que apresentam baixos valores de atividade biológica, possuem seus grupos substituintes orientados para as regiões desfavoráveis.

Resultados semelhantes foram obtidos ao se considerar as análises realizadas para as séries contendo 23 e 22 compostos, séries D e E.

Os resultados obtidos com o emprego da metodologia *CoMFA* e os obtidos nas análises de *QSAR* clássico são muito semelhantes, dentro do erro experimental e das limitações de cada abordagem. De modo análogo, foi possível verificar que os grupos substituintes na posição 5 do anel piridínico não podem ser relativamente muito volumosos, o que diminuiria os valores da atividade biológica dos compostos. No entanto, embora complementares, uma grande diferença entre estas duas metodologias é que através dos mapas de contorno do *CoMFA* é possível verificar regiões favoráveis e desfavoráveis para se introduzir substituições. Adicionalmente, é possível identificar as características físico-químicas que os grupos substituintes devem possuir como, por exemplo, a necessidade de grupos eletronegativos próximos à região 5 do anel piridínico.

3.8 Estudos de *CoMSIA*

Os resultados obtidos através da metodologia *CoMSIA* aplicada para as séries A (n=40), B (n=39), C (n=30), D (n=23) e E (n=22) estão apresentados nas tabelas e nos gráficos das figuras apresentados a seguir. Os gráficos foram construídos com os valores de pIC_{50} preditos através dos modelos *CoMSIA*, gerados para cada uma das cinco séries (A-E), para cada um dos três grupos de treinamento e teste (I-III) e considerando-se os três alinhamentos utilizados (R_A , R_B e S_C). Os valores de pIC_{50} preditos estão no APÊNDICE, tabelas de XVI a XXX.

Tabela 26. Resultados obtidos através do emprego da metodologia *CoMSIA*, para o grupo de treinamento, empregando-se a série de 40 derivados (série A) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

	Modelo 1			Modelo 2			Modelo 3		
	(grupo treinamento IA)			(grupo treinamento IIA)			(grupo treinamento IIIA)		
	Rígido	SEAL		Rígido	SEAL		Rígido	SEAL	
	R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
Número de Componentes	4	5	6	2	3	6	4	5	5
Q^2_{cv}	0,547	0,482	0,453	0,447	0,416	0,539	0,606	0,498	0,517
s_{press}	0,772	0,842	0,884	0,712	0,746	0,704	0,734	0,846	0,830
r^2	0,968	0,982	0,986	0,837	0,911	0,974	0,985	0,993	0,993
S	0,205	0,157	0,141	0,387	0,291	0,167	0,141	0,103	0,098
teste F	188,812	262,635	271,789	69,130	88,881	144,353	422,206	639,689	704,476
Contribuições									
Estérica	0,138	0,142	0,135	0,178	0,170	0,162	0,115	0,113	0,112
Eletrostática	0,221	0,248	0,253	0,198	0,262	0,284	0,168	0,204	0,200
Hidrofóbica	0,258	0,287	0,295	0,256	0,284	0,298	0,247	0,289	0,275
Doadora	0,253	0,161	0,149	0,226	0,070	0,064	0,300	0,158	0,163
Aceptora	0,130	0,162	0,168	0,143	0,214	0,193	0,171	0,236	0,249

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 72. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMSIA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 (1RA, 1RB e 1SC) (grupo de treinamento IA), modelo 2 (2RA, 2RB e 2SC) (grupo de treinamento IIA) e, modelo 3 (3RA, 3RB e 3SC) (grupo de treinamento IIIA) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 40 compostos (série A).

Nestes gráficos, as letras R_A , R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B – alinhamento rígido (alinhamento utilizando $N(1)$ e $C(2)$ do anel da piridino e os átomos $C=NNC=SN$ do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Tabela 27. Resultados obtidos através do emprego da metodologia *CoMSIA*, para o grupo de treinamento, empregando-se a série de 39 derivados (série B) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

	Modelo 1			Modelo 2			Modelo 3		
	(grupo treinamento IB)			(grupo treinamento IIB)			(grupo treinamento IIIB)		
	Rígido	SEAL		Rígido	SEAL		Rígido	SEAL	
	R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
Número de Componentes	4	6	6	5	4	4	4	5	4
Q^2_{cv}	0,619	0,580	0,518	0,414	0,394	0,499	0,609	0,478	0,491
s_{press}	0,700	0,768	0,822	0,754	0,751	0,683	0,722	0,852	0,824
r^2	0,974	0,993	0,988	0,952	0,938	0,943	0,987	0,993	0,989
S	0,182	0,101	0,131	0,212	0,240	0,231	0,130	0,098	0,119
teste F	227,056	503,556	294,561	117,728	91,215	98,899	469,247	666,472	561,850
Contribuições									
Estérica	0,126	0,126	0,116	0,163	0,147	0,138	0,095	0,095	0,094
Eletrostática	0,239	0,270	0,274	0,257	0,288	0,299	0,180	0,214	0,207
Hidrofóbica	0,252	0,294	0,288	0,254	0,271	0,272	0,235	0,287	0,269
Doadora	0,260	0,166	0,156	0,208	0,071	0,074	0,307	0,162	0,175
Aceptora	0,124	0,145	0,167	0,118	0,223	0,217	0,187	0,242	0,254

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 73. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMSIA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 (1RA, 1RB e 1SC) (grupo de treinamento IB), modelo 2 (2RA, 2RB e 2SC) (grupo de treinamento IIB) e, modelo 3 (3RA, 3RB e 3SC) (grupo de treinamento IIIB) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 39 compostos (série B).

Nestes gráficos, as letras R_A , R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Tabela 28. Resultados obtidos através do emprego da metodologia *CoMSIA*, para o grupo de treinamento, empregando-se a série de 30 derivados (série C) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

	Modelo 1			Modelo 2			Modelo 3		
	(grupo treinamento IC)			(grupo treinamento IIC)			(grupo treinamento IIIC)		
	Rígido	SEAL		Rígido	SEAL		Rígido	SEAL	
Número de Componentes	R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
	3	3	3	4	5	5	5	6	4
Q^2_{cv}	0,237	0,188	0,219	0,360	0,429	0,409	0,633	0,658	0,674
s_{press}	0,517	0,533	0,523	0,503	0,489	0,497	0,359	0,357	0,329
r^2	0,876	0,882	0,935	0,945	0,970	0,968	0,982	0,992	0,975
S	0,208	0,203	0,151	0,148	0,113	0,115	0,080	0,056	0,090
teste F	44,697	47,352	90,668	76,700	108,757	103,704	185,019	319,380	178,871
Contribuições									
Estérica	0,181	0,170	0,163	0,214	0,228	0,214	0,136	0,132	0,136
Eletrostática	0,322	0,332	0,311	0,285	0,308	0,304	0,351	0,398	0,404
Hidrofóbica	0,206	0,240	0,299	0,289	0,322	0,316	0,180	0,213	0,175
Doadora	0,064	0,042	0,053	0,071	0,019	0,035	0,088	0,028	0,041
Aceptora	0,228	0,216	0,173	0,142	0,123	0,132	0,245	0,230	0,243

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 74. Gráficos dos valores de pIC₅₀ calculados *versus* observados, sendo que os pIC₅₀ calculados foram obtidos através dos modelos gerados pelo método *CoMSIA*, utilizando-se três diferentes modos de alinhamento (R_A, R_B e S_C): modelo 1 (1R_A, 1R_B e 1S_C) (grupo de treinamento IC), modelo 2 (2R_A, 2R_B e 2S_C) (grupo de treinamento IIC) e, modelo 3 (3R_A, 3R_B e 3S_C) (grupo de treinamento IIIC) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 30 compostos (série C).

Nestes gráficos, as letras R_A, R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Tabela 29. Resultados obtidos através do emprego da metodologia *CoMSIA*, para o grupo de treinamento, empregando-se a série de 23 derivados (série D) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

	Modelo 1			Modelo 2			Modelo 3		
	(grupo treinamento ID)			(grupo treinamento IID)			(grupo treinamento IIID)		
	Rígido	SEAL		Rígido	SEAL		Rígido	SEAL	
Número de Componentes	R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
	4	4	4	5	5	5	4	4	5
Q^2_{cv}	0,662	0,652	0,715	0,591	0,642	0,638	0,644	0,674	0,705
s_{press}	0,433	0,439	0,398	0,546	0,511	0,514	0,511	0,489	0,484
r^2	0,988	0,987	0,987	0,993	0,993	0,991	0,991	0,990	0,996
S	0,080	0,084	0,085	0,073	0,072	0,079	0,081	0,085	0,055
teste F	275,245	252,744	244,287	325,197	332,941	279,268	357,442	327,220	637,365
Contribuições									
Estérica	0,194	0,194	0,198	0,258	0,245	0,243	0,221	0,212	0,223
Eletrostática	0,317	0,326	0,339	0,300	0,323	0,330	0,335	0,368	0,367
Hidrofóbica	0,196	0,191	0,189	0,255	0,254	0,251	0,240	0,234	0,217
Doadora	0,028	0,015	0,026	0,032	0,024	0,030	0,031	0,016	0,023
Aceptora	0,265	0,274	0,248	0,155	0,155	0,145	0,173	0,171	0,169

⁽¹⁾ R_A – alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 75. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMSIA*, utilizando-se três diferentes modos de alinhamento (R_A, R_B e S_C): modelo 1 (1R_A, 1R_B e 1S_C) (grupo de treinamento ID), modelo 2 (2R_A, 2R_B e 2S_C) (grupo de treinamento IID) e, modelo 3 (3R_A, 3R_B e 3S_C) (grupo de treinamento IIID) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 23 compostos (série D).

Nestes gráficos, as letras R_A, R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Tabela 30. Resultados obtidos através do emprego da metodologia *CoMSIA*, para o grupo de treinamento, empregando-se a série de 22 derivados (série E) da 2-piridino carboxaldeído tiossemicarbazona, retirados da literatura selecionada (French *et al.* 1974), utilizando-se diferentes procedimentos de alinhamento: rígido (R_A ou R_B)⁽¹⁾ e, através da similaridade dos campos estéricos e eletrostáticos (S_C) gerada pelo programa SEAL⁽²⁾.

	Modelo 1			Modelo 2			Modelo 3		
	(grupo treinamento IE)			(grupo treinamento IIE)			(grupo treinamento IIIE)		
	Rígido	SEAL		Rígido	SEAL		Rígido	SEAL	
Número de Componentes	R_A	R_B	S_C	R_A	R_B	S_C	R_A	R_B	S_C
	6	6	6	6	5	5	5	5	5
Q^2_{cv}	0,720	0,692	0,791	0,538	0,565	0,626	0,691	0,690	0,724
s_{press}	0,348	0,364	0,300	0,512	0,473	0,439	0,423	0,423	0,399
r^2	0,997	0,998	0,998	0,998	0,992	0,991	0,998	0,998	0,997
s	0,033	0,030	0,029	0,037	0,063	0,068	0,036	0,034	0,044
teste F	642,329	783,114	868,704	697,021	280,527	243,081	988,070	1126,444	648,430
Contribuições									
Estérica	0,153	0,152	0,162	0,231	0,220	0,219	0,190	0,186	0,192
Eletrostática	0,396	0,408	0,418	0,378	0,378	0,393	0,406	0,431	0,421
Hidrofóbica	0,161	0,159	0,146	0,213	0,213	0,205	0,163	0,168	0,163
Doadora	0,035	0,023	0,037	0,030	0,024	0,032	0,035	0,021	0,029
Aceptora	0,255	0,258	0,236	0,149	0,165	0,151	0,207	0,194	0,194

⁽¹⁾ R_A - alinhamento rígido ou átomo à átomo (alinhamento utilizando apenas o anel da piridino) e R_B - alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e,

⁽²⁾ S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Figura 76. Gráficos dos valores de pIC_{50} calculados *versus* observados, sendo que os pIC_{50} calculados foram obtidos através dos modelos gerados pelo método *CoMSIA*, utilizando-se três diferentes modos de alinhamento (R_A , R_B e S_C): modelo 1 ($1R_A$, $1R_B$ e $1S_C$) (grupo de treinamento IE), modelo 2 ($2R_A$, $2R_B$ e $2S_C$) (grupo de treinamento IIE) e, modelo 3 ($3R_A$, $3R_B$ e $3S_C$) (grupo de treinamento IIIE) incluindo-se os compostos do grupo de treinamento e do grupo de teste para a série de 22 compostos (série E).

Nestes gráficos, as letras R_A , R_B e S_C representam o método utilizado no alinhamento, sendo:

R_A – alinhamento rígido (alinhamento utilizando apenas o anel da piridino)

R_B – alinhamento rígido (alinhamento utilizando N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona)

S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$)

Análises realizadas para a série de 40 compostos (série A, n=40)

Para o modelo 1, grupo de treinamento IA, Tabela 26, as análises *PLS* com autovalidação forneceram um valor de $Q^2_{cv} = 0,547$ ($s_{press} = 0,772$), para o alinhamento rígido considerando apenas o anel da piridino (R_A), $Q^2_{cv} = 0,482$ ($s_{press} = 0,842$), para o alinhamento rígido utilizando N(1) e C(2) da piridino e os átomos C=NNC=SN, do fragmento da tiossemicarbazona (R_B), e $Q^2_{cv} = 0,453$ ($s_{press} = 0,884$), para o alinhamento através do programa SEAL (S_C), utilizando os seguintes valores dos parâmetros: $\alpha = 0,30$; $W_E = 1,0$, $W_S = 1,0$ e $W_H = 0,0$. Os valores de Q^2_{cv} obtidos através da metodologia *CoMFA* são menores em relação aos obtidos através do emprego da abordagem *CoMSIA*. Conseqüentemente, os valores de s_{press} dos modelos *CoMSIA* são menores em relação aos obtidos nos modelos *CoMFA*.

Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 4, 5 e 6 componentes principais, respectivamente. A análise *PLS* de sem autovalidação com o mesmo número de componentes principais resultou em valores de $r^2 = 0,968$ ($s = 0,205$), para o alinhamento (R_A), de $r^2 = 0,982$ ($s = 0,157$) para o alinhamento (R_B) e de $r^2 = 0,986$ ($s = 0,141$) para o alinhamento (S_C). Estes modelos explicam 97%, 98% e 99%, respectivamente, da variância da atividade inibitória da enzima ribonucleotídeo difosfato redutase de células tumorais H. ep-2 dos compostos utilizados no grupo de treinamento.

A contribuição das interações estéricas foi de aproximadamente 14%, a contribuição eletrostática variou entre 22% a 25%, a contribuição hidrofóbica variou entre 26% a 30%, a contribuição doadora de ligações de hidrogênio variou entre 15% a 25% e a aceptora de ligações de hidrogênio variou entre 13% a 17%. Para os modelos *CoMSIA* gerados, as contribuições eletrostáticas são maiores do que as contribuições estéricas, ao contrário dos resultados obtidos para os modelos *CoMFA*. Para estes modelos, as contribuições hidrofóbicas seguidas da contribuição eletrostática foram as que apresentaram os maiores valores de contribuições para os modelos. Estes resultados demonstram a importâncias da inclusão destes dois campos para o modelo que descreve a atividade biológica dos compostos.

Os gráficos da figura 72 (1 R_A , 1 R_B e 1 S_C), mostram os valores de pIC_{50} calculados *versus* observados pelo modelo 1, grupo de treinamento IA, utilizando-se os alinhamentos (R_A), (R_B) e (S_C), respectivamente. O composto 4 (1'-CH₃, $pIC_{50} = 5,82$) (APÊNDICE, Tabela XVI) pode ser considerado um “outlier”, pois o valor da atividade biológica predito pelo modelo é significativamente diferente do valor observado experimentalmente. A razão para que isto ocorra pode ser explicada de modo análogo como foi discutido anteriormente nos resultados da abordagem *CoMFA*. De uma forma geral, a predição dos valores da

atividade biológica obtidas pelos modelos *CoMFA* e *CoMSIA* para os grupos de treinamento e de teste são muito semelhantes, dentro do erro experimental.

Para o modelo 2, grupo de treinamento IIA, Tabela 26, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,447$ ($s_{press}=0,712$), $Q^2_{cv} = 0,416$ ($s_{press}=0,746$) e $Q^2_{cv} = 0,539$ ($s_{press}=0,704$) para os alinhamentos (R_A), (R_B) e (SC), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 2, 3 e 6 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2 = 0,837$ ($s = 0,387$), $r^2 = 0,911$ ($s = 0,291$) e $r^2 = 0,974$ ($s = 0,167$) para os alinhamentos (R_A), (R_B) e (SC), respectivamente. Estes modelos explicam 84%, 91% e 97%, respectivamente, da variância da atividade biológica.

As contribuições da interação estérica variou entre 16% a 18%, da eletrostática variou entre 20% a 28%, da hidrofóbica variou entre 26% a 30%, da doadora de ligações de hidrogênio variou entre 6% a 23% e da aceptora de ligações de hidrogênio variou entre 14% a 21%. Estes valores são semelhantes aos obtidos para o modelo 1.

Os gráficos $2R_A$, $2R_B$ e, $2SC$ da figura 72 apresentam a correlação entre os valores calculados da atividade biológica *versus* os observados para o modelo2. O composto 5 (2'-CH₃, pIC₅₀ = 2,28) (APÊNDICE, Tabela XVII) foi o que apresentou a maior diferença entre o valor de pIC₅₀ calculado e observado. Como discutido anteriormente não há nenhum composto estruturalmente análogo a este no grupo de treinamento.

Para o modelo 3, grupo de treinamento IIIA, Tabela 26, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,606$ ($s_{press}=0,734$), $Q^2_{cv} = 0,498$ ($s_{press}=0,846$) e $Q^2_{cv} = 0,517$ ($s_{press}=0,830$) para os alinhamentos (R_A), (R_B) e (SC). Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 4, 5 e 5 componentes principais, respectivamente. Comparando-se com os resultados obtidos nos modelos *CoMFA*, os modelos *CoMSIA* para estes grupos de treinamento forneceram valores de Q^2_{cv} maiores, indicando desta forma um maior poder de predição da atividade biológica. As análises *PLS* sem autovalidação com o mesmo número de componentes principais resultaram nos seguintes valores estatísticos: $r^2 = 0,985$ ($s = 0,141$), $r^2 = 0,993$ ($s = 0,103$) e $r^2 = 0,993$ ($s = 0,098$) para os alinhamentos (R_A), (R_B) e (SC). Estes modelos explicam 98%, 99% e 99%, respectivamente, da variância da atividade biológica.

A contribuição da interação estérica foi de aproximadamente 11%, da eletrostática variou entre 17% a 20%, da hidrofóbica variou entre 25% a 29%, da doadora de ligações de hidrogênio variou entre 16% a 30% e da aceptora de ligações de hidrogênio variou entre 17%

a 25%. Estes resultados são semelhantes ao obtidos para os modelos 1 e 2, para esta série de compostos.

A análise dos gráficos $3R_A$, $3R_B$ e $3S_C$ apresentados na figura 72 indicam que os modelos *CoMSIA* gerados apresentam a predição interna significativa, para o grupo de treinamento. Para o grupo de teste, os compostos 15 (6-CH₃, $pIC_{50} = 5,11$) e 28 (5-OC₂H₄N(CH₃)₂, $pIC_{50} = 4,62$) (APÊNDICE, Tabela XVIII) foram os que apresentaram as maiores diferenças entre os valores de pIC_{50} calculados e observados. Os motivos são análogos aos apresentados anteriormente nas abordagens *CoMFA*.

Análises realizadas para a série de 39 compostos (série B, n=39)

Assim como nas abordagens *CoMFA*, o composto 33 (5-n-OOCC₁₅H₃₁) foi retirado das análises com o intuito de obter uma série estruturalmente mais homogênea.

Para o modelo 1, grupo de treinamento IB, Tabela 27, as análises *PLS* de autovalidação forneceram os respectivos valores estatísticos: $Q^2_{cv} = 0,619$ ($s_{press}=0,700$), $Q^2_{cv} = 0,580$ ($s_{press}=0,768$) e $Q^2_{cv} = 0,518$ ($s_{press}=0,822$) para os alinhamentos (R_A), (R_B) e (S_C). Em relação ao estudo considerando a série de 40 compostos, observou-se, de uma forma geral, valores dos parâmetros estatísticos Q^2_{cv} e s_{press} , significativamente melhores, sendo que o primeiro teve um pequeno aumento e o segundo uma pequena diminuição. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 4, 6 e 6 componentes principais, respectivamente. As análises *PLS* sem autovalidação com o mesmo número de componentes principais resultaram nos seguintes valores estatísticos: $r^2 = 0,974$ ($s = 0,182$), $r^2 = 0,993$ ($s = 0,101$) e $r^2 = 0,988$ ($s = 0,131$) para os alinhamentos (R_A), (R_B) e (S_C). Estes modelos explicam 97%, 99% e 99%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica variou entre 12% a 13%, da eletrostática variou entre 24% a 27%, da hidrofóbica variou entre 25% a 29%, da doadora de ligações de hidrogênio variou entre 16% a 26% e da aceptora de ligações de hidrogênio variou entre 12% a 17%. As contribuições dos campos na análise da série com os 40 compostos para o modelo 1 foram semelhantes.

A análise dos gráficos $1R_A$, $1R_B$ e $1S_C$ da figura 73 mostra, de modo análogo a *CoMFA*, que o valor da atividade biológica do composto 4 (1'-CH₃, $pIC_{50} = 5,82$) (APÊNDICE, Tabela XIX) não foi predito corretamente.

Para o modelo 2, grupo de treinamento IIB, Tabela 27, as análises *PLS* com autovalidação resultaram os seguintes valores estatísticos: $Q^2_{cv} = 0,414$ ($s_{press}=0,754$) A), $Q^2_{cv} = 0,394$ ($s_{press}=0,751$) e $Q^2_{cv} = 0,499$ ($s_{press}=0,683$) para os alinhamentos (R_A), (R_B) e (S_C).

Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 5, 4 e 4 componentes principais, respectivamente. As análises *PLS* sem autovalidação com o mesmo número de componentes principais resultaram nos seguintes valores estatísticos: $r^2 = 0,952$ ($s = 0,212$), $r^2 = 0,938$ ($s = 0,240$) e $r^2 = 0,943$ ($s = 0,231$) para os alinhamentos (R_A), (R_B) e (S_C). Estes modelos explicam 95%, 94% e 94%, respectivamente, da variância da atividade biológica. A contribuição das interações estéricas variou entre 14% a 16%, a contribuição eletrostática variou entre 26% a 30%, a contribuição hidrofóbica variou entre 25% a 27%, a contribuição doadora de ligações de hidrogênio variou entre 7% a 21% e a contribuição aceptora de ligações de hidrogênio variou entre 12% a 22%.

A análise dos gráficos $2R_A$, $2R_B$ e $2S_C$ da figura 73 mostram a presença de um “outlier”. O composto 5 ($2'$ -CH₃, $pIC_{50} = 2,28$) (APÊNDICE, Tabela XX) de modo análogo não apresentou o valor de sua atividade biológica corretamente predito.

Para o modelo 3, grupo de treinamento IIIB, Tabela 27, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos para os alinhamentos (R_A), (R_B) e (S_C), respectivamente: $Q^2_{cv} = 0,609$ ($s_{press}=0,722$); $Q^2_{cv} = 0,478$ ($s_{press}=0,852$) e $Q^2_{cv} = 0,491$ ($s_{press}=0,824$). Os valores de Q^2_{cv} , na autovalidação, foram obtidos utilizando 4, 5 e 4 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2 = 0,987$ ($s = 0,130$), $r^2 = 0,993$ ($s = 0,098$) e $r^2 = 0,989$ ($s = 0,119$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Todos estes três modelos explicam 99% da variância da atividade biológica.

A contribuição da interação estérica foi de aproximadamente 9%, da eletrostática variou entre 18% a 21%, da hidrofóbica variou entre 24% a 29%, da doadora de ligações de hidrogênio variou entre 16% a 31% e da aceptora de ligações de hidrogênio variou entre 19% a 25%. Para esta série de compostos, as interações hidrofóbicas e as eletrostáticas foram as que mais contribuíram para os modelos.

Apesar dos gráficos $3R_A$, $3R_B$ e $3S_C$, da figura 73 apresentarem uma boa predição interna, alguns compostos não puderam ter os seus valores de pIC_{50} corretamente preditos. A diferença entre os valores da atividade biológica calculada e observada foi maior para os compostos 15 (6 -CH₃) e 28 (5 -OC₂H₄N(CH₃)₂) (APÊNDICE, Tabela XXI), por motivos já discutidos previamente.

Análises realizadas para a série de 30 compostos (série C, n=30)

Para o modelo 1, grupo de treinamento IC, Tabela 28, as análises *PLS* de autovalidação forneceram os seguintes valores dos parâmetros estatísticos: $Q^2_{cv} = 0,237$ ($s_{press}=0,517$), $Q^2_{cv} = 0,188$ ($s_{press}=0,533$) e $Q^2_{cv} = 0,219$ ($s_{press}=0,523$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 3 componentes principais em todos os grupos de treinamento. As análises *PLS* sem autovalidação com o mesmo número de componentes principais resultaram nos seguintes valores estatísticos: $r^2 = 0,876$ ($s = 0,208$), $r^2 = 0,882$ ($s = 0,203$) e $r^2 = 0,935$ ($s = 0,151$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os dois primeiros modelos explicam 88% da variância da atividade biológica e o terceiro modelo explica 94% da variância da atividade biológica. A contribuição da interação estérica foi de aproximadamente 14%, da eletrostática variou entre 31% a 33%, da hidrofóbica variou entre 21% a 30%, da doadora de ligações de hidrogênio variou entre 4% a 6% e da aceptora de ligações de hidrogênio variou entre 17% a 23%. Para estas análises a contribuição eletrostática foi maior do que a contribuição hidrofóbica.

Uma análise dos gráficos $1R_A$, $1R_B$ e $1S_C$ da figura 74 mostra que alguns compostos não tiveram os valores da atividade biológica preditos pelos modelos (APÊNDICE, Tabela XXII). E, as explicações são respectivamente análogas às apresentadas nas abordagens *CoMFA*.

Para o modelo 2, grupo de treinamento IIC, Tabela 28, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,360$ ($s_{press}=0,503$), $Q^2_{cv} = 0,429$ ($s_{press}=0,489$) e $Q^2_{cv} = 0,409$ ($s_{press}=0,497$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 4, 5 e 5 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram em: $r^2 = 0,945$ ($s = 0,148$); $r^2 = 0,970$ ($s = 0,113$) e $r^2 = 0,968$ ($s = 0,115$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 95%, 97% e 97%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica variou entre 21% a 23%, da eletrostática variou entre 29% a 31%, da hidrofóbica variou entre 29% a 32%, da doadora de ligações de hidrogênio variou entre 2% a 7% e da aceptora de ligações de hidrogênio variou entre 12% a 14%.

Os gráficos $2R_A$, $2R_B$ e $2S_C$ da figura 74 mostram os valores da atividade biológica calculados *versus* observados por este modelo. Os gráficos $2R_A$ e $2R_B$ apresentam uma boa predição interna do modelo, mas uma predição não significativa dos valores de pIC_{50} para os

compostos incluídos no grupo de teste. Adicionalmente, o gráfico 2S_C mostra que tanto a predição interna bem como a predição externa não é significativa.

Para o modelo 3, grupo de treinamento IIIC, Tabela 28, as análises *PLS* de autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,633$ ($s_{press}=0,359$); $Q^2_{cv} = 0,658$ ($s_{press}=0,357$) e $Q^2_{cv} = 0,674$ ($s_{press}=0,329$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 5, 6 e 4 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram nos seguintes valores estatísticos: $r^2 = 0,982$ ($s = 0,080$); $r^2 = 0,992$ ($s = 0,056$) e $r^2 = 0,975$ ($s = 0,090$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 98%, 99% e 98%, respectivamente, da variância da atividade biológica. A contribuição da interação estérica foi de aproximadamente 13%, da eletrostática variou entre 35% a 40%, da hidrofóbica variou entre 18% a 21%, da doadora de ligações de hidrogênio variou entre 3% a 9% e da aceptora de ligações de hidrogênio foi de aproximadamente 24%.

Para esta série de 30 compostos a contribuição do fator eletrostático foi o que demonstrou ser o mais importante para a atividade biológica dos compostos, seguido da contribuição hidrofóbica e da aceptora de ligações de hidrogênio.

A análise dos gráficos 3R_A, 3R_B e 3S_C apresentados na figura 74 mostra um alto poder de predição interno. Entretanto, os compostos 15 (6-CH₃, pIC₅₀ = 5,11) e 28 (5-OC₂H₄N(CH₃)₂, pIC₅₀ = 4,62) (APÊNDICE, Tabela XXIV) não tiveram os valores das suas atividade biológica preditos de forma correta. Os motivos destas discrepâncias são análogos aos apresentados previamente nas abordagens *CoMFA*.

Análises realizadas para a série de 23 compostos (série D, n=23)

Para o modelo 1, grupo de treinamento ID, Tabela 29, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,720$ ($s_{press}=0,348$); $Q^2_{cv} = 0,692$ ($s_{press}=0,364$) e $Q^2_{cv} = 0,791$ ($s_{press}=0,300$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 6 componentes principais nas três análises. As análises *PLS* sem autovalidação, empregando o mesmo número de componentes principais, forneceram os seguintes resultados estatísticos: $r^2 = 0,997$ ($s = 0,033$); $r^2 = 0,998$ ($s = 0,030$) e $r^2 = 0,998$ ($s = 0,029$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes três modelos explicam praticamente 100% da variância da atividade biológica. A contribuição da interação estérica foi de aproximadamente 15%, da

eletrostática foi de aproximadamente 40%, da hidrofóbica foi de aproximadamente 16%, da doadora de ligações de hidrogênio variou entre 2% a 4% e da aceptora de ligações de hidrogênio variou entre 24% a 26%.

Os gráficos 1R_A, 1R_B e 1S_C apresentados na figura 75 mostram os valores da atividade biológica calculados *versus* observados. Observa-se uma alta correlação entre os valores de pIC₅₀ calculados e observados, com exceção de um “outlier”. O composto 28 (5-OC₂H₄N(CH₃)₂, pIC₅₀ = 4,62) (APÊNDICE, Tabela XXV), como discutido anteriormente, por não possuir análogo estrutural no grupo de treinamento não pôde ter seu valor de sua atividade biológica predito.

Para o modelo 2, grupo de treinamento IID, Tabela 29, as análises *PLS* com autovalidação forneceram os seguintes resultados estatísticos: $Q^2_{cv} = 0,538$ ($s_{press}=0,512$); $Q^2_{cv} = 0,565$ ($s_{press}=0,473$) e $Q^2_{cv} = 0,626$ ($s_{press}=0,439$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 6, 5 e 5 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram em: $r^2 = 0,998$ ($s = 0,037$); $r^2 = 0,992$ ($s = 0,063$) e $r^2 = 0,991$ ($s = 0,068$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam praticamente 100% da variância da atividade biológica. A contribuição da interação estérica foi de ~22%, da eletrostática variou entre 38% a 40%, da hidrofóbica foi de ~21%, da doadora de ligações de hidrogênio foi de ~3% e da aceptora de ligações de hidrogênio foi de ~15%.

Os gráficos 2R_A, 2R_B e 2S_C apresentados na figura 75 mostram um alto e significativo poder de predição tanto interna como externa para o modelo

Para o modelo 3, grupo de treinamento IIID, Tabela 29, as análises *PLS* com autovalidação forneceram os seguintes resultados estatísticos: $Q^2_{cv} = 0,691$ ($s_{press}=0,423$), $Q^2_{cv} = 0,690$ ($s_{press}=0,423$) e $Q^2_{cv} = 0,724$ ($s_{press}=0,399$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 5 componentes principais nas três análises. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram em: $r^2 = 0,998$ ($s = 0,036$), $r^2 = 0,998$ ($s = 0,034$) e $r^2 = 0,997$ ($s = 0,044$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam praticamente 100% da variância da atividade biológica. A contribuição da interação estérica foi de ~19%, da eletrostática variou entre 41% a 43%, da hidrofóbica foi de ~16%, da contribuição doadora de ligações de hidrogênio foi de ~3% e da contribuição aceptora de ligações de hidrogênio foi de ~20%.

Os gráficos $3R_A$, $3R_B$ e $3S_C$ da figura 75 mostram que os valores de pIC_{50} dos compostos do grupo de teste foram preditos de forma significativa. Apenas os compostos 26 (5-OCF₃, $pIC_{50} = 5,60$) e 38 (5-NHCOCH₃, $pIC_{50} = 5,92$) (APÊNDICE, Tabela XXVII) tiveram os valores da atividade biológica significativamente diferentes dos valores observados, dentro do erro experimental.

Análises realizadas para a série de 22 compostos (série E, n=22)

Para o modelo 1, grupo de treinamento IE, Tabela 30, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,662$ ($s_{press}=0,433$), $Q^2_{cv} = 0,652$ ($s_{press}=0,439$) e $Q^2_{cv} = 0,715$ ($s_{press}=0,398$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação, foram obtidos utilizando 4 componentes principais nas três análises. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram em: $r^2 = 0,988$ ($s = 0,080$), $r^2 = 0,987$ ($s = 0,439$) e $r^2 = 0,987$ ($s = 0,085$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes três modelos explicam 99% da variância da atividade biológica. A contribuição da interação estérica foi de ~20%, da eletrostática variou entre 32% a 34%, da hidrofóbica variou entre 19% a 20%, da doadora de ligações de hidrogênio variou entre 2% a 3% e da aceptora de ligações de hidrogênio variou entre 25% a 27%.

A análise dos gráficos $1R_A$, $1R_B$ e $1S_C$ da figura 76 apresentam um “outlier”. O composto que não teve o valor da sua atividade biológica corretamente predito é o 28 (5-OC₂H₄N(CH₃)₂, $pIC_{50} = 4,62$) (APÊNDICE, Tabela XXVIII).

Para o modelo 2, grupo de treinamento IIE, Tabela 30, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,591$ ($s_{press}=0,546$), $Q^2_{cv} = 0,642$ ($s_{press}=0,511$) e $Q^2_{cv} = 0,638$ ($s_{press}=0,514$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação, foram obtidos utilizando 5 componentes principais nas três análises. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram em: $r^2 = 0,993$ ($s = 0,073$), $r^2 = 0,993$ ($s = 0,072$) e $r^2 = 0,991$ ($s = 0,079$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam 99% da variância da atividade biológica. A contribuição da interação estérica variou entre 24% a 26%, da eletrostática variou entre 30% a 33%, da hidrofóbica foi de ~25%, da doadora de ligações de hidrogênio foi de ~3% e da aceptora de ligações de hidrogênio foi de ~15%.

Os gráficos $2R_A$, $2R_B$ e $2S_C$ da figura 76 mostram que a predição interna e a externa foram realizadas com sucesso para este modelo.

Para o modelo 3, grupo de treinamento IIIE, Tabela 30, as análises *PLS* com autovalidação forneceram os seguintes valores estatísticos: $Q^2_{cv} = 0,644$ ($s_{press}=0,511$), $Q^2_{cv} = 0,674$ ($s_{press}=0,489$) e $Q^2_{cv} = 0,705$ ($s_{press}=0,484$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Os valores de Q^2_{cv} , na autovalidação foram obtidos utilizando 4, 4 e 5 componentes principais, respectivamente. As análises *PLS* sem autovalidação, com o mesmo número de componentes principais, resultaram em: $r^2 = 0,991$ ($s = 0,081$), $r^2 = 0,990$ ($s = 0,085$) e $r^2 = 0,996$ ($s = 0,055$) para os alinhamentos (R_A), (R_B) e (S_C), respectivamente. Estes modelos explicam de 99% a 100% da variância da atividade biológica. A contribuição da interação estérica foi de ~22%, da eletrostática variou entre 34% a 37%, da hidrofóbica variou entre 22% a 24%, da doadora de ligações de hidrogênio foi de ~2% e da acceptora de ligações de hidrogênio foi de ~17%.

Para esta série de 22 compostos o fator eletrostático foi o que mais contribuiu para o valor de pIC_{50} dos compostos, seguido do estérico, hidrofóbico e do acceptor de ligações de hidrogênio.

Os gráficos $3R_A$, $3R_B$ e $3S_C$ da figura 76 mostram a presença de um “outlier”. O valor da atividade biológica para o composto 26 (5-OCF₃, $pIC_{50} = 5,60$) (APÊNDICE, Tabela XXX) não pode ser predito pelo modelo.

De uma maneira geral, os resultados estatísticos dos modelos gerados por *CoMSIA* foram significativamente melhores em relação aos modelos gerados através da abordagem *CoMFA*. Este fato pode ser explicado considerando-se dois fatores:

- i) Nas análises *CoMFA* as interações são calculadas através de uma função do tipo gaussiana, não promovendo cortes bruscos nos valores das energias de interação.
- ii) A inclusão das contribuições hidrofóbicas, doadoras e acceptoras de ligação de hidrogênio permitem que os modelos gerados sejam mais consistentes com as interações realmente observadas entre o ligante e o seu receptor. Na abordagem *CoMFA*, por outro lado, apenas as interações estéricas e eletrostáticas são consideradas.

A figura a seguir mostra os mapas de contornos do *CoMSIA* para um dos modelos estatisticamente mais significativo, gerados através das análises *PLS*, empregando o modelo 1B, para a série de 39 compostos, com as moléculas alinhadas de forma rígida, utilizando-se N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona (R_B), tabela 26.

Figura 77. Mapas de contorno gerados por *CoMSIA* (modelo 1, grupo de treinamento 1B, com 39 compostos (série B) alinhados pelo N(1) e C(2) do anel da piridino e os átomos C=NNC=SN do fragmento da tiossemicarbazona (R_B) – Tabela 26) para os derivados da α -(N)-piridino carboxaldeído tiossemicarbazona.

(a) Mapas de contorno eletrostático: as regiões dos poliedros verdes indicam que a presença de grupos com densidade negativa aumentam o valor da atividade biológica enquanto que as regiões dos poliedros vermelhos indicam que a presença de grupos com densidade positiva são favoráveis à atividade.

(b) Mapas de contorno estérico: as regiões dos poliedros verdes indicam que a presença de grupos relativamente volumosos favorece a atividade biológica e as regiões dos poliedros vermelhos indicam que a presença de grupos relativamente volumosos desfavorece a atividade biológica.

(c) Mapas de contorno hidrofóbico: as regiões dos poliedros verdes indicam que a presença de grupos hidrofóbicos aumenta os valores da atividade biológica e as regiões dos poliedros vermelhos indicam que grupos hidrofílicos aumentam os valores da atividade.

(d) Mapas de contorno de doadores de ligação de hidrogênio: as regiões dos poliedros verdes indicam que a presença de grupos doadores de ligação de hidrogênio aumenta os valores da atividade dos compostos e nas regiões dos poliedros vermelhos estes grupos diminuem os valores da atividade biológica.

(e) Mapas de contorno de aceptores de ligação de hidrogênio: as regiões dos poliedros verdes indicam que a presença de grupos aceptores de ligação de hidrogênio aumentam os valores da atividade dos compostos e nas regiões dos poliedros vermelhos estes grupos diminuem os valores da atividade dos compostos. Um dos compostos mais ativos, composto 1 ($pIC_{50}=6,55$) está situado no centro.

Através da inspeção visual da figura 77(a) pode-se sugerir que o valor da atividade biológica dos compostos aumentará se grupos com densidade de carga negativa estiverem próximos às regiões dos poliedros verdes (80% da contribuição) e/ou se grupos com densidade de carga positiva estiverem próximos às regiões dos poliedros vermelhos (20% da contribuição). De modo análogo para os modelos gerados por *CoMFA*, a introdução de substituintes eletronegativos na posição 5 do anel piridínico deve aumentar os valores da atividade biológica dos compostos. Adicionalmente, o átomo de carbono da posição 5 do anel piridínico apresentando uma densidade de carga positiva, devido ao efeito indutivo retirador de elétrons do substituinte, também contribuirá para a atividade biológica. Nestas análises, os

mapas de contorno eletrostáticos não mostraram poliedros vermelhos próximo ao nitrogênio da piridino, fato que está de acordo com os dados da literatura (French & Blanz-Jr. 1966).

Na figura 77(b) observa-se um aumento no valor da atividade biológica dos compostos quando grupos substituintes relativamente volumosos ocupam as regiões dos poliedros verdes (80% da contribuição) e/ou um decréscimo na atividade quando grupos substituintes volumosos ocupam as regiões dos poliedros vermelhos (20% da contribuição). Como foi observado previamente nos correspondentes modelos gerados por *CoMFA*, grupos substituintes relativamente pequenos nas posições 3, 4 e 5, do anel piridínico, deverão aumentar os valores da atividade inibitória dos compostos.

Observando-se a figura 77(c) verifica-se que os valores da atividade dos compostos podem ser aumentados com grupos hidrofóbicos posicionados nas regiões dos poliedros verdes (80% da contribuição) e/ou grupos hidrofílicos posicionados nas regiões dos poliedros vermelhos (20% da contribuição). A introdução de substituintes hidrofóbicos na posição 5 do anel piridínico e de substituintes hidrofílicos na posição 6 do anel piridínico e no NH_2 , do fragmento da tiossemicarbazona, devem aumentar os valores de pIC_{50} dos compostos. Por exemplo, o composto 14 ($5\text{-C}_2\text{H}_5$, $\text{pIC}_{50} = 6,66$) que possui um substituinte hidrofóbico é mais ativo em relação ao composto 27 ($5\text{-OC}_2\text{H}_5$, $\text{pIC}_{50} = 6,07$) que possui um substituinte menos hidrofóbico.

A inspeção visual da figura 77(d) sugere um aumento nos valores da atividade biológica quando grupos doadores de ligação de hidrogênio estão localizados nas regiões dos poliedros verdes (80% da contribuição) e/ou uma diminuição na atividade biológica quando estes grupos estão localizados nas regiões dos poliedros vermelhos (20% da contribuição). Na figura 77(e), observa-se um aumento no valor de pIC_{50} dos compostos quando grupos aceptores de ligação de hidrogênio estão localizados nas regiões dos poliedros verdes (80% da contribuição) e/ou um decréscimo na atividade quando estes grupos estão localizados nas regiões dos poliedros vermelhos (20% da contribuição). Grupos substituintes aceptores de ligação de hidrogênio próximos ao nitrogênio do anel piridínico devem aumentar o valor de pIC_{50} . Verifica-se, deste modo, uma concordância com os dados da literatura (French & Blanz 1966; Sartorelli *et al.* 1976). Nesta, sugere-se que a formação do complexo entre o átomo de ferro e o derivado da tiossemicarbazona é importante para que se observe a inibição da enzima RNR por estes compostos. O composto 28 ($5\text{-OC}_2\text{H}_4\text{N}(\text{CH}_3)_2$) que apresenta uma baixa atividade biológica ($\text{pIC}_{50}=4,62$) possui um grupo aceptor de ligação de hidrogênio em uma região desfavorável.

Os mapas de contorno estérico e eletrostático gerados por *CoMSIA* e por *CoMFA* estão em boa concordância. Observando-se a região correspondente ao campo aceptor de ligações de hidrogênio, pode-se verificar que a presença de átomos/grupos com esta característica próxima à região do N1, do anel piridínico e do enxofre favorecem a atividade dos compostos. Na região do poliedro verde, figura 77(e), deve-se esperar a presença do átomo de ferro que poderá ser complexado e que causará a inativação da enzima.

3.9 Proposição de novos compostos

Através das análises dos modelos *CoMFA* e *CoMSIA* gerados neste estudo, verificou-se que as abordagens *CoMSIA*, de uma forma geral, forneceram os valores relativamente mais significativos para os parâmetros estatísticos. Foram obtidos maiores valores do coeficiente de correlação, r^2 ; e conseqüentemente, menores valores do desvio padrão na predição dos erros, s , altos valores de (F) do teste de Fischer, indicando que os modelos obtidos são robustos, e valores significantes do quadrado do coeficiente de correlação da validação cruzada, Q^2_{cv} e conseqüentemente, menores valores do desvio padrão da autovalidação, s_{press} . Adicionalmente, foi possível constatar que, de uma maneira geral, a substituição na posição 5 do anel piridínico ocasiona uma melhoria nos valores de pIC_{50} do composto.

Esta observação pode ser explicada considerando-se que a maioria dos compostos, nos grupos de treinamento apresenta substituintes nesta posição. Uma análise dos dados apresentados na tabela 1, indica que as substituições introduzidas no fragmento da tiossemicarbazona não aumenta o valor da atividade biológica dos compostos, ao contrário, observa-se que os valores da atividade biológica diminuem.

Considerando-se os fatores antes mencionados, selecionou-se um dos modelos mais significativos, gerados por *CoMSIA*, ou seja, modelo obtido para a série de compostos substituídos apenas na posição 5 do anel piridínico, visando prever novas estruturas de compostos, ainda não estudados até o presente. Para este estudo foi selecionado o modelo 1, gerado por *CoMSIA*, para a série de 22 compostos (série E), grupo de treinamento IE, com o alinhamento realizado através dos campos estéricos e eletrostáticos (S_C).

De modo análogo, para os modelos gerados por *QSAR* clássico, foi escolhido um dos modelos mais significativos obtidos para a série de 22 compostos, expresso pela equação 244. Esta apresenta valor de $Q^2_{ass}=0,056$, sendo desta forma uma equação estatisticamente válida, como discutido anteriormente. Esta escolha foi realizada, considerando-se de modo análogo, os valores dos parâmetros estatísticos obtidos.

$$pIC_{50} = + 0.281 (\pm 0.24) Jhetv - 0.195 (\pm 0.086) SP13 + 0.135 (\pm 0.098) RDF050v - 0.204 (\pm 0.11) RDF065p + 7.280 (\pm 0.74)$$

(n = 22; r = 0.942; s = 0.217; F = 33.209; $Q^2_{cv} = 0.812$; $s_{press} = 0.279$) Equação 244

Deste modo, a tabela a seguir apresenta uma relação de possíveis substituintes propostos neste trabalho e, os respectivos valores para a atividade biológica na inibição da enzima RNR de células tumorais calculados pelos modelos selecionados.

Tabela 31. Série selecionada de grupos substituintes sugeridos por esta tese e pelos modelos gerados por *QSAR* e por *CoMSIA*, respectivamente, propostos nesta tese e os respectivos valores de pIC_{50} preditos através do modelo *QSAR* expresso pela equação 244 e pelo modelo 1E (*CoMSIA*), alinhamento S_C , para a série da 2-piridino carboxaldeído tiossemicarbazona.

Grupo Substituintes	Valores de pIC_{50} preditos pelo modelo <i>QSAR</i> proposto ^a	Valores de pIC_{50} preditos pelo modelo <i>CoMSIA</i> ^b proposto
5-CH ₂ OH	6,64	5,94
5-CH(CH ₃) ₂	6,40	6,38
5-CH(OCH ₃) ₂	5,18	5,24
5-CH ₂ CH ₂ CH ₃	6,67	6,22
5-CH ₂ CH ₂ Cl	7,14	6,32
5-CH ₂ COOH	6,43	5,61
5-CH ₂ N(CH ₃) ₂	6,61	5,91
5-CH ₂ NH ₂	6,36	6,14
5-CN	5,92	5,87
5-COOH	6,18	5,50
5-NH ₂	7,16	6,25

^a Valores preditos pelo modelo *QSAR* expresso pela equação 244

^b Valores preditos pelo modelo *CoMSIA* modelo 1, grupo de treinamento IE, alinhamento SEAL, para a série de 22 compostos

Tabela 32. Valores dos descritores DRAGON incluídos no modelo de *QSAR* expresso pela equação 244 para a série de substituintes selecionados nesta tese, visando a proposição de novos derivados para a série da 2-piridino carboxaldeído tiossemicarbazona.

Grupo Substituinte	Descritor Dragon			
	Jhetv	SP13	RDF050v	RDF06
CH ₂ OH	2,947	7,482	1,796	1,092
CH(CH ₃) ₂	2,945	7,911	2,892	3,796
CH(OCH ₃) ₂	2,166	8,735	8,764	4,767
CH ₂ CH ₂ CH ₃	2,040	8,639	10,396	2,002
CH ₂ CH ₂ Cl	2,040	8,600	1,768	2,448
CH ₂ COOH	2,901	8,553	2,561	1,703
CH ₂ N(CH ₃) ₂	2,073	8,440	1,537	6,120
CH ₂ NH ₂	2,041	7,437	8,451	2,004
CN	2,999	7,585	1,233	0,817
COOH	3,033	7,907	1,598	0,957
NH ₂	2,022	7,021	1,140	1,355

A seguir, os valores da atividade biológica preditos pelo modelo *QSAR* expresso pela equação 244, foram correlacionados com os valores da atividade biológica preditos pelo modelo 1 (*CoMSIA*), gerado para a série de 22 compostos, utilizando o grupo de treinamento 1E e o alinhamento sendo realizado através da similaridade dos campos estéricos e eletrostáticos (SEAL).

O gráfico dos valores de pIC_{50} preditos pelo modelo *QSAR* em função dos valores de pIC_{50} preditos pelo modelo *CoMSIA* está apresentado na figura a seguir.

$$n=11, r=0,76, SD=0,25$$

Figura 78 – Gráfico dos valores de pIC_{50} preditos pelo modelo *QSAR* expresso pela equação 244 em função dos correspondentes preditos pelo modelo *CoMSIA* (modelo 1, para a série de 22 compostos, grupo de treinamento 1E, com o alinhamento realizado através dos campos estéricos e eletrostáticos (SEAL), para a série da 2-piridino carboxaldeído tiossemicarbazona.

Ao se observar o gráfico da figura 78 pode-se notar uma correlação significativa entre os valores da atividade biológica preditos pelos modelos empregados de *QSAR* e *CoMSIA*.

Os compostos apresentando 5-NH₂ e 5-CH₂CH₂Cl como substituintes apresentam os maiores valores da atividade biológica nos dois modelos empregados para a predição. Ainda mais, considerando-se os mapas de contorno gerados por *CoMFA* e por *CoMSIA*, para a série da 2-piridino carboxaldeído tiossemicarbazona observa-se que substituintes na posição 5 não podem ser relativamente muito volumosos e devem possuir uma densidade de carga negativa, requisitos preenchidos por estes dois substituintes.

Desta forma, estes dois compostos parecem apresentar características promissoras como inibidores da enzima RNR de células tumorais e, potenciais inibidores da RNR do *M. tuberculosis*.

3.10 Estudos de modelagem por homologia da proteína

A modelagem da enzima ribonucleotídeo redutase, do *M. tuberculosis* foi realizada a partir da sua estrutura primária. Esta foi obtida pela procura sistemática da sequência primária no banco de dados do National Center for Biotechnology Information (NCBI) e, foi feita ajustando-se o campo de busca para “proteínas” e, incluindo-se “*Mycobacterium tuberculosis ribonucleotide reductase*” como palavras chaves.

Nesta primeira etapa, foram obtidas 35 sequências possíveis, das quais apenas 2 foram selecionadas para a modelagem, sendo elas a subunidade menor R2-1 (cadeia beta) da ribonucleotídeo redutase (código AAB81405) com 321 resíduos (Yang *et al.* 1997) e, a subunidade menor R2-2 (cadeia beta) da ribonucleotídeo redutase (código AAB81406) com 324 resíduos (Yang *et al.* 1997). Os sítios ativos contendo os átomos de ferro se encontram nestas duas subunidades (Yang *et al.* 1997; Liu *et al.* 1998). Além destas, observou-se apenas uma sequência da subunidade maior (cadeia alfa), a R1, (código AAA6444) (Yang *et al.* 1994) e uma possível sequência da segunda subunidade. Estas subunidades maiores não foram incluídas neste estudo de modelagem, considerando-se a incerteza associada à umas destas estruturas.

Para cada subunidade foram observadas as sequências de aminoácidos, descritas a seguir.

i) Subunidade menor R2-1 do *M. tuberculosis* (código **AAB81405**)

MTGKLVERVHAINWNRLDAKDLQVWERLTGNFWLPEKIPLSNDLASWQTLSSTEQ
QTTIRVFTGLTLLDTAQATVGAVAMIDDAVTPHEEAVLTNMAFMESVHAKSYSSIFS
TLCSTKQIDDAFDWSEQNPYLQRKAQIIVDYYRGDDAQAQRSSVMLESFLFYSYFLP
MYWSSRGKLTNTADLIRLIIRDEAVHGYIYGYKQQRGLADLTDAERADHREYTCCELL
HTLYANEIDYAHDLYDELGWTDDVLPYMRYNANKALANLGYQPAFDRDTCQVNPA
VRAALDPGAGENHDFFSGSGSSYVMGTHQPTTDDTDWDF

ii) Subunidade menor R2-2 do *M. tuberculosis* (código **AAB81406**)

MTGNAKLIDRVSAINWNRLQDEKDAEVWDRLTGNFWLPEKVPVSNIPSWGTLTAG
EKQLTMRVFTGLTMLDTIQGTVGAVSLIPDALTPHEEAVLTNIAFMESVHAKSYSQIF
STLCSTAEIDDAFRWSEENRNLQRKAEIVLQYYRGDEPLKRKVASTLLESFLFYSGFY
LPMYWSSRAKLTNTADMIRLIIRDEAVHGYIYGYKFQRGLALVDDVTRAEKDYTYE
LLFELYDNEVEYTQDLYDEVGLTEDVKKFLRYNANKALMNLGYEALFPRDETDVNP
AILSALSPNADENHDFFSGSGSSYVIGKAVVTEDDDWDF

Numa segunda etapa, a partir das sequências primárias foi realizada a procura das sequências secundárias homólogas no banco de dados do SWISS-PROT (Appel *et al.* 1994).

Como resultado foram encontradas 100 seqüências homólogas da ribonucleotídeo redutase do *M. tuberculosis*.

Cada seqüência secundária foi analisada minuciosamente, com o objetivo de se determinar quais seriam as relevantes. A seleção da seqüência foi realizada considerando-se dois critérios adicionais (i) apresentar pelo menos 25-30% de homologia com a seqüência da RNR do *M. tuberculosis*. (ii) apresentar estruturas tridimensionais depositadas em banco de dados específicos.

A necessidade de no mínimo de 25% de homologia se justifica, considerando-se que desta forma se assegura uma grande probabilidade das duas proteínas terem suas estruturas 3D semelhantes (Sander & Schneider 1991).

Nesta etapa, foram encontradas duas seqüências homólogas à RNR do *M. tuberculosis* que apresentam tanto um valor alto de homologia bem como apresentam estruturas 3D elucidadas e, os arquivos correspondentes depositados no PDB. Estes são identificados pelo correspondente número de arquivo, a saber: O69274 e P17424 com sequenciamentos e determinações estruturais realizados, respectivamente, para *C. ammoniagenes* (Fieschi *et al.* 1998) e para *S. typhimurium*. (Jordan *et al.* 1994). Comparando-se as seqüências de aminoácidos das RNRs do *M. tuberculosis* e do *C. ammoniagenes*, verifica-se que ambas apresentam 65,3% de identidade para a cadeia R2-1 e 75,9% para a cadeia R2-2. E, comparando-se as seqüências de aminoácidos das RNR do *M. tuberculosis* e da *S. typhimurium* verifica-se que ambas apresentam 64,7% de identidade para a cadeia R2-1 e 71,9% para a cadeia R2-2.

Os códigos dos arquivos extraídos do PDB para a modelagem da proteína foram, 1KGN e 1R2F, correspondentes aos dados cristalográficos das enzimas ribonucleotídeo redutases, respectivamente, do *C. ammoniagenes* (Hogbom *et al.* 2002) e da *S. typhimurium* (Eriksson *et al.* 1988).

As estruturas cristalográficas da RNRs da *Salmonella typhimurium* (código PDB 1R2F) e do *Corynebacterium ammoniagenes* (código PDB 1KGN) foram obtidas, respectivamente, com resoluções de 2,10Å e de 1,85Å. As determinações foram realizadas utilizando os valores de dimensões da cela e, de ângulos, seguintes: $a = 58,49$; $b = 71,74$ e $C = 96,07$ e, de $\alpha = 90,00$; $\beta = 95,01$ e $\gamma = 90,00$, para a primeira estrutura, enquanto para a segunda estes foram de $a = 50,81$; $b = 90,71$ e $c = 136,81$ e, de $\alpha = 90,00$; $\beta = 91,24$ e $\gamma = 90,00$.

Em seguida, foram alinhadas as seqüências, respectivamente extraídas do PDB com as do *M. tuberculosis*, acessando o site <http://www.ncbi.nlm.nih.gov/BLAST> e utilizando

programa disponível (Tatusova & Madden 1999) que compara a identidade entre duas seqüências de aminoácidos. Os resultados observados estão apresentados a seguir.

1 Ribonucleotídeo Redutase do *M. tuberculosis* – Cadeia A (R2-1)

2 Ribonucleotídeo Redutase do *C. Ammoniaenes* – Cadeia A (1KGN)

Identidade = 205/314 (65%), Positivos = 255/314 (81%)

```
Query: 9  VHAINWNRLLDKDLQVWERLTGNFWLPEKIPLSNDLASWQTLSSTEQQTIRVFTGLTL 68
          V AINWN + D KDL+VW+RLTGNFWLPEKIP+SND+ SW ++ EQ T+RVFTGLTL
Sbjct: 16  VKAINWNVIPDEKDLEVDRLTGNFWLPEKIPVSNDIQSWNKMTPEQLATMRVFTGLTL 75

Query: 69  LDTAQATVGAVAMIDDAVTPHEEAVLTNMAFMESVHAKSYSSIFSTLCSTKQIDDAFDWS 128
          LDT Q TVGA++++ DA T HEEAV TN+AFMESVHAKSYS+IF TL ST QI++AF WS
Sbjct: 76  LDTIQGTVGAIISLLPDAETMHEEAVYTNIAMFESVHAKSYSNIFMTLASTPQINEAFRWS 135

Query: 129  EQNPYLQRKAQIIVDYYRGDDAQAQR-SSVMLESFLFYSGFYLPYWSSRGKLTNTADLI 187
          E+N LQRKA+II+ YY GDD ++ +S +LESFLFYSGFYLPY SSR KLTNTAD+I
Sbjct: 136  EENENLQRKAKIIMSYNGDDPLKKKVASTLLESFLFYSGFYLPYMLSSRAKLTNTADII 195

Query: 188  RLIIRDEAVHGYIIGYKQRLADLTDAERADHREYTCCELLHTLYANEIDYAHDLDELG 247
          RLIIRDE+VHGYIIGYK Q+G+ L++AE+ +++ YT +L++ LY NEI+Y D+YD+LG
Sbjct: 196  RLIIRDESVMHGYIIGYKYQQGVKKLSEAEQEEYKAYTFDLMYDLYENEIEYTEDIYDDL 255

Query: 248  WTDDVLPYMRYNANKALANLGYQPAFDRDTCQVNPVRAALDPGAGENHDFFSGSGSSSYV 307
          WT+DV ++RYNANKAL NLGY+ F D +V+PA+ ++L P A ENHDFFSGSGSSSYV
Sbjct: 256  WTEDVKRFLRYNANKALNNLGYEGLFPTDETKVSPAILSSLSPNADENHDFFSGSGSSSYV 315

Query: 308  MGTHQPTTDDTDWDF 321
          +G + TTD DWDF
Sbjct: 316  IGKAEDTTDDDWDF 329
```

1 Ribonucleotídeo Redutase do *M. tuberculosis* – Cadeia B (R2-2)

2 Ribonucleotídeo Redutase do *C. Ammoniaenes* – Cadeia B (1KGN)

Identidade = 241/320 (75%), Positivos = 278/320 (86%)

```
Query: 5  AKLIDRVSAINWNRLQDEKDAEVWDRLTGNFWLPEKVPVSNNDIPSWGTLTAGEKQLTMRV 64
          A D V AINWN + DEKD EVWDRLTGNFWLPEK+PVSNNDI SW +T E+ TMRV
Sbjct: 10  ANHTDPVKAINWNVIPDEKDLEVDRLTGNFWLPEKIPVSNDIQSWNKMTPEQLATMRV 69

Query: 65  FTGLTMLDTIQGTVGAVSLIPDALTPHEEAVLTNIAFMESVHAKSYSQIFSTLCSTAEID 124
          FTGLT+LDTIQGTVGA+SL+PDA T HEEAV TNIAFMESVHAKSYS IF TL ST +I+
Sbjct: 70  FTGLTLLDTIQGTVGAIISLLPDAETMHEEAVYTNIAMFESVHAKSYSNIFMTLASTPQIN 129

Query: 125  DAFRWSEENRNLQRKAEIVLQYYRGDEPLKRVASTLLESFLFYSGFYLPYWSSRAKLT 184
          +AFRWSEEN NLQRKA+I++ YY GD+PLK+KVASTLLESFLFYSGFYLPY SSRKLT
Sbjct: 130  EAFRWSEENENLQRKAKIIMSYNGDDPLKKKVASTLLESFLFYSGFYLPYMLSSRAKLT 189

Query: 185  NTADMIRLIIRDEAVHGYIIGYKFQRLALVDDVTRAEKLDYTYELLFELYDNEVEYTD 244
          NTAD+IRLIIRDE+VHGYIIGYK+Q+G+ + + + E K YT++L+++LY+NE+EYTD
Sbjct: 190  NTADIIRLIIRDESVMHGYIIGYKYQQGVKKLSEAEQEEYKAYTFDLMYDLYENEIEYTED 249

Query: 245  LYDEVGLTEDVKKFLRYNANKALMNLGYEALFPRDETVDNPAISLSPNADENHDFFSG 304
          +YD++G TEDVK+FLRYNANKAL NLGYE LFP DET V+PAISL+LSPNADENHDFFSG
Sbjct: 250  IYDDLGWTEDEVKFLRYNANKALNNLGYEGLFPTDETKVSPAILSSLSPNADENHDFFSG 309
```


Query: 305 SGSSYVIGKAVVTEDDDWDF 324
 SGSSYVIGKA T DDDWDF
 Sbjct: 310 SGSSYVIGKAEDTTDDWDF 329

1 Ribonucleotídeo Redutase do *M. tuberculosis* – Cadeia A (R2-1)

2 Ribonucleotídeo Redutase da *S. typhimurium* – Cadeia A (1R2F)

Identidade = 205/317 (64%), Positivos = 247/317 (77%)

Query: 6 VERVHAINWNRLLDADKDLQVWERLTGNFWLPEKIPLSNDLASWQTLSSSTEQQTTIRVFTG 65
 + R+ AINWN++ D KDL+VW RLT NFWLPEK+PLSND+ +WQTL+ EQQ TIRVFTG
 Sbjct: 3 LSRISAINWNKIQDDKDLEVNRLTSNFWLPEKVPLSNDIPAWQTLAAEQQLTIRVFTG 62

Query: 66 LTLLDTAQATVGAVAMIDDAVTPHEEAVLTNMAFMESVHAKSYSSIFSTLCSTKQIDDAF 125
 LTLLDT Q GA +++ DA+TPHEEAVL+N++FME+VHA+SYSSIFSTLC TK++D A+
 Sbjct: 63 LTLLDTIQNIAGAPSLMADAITPHEEAVLSNISFMEAVHARSYSSIFSTLCQTKVEDAAY 122

Query: 126 DWSEQNPYLQRKAQIIVDYYRGDDAQQR-SSVMLESFLFYSGFYLPYMWSSRGKLTNTA 184
 WSE+NP LQRKAQII+ +Y D+ ++ +SV LESFLFYSGF+LPMY+SSRGKLTNTA
 Sbjct: 123 AWSEENPPLQRKAQIILAHYVSDEPLKKKIASVFLESFLFYSGFWLPMYFSSRGKLTNTA 182

Query: 185 DLIRLIIRDEAVHGYIIGYKQCRGLADLTDAERADHREYTCCELLHTLYANEIDYAHDLVD 244
 DLIRLIIRDEAVHGYIIGYK Q L L+ ER + + + +LL LY NEI Y LY
 Sbjct: 183 DLIRLIIRDEAVHGYIIGYKYQIALQKLSAIEREELKLFALDLLMELYDNEIRYTEALYA 242

Query: 245 ELGWTDDVLPYMRYNANKALANLGYQPAFDRDTCQVNPAPRAALDPGAGENHDDFFSGSGS 304
 E GW +DV ++ YNANKAL NLGY+ F + VNPA+ AAL P A ENHDDFFSGSGS
 Sbjct: 243 ETGWVNDVKAFLCYNANKALMNLGYEALFPPEMADVNPAILAALSPNADENHDDFFSGSGS 302

Query: 305 SYVMGTHQPTTDDTDWDF 321
 SYVMG T D DW+F
 Sbjct: 303 SYVMGKTVETEDEDWNF 319

1 Ribonucleotídeo Redutase do *M. tuberculosis* – Cadeia B (R2-2)

2 Ribonucleotídeo Redutase da *S. typhimurium* – Cadeia B (1R2F)

Identidade = 228/317 (71%), Positivos = 268/317 (84%)

Query: 8 IDRVSAINWNRLQDEKDAEVDRLTGNFWLPEKVPVSNDIPSWGTLTAGEKQLTMRVFTG 67
 + R+SAINWN++QD+KD EVW+RLT NFWLPEKVP+SNDIP+W TL+A E+QLT+RVFTG
 Sbjct: 3 LSRISAINWNKIQDDKDLEVNRLTSNFWLPEKVPLSNDIPAWQTLAAEQQLTIRVFTG 62

Query: 68 LTMLDTIQGTVGAVSLIPDALTPHEEAVLTNIAFMESVHAKSYSQIFSTLCSTAEIDDAF 127
 LT+LDTIQ GA SL+ DA+TPHEEAVL+NI+FME+VHA+SYS IFSTLC T E+D A+
 Sbjct: 63 LTLLDTIQNIAGAPSLMADAITPHEEAVLSNISFMEAVHARSYSSIFSTLCQTKVEDAAY 122

Query: 128 RWSEENRNLQRKAEIVLQYYRGDEPLKRVASTLLESFLFYSGFYLPYMWSSRAKLTNTA 187
 WSEEN LQRKA+I+L +Y DEPLK+K+AS LESFLFYSGF+LPMY+SSR KLTNTA
 Sbjct: 123 AWSEENPPLQRKAQIILAHYVSDEPLKKKIASVFLESFLFYSGFWLPMYFSSRGKLTNTA 182

Query: 188 DMIRLIIRDEAVHGYIIGYKFQRLALVDDVTRAELKDYTYELLFELYDNEVEYTDQDLYD 247
 D+IRLIIRDEAVHGYIIGYK+Q L + + R ELK + +LL ELYDNE+ YT+ LY
 Sbjct: 183 DLIRLIIRDEAVHGYIIGYKYQIALQKLSAIEREELKLFALDLLMELYDNEIRYTEALYA 242

Query: 248 EVGLTEDVKKFLRYNANKALMNLGYEALFPRDETVDNPAILSALSPNADENHDDFFSGSGS 307
 E G DVK FL YNANKALMNLGYEALFP + DVNPAIL+ALSPNADENHDDFFSGSGS
 Sbjct: 243 ETGWVNDVKAFLCYNANKALMNLGYEALFPPEMADVNPAILAALSPNADENHDDFFSGSGS 302

Query: 308 SYVIGKAVVTEDDDWDF 324
 SYV+GK V TED+DW+F
 Sbjct: 303 SYVMGKTVETEDEDWNF 319

Esta etapa de alinhamento permite verificar os resíduos que são estruturalmente equivalentes ao se considerar as características estruturais comuns. Os espaços vazios apresentados nas seqüências de alinhamento são os denominados “*gaps*” e, geralmente expressam as regiões das alças. Pode-se observar que devido a grande identidade entre as seqüências de aminoácidos a quantidade de “*gaps*” é relativamente baixa.

Considerando-se os resultados obtidos, verificou-se a homologia entre a proteína problema, que se quer determinar a estrutura tridimensional e, a do molde, cuja estrutura tridimensional já foi elucidada. Desta forma, os arquivos extraídos do PDB, respectivamente, para o *C. ammoniagenes* (1KGN) e para a *S. typhimurium* (1R2F) foram utilizados para a determinação da estrutura tridimensional da RNR do *M. tuberculosis*. No PDB não foram encontradas as estruturas tridimensionais das cadeias alfa em conjunto com as cadeias betas.

Apenas as cadeias betas foram modeladas, considerando-se que: (i) o sítio de ferro está justamente nesta cadeia; (ii) a formação e a estabilidade do radical tirosil dependem da presença do centro de ferro (Barlow *et al.* 1983; Sahlin *et al.* 1987) e, (iii) não havia informações sobre a estrutura primária de uma das cadeias alfa.

Para a realização da modelagem estrutural da proteína foi empregado o programa WHATIF (Vriend 1990).

Após a obtenção das estruturas de raios-X das RNRs da *S. typhimurium* e do *C. ammoniagenes* no formato PDB, ambas foram submetidas à averiguação de possíveis erros nas estruturas, nos arquivos de dados. Antes, porém, as moléculas de água foram retiradas dos arquivos.

O programa WHATIF automaticamente confere as informações contidas nos arquivos e classifica as informações em, respectivamente, notas, erros e advertências. As notas (observações) apresentam as informações sobre os dados das estruturas que estão sendo avaliadas, os erros indicam os problemas encontrados nas estruturas e as advertências apontam os possíveis erros ou falhas existentes nos modelos averiguados. Nesta etapa, o programa confere as estruturas e compara os dados com as informações contidas em um banco de dados pré-existente. Neste banco de dados, alguns valores são obtidos através do valor médio, não expressando todos os valores reais e possíveis. Desta forma, algumas informações sobre as estruturas 3D podem estar em discordância com os valores utilizados pelo programa WHATIF. Deve-se considerar que as estruturas 3D obtidas no banco de dados

expressam apenas uma única posição da proteína, obtidas em apenas um dado momento. No entanto, sabe-se que as proteínas não são sistemas estáticos e estão em constante movimento (Brooks *et al.* 1988).

Para a estrutura da *S. typhimurium* (1R2F) as advertências apontadas pelo programa WHATIF foram:

- i) Altos desvios no valor do ângulo diedro, foram observados sendo igual a 5,62, podendo indicar um refinamento sobreestimado, considerando-se que para estruturas bem refinadas os valores do desvio são $\sim 1,0$.
- ii) Falta dos átomos de oxigênios nos carbonos terminais da prolina 288 (cadeia A) e da leucina 286 (cadeia B). Este fato ocorreu por se tratar dos dois últimos fragmentos das cadeias A e B, respectivamente.
- iii) Possível problema no escalonamento das células. Dentre os vários problemas pode-se destacar o uso de cela diferente no refinamento e no cálculo da melhor cela; o valor de lambda utilizado para os dados sincrotron estavam descalibrados ou, os dados não foram corrigidos para um significativo movimento térmico anisotrópico. O programa automaticamente ajusta novos valores para as celas.
- iv) Amplitudes de dobramentos não usuais da prolina. Geralmente os valores da amplitude se situam entre 0,20 e 0,45. Nesta estrutura algumas prolinas estão com valores variando entre 0,47 a 0,50, indicando uma desordem entre a forma de dois anéis.
- v) Fases de dobramentos não usuais da prolina. Os anéis da prolina geralmente estão na denominada conformação de envelope, com o átomo de carbono gama acima do plano do anel ($\Phi=+72^\circ$) ou conformação de meia cadeira com o átomo de carbono gama abaixo e o carbono beta acima do plano do anel ($\Phi=90^\circ$). Se os valores ficarem muito distantes há indícios de uma conformação errada. A estrutura apresentou os ângulos gama da PRO(147)A= -108° , da PRO(280)A= -108° e da PRO(147)B= -108° .
- vi) Alguns ângulos torsionais apresentaram valores não esperados, sendo que o número de resíduos problemáticos não ultrapassou 5%, valor esperado. No total apenas 6 aminoácidos possuem o ângulo Phi/Psi não esperado.
- vii) Ângulos Omega torsionados. Os valores dos ângulos Omega para trans-peptídeos em uma estrutura seguem uma distribuição gaussiana com um valor médio de $+178^\circ$ e, um desvio padrão de $\sim 5,5^\circ$. Nesta estrutura obteve-se um desvio padrão de 1,06 indicando uma forte torção angular.

- viii) Ambiente de dobramento não esperado. Um valor baixo pode significar um dobramento pobre, contato com o cristal, contato com um co-fator ou, o resíduo está posicionado no sítio ativo. Os resíduos que apresentaram problemas foram a GLN (15)B, a GLN (15)A, a ILE (83)A e o GLU (127)B.
- ix) Alguns fragmentos de aminoácidos apresentaram problema nos grupos doadores de ligações de hidrogênio. Em estruturas bem refinadas não se encontra este problema.

Para a estrutura da *S. typhimurium* (1R2F) os erros apontados pelo programa WHATIF foram:

- i) Alguns átomos possuem uma distância interatômica mais curta do que a esperada.
- ii) Movimento da cadeia lateral da histidina, da asparagina e da glutamina. Alguns destes aminoácidos apresentam orientação da cadeia lateral diferente dos dados de entrada. Estas cadeias podem ficar mais estáveis através da rotação de 180° do grupo terminal para se realizar interações de hidrogênio. Os aminoácidos que apresentaram problemas foram: GLN(53)A, ASN(71)A, ASN(128)A, GLN(132)A, ASN(232)A, ASN(264)A, ASN(279)A, ASN(29)B, GLN(53)B, GLN(70)B, ASN(71)B, ASN(128)B, GLN(136)B, ASN(180)B, GLN(208)B, ASN(232)B, ASN(259)B e ASN(279)B.

Para a estrutura do *C. ammoniagenes* (1KGN) as advertências apontadas pelo programa WHATIF foram:

- i) Problema de nomenclatura com a arginina 196A, que possui o N-H-1 e N-H-2 trocados.
- ii) As tirosinas listadas não possuem o valor de chi-2 entre -90,0 e 90,0. TYR(173)A, TYR(231)A, TYR(5)B, TYR(181)B e TYR(231)B.
- iii) As fenilalaninas listadas não possuem o valor de chi-2 entre -90,0 e 90,0. PHE(233)B e PHE(263)B.
- iv) Os ácidos aspárticos listados não possuem o valor de chi-2 entre -90,0 e 90,0. ASP(234)A, ASP(284)A e ASP(91)B.
- v) Os ácidos glutâmicos listados não possuem o valor de chi-2 entre -90,0 e 90,0. GLU(7)A, GLU(136)B e GLU(202)B.
- vi) Altos desvios no valor do ângulo diedro, como observado para a *S. typhimurium* (1R2F) (item i).

- vii) Ausência dos átomos de oxigênios nos carbonos terminais da serina 297 das cadeias A e B.
- viii) Possível problema no escalonamento das células, como observado para *S. typhimurium* (1R2F) (item iii).
- ix) Ângulos de ligação não usuais para TYR(173)A e HIS(111)B.
- x) Fases de dobramentos não usuais da prolina. Os anéis da prolina geralmente estão na denominada conformação de envelope, com o átomo de carbono gama acima do plano do anel ($\Phi=+72^\circ$) ou conformação de meia cadeira com o átomo de carbono gama abaixo e o carbono beta acima do plano do anel ($\Phi=90^\circ$). Se os valores ficarem muito distantes há indícios de uma conformação errada. A estrutura apresentou o ângulo gama da PRO(179)A= -108° .
- xi) Ângulo torsional não esperado. A ARG(185)A possui o valor do Z-score= $-2,41$, considerado um valor pobre. Isto significa que pode haver resíduos deste aminoácido em posições desfavoráveis.
- xii) Ângulos Omega torsionados, como observado para a *S. typhimurium* (1R2F) (item vii). Para esta estrutura obteve-se um desvio padrão de 1,89 indicando uma forte torção angular.
- xiii) Alguns fragmentos possuem hidrogênio mas não são doadores de ligações de hidrogênio.
- xiv) Alguns fragmentos não são aceptores de ligação de hidrogênio.

Para a estrutura do *C. ammoniagenes* (1KGN) os erros apontados pelo programa WHATIF foram:

- i) Alguns átomos possuem uma distância interatômica mais curta do que a esperada.
- ii) Movimento da cadeia lateral da histidina, asparagina e glutamina, como observado para a *S. typhimurium* (1R2F) (item ii). Os aminoácidos que apresentaram problemas foram: ASN(190)A, ASN(273)A, ASN(11)B, ASN(39)B, ASN(50)B, ASN(56)B, ASN(103)B, GLN(127)B, ASN(140)B, ASN(153)B e ASN(274)B.

Após a determinação dos erros correspondentes aos movimentos das cadeias laterais da histidina, asparagina e glutamina, estes foram verificados visualmente com o emprego do programa Sybyl. Os fragmentos dos aminoácidos que apresentaram problemas estão situados em regiões superficiais da enzima. Desta forma, foram considerados os dados originais dos arquivos extraídos do PDB, pois não deverão afetar significativamente o modelo a ser

construído. Um possível erro nesta etapa pode ser atribuído a nomeação dos átomos nestes fragmentos. No caso da histidina, pode ter ocorrido uma troca entre os nitrogênios do aminoácido e no caso da asparagina e da glutamina pode ter ocorrido uma troca entre o nitrogênio da amina terminal e o oxigênio da carbonila. Esta inversão é comum nos dados de cristalografia de raios-X considerando que o método determina apenas a posição dos átomos mais pesados.

Os gráficos de Ramachandran obtidos para a estrutura cristalográfica da *S. typhimurium* (1R2F) cadeias A e B e para o *C. ammoniagenes* (1KGN) cadeias A e B estão mostrados nas figuras a seguir.

Figura 79. Gráficos de Ramachandran obtidos para as cadeias A e B da ribonucleotídeo redutase da (a) *S. typhimurium* (1R2F) e do (b) *C. ammoniagenes* (1KGN). Os sinais de X representam as glicinas, os □ representam as prolínas e os sinais de + representam os outros resíduos.

A existência de vários sinais de + fora da área de contorno implica em uma estrutura mal refinada. Como pode ser observado, em todos os gráficos de Ramachandran, existem poucos pontos fora da área de contorno, indicando desta forma que o modelo possui um bom refinamento.

Na etapa seguinte, as seqüências das proteínas foram convertidas para o formato PIR, muito semelhantes ao formato BLAST. Esta conversão se faz necessária para que o programa WHATIF realize a modelagem das proteínas. Esta conversão foi realizada com o emprego do programa Sybyl.

A estrutura tridimensional da proteína molde e o seu respectivo arquivo no formato PIR foram adicionados no programa WHATIF juntamente com a seqüência de aminoácidos a

ser modelada. Automaticamente o programa vai construindo um novo modelo. Primeiramente a modelagem foi efetuada utilizando-se como molde a estrutura da *S. typhimurium* e posteriormente a estrutura do *C. ammoniagenes*. Nesta etapa foram modeladas as regiões estruturalmente conservadas e as regiões de alças. A modelagem das regiões de alças é uma das etapas mais importante na modelagem por homologia. Neste trabalho, a predição das regiões de alças foi realizada através da comparação com os fragmentos homólogos das estruturas moldes. Considerando a alta homologia existente entre os moldes e a estrutura a ser modelada, foram obtidos bons resultados nesta etapa. Assim como as regiões de alças, as cadeias laterais também foram modeladas a partir das informações contidas nas estruturas utilizadas como molde.

Após a modelagem da proteína foram efetuados os estudos para a validação dos modelos obtidos. Esta etapa é muito importante e visa avaliar a qualidade do empacotamento global da proteína, os possíveis erros estruturais e os parâmetros estereoquímicos. O modelo obtido depende diretamente da estrutura molde que foi utilizada, sendo que a resolução desta deve ser superior a 2Å e o fator-R inferior a 20%. Neste estudo, as estruturas das proteínas moldes foram obtidas com uma resolução de 2,10Å e 1,85Å e fator-R de 22,6% e 15,8%, respectivamente para a *S. typhimurium* e para o *C. ammoniagenes*.

Em seguida, utilizando o programa WHATIF, foram avaliados os erros referentes aos valores dos ângulos; aos valores dos comprimentos das ligações, aos valores dos ângulos torsionais, aos doadores e aceptores de ligações de hidrogênio, às nomenclaturas das cadeias e dos átomos, às conformações das cadeias laterais da histidina, glutamina e asparagina, às planaridades das cadeias laterais, à simetria, ao empacotamento da prolina, aos contatos atômico direcional, às ocupações atômicas e às rotações da ligação peptídica.

Para o modelo obtido utilizando-se o dado cristalográfico da ribonucleotídeo redutase elucidado para a *S. typhimurium* os erros encontrados na averiguação foram:

- i) O ângulo da cela está fora do intervalo de 25° a 155°. Provavelmente um ou mais valores podem estar incorretos. Este erro pode ter sido gerado devido a discrepância entre os valores considerados na medida experimental e os valores que estão no arquivo do PDB. Neste caso, o programa automaticamente sugere os valores corretos para serem utilizados.

A ribonucleotídeo redutase modelada com o dado cristalográfico do *C. ammoniagenes* apresentou os seguintes erros:

- i) O coeficiente de Matthews apresentou um valor muito alto. Este coeficiente é definido como sendo a densidade da estrutura da proteína em Ângstrons ao cubo por Dalton. Este erro geralmente ocorre ao se fornecer valor errados para a coordenada Z.
- ii) Algumas histidinas, asparaginas e glutaminas apresentaram problemas na orientação das suas respectivas cadeias laterais. Como discutido anteriormente, este erro é decorrente da interpretação dos dados de cristalografia de raios-X que elucida apenas a posição dos átomos mais pesados. Uma rotação de 180° nestas cadeias laterais irá fornecer uma conformação energeticamente mais estável.

Os erros apontados pelo programa WHATIF, nas estruturas dos modelos das RNR propostas, são análogos aos verificados pela análise das estruturas retiradas diretamente do PDB, utilizadas como molde.

Em seguida, as estruturas dos modelos da proteína RNR do *M. tuberculosis* propostos foram otimizados através da mecânica molecular com o emprego do programa SYBYL. Nesta etapa, o número de ciclos da otimização foi restringido para que apenas as interações desfavoráveis fossem minimizadas sem correr o risco de que a proteína se desviasse excessivamente de sua conformação. Para esta minimização adotou-se apenas 200 ciclos.

Adicionalmente foram averiguados os gráficos de Ramachandran para os dois modelos propostos.

Figura 80. Gráficos de Ramachandran para as proteínas modeladas utilizando-se as estruturas cristalográficas das RNRs da (a) *S. typhimurium* e do (b) *C. ammoniagenes* como moldes. As linhas no gráfico indicam as áreas preferidas. Os sinais de X representam as glicinas, os □ representam as prolinas e os sinais de + representam os outros resíduos.

Observa-se através dos gráficos de Ramachandran que poucos resíduos estão em áreas energeticamente desfavoráveis, isto é, fora das linhas, indicando a boa qualidade dos modelos propostos.

Alinhando-se as geometrias dos modelos propostos com os dos respectivos moldes, pode-se obter o valor do desvio quadrático médio (RMS). Os valores dos RMS determinados entre as estruturas propostas e as respectivas estruturas 3D da RNR da *S. typhimurium* (1R2F) e do *C. ammoniagenes* (1KGN) foram iguais a 0,15Å e 0,56 Å. As figuras a seguir mostram os alinhamentos efetuados entre os modelos obtidos para *M. tuberculosis* e as estruturas 3D utilizadas como molde, respectivamente.

Figura 81. Alinhamentos resultantes entre os modelos propostos para a RNR do *M. tuberculosis*, através da modelagem por homologia, e os seus respectivos moldes. As estruturas em cor azul indicam os modelos propostos e as estruturas em vermelho indicam os moldes utilizados: (a) estrutura 3D da *S. typhimurium* (código PDB 1R2F) e (b) estrutura 3D do *C. ammoniagenes* (código PDB 1KGN).

A seguir, os modelos propostos da enzima RNR, obtidos através da modelagem por homologia, foram comparados através do valor desvio quadrático médio (RMS) com a estrutura cristalográfica da cadeia beta (subunidade menor) da RNR do *M. tuberculosis*, recentemente elucidada (Uppsten *et al.* 2004). A estrutura cristalográfica da enzima RNR, disponível no PDB (código 1UZR), é formada por três monômeros na unidade assimétrica, sendo um dímero e um monômero, formando uma interação dimérica. A forma do dímero R2 obtido da *E. coli* pode ser comparada a um coração, onde cada monômero forma um dos lobos e duas fitas β formam a extremidade do coração. Estas duas fitas β estão ausentes em todas as estruturas R2 elucidadas até o presente momento, incluindo a R2F-2 do *M. tuberculosis*.

Os valores observados para o RMS entre o modelo da RNR do *M. tuberculosis*, proposto na modelagem por homologia, com os dados cristalográficos da RNR do *M. tuberculosis* foram de 1,04Å e de 0,83Å, considerando-se como moldes os dados

cristalográficos, respectivamente, da *S. typhimurium* e do *C. ammoniagenes*. As figuras a seguir mostram os alinhamentos realizados entre as estruturas das proteínas modeladas e a estrutura da cadeia beta (subunidade menor) da RNR do *M. tuberculosis* obtida experimentalmente (Uppsten *et al.* 2004).

Figura 82. Alinhamento resultante entre as estruturas da RNR do modelo proposto através da modelagem por homologia (estrutura em cor azul), utilizando-se como molde os dados cristalográficos da RNR da *S. typhimurium* (código PDB 1R2F), e a RNR do *M. tuberculosis* elucidada recentemente (código PDB 1UZR), (estrutura em cor vermelha).

Figura 83. Alinhamento resultante entre as estruturas da RNR do modelo proposto através da modelagem por homologia (estrutura em cor azul), utilizando-se como molde os dados cristalográficos da RNR do *C. ammoniagenes* (código PDB 1KGN), e a RNR do *M. tuberculosis* elucidada recentemente (código PDB 1UZR), (estrutura em cor vermelha).

As regiões em cor vermelha, apresentada nas figuras 83 e 84, sem o correspondente alinhamento, pertence ao monômero, enquanto que as outras regiões em vermelho com o correspondente alinhamento em azul, pertencem ao dímero.

Os dois átomos de ferro na R2F-2 estão ligados por duas histidinas (His106 e His200), um aspartato (Asp72), três glutamatos (Glu103, Glu163 e Glu197) e, uma molécula de água.

Nas proximidades deste sítio também se encontra uma tirosina (Tyr110). As moléculas de glutamato 103 e 107 estão ligadas com os dois átomos de ferro através de uma ponte e o glutamato 163 se liga de forma bidentada ao Fe2. A molécula de água liga-se apenas ao Fe2 tornando-o hexa-coordenado, com forma piramidal-quadrada, ao contrário do Fe1 que possui apenas quatro ligantes, criando uma coordenação tetraédrica. Esta coordenação é idêntica na estrutura R2F reduzida da *S. typhimurium* ao passo que na estrutura R2F reduzida do *C. ammoniagenes* não existe a molécula de água ligando-se ao Fe2. A figura a seguir mostra o centro de ferro reduzido e os aminoácidos coordenados ao ferro observado no modelo utilizando-se como molde a estrutura cristalográfica da *S. typhimurium*. Resultados semelhantes foram obtidos ao se utilizar a estrutura cristalográfica do *C. ammoniagenes* como molde.

Figura 84. Esquema do sítio de ferro reduzido na R2F-2 para a RNR do *M. tuberculosis*, respectivamente, (a) modelada por homologia, utilizando-se a estrutura cristalográfica da *S. typhimurium* (código PDB 1R2F) e (b) obtida experimentalmente através da cristalografia de raios-X (Uppsten *et al.* 2004).

Verifica-se pela inspeção visual da figura 85 que o sítio dos átomos de ferro, do modelo proposto, apresenta uma boa concordância com os dados de cristalografia de raios-X.

As duas histidinas (His106 e His200), o aspartato (Asp72) e os dois glutamatos (Glu103 e Glu197) estão em posições análogas na estrutura do modelo proposto e na estrutura obtida experimentalmente (Uppsten *et al.* 2004). Apenas o glutamato 163 está ligeiramente deslocado de sua posição, em comparação com os dados experimentais, embora interagindo da mesma forma o átomo de Fe2. Considerando-se que as moléculas de água foram retiradas das análises, não foi possível verificar as interações que a mesma realiza no sítio do átomo de ferro.

O radical tirosina é de grande importância para a enzima RNR. Estudos realizados com a RNR de mamíferos demonstraram que a inibição desta enzima, utilizando a

1-formilisoquinolina tiossemicarbazona (IQ-1), ocorre no radical livre da tirosina, situada na subunidade M2 (Thelander & Gräslund 1983). O radical é destruído pelo complexo do IQ-1 com o ferro em uma reação que necessita oxigênio. A necessidade do oxigênio demonstra que o mecanismo de inibição não se deve apenas à remoção do átomo de ferro pelo ligante de seu sítio ativo na enzima.

O radical tirosina está localizado em um sítio hidrofóbico, próximo aos átomos de ferro. O estudo de cristalografia de raios-X da R2F-2 do *M. tuberculosis* (Uppsten *et al.* 2004) mostra que o radical tirosina, Tyr110, está situado a uma distância de 6,8Å do átomo de ferro mais próximo. Para o modelo proposto, considerando-se como molde a estrutura cristalográfica da *S. typhimurium*, verifica-se que a distância entre o radical tirosina e o átomo de ferro mais próximo é de 6,5Å. Considerando-se o modelo proposto, utilizando-se a estrutura cristalográfica do *C. ammoniagenes*, esta distância é menor, 5,2Å. Novamente, pode-se observar que os modelos obtidos estão em boa concordância com os resultados obtidos experimentalmente, principalmente se for considerado a estrutura do modelo proposto a partir dos dados da *S. typhimurium*.

De acordo com dados de literatura, o mecanismo de inibição da enzima RNR de células tumorais pode ocorrer tanto através da coordenação do átomo de ferro na enzima bem como através da formação do complexo ligante-ferro que irá interagir posteriormente com a enzima (Finch *et al.* 1999).

Na tentativa de se sugerir um possível mecanismo para a inibição da RNR do *M. tuberculosis*, para esta classe de compostos, o sítio de ferro modelado foi minuciosamente analisado. Como resultado, sugere-se que um dos possíveis mecanismos envolveria a formação do complexo ligante-ferro no sítio do átomo de ferro. Desta forma, avaliou-se a ocupação estérica, através do valor do volume de van der Waals, nas proximidades dos dois átomos de ferro. A figura a seguir ilustra a distribuição estérica no sítio dos átomos de ferro para o modelo proposto por homologia da enzima RNR do *M. tuberculosis*, utilizando-se os dados cristalográficos da RNR da *S. typhimurium*, como molde. Resultados análogos foram obtidos considerando-se o modelo da RNR obtido através da homologia, utilizando-se os dados cristalográficos da RNR do *C. ammoniagenes*, como molde.

Figura 85. Representação esquemática da distribuição das características estéricas no sítio dos átomos de ferro para o modelo proposto por homologia da RNR do *M. tuberculosis*, empregando-se os dados cristalográficos da *S. typhimurium*, como molde. Os átomos de ferro estão apresentados na cor amarela.

Neste contexto, em uma estação de trabalho, submeteu-se à rotações e à translações os aminoácidos presentes na região dos átomos de ferro, do modelo proposto da RNR do *M. tuberculosis*. A inspeção visual dos resultados sugere que a região dos átomos de ferro é estericamente impedida. Ou seja, a série de derivados da tiossemicarbazona, estudada neste trabalho, é relativamente volumosa para poder se alocar neste sítio e complexar com os átomos de ferro, inibindo a enzima RNR. Deve-se ressaltar que estes resultados foram obtidos para modelos gerados nas condições estabelecidas nesta tese, considerando sistemas estáticos. Estes resultados indicam que a formação do complexo ligante-ferro não ocorrerá dentro do sítio do átomo de ferro.

Neste estudo, infelizmente, não foi possível realizar a modelagem das subunidades maiores da enzima RNR do *M. tuberculosis*, devido a indisponibilidade ou inexistência de uma de suas seqüências, bem como de informações de cristalografia de raios-X de proteínas homólogas. Estas limitações, impossibilitaram a realização dos estudos de *docking*, pois a região do radical tirosina envolve um fragmento da subunidade maior.

3.11 Preparação dos compostos e ensaios biológicos

3.11.1 Preparação do cloridrato de tiossemicarbazida

Procedendo-se de acordo com o método descrito no item 2.9.1, após a total dissolução de 5,99 g (75,8 mmol) de tiossemicarbazida em uma solução etanol-água (1:1) esta foi submetida à passagem de gás clorídrico por 30 minutos e deixada em repouso no equipamento por mais 30 minutos. O produto desta reação foi cristalizado de uma solução etanol-água (1:1). Foram obtidos 3,49g de um material sólido branco.

$$\eta=39,75 \%$$

$$p.f._{exp}=149^{\circ}-151^{\circ}$$

$$p.f._{lit}=175^{\circ}-177^{\circ} \text{ (West 1971-1972)}$$

A análise elementar, RMN ^1H e RMN ^{13}C não foram realizadas para este composto intermediário.

3.11.2 Preparação do benzenocarboxaldeidotiossemicarbazona

Procedendo-se de acordo com o método descrito no item 2.9.2, partiu-se de 10,0 mmol de benzaldeído diluído em 20,0 mL de etanol tratado e seco e de 10,0 mmol de cloridrato de tiossemicarbazida dissolvida em 40 mL de água destilada. O produto da reação foi destilado de etanol-água. Foram obtidos 0,63g de um material sólido branco e cristalino.

$$\eta=37,7\%$$

$$p.f._{exp}=160-161^{\circ}\text{C}$$

$$p.f._{lit}=160^{\circ}\text{C} \text{ (Sah \& Daniels 1950)}$$

Análise Elementar:

Calculada: C(%) 53,61 H(%) 5,06 N(%) 23,44

Experimental: C(%) 53,91 H(%) 5,25 N(%) 23,49

RMN ^1H (DMSO; ref. TMS):

δ (ppm): 7,38 a 8,22 (m; ArH; NH₂; HC=N; 8H) e 11,44 (s; NH; 1H)

3.11.3 Preparação da 2-piridino-carboxaldeído tiossemicarbazona

Procedendo-se de acordo com o método descrito no item 2.9.3, partiu-se de 10,0 mmol de 2-piridino carboxaldeído diluído em 20,0 mL de etanol tratado e seco e de 10,0 mmol de cloridrato de tiossemicarbazida dissolvida em 40 mL de água destilada. O produto da reação foi destilado de etanol-água. Foram obtidos 0,87g de um material sólido amarelo.

$$\eta=48,3\%$$

$$p.f._{exp}=119-121^{\circ}C$$

$$p.f._{lit}=197-198^{\circ}C \text{ (Balaban } et al. 2003)$$

Análise Elementar:

Calculada: C(%) 46,65 H(%) 4,47 N(%) 31,09

Experimental: C(%) 46,27 H(%) 4,37 N(%) 31,83

RMN 1H (DMSO; ref. TMS):

δ (ppm): 7,80 a 8,80 (m; ArH; NH₂; HC=N; 7H) e 12,17 (s; NH; 1H)

3.11.4 Ensaios biológicos

Os ensaios microbiológicos foram realizados pela Dra. Ana Olívia de Souza sob orientação do Prof. Dr. Célio Lopez Silva, no Departamento de Bioquímica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

Os dois compostos por nós preparados, o benzenocarboxaldeído tiossemicarbazona (HBTSC) e a 2-piridino carboxaldeído tiossemicarbazona (H2pyTSC) foram ensaiados contra o *M. tuberculosis* H37Rv. Neste estudo foram avaliados os valores da concentração inibitória mínima (CIM), sendo que o *M. tuberculosis* H37Rv foi cultivado em meio de cultura na ausência de antibiótico e na presença das amostras em teste. A cultura na ausência de antibiótico foi considerada como 100% de crescimento bacteriano e serviu como controle.

Figura 86. Gráfico da correlação entre a concentração dos compostos, em $\mu\text{mol.L}^{-1}$ e (a) viabilidade bacteriana, em porcentagem e (b) morte celular, em porcentagem, avaliados para o *M. tuberculosis* H37Rv.

Na figura 86, pode-se observar que o composto H2pyTSC é mais ativo em relação ao composto HBTSC. Estes resultados concordam com a literatura (French & Blanz 1966). Para os compostos denominados HBTSC e H2pyTSC verifica-se que foram necessárias doses maiores, respectivamente, que $800\mu\text{mol.L}^{-1}$ e $310\mu\text{mol.L}^{-1}$, para que se observe a morte celular de 100% do *M. tuberculosis*. E, os valores de IC_{50} (concentração do composto capaz de inibir 50% do crescimento bacteriano) observados para HBTSC e H2pyTSC são iguais a $332\mu\text{mol.L}^{-1}$ e $104\mu\text{mol.L}^{-1}$, respectivamente.

O fato destes dois compostos, por nós sintetizados, apresentarem atividade frente o *M. tuberculosis* nos pareceu muito promissor pois os testes de atividade foram realizados *in vivo*. Assim, pretendíamos, a seguir, verificar o seu comportamento no sistema isolado. Com este objetivo e, financiamento do projeto Instituto do Milênio (coordenado pelo Prof. Célio Lopes) a Dra. Ana Olívia estagiou nos laboratórios do Prof. Harvey Hubin, na Universidade da Pensilvânia, especialista na área (Elleingand *et al.* 1998; Avarbock *et al.* 1999; Dawes *et al.* 2003; Uppsten *et al.* 2004; Weinstein *et al.* 2005). Lá se buscou adquirir os conhecimentos necessários para o isolamento da RNR para depois poder realizar os testes *in vitro*. Entretanto, foram enormes as dificuldades encontradas tanto para se isolar como para padronizar o teste de inibição da enzima, tornando-o, até o presente, inviável. Alternativamente, foram realizados os ensaios *in vivo*, determinado-se os valores de CIM para dois compostos.

4. CONCLUSÕES

Através da análise de *QSAR* clássico aplicada a série de 40 derivados do α -(N)-piridino carboxaldeído tiossemicarbazonas substituídas, selecionados da literatura, com atividade inibitória frente à enzima RNR de células H.Ep.-2 (carcinoma humano da epiderme), constatamos que:

1. O emprego das análises *PLS* para a pré-seleção dos descritores obtidos com o emprego do programa DRAGON auxiliou a determinação dos descritores significativamente mais importantes para descrever a atividade biológica desta classe de compostos.
2. A utilização de três grupos de treinamento e dos respectivos grupos de teste auxiliou a obtenção de modelos *QSAR* robustos.
3. A utilização do Q^2 assintótico auxiliou a determinação dos modelos que possuem poder de predição.
4. Foi verificado uma grande contribuição do efeito estérico para atividade inibitória da RNR de células tumorais, para esta classe de compostos. A atividade biológica depende do tamanho do composto, sendo que este não pode possuir grupos substituintes relativamente volumosos.
5. Foi verificado que a presença de substituintes muito polares diminuem o valor da atividade biológica destes compostos.
6. Os resultados destas análises mostraram a importância das distâncias entre alguns átomos para a atividade dos compostos. Estes resultados não foram observados nos modelos de *QSAR-3D*.

Os dois métodos de alinhamentos empregados nesta tese, forneceram resultados equivalentes. Em ambos os casos, os compostos se alinharam através dos átomos pesados da piridino e dos pertencentes ao fragmento da tiossemicarbazonas.

Através das abordagens *CoMFA* e *CoMSIA* empregadas para a série de 40 derivados do α -(N)-piridino carboxaldeído tiossemicarbazonas substituídas, com atividade inibitória frente à enzima RNR de células H.Ep.-2 (carcinoma humano da epiderme), constatamos que:

CoMFA

1. Para todas as análises realizadas, a contribuição do fator estérico (~60-70%) para o valor da atividade inibitória da enzima RNR de células tumorais demonstrou ser maior do que a contribuição do fator eletrostático (~30-40%).

2. Assim como os resultados obtidos no *QSAR* clássico, os grupos substituintes não podem ser relativamente volumosos para que os compostos apresentem uma altos valores de atividade biológica.
3. Grupos com grande densidade eletrônica na posição 5 do anel piridínico devem apresentar um alto valor de atividade biológica.
4. Os mapas de contorno gerados pelos modelos CoMFA mostram as regiões permitidas e as regiões proibidas para a presença de grupos substituintes. Adicionalmente, através destes mapas de contorno foi possível verificar as características eletrônicas e estruturais necessárias para aumentar o valor da atividade biológica dos compostos estudados.

CoMSIA

1. Como observado através dos resultados das análises CoMFA, os resultados das análises CoMSIA mostram que a presença de grupos com grande densidade eletrônica na região da posição 5 do anel piridínico deve favorecer a atividade dos compostos.
2. O átomo de nitrogênio do anel piridínico é importante para a atividade inibitória da RNR. Este, juntamente com o enxofre e o átomo de nitrogênio, ligado diretamente ao C1' do fragmento da tiossemicarbazona, devem formar o complexo tridentado com o átomo de ferro acarretando a inativação da enzima.
3. Os grupos substituintes do anel piridínico não devem ser relativamente volumosos, o que diminuiria o valor da atividade biológica.
4. Grupos substituintes hidrofóbicos na posição 5 do anel piridínico e hidrofílicos na região do -NH₂, do fragmento da tiossemicarbazona, devem aumentar o valor da atividade biológica dos compostos.
5. A presença de grupos aceptores de ligação de hidrogênio na região do átomo de nitrogênio, do anel piridínico, e do átomo de enxofre, do fragmento da tiossemicarbazona, devem aumentar o valor da atividade biológica dos compostos. Estes resultados estão em boa concordância com os dados da literatura.

A análise dos resultados obtidos através dos modelos *QSAR/QSAR-3D* gerados forneceu o perfil das características físico-químicas necessárias para que os compostos tenham altos valores de atividade inibitória da enzima RNR de células tumorais. Baseando-se nestes resultados, foram propostos “*in silico*” novos grupos substituintes que atendessem tais

características. Como uma primeira abordagem, estes compostos tiveram os valores das respectivas atividades biológicas preditas através dos melhores modelos obtidos nos estudos de *QSAR* clássico e de *CoMSIA*.

O gráfico dos valores de pIC50 preditos pelo modelo de *QSAR* clássico *versus* preditos pelo modelo CoMSIA apresentou uma correlação significativa, indicando que estes compostos podem ser promissores inibidores da RNR de células tumorais e, possíveis inibidores da RNR do *M. tuberculosis*.

Adicionalmente, os dois compostos por nós preparados foram ensaiados, na Faculdade de Medicina de Ribeirão Preto – USP, pela Dra. Ana Olívia de Souza e sob a responsabilidade do Prof. Dr. Célio Lopez Silva, contra o *M. tuberculosis* H37Rv, e apresentaram atividade *in vivo*.

Os estudos de modelagem da proteína, mais especificamente da cadeia beta (subunidade menor) da RNR do *M. tuberculosis*, adotando-se como moldes as estruturas cristalográficas da *S. typhimurium* e do *C. ammoniagenes* foram realizados com sucesso. Os valores do RMS entre as estruturas dos dois modelos obtidos e a estrutura cristalográfica da RNR do *M. tuberculosis*, foram de 1.04Å e 0,83Å, respectivamente. Uma análise detalhada do sítio dos átomos de ferro demonstrou que os compostos não podem se complexar com o átomo de ferro especificamente neste local devido à existência de impedimento estérico. A falta de informações da cadeia alfa (subunidade maior) impediu a realização dos estudos de *docking*.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Abagyan, R., Totrov, M. *J. Mol. Biol.* **1994**, *235*, 983–1002.
- Abagyan, R., Totrov, M., Kuznetsov, D. *J. Comput. Chem.* **1994**, *15*, 488–506.
- Aloy, P., Ceulemans, H., Stark, A., Russell, R. B. *J. Mol. Biol.* **2003**, *332*, 989–998.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403–410.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
- Amaral, A. T. do; Oliveira, A. C.; Neidlein, R.; Gallacci, M.; Caprara, L.; Miyazaki, Y. *Eur. J. Med. Chem.* **1997**, *32*, 433–443.
- Amaral, A. T. do, Miyazaki, Y., Caprara, L., Oliveira, A.C., in *Rational Approaches to the Design of Bioactive Compound*, Amsterdam, Elsevier, 1991.
- Amaral, A. T. do, Miyazaki, Y., Capobianco, G., Frey, B. Y., Stachissini, A. S., Miguel, S. R. M., Kawabe, A. B., Tavares, L. C., Gallacci, M., Caprara, L., Oliveira, in *Trends in QSAR and Molecular Modelling* WERMUTH, C. G., Ed., Leiden, ESCOM, 1993.
- Appel, R. D., Bairoch, A. Hochstrasser, D. F. *Trends Biochem. Sci.* **1994**, *19*, 258–260.
- Armarego, W. L. F, Perrin, D. D. in: *Purification of Laboratory Chemicals*, 4th Ed., Great Britain, Elsevier, 1996.
- Aszódi, A., Taylor, W. R. *Fold. Des.* **1996**, *1*, 325–334.
- Auchus, R. J., Miller, W. L. *Mol. Endocrinol.* **1999**, *13*, 1169–1182.
- Avarbock, D., Salem, J., Li, L. S., Wang, Z. M., Rubin, H. *Gene* **1999**, *233*, 261–269.
- Bajorath, J. *Nat. Rev. Drug Discovery* **2002**, *1*, 882–894.
- Baker, J. *J. Comput. Chem.* **1986**, *7*, 385–395.
- Balaban, A., Sekerci, M., Erk, B. *Syn. React. Inorg. Met.* **2003**, *33*, 1775–1786.
- Baldwin, S. L., D'Souza, C., Roberts, A. D., Kelly, B. P., Frank, A. A., Lui, M. A., Ulmer, J. B., Huygen, K., McMurray, D. M., Orme, I. M. *Infect. Immun.* **1998**, *66*, 2951–2959.
- Bardou, F., Raynaud, C., Ramos, C., Lanéelle, M. A., Lanéelle, G. *Microbiology* **1998**, *144*, 2539–2544.
- Barlow, T., Eliasson, R., Platz, A., Reichard, P., Sjöberg, B. M. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 1492–1495.
- Baroni, R.V. “Aspectos estruturais, mecanismo de formação e atividade antimicrobiana de benzilidenotiossemocarbazonas 2- ou 4-, potencialmente quimioterápicas”. 1987. Dissertação de Mestrado - Faculdade de Ciências Farmacêuticas, USP, São Paulo.
- Barton, G. J., Sternberg, J. E. *J. Mol. Biol.* **1987**, *198*, 327–337.
- Behr, M. A., Small, P. M. *Nature* **1997**, *389*, 133–134.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., Wheeler, D. L. *Nucleic Acids Res.* **2000**, *28*, 15–18.
- Bernadou, J., Nguyen, M., Meunier, B. *Ann. Pharm. Fr.* **2001**, *59*, 331–337.
- Bernstein, F. C., Koetzle, T. F., Williams, G. F., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. *Mol. Biol.* **1977**, *112*, 535–542.
- Blaney, J. M., Dixon, J. S. *Perspect. Drug Discov.* **1993**, *1*, 301–319.
- Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., Thornton, J. M. *Nature* **1987**, *326*, 347–352.
- Bodiguel, J., Nagy, J. M., Brown, K. A., Jamart-Gregoire, B. *J. Am. Chem. Soc.* **2001**, *123*, 3832–3883.
- Böhm, H.-J. *J. Comput. Aid. Mol. Des.* **1994**, *8*, 243–256.
- Böhm, H.-J. *J. Comput. Aid. Mol. Design* **1992**, *6*, 593–606.

- Böhm, H.-J., Klebe, G., Kubinyi, H. in *Wirkstoffdesign*, Heidelberg, Berlin, Oxford, Spektrum Akademischer Verlag, 1996. Böhm, M., Stürzebecher, J., Klebe, G. *J. Med. Chem.* **1999**, *42*, 458-477.
- Bostrom, J., Bohm, M., Gundertofte, K., Klebe, G. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1020-1027.
- Bottari, B., Maccari, R., Monforte, F., Ottanà, R., Rotondo, E., Vigorita, M. G. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 657-660.
- Bouzida, D., Rejto, P. A., Arthurs, S., Colson, A. B., Freer, S. T., Gehlhaar, D. K., Larson, V., Luty, B. A., Rose, P. W., Verkhivker, G. M. *Int. J. Quant. Chem.* **1999**, *72*, 73-84.
- Bower, M. J., Cohen, F. E., Dunbrack, Jr., R. L. *J. Mol. Biol.* **1997**, *267*, 1268-1282.
- Bowie, J. U., Lüthy, R., Eisenberg, D. *Science* **1991**, *253*, 164-171.
- Brandli, O., Haegi, V., Villiger, B., Bohn, W., Baumann, H. R., Zach, R. *Schweiz. Med. Wochenschr.* **1989**, *119*, 299-305.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187-217.
- Brooks, C. L., Karplus, M., Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, New York, John Wiley & Sons, 1988.
- Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Hill, T. C. *J. Mol. Biol.* **1969**, *42*, 65-86.
- Bruccoleri, R. E., Karplus, M. *Biopolymers* **1987**, *26*, 137-168.
- Burke, B. J., Dunn, W. J., 3rd, Hopfinger, A. J. *J. Med. Chem.* **1994**, *37*, 3775-3788.
- Burke, B. J., Hopfinger, A. J. in *3D QSAR in Drug Design. Theory, Methods and Application* H. Kubinyi, Ed., Leiden, ESCOM, 1993.
- Bystroff, C., Oatley, S. J., Kraut, S. J. *Biochemistry* **1990**, *29*, 3263-3277.
- Carlson, H. A., McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213-218.
- Carrupt, P. A., Gaillard, P., Billois, F., Weber, P., Testa, B., Meyer, C., Pérez, S. in *Lipophilicity in Drug Action and Toxicology*; V. Pliska; B. Testa; H. van de Waterbeemd, Eds., Weinheim, John Wiley & Sons, 1996.
- Ceneviva, M. P. B. "Tiossemicarbazonas Heterocíclicas Potencialmente Antimaláricas: Aspectos Estruturais e Cinética de Formação". 1984. Tese de Livre-Docência – Faculdade de Ciências Farmacêuticas, USP, São Paulo.
- Chen, Z. W. *Curr. HIV Res.* **2004**, *2*, 351-355.
- Cheng, E., Haiduke, R. L. A., Pires, R., Ishiki, H. M., Bruns, R. E., Amaral, A. T. in *Proceedings of 14th European Symposium on QSAR: Designing Drugs: processes, problems and solutions*, Bournemouth, England, 2002.
- Cho, S. J., Tropsha, A. *J. Med. Chem.* **1995**, *38*, 1060-1066.
- Chothia, C., Lesk, A. *EMBO J.* **1986**, *5*, 823-826.
- Chothia, C., Lesk, A. M., Levitt, M., Amit, A. G., Mariuzza, R. A., Philips, S. E. V., Poljak, R. J. *Science* **1986**, *233*, 755-758.
- Claessens, M., Custem, E. V., Lasters, I., Wodak, S. *Protein Eng.* **1989**, *2*, 335-345.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A. *J. Am. Chem. Soc.*, **1995**, *117*, 5179-5197.
- Cory, J. G. *Adv. Enzyme Regul.* **1988**, *27*, 437-455.
- Cory, J. G., Downes, D. I., Cory, A. H., Schaper, K. J., Seydel, J. K. *Anticancer Res.* **1994**, *14*, 875-880.
- Cramer III, R. D., Patterson, D. E., Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
- Creasey, W. A., Agrawal, K. C., Capizzi, R. L., Stinson, K. K., Sartorelli, A. C. *Cancer Research* **1972**, *32*, 565-572.
- Cregut, D., Liautard, J. P., Chiche, L. *Protein Eng.* **1994**, *7*, 1333-1344.
- Crum-Brown, A., Fraser, T. R. *Trans. Roy. Soc. Edinburgh* **1868-69**, *25*, 151-203, 693-739.

- D'Alfonso, G., Tramontano, A., Lahm, A. *J. Struct. Biol.* **2001**, *134*, 246-256.
- Dang, P., Madam, A. K. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1162-1166.
- Dawes, S. S., Warner, D. F., Tsenova, L., Timm, J., McKinney, J. D., Kaplan, G., Rubin, H., Mizrahi, V. *Infect. Immun.* **2003**, *71*, 6124-6131.
- Debnath, A. K. *J. Med. Chem.* **1999**, *42*, 249-259.
- DeConti, R. C., Toftness, B. R., Agrawal, K. C., Tomchick, R., Mead, J. A. R., Bertino, J. R., Sartorelli, A. C., Crasey, W. A. *Cancer Research* **1972**, *32*, 1455-1462.
- DePristo, M. A., De Bakker, P. I., Shetty, R. P., Blundell, T. L. *Protein Sci.* **2003**, *12*, 2032-2046.
- Dewar, M. J. S., Yuan, Y.-C. *Inorg. Chem.* **1990**, *29*, 3881-3890.
- Dewar, M. J. S., Zoebisch, E. G. *J. Mol. Struct. (Theochem)* **1988**, *49*, 1-21.
- Dewar, M. J. S., Zoebisch, E. J., Healy, E. F., Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
- Di Nola, A., Roccatano, D., Berendsen, H. J. C. *Proteins* **1994**, *19*, 174-182.
- Di Paolo, T. *J. Pharm. Sci.* **1978a**, *67*, 564-566.
- Di Paolo, T. *J. Pharm. Sci.* **1978b**, *67*, 566-568.
- Domenech, R. G., March, F. J., Soler, R. M., Galvez, J., Antón-Fos, G. M., de Julián-Ortiz, J. V. *Quant. Struct. Act. Relat.* **1996**, *15*, 201-207.
- Doreleijers, J. F., Rullmann, J. A., Kaptein, R. *J. Mol. Biol.* **1998**, *281*, 149-164.
- DRAGON - <http://www.disat.unimib.it/chm/Dragon.htm> acessado em 16/08/2003.
- Dunbrack, R. L., Karplus, M. *J. Mol. Biol.* **1993**, *230*, 543-574.
- Eklund, H., Uhlin, U., Färnegårdh, M., Logan, D. T., Nordlund, P. *Prog. Biophys. Mol. Biol.* **2001**, *77*, 177-268.
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., Mee, R. P. *J. Comput. Aid. Mol. Des.* **1997**, *11*, 425-445.
- Eliasson, R., Fontecave, M., Jörnvall, H., Krook, M., Pontis, E., Reichard, P. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 3314-3318.
- Eliasson, R., Pontis, E., Jordan, A., Reichard, P. *J. Biol. Chem.* **1996**, *271*, 26582-26587.
- Elleingand, E., Gerez, C., Un, S., Knupling, M., Lu, G., Salem, J., Rubin, H., Sauge-Merle, S., Laulhere, J. P., Fontecave, M. *Eur. J. Biochem.* **1998**, *258*, 485-490.
- Eriksson, M., Jordan, A., Eklund, H. *Biochemistry*, **1988**, *37*, 13359-13369.
- Ewing, T. J. A., Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175-1189.
- Fieschi, F., Torrents, E., Touloukhouva, L., Jordan, A., Hellman, U., Barbe, J., Gibert, I., Karlsson, M., Sjöberg, B. M. *Biol. Chem.* **1998**, *273*, 4329-4337.
- Fieschi, F., Torrents, E., Touloukhouva, L., Jordan, A., Hellman, U., Barbe, J., Gibert, I., Karlsson, M., Sjöberg, B.-M. *Biol. Chem.* **1998**, *273*, 4329-4337.
- Finch, R. A., Liu, M.-C., Cory, A. H., Cory, J. G., Sartorelli, A. C. *Advan. Enzyme Regul.* **1999**, *33*, 3-12.
- Fine, R. M., Wang, H., Shenkin, P. S., Yarmush, D. L., Levinthal, C. *Proteins* **1986**, *1*, 342-362.
- Fischer, D., Rice, D., Bowie, J. U., Eisenberg, D. *FASEB J.* **1996**, *10*, 126-136.
- Fiser, A. S., Sali, A. *Methods Enzymol.* **2003**, *374*, 461-491.
- Fiser, A. S., Do, R. K., Sali, A. *Protein Sci.* **2000**, *9*, 1753-1773.
- Fontecave, M., Nordlund, P., Eklund, H., Reichard, P. *Adv. Enzymol.* **1992**, *65*, 147-183.
- Forster, M. *J. Micron* **2002**, *33*, 365-384.
- Free Jr., S. M., Wilson, J. W. *J. Med. Chem.* **1964**, *7*, 395-399.
- French, F.A., Blanz-Jr., E.J. *J. Med. Chem.* **1966**, *9*, 585-589.
- French, F.A., Blanz-Jr., E.J., *J. Med. Chem.* **1974**, *17*, 172-181.
- Fühner, H., Neubauer, E. *Arch. Exp. Path. Pharm.* **1907**, *56*, 333-345.

- Fujita, T., Ban, T. *J. Med. Chem.* **1971**, *14*, 148-152.
- Fulerand, P., Berge, G., Noel, A., Chevallet, P., Castel, J., Orzalesi, H. *Eur. J. Med. Chem.* **1978**, *13*, 177-182.
- Gabb, H. A., Jackson, R. M., Sternberg, M. J. E. *J. Mol. Biol.* **1997**, *272*, 106-120.
- Gasteiger, J., Marsili, M. *Tetrahedron* **1980**, *36*, 3219-3228.
- Gehlhaar, D. K., Moerder, K. E., Zichi, D., Sherman, C. J., Ogden, R. C., Freer, S. T. *J. Med. Chem.* **1995**, *38*, 466-472.
- Gerczei, T., Keseru, G. M., Naray-Szabo, G. *J. Mol. Graph. Model.* **2000**, *18*, 7-17, 57-8.
- Ginsberg, A. M. *Bull. World Health Organ.* **2002**, *80*, 483-488.
- Godzik, A., Kolinski, A., Skolnick, J. *J. Mol. Biol.* **1992**, *227*, 227-238.
- Golbraikh, A., Tropsha, A. *J. Mol. Graph. Model.* **2002**, *20*, 269-276.
- Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849-857.
- Goodsell, D. S., Olson, A. J. *Proteins* **1990**, *8*, 195-202.
- Gordin, F., Chaisson, R. E., Matts, J. P., Miller, C., de Lourdes Garcia, M., Hafner, R., Valdespino, J. L., Coberly, J., Schechter, M., Klukowicz, A. J., Barry, M. A., O'Brien, R. J. *JAMA* **2000**, *283*, 1445-1450.
- Greer, J. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 3393-3397.
- Greer, J. *Proteins: Struct., Funct., Genet.* **1990**, *7*, 317-334.
- Gupta, S.P., Prabhakar, Y.S., Bindal, M.C., Handa, A., *Arzneim.-Forsch.* **1984**, *34*, 147-150.
- Halperin, I., Ma, B., Wolfson, H., Nussinov, R. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 409-443.
- Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96-103.
- Hammett, L. P. *Physical Organic Chemistry. Reaction Rates, Equilibria, and Mechanisms*, New York, McGraw-Hill, 1970.
- Hansch, C. Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, New York, John Wiley, 1979.
- Hansch, C., Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616-1626.
- Hansch, C., Lien, E. J. *Biochem. Pharmacol.* **1968**, *17*, 709-720.
- Hansch, C., Steward, A. R., Anderson, S. M., Bentley, D. L. *J. Med. Chem.* **1968**, *11*, 1-11.
- Hansch, C., Yoshimoto, M. *J. Med. Chem.* **1974**, *17*, 1160-1167.
- Havel, T. F., Snow, M. E. *J. Mol. Biol.* **1991**, *217*, 1-7.
- Helden, P. D. van *IUBMB Life* **2002**, *53*, 219-223.
- Hizel, K., Maral, I., Karakus, R., Aktas, F. *Clin. Microbiol. Infect.* **2004**, *10*, 980-983.
- Hodsdon, M. E., Ponder, J. W., Cistola, D. P. *J. Mol. Biol.* **1996**, *264*, 585-602.
- Hogbom, M., Huque, Y., Sjoberg, B. M., Nordlund, P. *Biochemistry* **2002**, *41*, 1381-1389.
- Holmgren, A. *J. Biol. Chem.* **1989**, *264*, 13963-13966.
- Höltje, H.-D., Folkers, G. *Methods and Principles in Medicinal Chemistry*; R. Mannhold, H. Kubinyi, H. Timmerman, Eds., New York, VCH, 1996.
- Hoof, R. W., Vriend, G., Sander, C., Abola, E. E. *Nature* **1996**, *381*, 272.
- Ishiki, H. M. "Relações Quantitativas Estrutura-Atividade (QSAR/QSAR-3D) de Dipiridodiazepinonas Inibidoras da Transcriptase Reversa do Vírus da Imunodeficiência Humana do Tipo I (HIV-1)". 1999. Dissertação de Mestrado – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, USP, Ribeirão Preto.
- Ishiki, H. M.; Amaral, A. T. do. *Three-Dimensional Quantitative Structure-Activity Relationship Study of Antitumor 2-Formylpyridine Thiosemicarbazones Derivatives as Inhibitors of Ribonucleotide Reductase*. Em fase final de redação.

- Ishiki, H.M, Galembeck, S.E & Amaral, A.T.do in *Proceedings of 13th European Symposium on Quantitative Structure-Activity Relationships: Rational Approaches to Drug Design*, The Netherlands, ESCOM, Science Publishers B.V., 2001.
- Itai, A., Tomioka, N., Yamada, M., Inoue, A., Kato, Y. in *Drug Design: Theory, Methods and Applications*; H. Kubinyi, Ed., Leiden, ESCOM, 1993.
- Jackson, R. M., Gabb, H. A., Sternberg, M. J. E. *J. Mol. Biol.* **1998**, *276*, 265–285.
- Jawahar, M. S. *Indian J. Med. Res.* **2004**, *120*, 398–417.
- Jiang, F., Kim, S. H. *J. Mol. Biol.* **1991**, *219*, 79–102.
- Jindani, A., Nunn, A. J., Enarson, D. A. *Lancet* **2004**, *364*, 1244–1251.
- Johnson, M. A., Hoog, C., Pinto, B. M. *Biochem.* **2003**, *42*, 1842–1853.
- Johnson, M. S., Srinivasan, N., Sowdhamini, R., Blundell, T. L. *Crit. Rev. Biochem. Mol. Biol.* **1994**, *29*, 1–68.
- Jones, D. T., Taylor, W. R., Thornton, J. M. *Nature* **1992**, *358*, 86–89.
- Jones, G., Willett, P., Glen, R. C., Leach, A. R., Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748.
- Jones, T. H., Thirup, S. *EMBO J.* **1986**, *5*, 819–822.
- Jordan A., Reichard, P. *Annu. Rev. Biochem.* **1998**, *67*, 71–98.
- Jordan, A., Aragall, E., Gibert, I., Barbé, J. *Molecular Microbiol.* **1996**, *19*, 777–790.
- Jordan, A., Gibert, I., Barbe, J. *Gene.* **1995**, *167*, 75–79.
- Jordan, A., Gibert, I., Barbe, J. *J. Bacteriol.* **1994**, *176*, 3420–3427.
- Jorgensen, W. L., Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- Judson, R. S., Jaeger, E. P., Treasurywala, A. M. *Theochem.* **1994**, *114*, 191–206.
- Jung, S. H., Pastan, I., Lee, B. *Proteins* **1994**, *19*, 35–47.
- Kabsch, W. *Acta Cryst.* **1976**, *A32*, 922–923.
- Katritzky, A. R., Gordeeva, E. V., *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835–857.
- Kazmierkiewicz, R., Liwo, A., Scheraga, H. A. *Biophys. Chem.* **2003**, *100*, 261–280.
- Kearsley, S. K., Smith, G. M. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615–633.
- Kier, L. B., Hall, L. H. in *Molecular Connectivity in Chemistry and Drug Research*, New York, Academic Press, 1976.
- Kier, L. B., Hall, L. H., Murray, W. J. *J. Pharm. Sci.* **1975**, *64*, 1974–1977.
- Kier, L. B., Hall, L. H., Murray, W. J. *J. Pharm. Sci.* **1976**, *65*, 1806–1809.
- Klebe, G. *J. Mol. Med.* **2000**, *78*, 269–281.
- Klebe, G., Abraham, U., Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130–4146.
- Klopman, G., Fercu, D., Jacob, J. *Chem. Phys.* **1996**, *204*, 181–193.
- Kollman, P. A. AMBER 5.0, University of California, San Francisco, 1996.
- Kubinyi, H. – Programa BILIN versão 1994b.
- Kubinyi, H. *J. Med. Chem.* **1976**, *19*, 587–600.
- Kubinyi, H. *J. Med. Chem.* **1977**, *20*, 625–629.
- Kubinyi, H. *Farmaco [Sci]* **1979a**, *34*, 248–276.
- Kubinyi, H. *Fortschr. Arzneimittelforsch.* **1979b**, *23*, 97–198.
- Kubinyi, H. in *3D QSAR in Drug Design: Theory, Methods and Applications*, Leiden, ESCOM, 1993a.
- Kubinyi, H. in *QSAR: Hansch Analysis and Related Approaches*, Weinheim, VCH, 1993b.
- Kubinyi, H. in *Burger's Medicinal Chemistry and Drug Discovery, Volume 1: Principles and Practice* M. E. Wolff, Ed., New York, Wiley Publishers 1994a.
- Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.

- Lam, P. Y. S., Jadhav, P. K., Eyermann, C. J., Hodge, C. N., Ru, Y., Bacheler, L. T., Meek, J. L., Otto, M. J., Rayner, M. M., Wong, Y. N., Chang, C. H., Weber, P. C., Jackson, D. A., Sharpe, T. R., Erickson-Viitanen, S. *Science* **1994**, *263*, 380–384.
- Larsson, A., Sjöberg, B. M. *Embo. J.* **1986**, *5*, 2037–2040.
- Laskowski, R. A. *Nucleic Acids Res.* **2001**, *29*, 221–222.
- Laskowski, R. A., Chistyakov, V. V., Thornton, J. M. *Nucleic Acids Res.* **2005**, *33*, 266–268.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- Laskowski, R. A., MacArthur, M. W., Thornton, J. M. *Curr. Opin. Struct. Biol.* **1998**, *5*, 631–639.
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477–486.
- Leach, A. R. in *Molecular Modelling: Principles and Applications* Cingapura, Longman, 1996.
- Leach, A. R., Smellie, A. S. *J. Chem. Inform. Comput. Sci.* **1992**, *32*, 379–385.
- Lee, E., Holzman, R. S. *Clin. Infect. Dis.* **2002**, *34*, 365–370.
- Leo, A. *J. Chem. Rev.* **1993**, *93*, 1281–1306.
- Levitt, M. *J. Mol. Biol.* **1992**, *226*, 507–533.
- Lewis, D. F. V., Lake, B. G., Dickins, M., Goldfarb, P. S. *Xenobiotica* **2004**, *34*, 549–569.
- Lindberg, W., Persson, J.-A., Wold, S. *Anal. Chem.* **1983**, *55*, 643–648.
- Liu, A., Pötsch, S., Davydov, A., Barra, A.-L., Rubin, H., Gräslund, A. *Biochemistry* **1998**, *37*, 16369–16377.
- Liu, J., Rost, B. *Nucleic Acids Res.* **2004**, *32*, 3522–3530.
- Liu, M., Wang, S. *J. Comput. Aid. Mol. Des.* **1999**, *13*, 435–451.
- Liu, M.-C., Lin, T.-S., Gory, J. G., Cory, A. H., Sartorelli, A. C. *J. Med. Chem.* **1996**, *39*, 2586–2593.
- Liu, Z., Jiang, L., Gao, Y., Liang, S., Chen, H., Han, Y., Lai, L. *Proteins* **2003**, *50*, 49–62.
- Livingstone, D. in *Data Analysis for Chemists* New York, Oxford University Press Inc., 1995.
- Lopez de Compadre, R. L., Compadre, C. M., Castillo, R., Dunn III, W. J. *Eur. J. Med. Chem.* **1983**, *18*, 569–571.
- Lycksell, P. O., Ingemarson, R., Davis, R., Gräslund, A., Thelander, L. *Biochem.* **1994**, *33*, 2838–2842.
- Maccari, R., Ottana, R., Bottari, B., Rotondo, E., Vigorita, M. G. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5731–5733.
- Makino, S., Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1812–1825.
- Malvezzi, A. “Estudo Comparativo de Parâmetros Hidrofóbicos e relacionados à ionização, de Série de Derivados da Procaina com Atividade Bloqueadora Neuromuscular”. 2003. Dissertação de Mestrado – Instituto de Química, USP, São Paulo.
- Malvezzi, A., Policastro, D. S., Da Silva, W. C. F. N., Amaral A. T. in *Proceedings of the second log P symposium Lipophilicity in Drug Disposition*, Lausanne, Suíça, 2001.
- Mandel, G. L. e Sande, M. in *Goodman and Gilman's the Pharmacological Basis of therapeutics*. 8. ed., A. Goodman Gilman, T. W. Rall, A. S. Nies, P. Taylor, Ed., New York, Pergamon Press, 1990.
- Mangoni, R., Roccatano, D., Di Nola, A. *Proteins* **1999**, *35*, 153–162.
- Marsili, M., Gasteiger, J. *Croat. Chem. Acta* **1980**, *53*, 601–614.
- Martin, A. C. R., Cheetham, J. C., Rees, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9268–9272.
- Martin, Y. C. in *3D QSAR in Drug Design: Recent Advances* H. Kubinyi, G. Folkers, Y. C. Martin, Eds, Great Britain, Kluwer Academic Publishers, 1998.
- Mas, M. T., Smith, K. C., Yarmush, D. L., Aisaka, K., Fine, R. M. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 483–498.
- Mattioni, B. E., Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 94–102.

- McShane, H., Brookes, R., Gilbert, S. C., Hill, A. V. *Infect. Immun.* **2001**, *69*, 681-686.
- McShane, H., Pathan, A. A., Sander, C. R., Keating, S. M., Gilbert, S. C., Huygen, K., Fletcher, H. A., Hill, A. V. *Nat. Med.* **2004**, *10*, 1240-1244.
- Mdluli, K., Slayden, R. A., Zhu, Y., Ramaswamy, S., Pan, X., Mead, D., Crane, D. D., Musser, J. M., Barry III, C. E. *Science* **1998**, *280*, 1607-1610.
- Methaphorics LLC, Mission Viejo, CA 92691.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., *J. Chem. Phys.* **1953**, *21*, 1087-1092.
- Meyer, H. *Arch. Exp. Path. Pharm.* **1899**, *42*, 109-118.
- Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., Wlodawer, A. *Science* **1989**, *246*, 1149-1152.
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
- Morris, G. M., Goodsell, D. S., Huey, R., Olson, A. J. *J. Comput. Aid. Mol. Des.* **1996**, *10*, 293-304.
- Moult, J., James, M. N. *Proteins* **1986**, *1*, 146-163.
- Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791-804.
- Nakajima, N., Higo, J., Kidera, A., Nakamura, H. *Chem. Phys. Lett.* **1997**, *278*, 297-301.
- Needleman, S. B., Wunsch, C. D. *J. Mol. Biol.* **1970**, *48*, 443-453.
- Némethy, G., Gibson, K. D., Palmer, K. A., Yoon, C. N., Paterlini, G., Zagari, A., Rumsey, S., Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472-6484.
- Neves, P. J., da Costa, J. B., Ndiya, P. M., Carneiro, J. W. M. *Química Nova*, **1998**, *21*, 709-713.
- Nguyen, M., Quemard, A., Broussy, S., Bernadou, J., Meunier, B. *Antimicrob. Agents Chemother.* **2002**, *46*, 2137-2144.
- Nicholls, A., Honig, B. *J. Comput. Chem.* **1991**, *12*, 435-445.
- Nordlund, P., Sjöberg, B. M., Eklund, H. *Nature* **1990**, *345*, 59-598.
- Nordlund, P., Eklund, H. *J. Mol. Biol.* **1993**, *232*, 123-164.
- Norinder, U. in *Drug Discovery and Design*, H. Kubinyi, G. Folkers, Y. C. Martin, Eds., Great Britain, Kluwer Academic Publishers, 1998.
- Ojcius, D. *Nat. Rev. Microbiol.* **2004**, *2*, 858.
- OMS – <http://www.who.int> acessado em 10/02/2005.
- Oprea, T. I., Waller, C. L., Marshall, G. R. *J. Med. Chem.* **1994**, *37*, 2206-2215.
- Overton, E. *Studien uber die Narkose, zugleich ein Beitrag zur allgemeiner Pharmakologie*, Gustav Fischer, Jena, 1901.
- Overton, E. *Studien über die Narkose, zugleich ein Beitrag zur allgemeinen Pharmakologie* (Tradução em inglês: “*Studies on Narcosis Charles Ernest Overton*”), R. L. Lopnick, Ed., London, Chapman and Hall, 1991.
- Padovani, D., Thomas, F., Trautwein, A. X., Mulliez, E., Fontecave, M. *Biochemistry* **2001**, *40*, 6713-6719.
- Pak, Y., Wang, S. *J. Phys. Chem. B* **2000**, *104*, 354-359.
- Pang, Y. P., Kozikowski, A. P. *J. Comput. Aid. Mol. Des.* **1994**, *8*, 669-681.
- Pearlman, R. S., Smith, K. M. *Persp. Drug Disc. Des.* **1998**, *9/10/11*, 339-353.
- Pearson, W. R. *Methods Enzymol.* **1990**, *183*, 63-98.
- Peitsch, M. C. *Bioinformatics* **2002**, *18*, 934-938.
- Peitsch, M. C. in *Practical Applications of Computer-Aided Drug Design*, P. S. Charifson, Ed., New York, Marcel Dekker, 1997.
- Perola, E., Xu, K., Kollmeyer, T. M., Kaufmann, S. H., Prendergast, F. G., Pang, Y. P. *J. Med. Chem.* **2000**, *43*, 401-408.

- Peterson, R. W., Dutton, P. L., Wand, A. J. *Protein Sci.* **2004**, *13*, 735-751.
- Pierattelli, R., Banci, L., Eady, N. A., Bodiguel, J., Jones, J. N., Moody, P. C., Raven, E. L., Jamart-Gregoire, B., Brown, K. A. *J. Biol. Chem.* **2004**, *279*, 39000-39009.
- Pires, J. R. "Relações Quantitativas entre a Estrutura Química de Derivados Nitrofurânicos e a Atividade Antimicrobiana contra *Caulobacter crescentus* e *Staphylococcus aureus*". 1998. Tese de Doutorado - Instituto de Química, USP, São Paulo.
- Pires, J. R.; Saito, C.; Gomes, S. L.; Giesbrecht, A. M.; Amaral, A. T. D. *J. Med. Chem.* **2001**, *44*, 3673-3681.
- Platt, J. R. *J. Phys. Chem.* **1952**, *56*, 328-336.
- Przybylski, D., Rost, B. *J. Mol. Biol.* **2004**, *341*, 255-269.
- Raichurkar, A. V., Kulkarni, V. M. *J. Med. Chem.* **2003**, *46*, 4419-4427.
- Ramachandran, G. N., Sasisekharan, V. *Adv. Prot. Chem.* **1968**, *23*, 283-438.
- Raminelli, C., "Estudo da relações quantitativa entre a estrutura química e atividade citotóxica de séries de derivados de bases de Mannich". 2001. Dissertação de Mestrado - Instituto de Química, USP, São Paulo.
- Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
- Randic, M., Wilkins, C. L. *Chem. Phys. Lett.* **1979**, *63*, 332-336.
- Rando, D. G.; Sato, D. N.; Siqueira, L.; Malvezzi, A.; Leite, C. Q. F.; do Amaral, A. T.; Ferreira, E. I.; Tavares, L. C. *Bioorg. Med. Chem.* **2002**, *10*, 557-560.
- Rarey, M., Kramer, B., Lengauer, T. *J. Comput. Aid. Mol. Des.* **1997**, *11*, 369-384.
- Rarey, M., Kramer, B., Lengauer, T., Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470-489.
- Reichard, P. *Science* **1993**, *260*, 1773-1777.
- Rhyu, K. B., Patel, H. C., Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 771-778.
- Richard, A. J., Kier, L. B. *J. Pharm. Sci.* **1981**, *69*, 124-126.
- Richardson, B. J. *Medical Times and Gazette* **1869**, *2*, 703.
- Richet, M. C. *Compt. Rend. Soc. Biol.* **1893**, *45*, 775-776.
- Riera, J., Robb, F. T., Weiss, R., Fontecave, M. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 475-478.
- Rohl, C. A., Strauss, C. E., Chivian, D., Baker, D. *Proteins* **2004**, *55*, 656-677.
- Rook, G. A. W., Pando, R. H. *Annu. Rev. Microbiol.* **1996**, *50*, 259-284.
- Rost, B., Sander, C. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 113-136.
- Rost, B., Sander, C. *Methods Mol. Biol.* **2000**, *143*, 71-95.
- Rozwarski, D. A., Grant, G. A., Barton, D. H. R., Jacobs Jr., W. R., Sacchettini, J. C. *Science* **1998**, *279*, 98-102.
- Sah, P.P.T., Daniels, T. C. *Rec. Trav. Chim. Pays-Bas-Journal* **1950**, *69*, 1545-1546.
- Sahlin, M., Petersson, L., Graslund, A., Ehremberg, A., Sjöberg, B. M., Thelander, L. **1987**, *6*, 5541-5548.
- Sahlin, M., Sjöberg, B. M., Backes, G., Loehr, T., Sanders-Loehr, J. *Biochem. Biophys. Res. Commun.* **1990**, *167*, 813-818.
- Sali, A. *Curr. Opin. Biotechnol.* **1995**, *6*, 437-451.
- Sali, A., Blundell, T. L. *J. Mol. Biol.* **1993**, *134*, 779-815.
- Sali, A., Overington, J. P. *Protein Sci.* **1994**, *3*, 1582-1596.
- Sali, A., Potterton, L., Yuan, F., Vlijmen, H., Karplus, M. *Proteins* **1995**, *23*, 318-326.
- Sánchez, R., Sali, A. *Curr. Opin. Struct. Biol.* **1997a**, *7*, 206-214.
- Sánchez, R., Sali, A. *J. Comput. Phys.* **1999**, *151*, 388-401.
- Sánchez, R., Sali, A. *Proteins (Suppl.)* **1997b**, *1*, 50-58.
- Sander, C., Schneider, R. *Proteins* **1991**, *9*, 56-68.
- Sanderson, R. T. *Chemical Bonds and Bond Energy*, New York, Academic Press, 1976.

- Santha, T., Rehman, F., Mitchison, D. A., Sarma, G. R., Reetha, A. M., Prabhaker, R. *Trop. Med. Int. Health* **2004**, *9*, 551-558.
- Santos Filho, O. A., de Alencastro, R. B. *Química Nova* **2003**, *2*, 253-259.
- Sartorelli, A. C., Agrawal, K. C., Tsiftoglou, A. S., Moore, E. C. *Adv. Enzyme Regul.* **1976**, *15*, 117-139.
- Sattar, R., Ali, S. A., Kamal, M., Khan, A. A., Abbasi, A. *Biochem. Biophys. Res. Commun.* **2004**, *314*, 755-765.
- Sauder, J. M., Arthur, J. W., Dunbrack, R. L. *Proteins* **2000**, *40*, 6-22.
- Schaper, K.-J., Seydel, J. K., Rosenfeld, M., Kazda, J. *Lepr. Rev.* **1986**, *57*, 254-264.
- Scotti, M. T. "Estudo de Propriedades Físico-Químicas e de Critérios para Obtenção e Validação de Modelos QSAR para Séries de Análogos de Semicarbazonas com Atividade Antichagástica, Retiradas da Literatura". 2005. Dissertação de Mestrado – Instituto de Química, USP, São Paulo.
- Seydel, J. K., Schaper, K. J. *Pharm. Unserer. Zeit.* **1995**, *24*, 313-23.
- Shafferhans, A., Klebe, G. J. *Mol. Biol.* **2001**, *307*, 407-27.
- Shen, Y., Shen, L., Sehgal, P., Huang, D., Qiu, L., Du, G., Letvin, N.L., Chen, Z.W. *J. Virol.* **2004**, *78*, 14023-14032.
- Shoichet, B. K., Kuntz, I. D. *Chem. Biol.*, **1996**, *3*, 151-156.
- Silva, C. L., Bonato, V. L. D., Lima, V. M. F. *Braz. J. Med. Biol. Res.* **1999**, *32*, 231-234.
- Sippl, M. J. *Proteins* **1993**, *17*, 355-362.
- Siqueira, L. J. A. "Relação quantitativa entre a estrutura química e o bloqueio da transmissão neuromuscular para série de brometos de [2-(4-benzamido)etil] benzildimetilamônio para-substituídos". 2001. Dissertação de Mestrado - Instituto de Química, USP, São Paulo.
- Skinner, M.A., Parlane, N., McCarthy, A., Buddle, B.M. *Immunology* **2003**, *108*, 548-555.
- Smith, T. F., Waterman, M. S. *J. Mol. Biol.* **1981**, *147*, 195-197.
- Sousa, W. A. "Estudo da Relação Entre a Estrutura Química e a toxicidade Aguda em Série de Cloreto de [2-(Benzoiloxi)Etil]Dimetilamônio Meta-Substituídos, com Atividade Anestésica Local". 1997. Dissertação de Mestrado - Instituto de Química, USP, São Paulo.
- Srinivasan, S., March, C. J., Sudarsanam, S. *Protein Sci.* **1993**, *2*, 277-289.
- Stewart, J. J. P. *J. Comput. Aid. Mol. Des.* **1990**, *4*, 1-45.
- Stryer, L. *Biochemistry*, 4th Edition, New York, Freeman & Company, 1995.
- Stubbe, J. A. *Adv. Enzymol.* **1990**, *63*, 349-419.
- Stubbe, J. A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2723-2724.
- Sun, X. Y.; Eliasson, R., Pontis, E., Andersson, J., Buist, G., Sjöberg, B. M., Reichard, P. *J. Biol. Chem.* **1995**, *270*, 2443-2446.
- Svozil, D., Lohninger, H. (<http://www.lohninger.com/topix.html>)
- Swain, C. G., Lupton, E. C. Jr. *J. Am. Chem. Soc.* **1968**, *90*, 4328-4337.
- SYBYL Molecular Modeling System, version 6.9. Tripos Associates Inc., 1699 South Hanley Rd, St. Louis, MO 63144.
- Taft, R. W. Jr. in *Steric Effects in Organic Chemistry*, M. S. Newman, Ed., New York, Academic Press, 1971.
- Takeda-Shitaka, M., Takaya, D., Chiba, C., Tanaka, H., Umeyama, H. *Curr. Med. Chem.* **2004**, *11*, 551-558.
- Tanghe, A., Content, J., Van Vooren, J.P., Portaels, F., Huygen, K. *Infect. Immun.* **2001**, *69*, 3041-3047.
- Tatusova, T. A., Madden, T. L. *FEMS Microbiol Lett.* **1999**, *174*, 247-250.
- Tauer, A., Benner, S. A. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 53-58.

- Tavares, L. C. "Efeitos de substituintes sobre a polaridade do grupo carbonila e a atividade anestésica local de N,N-[(dimetilamino)metil] benzamidas para-substituídas". 1987. Dissertação de Mestrado – Instituto de Química, USP, São Paulo.
- Tavares, L. C. "Relações quantitativas entre a estrutura química e a atividade antimicrobiana de análogos à nifuroxazida". 1993. Tese de Doutorado - Instituto de Química, USP, São Paulo.
- Tavares, L.; Amaral, A. T. *Rev. Farm. Bioquim. Univ. S. Paulo* **1997**, *33*, 123-129.
- Taylor, J.S., Burnett, R.M. *Proteins* **2000**, *41*, 173-191.
- Taylor, R. D., Jewbury, P. J., Essex, J. W. A. *J. Comput. Aid. Mol. Des.* **2002**, *16*, 151-166.
- Taylor, W. R. *J. Mol. Evol.* **1988**, *28*, 161-169.
- Thelander, L., Gräslund, A. in *Metal Ions in Biological Systems* Vol.30, H. A. A. Sigel, Ed., New York, Marcel Dekker Inc., 1993.
- Thelander, L., Gräslund, A. *J. Biol. Chem.* **1983**, *258*, 4063-4066.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G. *Nucleic Acids Res.* **1997**, *25*, 4876-4882.
- Todeschini, R. *Anal. Chim. Acta* **1997**, *348*, 419-430.
- Todeschini, R., Consonni, V., Maiocchi, A. *Chemom. Intell. Lab. Syst.* **1999**, *46*, 13-29.
- Todeschini, R., Consonni, V., Mauri, A., Pavan, M. *Anal. Chim. Acta* **2004**, *515*, 199-208.
- Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*. Wiley-VCH: Weinheim, Germany, **2000**.
- Trinajstić, N. in *Chemical Graph Theory*, Florida, CRC Press, Inc., Boca Raton, 1992.
- Trosset, J. Y., Scheraga, H. A. *J. Comput. Chem.* **1999a**, *20*, 412-427.
- Trosset, J. Y., Scheraga, H. A. *J. Comput. Chem.* **1999b**, *20*, 244-252.
- Uppsten, M., Davis, J., Rubin, H. Uhlin, U. *FEBS Letters* **2004**, *569*, 117-122.
- van de Waterbeemd, H., Testa, B. *Adv. Drug. Res.* **1987**, *16*, 85-225.
- van Vlijmen, H. W. T., Karplus, M. *J. Mol. Biol.* **1997**, *267*, 975-1001.
- Vásquez, M. *Curr. Opin. Struct. Biol.* **1996**, *6*, 217-221.
- Verloop, A., Hoogenstraaten, W., Tipker, J. in *Drug Design* J. Ariens, Ed., New York, Academic Press, 1976.
- Vinter, J. G., Davis, A., Saunders, M. R. *J. Comput. Aid. Mol. Des.* **1987**, *1*, 31-51.
- Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163-172.
- Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., Teichmann, S. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 208-216.
- Vordermeier, H.M., Rhodes, S.G., Dean, G., Goonetilleke, N., Huygen, K., Hill, A.V., Hewinson, R.G., Gilbert, S.C. *Immunology*, **2004**, *112*, 461-470.
- Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52-56.
- Wadler, S., Makower, D., Clairmont, C., Lambert, P., Fehn, K., Sznol, M. *J. Clin. Oncol.* **2004**, *22*, 1553-1563.
- Waller, C. L., Marshall, G. R. *J. Med. Chem.* **1993**, *36*, 2390-2403.
- Waller, C.L., Oprea, T.I., Giolitti, A., Marshall, G. R. *J. Med. Chem.* **1993**, *36*, 4152-4160.
- Walters, W. P., Stahl, M. T., Murcko, M. A. *Drug Discovery Today*, **1998**, *3*, 160-178.
- Weinstein, E. A., Yano, T., Li, L. S., Avarbock, D., Avarbock, A., Helm, D., McColm, A. A., Duncan, K., Lonsdale, J. T., Rubin, H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4548-4553.
- Welch, W., Ruppert, J., Jain, A. N. *Chem. Biol.* **1996**, *3*, 449-462.
- West R. C. in *Handbook of Chemistry and Physics*. 52. ed., Cleveland, Chemical Rubber Co., 1971-1972.
- Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.

- Wiener, H. *J. Chem. Phys.* **1948**, *52*, 425-430.
- Wilson, C., Gregoret, L. M., Agard, D. A. *J. Mol. Biol.* **1993**, *229*, 996-1006.
- Wolf, Y. I., Grishin, N. V., Koonin, E. V. *J. Mol. Biol.* **2000**, *299*, 897-905.
- Wüthrich, K., Freyberg, V. B., Weber, C., Wider, G., Traber, R., Widmer, H., Braun, W. *Science*, **1991**, *254*, 953-954.
- Yang, F., Curran, S. C., Li, L. S., Avarbock, D., Graf, J. D., Chua, M. M., Lu, G., Salem, J. Rubin, H. *J. Bacteriol.* **1997**, *179*, 6408-6415.
- Yang, F., Lu, G., Rubin, H. *J. Bacteriol.* **1994**, *176*, 6738-6743.
- Zheng, Q., Kyle, D. J. *Proteins* **1996**, *24*, 209-217.
- Zheng, Q., Rosenfeld, R. J., Vajda, S., DeLisi, C. *Protein Sci.* **1993**, *2*, 1242-1248.
- Zhou, Z., Madura, J. D. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2167-2178.

6. APÊNDICE

Tabela 1. Valores de pIC_{50} observado e predito através dos modelos *CoMFA*, modelo 1 (grupo de treinamento IA), para a série de 40 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		Rigido		R _A ¹		S _C ³	
		Treino	Teste	Treino	Teste	Treino	Teste
01	6,55	5,80	5,88				5,76
02	2,70	2,86	2,64			2,87	
03	3,30	3,31	2,78			2,58	
04	5,82	3,94	3,59			4,65	
05	2,28	2,11	3,06			3,16	
06	5,52	5,76	5,95			5,80	
07	5,66	5,89	5,85			5,92	
08	4,96	5,03	5,36			5,39	
09	5,34	5,33	5,37			5,45	
10	5,74	5,72	5,79			5,81	
11	6,59	6,38	6,20			6,29	
12	6,57	6,48	6,38			6,21	
13	6,51	6,15	6,28			6,22	
14	6,66	6,40	6,19			6,20	
15	5,11	5,38	5,90			5,83	
16	5,89	5,97	5,96			6,26	
17	6,04	5,67	6,17			6,11	
18	5,44	5,36	5,82			5,88	
19	5,42	5,42	5,33			5,45	
20	5,92	6,01	5,85			5,69	
21	6,25	6,01	6,11			5,94	
22	6,30	5,99	6,14			6,04	
23	6,39	6,01	6,21			6,10	
24	5,62	5,82	6,12			5,90	
25	5,92	5,97	5,76			5,73	
26	5,60	5,75	5,91			5,91	
27	6,07	6,43	6,29			6,04	
28	4,62	4,51	4,75			4,92	
29	5,69	5,87	5,21			5,32	
30	5,44	5,44	5,23			5,51	
31	5,28	5,34	5,40			5,43	
32	5,17	5,31	5,68			5,56	
33	3,96	3,84	3,88			3,63	
34	5,30	5,47	5,14			5,19	
35	5,25	5,24	5,00			5,03	
36	5,24	5,22	5,29			5,31	
37	4,89	5,10	5,23			5,06	
38	5,92	5,94	5,51			5,36	
39	6,40	6,51	6,30			6,33	
40	5,68	5,61	5,67			5,58	
ΔpIC_{50}	4,38	4,40	2,49	3,74	2,70	3,75	1,57

¹R_A – alinhamento rígido utilizando apenas o anel da piridina;²R_B – alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (tossemicarbazona) e³S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_R=1,0$ e $W_H=0,0$).Tabela II. Valores de pIC_{50} observado e predito através dos modelos *CoMFA*, modelo 2 (grupo de treinamento IA), para a série de 40 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}						SEAL	
		R _A ¹		Rigido		R _B ²		S _C ³	
		Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
01	6,55	6,09		6,11		6,13		6,13	
02	2,70	3,17		2,65		2,69		2,69	
03	3,30	3,56		3,26		3,23		3,23	
04	5,82	5,64		5,94		5,95		5,95	
05	2,28		5,47		5,53		5,53		5,53
06	5,52	6,11		5,85		5,83		5,83	
07	5,66	5,94		5,68		5,72		5,72	
08	4,96	5,43		5,08		5,03		5,03	
09	5,34		5,77		5,03		5,22		5,22
10	5,74	5,92		5,72		5,66		5,66	
11	6,59	6,20		6,37		6,54		6,54	
12	6,57	6,28		6,63		6,64		6,64	
13	6,51	6,08		6,52		6,56		6,56	
14	6,66		5,82		6,17		6,14		6,14
15	5,11	5,92		5,62		5,40		5,40	
16	5,89	5,94		5,91		5,91		5,91	
17	6,04	6,06		5,94		5,99		5,99	
18	5,44		5,62		5,86		5,76		5,76
19	5,42	5,85		5,37		5,35		5,35	
20	5,92	6,13		5,82		5,94		5,94	
21	6,25	6,09		6,14		6,16		6,16	
22	6,30		6,08		6,18		6,26		6,26
23	6,39	6,03		6,20		6,24		6,24	
24	5,62	5,92		5,88		5,59		5,59	
25	5,92		5,78		5,58		5,66		5,66
26	5,60	5,77		5,59		5,67		5,67	
27	6,07	6,24		6,19		6,01		6,01	
28	4,62	4,93		4,47		4,54		4,54	
29	5,69	5,31		5,70		5,81		5,81	
30	5,44		5,84		4,64		4,87		4,87
31	5,28		5,22		5,33		5,48		5,48
32	5,17	5,12		5,07		5,11		5,11	
33	3,96	3,46		4,08		4,03		4,03	
34	5,30	5,06		5,26		5,32		5,32	
35	5,25	5,00		5,21		5,22		5,22	
36	5,24	5,04		5,11		5,08		5,08	
37	4,89	4,48		5,00		4,96		4,96	
38	5,92		5,42		5,21		5,09		5,09
39	6,40	5,99		6,41		6,43		6,43	
40	5,68	5,29		5,21		5,06		5,06	
ΔpIC_{50}	4,38	3,11	0,86	3,98	1,54	3,95	1,39	3,95	1,39

¹R_A – alinhamento rígido utilizando apenas o anel da piridina;²R_B – alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (tossemicarbazona) e³S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_R=1,0$ e $W_H=0,0$).

Tabela III. Valores de pIC_{50} observado e predito através dos modelos *ComFA*, modelo 3 (grupo de treinamento IIIA), para a série de 40 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL ³			
		R _A ¹		R _B ²		R _C ³		Teste	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,24		5,89				5,78	
02	2,70	2,67		2,96				2,73	
03	3,30	3,24		2,96				2,51	
04	5,82	5,93		4,86				5,25	
05	2,28	2,30		2,96				3,01	
06	5,52		6,07						5,75
07	5,66	5,81		5,65				5,83	
08	4,96	5,01		4,86				5,10	
09	5,34	5,20		4,77				5,05	
10	5,74	5,80		5,61				5,71	
11	6,59	6,64		6,20				6,26	
12	6,57		6,40						6,08
13	6,51	6,48		6,57				6,42	
14	6,66	6,45		6,45				6,43	
15	5,11		6,32						6,05
16	5,89	5,90		5,64				5,81	
17	6,04	6,06		6,37				6,16	
18	5,44	5,42		5,63				5,66	
19	5,42	5,72		5,29				5,29	
20	5,92		6,23						5,62
21	6,25	6,29		6,22				5,98	
22	6,30	6,25		6,28				6,12	
23	6,39	6,22		6,38				6,22	
24	5,62		6,18						5,95
25	5,92	6,11		6,15				6,09	
26	5,60		5,83						6,07
27	6,07	6,25		6,14				6,05	
28	4,62		5,75					5,97	
29	5,69		5,92					6,01	
30	5,44	5,56		5,67				5,66	
31	5,28	5,51		5,54				5,73	
32	5,17	5,13		5,67				5,75	
33	3,96	4,12		4,22				4,17	
34	5,30	5,17		5,59				5,74	
35	5,25		5,04						5,66
36	5,24	5,13		5,71				5,84	
37	4,89	4,79		5,11				5,12	
38	5,92	5,74		5,74				5,66	
39	6,40		6,18						6,36
40	5,68	5,58		5,93				5,60	
ΔpIC_{50}	4,38	4,34	1,36	3,90	0,87			3,92	0,74

¹R_A – alinhamento rígido utilizando apenas o anel da piridina;

²R_B – alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona; e

³R_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_M=0,0$).

Tabela IV. Valores de pIC_{50} observado e predito através dos modelos *ComFA*, modelo 1 (grupo de treinamento IB), para a série de 39 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL ³			
		R _A ¹		R _B ²		R _C ³		Teste	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55								
02	2,70	2,77		2,77				2,90	
03	3,30	3,37		3,11				3,02	
04	5,82		3,89						4,78
05	2,28	2,13		2,40				2,34	
06	5,52	5,82		5,69				5,58	
07	5,66		5,93						5,78
08	4,96	4,91		4,90				4,88	
09	5,34	5,13		4,94				4,93	
10	5,74		5,71						5,56
11	6,59	6,34		6,09				6,26	
12	6,57	6,45		6,36				6,28	
13	6,51		6,08						6,34
14	6,66	6,34		6,30				6,28	
15	5,11	5,53		5,70				5,59	
16	5,89	5,97		5,96				6,08	
17	6,04		5,66						6,04
18	5,44	5,41		5,64				5,62	
19	5,42	5,43		5,34				5,40	
20	5,92	6,08		5,88				5,94	
21	6,25		6,01						5,99
22	6,30	5,98		6,13				6,11	
23	6,39	6,00		6,23				6,20	
24	5,62	5,91		6,17				5,89	
25	5,92	5,97		5,97				6,13	
26	5,60	5,84		5,97				6,03	
27	6,07		6,54						6,20
28	4,62	4,52		4,69				4,83	
29	5,69	5,61		5,46				5,46	
30	5,44	5,61		5,56				5,62	
31	5,28	5,46		5,23				5,15	
32	5,17		5,51						5,93
33	3,96								
34	5,30	5,39		5,36				5,41	
35	5,25	5,14		5,20				5,19	
36	5,24	5,25		5,21				5,15	
37	4,89		5,57						5,48
38	5,92	5,85		5,79				5,76	
39	6,40	6,52		6,41				6,50	
40	5,68	5,66		5,93				5,84	
ΔpIC_{50}	4,38	4,39	2,65	4,01	2,87			4,16	1,56

¹R_A – alinhamento rígido utilizando apenas o anel da piridina;

²R_B – alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona; e

³R_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_M=0,0$).

Tabela V. Valores de pIC_{50} observado e predito através dos modelos *ComFA*, modelo 2 (grupo de treinamento IIB), para a série de 39 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	pIC_{50}					
	Observado	R _A		R _B		SEAL
		Treino	Teste	Treino	Teste	S _C
01	6,55	6,22		6,08		6,13
02	2,70	2,78		2,70		2,71
03	3,30	3,29		3,23		3,31
04	5,82	5,72		5,99		5,86
05	2,28		5,19		5,52	5,35
06	5,52	5,77		5,86		5,78
07	5,66	5,54		5,63		5,65
08	4,96	4,98		4,98		5,00
09	5,34		5,31		4,97	5,17
10	5,74	5,63		5,67		5,62
11	6,59	6,60		6,35		6,54
12	6,57	6,70		6,57		6,67
13	6,51	6,45		6,49		6,59
14	6,66		5,90		6,17	6,15
15	5,11	5,34		5,70		5,31
16	5,89	6,02		5,88		5,91
17	6,04	6,06		6,02		5,93
18	5,44		5,70		5,86	5,76
19	5,42	5,32		5,38		5,35
20	5,92	6,00		5,79		6,02
21	6,25	6,23		6,12		6,21
22	6,30		6,25		6,16	6,30
23	6,39	6,21		6,20		6,29
24	5,62	5,68		5,94		5,62
25	5,92		5,94		5,59	5,71
26	5,60	5,61		5,64		5,72
27	6,07	6,31		6,19		6,12
28	4,62	4,42		4,47		4,58
29	5,69	5,62		5,63		5,76
30	5,44		5,24		4,70	4,94
31	5,28		5,29		5,30	5,44
32	5,17	5,06		5,09		5,15
34	5,30	5,48		5,31		5,30
35	5,25	5,37		5,26		5,20
36	5,24	5,19		5,16		5,07
37	4,89	4,91		5,10		4,95
38	5,92		5,75		5,26	5,10
39	6,40	6,28		6,37		6,46
40	5,68		5,24		5,22	5,05
ΔpIC_{50}	4,38	3,92	1,06	3,87	1,47	3,96

¹R_A - alinhamento rígido utilizando apenas o anel da piridina;
²R_B - alinhamento rígido utilizando N(1) e Q(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (tossemicarbazona) e
³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela VI. Valores de pIC_{50} observado e predito através dos modelos *ComFA*, modelo 3 (grupo de treinamento IIB), para a série de 39 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	pIC_{50}					
	Observado	R _A		R _B		SEAL
		Treino	Teste	Treino	Teste	S _C
01	6,55	6,25		6,12		5,91
02	2,70	2,74		2,88		2,93
03	3,30	3,26		3,02		3,01
04	5,82	5,93		4,94		5,89
05	2,28	2,25		3,29		2,51
06	5,52		6,05		6,18	5,81
07	5,66	5,74		5,96		5,80
08	4,96	5,06		5,27		4,93
09	5,34	5,25		5,24		4,91
10	5,74	5,77		5,92		5,70
11	6,59	6,64		6,46		6,46
12	6,57		6,41		6,38	6,22
13	6,51	6,53		6,60		6,61
14	6,66	6,49		6,38		6,57
15	5,11		6,34		6,28	6,23
16	5,89	5,86		6,09		5,93
17	6,04	5,97		6,63		6,19
18	5,44	5,55		5,88		5,56
19	5,42	5,64		5,59		5,45
20	5,92		6,21		5,92	5,90
21	6,25	6,30		6,30		6,17
22	6,30	6,26		6,33		6,31
23	6,39	6,23		6,39		6,40
24	5,62		6,17		6,29	6,16
25	5,92	6,14		5,86		6,29
26	5,60		5,83		5,88	6,10
27	6,07	6,26		6,07		5,96
28	4,62		5,77		5,50	5,87
29	5,69	5,96		5,96		5,85
30	5,44	5,49		5,38		5,77
31	5,28	5,55		5,16		5,28
32	5,17	5,15		5,28		5,41
34	5,30	5,24		4,99		5,45
35	5,25		5,10		4,91	5,33
36	5,24	5,18		5,20		5,39
37	4,89	4,87		4,43		4,64
38	5,92	5,80		5,28		5,57
39	6,40		6,21		6,13	6,40
40	5,68	5,55		5,82		5,76
ΔpIC_{50}	4,38	4,39	1,31	3,75	1,47	4,10

¹R_A - alinhamento rígido utilizando apenas o anel da piridina;
²R_B - alinhamento rígido utilizando N(1) e Q(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (tossemicarbazona) e
³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela VII. Valores de pIC_{50} observado e predito através dos modelos *ComFA*, modelo 1 (grupo de treinamento C), para a série de 30 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		Rígido		R_A^1	R_B^2	Treinamento	Teste
		Teste	Treinamento				
01	6,55	6,21	6,23			6,36	6,22
11	6,59	6,69	6,29			6,40	
12	6,57	6,63	6,28			6,40	6,30
13	6,51	6,28	6,05			6,40	
14	6,66	6,59	5,91			6,40	
15	5,11	5,21	6,09			5,95	6,00
16	5,89	6,32	6,26			6,06	
17	6,04	6,09	6,33			5,41	
18	5,44	5,36	6,04			6,17	
19	5,42	6,17	6,20			5,79	6,17
20	5,92	5,87	6,07			5,99	
21	6,25	6,16	6,07			6,06	
22	6,30	6,22	6,02			6,06	
23	6,39	6,20	6,01			6,06	
24	5,62	5,57	5,90			5,64	
25	5,92	5,93	5,70			5,92	
26	5,60	5,77	5,77			5,81	
27	6,07	6,07	5,90			6,17	
28	4,62	4,64	5,16			4,50	5,15
29	5,69	5,21	5,43			5,36	
30	5,44	5,38	5,60			5,26	
31	5,28	5,27	5,48			5,26	
32	5,17	5,11	5,47			5,36	
34	5,30	5,46	5,23			5,20	
35	5,25	5,27	5,14			5,10	
36	5,24	5,17	5,18			5,39	5,53
37	4,89	5,37				5,96	
38	5,92	5,91	5,42			6,46	
39	6,40	6,34	5,82			5,94	
40	5,68	5,70	5,54			1,96	1,15
ΔpIC_{50}	2,04	2,05	1,11	1,19	0,87		

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina.

² R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona) e

³ SC - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_I=0,0$).

Tabela VIII. Valores de pIC_{50} observado e predito através dos modelos *ComFA*, modelo 2 (grupo de treinamento IC), para a série de 30 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		Rígido		R_A^1	R_B^2	Treinamento	Teste
		Teste	Treinamento				
01	6,55	6,29	6,21			6,19	
11	6,59	6,55	6,23			6,23	
12	6,57	6,53	6,24			6,20	
13	6,51	6,24	6,06			6,05	
14	6,66	5,99	5,81			5,80	
15	5,11	5,61	6,02			6,08	
16	5,89	5,88	6,27			6,25	
17	6,04	6,18	6,25			6,24	
18	5,44	6,10	6,18			6,18	
19	5,42	5,53	6,16			6,18	
20	5,92	6,05	6,09			6,09	
21	6,25	6,23	6,09			6,08	
22	6,30	6,27	6,04			6,03	
23	6,39	6,23	6,02			5,98	
24	5,62	5,67	5,93			5,83	
25	5,92	5,81	5,67			5,68	
26	5,60	5,78	5,78			5,77	
27	6,07	6,16	5,85			5,85	
28	4,62	4,52	5,18			5,20	
29	5,69	5,93	5,30			5,31	
30	5,44	5,14	5,65			5,70	
31	5,28	5,31	5,28			5,34	
32	5,17	5,87	5,62			5,55	
34	5,30	5,46	5,36			5,35	
35	5,25	5,39	5,22			5,32	
36	5,24	5,12	4,98			5,18	
37	4,89	4,85	5,45			5,00	
38	5,92	5,98	5,67			5,49	
39	6,40	5,93	5,58			5,67	
40	5,68	5,04	5,58			5,66	
ΔpIC_{50}	2,04	2,03	1,06	1,29	0,86	1,25	0,86

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina.

² R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona) e

³ SC - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_I=0,0$).

Tabela IX. Valores de pIC_{50} observado e predito através dos modelos CoMFA, modelo 3 (grupo de treinamento IIIc), para a série de 30 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}						SEAL		
		R _A ¹			R _B ²			S _C ³		
		Treinamento	Teste	Treinamento	Treinamento	Teste	Treinamento	Treinamento	Teste	Treinamento
01	6,55	6,42		6,39	6,39		6,59	6,59		
11	6,59	6,44		6,44	6,44		6,59	6,59		
12	6,57	6,56		6,57	6,57		6,61	6,61		
13	6,51	6,39		6,41	6,41		6,54	6,54		
14	6,66	6,31		6,34	6,34		6,66	6,66		
15	5,11		6,42			6,43		6,60	6,60	
16	5,89	5,85		5,84	5,84		5,80	5,80		
17	6,04		6,03			6,25		5,89	5,89	
18	5,44	5,45		5,42	5,42		5,48	5,48		
19	5,42	5,73		5,72	5,72		5,50	5,50		
20	5,92	6,16		6,10	6,10		5,92	5,92		
21	6,25	6,22		6,21	6,21		6,12	6,12		
22	6,30		6,24			6,19		6,15	6,15	
23	6,39		6,23			6,20		6,14	6,14	
24	5,62	5,99		5,89	5,89		5,63	5,63		
25	5,92	5,96		6,09	6,09		5,91	5,91		
26	5,60	5,88		5,85	5,85		5,72	5,72		
27	6,07	6,16		6,16	6,16		6,02	6,02		
28	4,62		5,69			5,80		5,55	5,55	
29	5,69	5,90		5,93	5,93		5,74	5,74		
30	5,44	5,40		5,34	5,34		5,40	5,40		
31	5,28		5,90			5,89		5,89	5,89	
32	5,17	5,40		5,45	5,45		5,08	5,08		
34	5,30		5,38			5,29		5,42	5,42	
35	5,25	5,20		5,10	5,10		5,27	5,27		
36	5,24	5,35		5,32	5,32		5,25	5,25		
37	4,89	4,92		4,93	4,93		4,94	4,94		
38	5,92	5,56		5,66	5,66		5,85	5,85		
39	6,40	6,27		6,33	6,33		6,39	6,39		
40	5,68	5,17		5,21	5,21		5,70	5,70		
ΔpIC_{50}	2,04	1,64	0,86	1,64	1,64	1,14	1,72	1,72	1,18	

¹R_A – alinhamento rígido utilizando apenas o anel da piridina;

²R_B – alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona) e

³S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela X. Valores de pIC_{50} observado e predito através dos modelos CoMFA, modelo 1 (grupo de treinamento ID), para a série de 23 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}						SEAL		
		R _A ¹			R _B ²			S _C ³		
		Treinamento	Teste	Treinamento	Treinamento	Teste	Treinamento	Treinamento	Teste	Treinamento
01	6,55	6,58		6,60	6,61		6,62	6,62		
13	6,51	6,58		6,60	6,61		6,50	6,50		
14	6,66	6,58		6,24	6,24		6,44	6,44		
20	5,92	5,87		6,03	6,03		6,00	6,00		
21	6,25	6,21		6,26	6,26		6,27	6,27		
22	6,30	6,31		6,29	6,29		6,34	6,34		
23	6,39		6,24			6,16		6,26	6,26	
24	5,62	5,55		5,76	5,76		5,59	5,59		
25	5,92		5,84			5,97		6,16	6,16	
26	5,60	5,78		5,74	5,74		5,63	5,63		
27	6,07	6,09		5,92	5,92		5,91	5,91		
28	4,62		5,56			5,54		5,65	5,65	
29	5,69	5,64		5,71	5,71		5,58	5,58		
30	5,44	5,41		5,19	5,19		5,40	5,40		
31	5,28	5,32		5,52	5,52		5,30	5,30		
32	5,17	5,15		5,45	5,45		5,25	5,25		
33	3,96	3,92		3,92	3,92		3,90	3,90		
34	5,30	5,36		5,29	5,29		5,42	5,42		
35	5,25	5,24		5,24	5,24		5,37	5,37		
36	5,24	5,25		5,23	5,23		5,27	5,27		
37	4,89	4,92		4,82	4,82		4,85	4,85		
38	5,92	5,92		5,88	5,88		5,92	5,92		
39	6,40		6,15			5,90		6,02	6,02	
ΔpIC_{50}	2,70	2,66	1,04	2,69	2,69	0,90	2,72	2,72	0,94	

¹R_A – alinhamento rígido utilizando apenas o anel da piridina;

²R_B – alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona) e

³S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XI. Valores de pIC_{50} observado e predito através dos modelos CoMFA, modelo 2 (grupo de treinamento IID), para a série de 23 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,49		6,61		6,61	
13	6,51	6,50		6,45		6,52	
14	6,66		6,00		6,04		6,06
20	5,92	5,99		6,18		6,23	
21	6,25	6,24		6,38		6,37	
22	6,30	6,31		6,41		6,36	
23	6,39	6,30		6,29		6,28	
24	5,62		5,94		6,04		5,92
25	5,92	5,92		5,90		5,98	
26	5,60	5,66		5,76		5,69	
27	6,07		6,25		6,21		6,12
28	4,62	4,60		4,83		4,88	
29	5,69	5,67		5,44		5,45	
30	5,44	5,41		5,19		5,21	
31	5,28		5,58		5,64		5,52
32	5,17	5,17		5,41		5,34	
33	3,96	3,94		4,02		4,04	
34	5,30		5,40		5,30		5,28
35	5,25	5,27		5,25		5,25	
36	5,24	5,25		5,20		5,19	
37	4,89	4,89		4,88		4,86	
38	5,92	5,95		5,70		5,71	
39	6,40	6,44		6,14		6,06	
ΔpIC_{50}	2,70	2,56	0,85	2,59	0,91	2,57	0,84

¹R_A - alinhamento rígido utilizando apenas o anel da piridina,

²R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_F=1,0$, $W_S=1,0$ e $W_T=0,0$).

Tabela XII. Valores de pIC_{50} observado e predito através dos modelos CoMFA, modelo 3 (grupo de treinamento IID), para a série de 23 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}						SEAL	
		R _A ¹		R _B ²		S _C ³		S _C ³	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,56		6,55		6,57		6,57	
13	6,51	6,61		6,63		6,60		6,60	
14	6,66	6,62		6,62		6,60		6,60	
20	5,92	5,91		5,86		5,95		5,95	
21	6,25		6,23		6,23		6,21		6,21
22	6,30	6,32		6,33		6,30		6,30	
23	6,39	6,29		6,24		6,26		6,26	
24	5,62	5,63		5,67		5,57		5,57	
25	5,92	5,94		6,01		5,95		5,95	
26	5,60		6,07		5,93		5,84		5,84
27	6,07	6,06		6,09		6,09		6,09	
28	4,62	4,63		4,61		4,61		4,61	
29	5,69	5,69		5,68		5,65		5,65	
30	5,44		5,82		5,60		5,59		5,59
31	5,28	5,29		5,31		5,32		5,32	
32	5,17	5,16		5,17		5,21		5,21	
33	3,96	3,97		3,96		3,93		3,93	
34	5,30	5,32		5,32		5,35		5,35	
35	5,25	5,21		5,23		5,27		5,27	
36	5,24		5,22		5,17		5,22		5,22
37	4,89	4,89		4,88		4,88		4,88	
38	5,92		5,65		5,52		5,52		5,52
39	6,40	6,40		6,34		6,40		6,40	
ΔpIC_{50}	2,70	2,65	1,01	2,67	1,06	2,67	0,99	2,67	0,99

¹R_A - alinhamento rígido utilizando apenas o anel da piridina,

²R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_F=1,0$, $W_S=1,0$ e $W_T=0,0$).

Tabela XIII. Valores de pIC_{50} observado e predito através dos modelos *CoMFA*, modelo 1 (grupo de treinamento IIE), para a série de 22 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,56	6,63	6,61	6,45	6,60	6,52
13	6,51	6,59	6,24			6,29	
14	6,66	6,58	6,04	6,04		6,10	
20	5,92	6,23	6,27	6,32		6,32	
21	6,25	6,33	6,31	6,34		6,34	
22	6,30	6,28	6,17	6,24		6,24	
23	6,39	5,62	5,78	5,70		5,70	
24	5,62	5,84	5,95	5,60		6,12	
25	5,92	5,70	5,74	5,89		5,88	
26	5,60	5,54	5,52	5,52		5,65	
27	6,07	5,71	5,82	5,77		5,77	
28	4,62	5,39	5,20	5,27		5,27	
29	5,69	5,27	5,44	5,32		5,32	
30	5,44	5,12	5,40	5,47		5,47	
31	5,28	5,32	5,29	5,29		5,29	
32	5,17	5,23	5,26	5,23		5,23	
34	5,30	4,91	4,75	4,79		4,79	
35	5,25	5,89	5,84	5,89		5,89	
36	5,24	6,18	5,87	5,91		5,91	
37	4,89	1,68	1,09	1,86	0,93	1,81	0,87
38	5,92						
39	6,40						
ΔpIC_{50}	2,04						

¹R_A - alinhamento rígido utilizando apenas o anel da piridina.

²R_B - alinhamento rígido utilizando apenas o anel da piridina e os átomos C=NNC=SN do fragmento da (tiosemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XIV. Valores de pIC_{50} observado e predito através dos modelos *CoMFA*, modelo 2 (grupo de treinamento IIE), para a série de 22 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,54	6,53	6,53		6,58	
13	6,51	6,55	6,50	6,51		6,51	
14	6,66	6,01	6,16			6,04	
20	5,92	5,93	5,91	5,91		6,15	
21	6,25	6,25	6,28	6,28		6,33	
22	6,30	6,33	6,37	6,33		6,33	
23	6,39	6,34	6,31	6,25		6,25	
24	5,62	5,92	5,96	5,88		5,88	
25	5,92	5,90	5,91	6,02		6,02	
26	5,60	5,61	5,63	5,62		5,62	
27	6,07	6,21	6,18	6,09		6,09	
28	4,62	4,64	4,64	4,83		4,83	
29	5,69	5,68	5,68	5,67		5,67	
30	5,44	5,42	5,42	5,09		5,09	
31	5,28	5,52	5,60	5,45		5,45	
32	5,17	5,15	5,17	5,34		5,34	
34	5,30	5,40	5,34	5,34		5,34	
35	5,25	5,25	5,22	5,30		5,30	
36	5,24	5,23	5,24	5,23		5,23	
37	4,89	4,91	4,90	4,86		4,86	
38	5,92	5,92	5,94	5,81		5,81	
39	6,40	6,41	6,43	6,13		6,13	
ΔpIC_{50}	2,04	1,91	0,81	1,89	0,84	1,75	0,75

¹R_A - alinhamento rígido utilizando apenas o anel da piridina.

²R_B - alinhamento rígido utilizando apenas o anel da piridina e os átomos C=NNC=SN do fragmento da (tiosemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XV. Valores de pIC_{50} observado e predito através dos modelos *CoMFA*, modelo 3 (grupo de treinamento III-E), para a série de 22 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		Rígido		R_A^2		S_C^2	
		Treino	Teste	Treino	Teste	Treino	Teste
01	6,55	6,52	6,54			6,57	
13	6,51	6,57	6,61			6,59	
14	6,66	6,66	6,64			6,61	
20	5,92	5,93	5,87			5,95	
21	6,25		6,23		6,24		6,22
22	6,30	6,34	6,34			6,30	
23	6,39	6,30	6,25			6,26	
24	5,62	5,64	5,67			5,57	
25	5,92	5,94	5,99			5,95	
26	5,60		6,15		5,98		5,84
27	6,07	6,07	6,08			6,08	
28	4,62	4,62	4,63			4,63	
29	5,69	5,69	5,68			5,70	
30	5,44	5,26	5,93		5,65		5,58
31	5,28	5,26	5,26			5,32	
32	5,17	5,17	5,17			5,21	
34	5,30	5,31	5,34			5,30	
35	5,25	5,23	5,26			5,24	
36	5,24		5,21		5,19		5,21
37	4,89	4,91	4,86			4,85	
38	5,92		5,65		5,55		5,54
39	6,40	6,39			6,35		6,41
ΔpIC_{50}	2,04	2,04	1,02	2,01	1,05	1,98	1,01

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina.

² R_A - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (isosemicarbazona) e

³ S_C - alinhamento utilizando o programa SEAL. ($\alpha=0,30$, $W_E=1,0$, $W_F=1,0$ e $W_H=0,0$).

Tabela XVI. Valores de pIC_{50} observado e predito através dos modelos *CoMSIA*, modelo 1 (grupo de treinamento IA), para a série de 40 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		Rígido		R_A^2		S_C^2	
		Treino	Teste	Treino	Teste	Treino	Teste
01	6,55		6,03		5,97		5,76
02	2,70	2,77		2,70		2,87	
03	3,30	3,25		3,32		2,58	
04	5,82		4,75		3,83		4,65
05	2,28	2,09		2,25		3,16	
06	5,52	5,67		5,68		5,80	
07	5,66		5,87		5,90		5,92
08	4,96	5,21		5,02		5,39	
09	5,34	5,28		5,25		5,45	
10	5,74		5,25		5,24		5,81
11	6,59	6,30		6,42		6,29	
12	6,57	6,27		6,42		6,21	
13	6,51		6,15		6,19		6,22
14	6,66	6,43		6,39		6,20	
15	5,11	5,77		5,60		5,83	
16	5,89	5,85		5,74		6,26	
17	6,04		6,27		6,26		6,11
18	5,44	5,49		5,38		5,88	
19	5,42	5,51		5,39		5,45	
20	5,92	5,87		5,88		5,69	
21	6,25		6,03		6,09		5,94
22	6,30	6,12		6,22		6,04	
23	6,39	6,24		6,42		6,10	
24	5,62	5,65		5,63		5,90	
25	5,92	5,99		6,04		5,73	
26	5,60	5,75		5,73		5,91	
27	6,07		6,07		6,13		6,04
28	4,62	4,64		4,45		4,92	
29	5,69	5,76		5,88		5,32	
30	5,44	5,26		5,41		5,51	
31	5,28	5,34		5,22		5,43	
32	5,17		5,32		5,41		5,56
33	3,96	4,11		4,02		3,63	
34	5,30	5,15		5,27		5,19	
35	5,25	5,06		5,18		5,03	
36	5,24	5,20		5,14		5,31	
37	4,89		5,13		5,05		5,06
38	5,92	6,06		6,02		5,36	
39	6,40	6,45		6,46		6,33	
40	5,68	5,77		5,80		5,58	
ΔpIC_{50}	4,38	4,36	1,52	4,21	2,43	3,75	1,57

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina.

² R_A - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (isosemicarbazona) e

³ S_C - alinhamento utilizando o programa SEAL. ($\alpha=0,30$, $W_E=1,0$, $W_F=1,0$ e $W_H=0,0$).

Tabela XVIII. Valores de ρIC_{90} observado e predito através dos modelos *CoMSIA*, modelo 2 (grupo de treinamento IIA), para a série de 40 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	ρIC_{90}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,07	5,99	6,13	2,69	6,13	2,69
02	2,70	3,01	2,79	3,40	3,23	3,40	3,23
03	3,30	3,42	3,40	5,72	5,95	5,95	5,35
04	5,82	5,47	4,88	4,95	5,83	5,83	5,35
05	2,28	5,93	5,81	5,66	5,72	5,72	5,03
06	5,52	5,91	5,66	4,89	4,93	4,93	5,22
07	5,66	5,40	5,39	5,38	5,66	5,66	5,22
08	4,96	5,99	5,99	5,75	5,99	5,99	5,76
09	5,34	5,59	5,59	5,43	5,92	5,92	5,35
10	5,74	6,15	6,37	6,54	6,64	6,64	6,14
11	6,59	6,11	6,05	6,08	6,16	6,16	6,14
12	6,57	6,08	6,08	6,24	6,24	6,24	6,26
13	6,51	6,06	6,06	6,12	6,12	6,12	6,26
14	6,66	6,08	6,08	6,12	6,12	6,12	6,26
15	5,11	6,08	6,08	6,12	6,12	6,12	6,26
16	5,89	5,86	5,99	5,99	5,99	5,99	5,99
17	6,04	5,99	5,99	5,99	5,99	5,99	5,99
18	5,44	5,60	5,60	5,76	5,76	5,76	5,76
19	5,42	5,82	5,82	5,92	5,92	5,92	5,92
20	5,92	6,04	6,04	6,08	6,08	6,08	6,08
21	6,25	6,18	6,18	6,12	6,12	6,12	6,12
22	6,30	6,39	6,39	6,22	6,22	6,22	6,22
23	6,39	6,09	6,09	5,82	5,82	5,82	5,82
24	5,62	5,75	5,75	5,78	5,78	5,78	5,78
25	5,92	6,04	6,04	5,72	5,72	5,72	5,72
26	5,60	5,88	5,88	6,01	6,01	6,01	6,01
27	6,07	5,88	5,88	4,89	4,89	4,89	4,89
28	4,62	5,25	5,25	5,91	5,91	5,91	5,91
29	5,69	5,47	5,47	5,74	5,74	5,74	5,74
30	5,44	5,40	5,40	5,41	5,41	5,41	5,41
31	5,28	5,08	5,08	5,24	5,24	5,24	5,24
32	5,17	3,62	3,62	3,54	3,54	3,54	3,54
33	3,96	4,98	4,98	5,39	5,39	5,39	5,39
34	5,30	4,93	4,93	5,34	5,34	5,34	5,34
35	5,25	5,13	5,13	5,41	5,41	5,41	5,41
36	5,24	4,81	4,81	4,75	4,75	4,75	4,75
37	4,89	5,36	5,36	5,17	5,17	5,17	5,17
38	5,92	5,91	5,91	6,29	6,29	6,29	6,29
39	6,40	5,47	5,47	5,55	5,55	5,55	5,55
40	5,68	3,29	3,29	1,19	1,19	1,19	1,19
$\Delta \rho\text{IC}_{90}$	4,38	3,29	3,29	3,58	3,58	3,58	3,58

¹R_A – alinhamento rígido utilizando apenas o anel da piridina;
²R_B – alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (tossemicarbazona) c
³S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_C=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XVIII. Valores de ρIC_{90} observado e predito através dos modelos *CoMSIA*, modelo 3 (grupo de treinamento IIIA), para a série de 40 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	ρIC_{90}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,34	6,29	6,29	5,78	6,29	5,78
02	2,70	2,69	2,72	2,72	2,73	2,72	2,73
03	3,30	3,33	3,24	3,24	2,51	3,24	2,51
04	5,82	5,85	5,96	5,96	5,25	5,96	5,25
05	2,28	2,21	2,32	2,32	3,01	2,32	3,01
06	5,52	6,00	6,00	6,01	5,75	6,01	5,75
07	5,66	5,86	5,66	5,66	5,83	5,66	5,83
08	4,96	5,20	5,08	5,08	5,10	5,20	5,10
09	5,34	5,21	5,19	5,19	5,05	5,34	5,05
10	5,74	5,46	5,62	5,62	5,71	5,74	5,71
11	6,59	6,48	6,62	6,62	6,26	6,59	6,26
12	6,57	6,41	6,41	6,47	6,08	6,57	6,08
13	6,51	6,38	6,41	6,41	6,42	6,51	6,42
14	6,66	6,48	6,56	6,56	6,43	6,66	6,43
15	5,11	6,41	6,41	6,52	6,05	5,11	6,05
16	5,89	5,92	5,76	5,76	5,81	5,89	5,81
17	6,04	6,10	6,14	6,14	6,16	6,04	6,16
18	5,44	5,54	5,42	5,42	5,66	5,44	5,66
19	5,92	5,72	5,51	5,51	5,29	5,92	5,29
20	5,92	6,10	6,14	6,14	5,62	5,92	5,62
21	6,25	6,24	6,31	6,31	5,98	6,25	5,98
22	6,30	6,29	6,37	6,37	6,12	6,30	6,12
23	6,39	6,36	6,48	6,48	6,22	6,39	6,22
24	5,62	6,06	6,12	6,12	5,95	5,62	5,95
25	5,92	6,06	6,10	6,10	6,09	5,92	6,09
26	5,60	6,13	6,13	6,11	6,07	5,60	6,07
27	6,07	6,09	6,06	6,06	6,05	6,07	6,05
28	4,62	5,79	5,79	5,83	5,97	4,62	5,97
29	5,69	5,67	5,67	5,52	6,01	5,69	5,52
30	5,44	5,47	5,46	5,46	5,66	5,44	5,66
31	5,28	5,33	5,28	5,28	5,73	5,28	5,73
32	5,17	5,24	5,18	5,18	5,75	5,17	5,75
33	3,96	3,99	3,98	3,98	4,17	3,96	4,17
34	5,30	5,24	5,25	5,25	5,74	5,30	5,74
35	5,25	5,16	5,16	5,19	5,66	5,25	5,66
36	5,24	5,30	5,20	5,20	5,84	5,24	5,84
37	4,89	4,69	4,91	4,91	5,12	4,89	5,12
38	5,92	6,00	5,95	5,95	5,66	5,92	5,66
39	6,40	6,24	6,24	6,24	6,36	6,40	6,36
40	5,68	5,65	5,73	5,73	5,60	5,68	5,60
$\Delta \rho\text{IC}_{90}$	4,38	4,27	4,30	4,30	3,92	4,38	3,92

¹R_A – alinhamento rígido utilizando apenas o anel da piridina;
²R_B – alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (tossemicarbazona) c
³S_C – alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_C=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XIX. Valores de pIC_{50} observado e predito através dos modelos *CoMSIA*, modelo 1 (grupo de treinamento IB), para a série de 39 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		Rígido		R_g^2	Teste	Treinamento	Teste
		R_A^1	R_B^2				
01	6,55						
02	2,70	2,74		2,69	5,98	2,90	5,61
03	3,30	3,14		3,39		3,02	
04	5,82						
05	2,28	2,25			3,62	2,34	4,78
06	5,52	5,66		2,28		5,58	
07	5,66		5,88	5,55	5,75		5,78
08	4,96	5,24		4,83		4,88	
09	5,34	5,37		5,25		4,93	
10	5,74		5,28		5,08		5,56
11	6,59	6,33		6,53		6,26	
12	6,57	6,31		6,54		6,28	
13	6,51		6,15		6,24		6,34
14	6,66	6,44		6,45		6,28	
15	5,11	5,68		5,30		5,59	
16	5,89	5,86		5,85		6,08	
17	6,04		6,29		6,42		6,04
18	5,44	5,45		5,48		5,62	
19	5,42	5,49		5,41		5,40	
20	5,92	5,88		5,88		5,94	
21	6,25		6,04		6,11		5,99
22	6,30	6,14		6,25		6,11	
23	6,39	6,28		6,48		6,20	
24	5,62	5,63		5,56		5,89	
25	5,92	5,98		5,98		6,13	
26	5,60	5,74		5,74		6,03	
27	6,07		6,13		6,32		6,20
28	4,62	4,45		4,54		4,83	
29	5,69	5,59		5,57		5,46	
30	5,44	5,39		5,50		5,62	
31	5,28	5,16		5,28		5,15	
32	5,17		5,54		5,75		5,93
34	5,30	5,34		5,28		5,41	
35	5,25	5,26		5,21		5,19	
36	5,24	5,25		5,25		5,15	
37	4,89		5,35		5,39		5,48
38	5,92	6,07		5,97		5,76	
39	6,40	6,46		6,51		6,50	
40	5,68	5,75		5,82		5,84	
ΔpIC_{50}	4,38	4,21	1,57	4,26	2,80	4,16	1,56

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina;

² R_B - alinhamento rígido utilizando apenas o anel da piridina e os átomos C=NNC=SN do fragmento da (isosemicarbazona) e

³ Sc - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XX. Valores de pIC_{50} observado e predito através dos modelos *CoMSIA*, modelo 2 (grupo de treinamento IIB), para a série de 39 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		Rígido		R_g^2	Teste	Treinamento	Teste
		R_A^1	R_B^2				
01	6,55	6,17					
02	2,70	2,72		2,62		2,71	
03	3,30	3,34		3,37		3,31	
04	5,82	5,88		5,89		5,86	
05	2,28		4,49		4,93		5,35
06	5,52	5,71		5,79		5,78	
07	5,66	5,55		5,61		5,65	
08	4,96	4,85		4,97		5,00	
09	5,34		4,87		4,97		5,17
10	5,74	5,40		5,62		5,62	
11	6,59	6,47		6,49		6,54	
12	6,57	6,43		6,40		6,67	
13	6,51	6,23		6,14		6,59	
14	6,66		6,30		6,11		6,15
15	5,11	5,76		5,89		5,31	
16	5,89	5,82		5,78		5,91	
17	6,04	6,23		6,30		5,93	
18	5,44		5,57		5,77		5,76
19	5,42	5,51		5,37		5,35	
20	5,92	5,97		5,96		6,02	
21	6,25	6,15		6,14		6,21	
22	6,30		6,26		6,22		6,30
23	6,39	6,40		6,37		6,29	
24	5,62	5,61		5,74		5,62	
25	5,92		5,81		5,76		5,71
26	5,60	5,79		5,69		5,72	
27	6,07	6,28		6,25		6,12	
28	4,62	4,54		4,52		4,58	
29	5,69	5,94		5,45		5,76	
30	5,44		5,35		5,15		4,94
31	5,28		5,47		5,27		5,44
32	5,17	5,13		5,21		5,15	
34	5,30	5,24		5,38		5,30	
35	5,25	5,18		5,34		5,20	
36	5,24	5,20		5,22		5,07	
37	4,89	4,93		4,92		4,95	
38	5,92		5,48		5,15		5,10
39	6,40	6,39		6,33		6,46	
40	5,68		5,33		5,44		5,05
ΔpIC_{50}	4,38	3,75	1,81	3,87	1,29	3,96	1,36

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina;

² R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (isosemicarbazona) e

³ Sc - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XXI. Valores de pIC_{50} observado e predito através dos modelos *CoMSIA*, modelo 3 (grupo de treinamento IIIB), para a série de 39 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treino	Teste	Treino	Teste	Treino	Teste
01	6,55	6,36		6,31		5,91	
02	2,70	2,66		2,71		2,93	
03	3,30	3,32		3,23		3,01	
04	5,82	5,79		5,96		5,89	
05	2,28	2,25		2,36		2,51	
06	5,52		6,01		6,02		5,81
07	5,66	5,83		5,66		5,80	
08	4,96	5,25		5,08		4,93	
09	5,34	5,30		5,23		4,91	
10	5,74	5,52		5,65		5,70	
11	6,59	6,51		6,64		6,46	
12	6,57		6,44		6,50		6,22
13	6,51	6,41		6,43		6,61	
14	6,66	6,51		6,59		6,57	
15	5,11		6,44		6,55		6,23
16	5,89	5,89		5,77		5,93	
17	6,04	6,08		6,13		6,19	
18	5,44	5,46		5,42		5,56	
19	5,42	5,69		5,51		5,45	
20	5,92		6,09		6,13		5,90
21	6,25	6,23		6,31		6,17	
22	6,30	6,29		6,38		6,31	
23	6,39	6,35		6,49		6,40	
24	5,62		6,09		6,62		6,16
25	5,92	6,09		6,07		6,29	
26	5,60		6,13		6,12		6,10
27	6,07	6,12		6,05		5,96	
28	4,62		5,79		5,79		5,87
29	5,69		5,65		5,45		5,85
30	5,44	5,48		5,45		5,77	
31	5,28	5,27		5,25		5,28	
32	5,17	5,21		5,15		5,41	
34	5,30	5,24		5,26		5,45	
35	5,25		5,16		5,20		5,33
36	5,24	5,27		5,20		5,39	
37	4,89	4,72		4,87		4,64	
38	5,92	6,05		5,92		5,57	
39	6,40		6,28		6,24		6,40
40	5,68	5,63		5,69		5,76	
ΔpIC_{50}	4,38	4,26	1,28	4,28	1,42	4,10	1,07

¹R_A - alinhamento rígido utilizando apenas o anel da piridina;

²R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (isosemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XXII. Valores de pIC_{50} observado e predito através dos modelos *CoMSIA*, modelo 1 (grupo de treinamento IC), para a série de 30 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treino	Teste	Treino	Teste	Treino	Teste
01	6,55		6,09		6,11		6,22
11	6,59	6,35		6,32		6,36	
12	6,57	6,33		6,34		6,40	
13	6,51		6,17		6,23		6,30
14	6,66	6,36		6,32		6,40	
15	5,11	5,69		5,59		5,95	
16	5,89		5,87		5,92		6,00
17	6,04	6,26		6,16		6,06	
18	5,44	5,37		5,37		5,41	
19	5,42	5,96		5,76		5,80	
20	5,92		5,98		5,79		6,17
21	6,25	6,05		6,09		5,99	
22	6,30	6,11		6,16		6,06	
23	6,39		6,17		6,25		6,06
24	5,62	5,52		5,48		5,64	
25	5,92	6,09		6,14		5,92	
26	5,60	5,73		5,70		5,81	
27	6,07	6,28		6,32		6,17	
28	4,62	4,63		4,67		4,50	
29	5,69		5,23		5,42		5,15
30	5,44	5,42		5,52		5,36	
31	5,28	5,19		5,15		5,31	
32	5,17	5,11		5,12		5,26	
34	5,30	5,28		5,25		5,36	
35	5,25	5,20		5,18		5,20	
36	5,24	5,17		5,11		5,10	
37	4,89		5,24		5,26		5,53
38	5,92	6,02		6,07		5,96	
39	6,40	6,54		6,51		6,46	
40	5,68	5,74		5,82		5,94	
ΔpIC_{50}	2,04	1,91	0,94	1,84	0,99	1,96	1,15

¹R_A - alinhamento rígido utilizando apenas o anel da piridina;

²R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (isosemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XXIII. Valores de pIC_{50} observado e predito através dos modelos CoMSIA, modelo 2 (grupo de treinamento IIC), para a série de 30 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}					SEAL		
		Rígido			R_b^2	Teste	S_c^1		
		R_A^1	Teste	Treinamento			Treinamento	Teste	
01	6,55	6,30		6,32			6,19		
11	6,59	6,57		6,60			6,23		
12	6,57	6,56		6,63			6,20		
13	6,51	6,29		6,33			6,05		
14	6,66		6,18			5,96		5,80	
15	5,11	5,49		5,28			6,08		
16	5,89	5,73		5,85			6,25		
17	6,04	6,05		6,08			6,24		
18	5,44		5,68			5,88		6,18	
19	5,42	5,41		5,45			6,18		
20	5,92	6,02		5,98			6,09		
21	6,25	6,23		6,20			6,08		
22	6,30	6,34		6,31			6,03		
23	6,39	6,48		6,42			5,98		
24	5,62	5,58		5,53			5,83		
25	5,92	5,93		5,96			5,68		
26	5,60		6,09			6,05		5,77	
27	6,07	6,22		6,22			5,85		
28	4,62	4,45		4,60			5,20		
29	5,69	5,68		5,66			5,31		
30	5,44	5,48		5,62			5,70		
31	5,28	5,31		5,18			5,34		
32	5,17		5,57			5,74		5,55	
34	5,30	5,26		5,29			5,35		
35	5,25		5,16			5,18		5,32	
36	5,24	5,15		5,16			5,18		
37	4,89	5,04		4,87			5,00		
38	5,92	5,97		6,01			5,49		
39	6,40		6,13			5,93		5,67	
40	5,68		5,16			5,13		5,66	
ΔpIC_{50}	2,04	2,12	1,02	2,03		0,92	1,25	0,86	

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina.

² R_b - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (iossemicarbazona) e

³ S_c - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XXIV. Valores de pIC_{50} observado e predito através dos modelos CoMSIA, modelo 3 (grupo de treinamento IIC), para a série de 30 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}					SEAL		
		Rígido			R_b^2	Teste	S_c^1		
		R_A^1	Teste	Treinamento			Treinamento	Teste	
01	6,55	6,46		6,45			6,59		
11	6,59	6,68		6,63			6,59		
12	6,57	6,63		6,60			6,61		
13	6,51	6,48		6,52			6,54		
14	6,66	6,64		6,68			6,66		
15	5,11		6,62			6,66		6,60	
16	5,89	5,89		5,90			5,80		
17	6,04		6,18			6,16		5,89	
18	5,44	5,43		5,43			5,48		
19	5,42	5,48		5,40			5,50		
20	5,92	5,89		5,88			5,92		
21	6,25	6,11		6,16			6,12		
22	6,30		6,17			6,27		6,15	
23	6,39		6,20			6,35		6,14	
24	5,62	5,65		5,67			5,63		
25	5,92	6,05		6,04			5,91		
26	5,60	5,72		5,64			5,72		
27	6,07	6,10		6,05			6,02		
28	4,62		5,79			5,79		5,55	
29	5,69	5,68		5,69			5,74		
30	5,44	5,41		5,47			5,40		
31	5,28		5,74			5,69		5,89	
32	5,17	5,12		5,20			5,08		
34	5,30		5,37			5,30		5,42	
35	5,25	5,31		5,24			5,27		
36	5,24	5,25		5,24			5,25		
37	4,89	4,93		4,89			4,94		
38	5,92	5,82		5,86			5,85		
39	6,40	6,40		6,43			6,39		
40	5,68	5,59		5,65			5,70		
ΔpIC_{50}	2,04	1,75	1,25	1,74		1,36	1,72	1,18	

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina.

² R_b - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da (iossemicarbazona) e

³ S_c - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_S=1,0$ e $W_H=0,0$).

Tabela XXV. Valores de pIC_{50} observado e predito através dos modelos CoMSIA, modelo 1 (grupo de treinamento ID), para a série de 23 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treino	Teste	Treino	Teste	Treino	Teste
01	6,55	6,51	6,50	6,52	6,51	6,49	6,47
13	6,51	6,59	6,57	6,57	6,55	6,55	
14	6,66	6,01	5,98	6,20	6,01	6,01	
20	5,92	6,20	6,28	6,28	6,26	6,26	
21	6,25	6,26	6,28	6,30	6,26	6,26	
22	6,30	6,26	6,28	6,30	6,26	6,26	
23	6,39	6,28	6,28	6,30	6,26	6,26	
24	5,62	5,68	5,68	5,68	5,70	5,70	
25	5,92	6,19	6,19	6,16	6,16	6,16	
26	5,60	5,71	5,75	5,75	5,65	5,65	
27	6,07	6,16	6,14	6,14	6,16	6,16	
28	4,62	5,70	5,76	5,70	5,75	5,75	
29	5,69	5,70	5,67	5,67	5,75	5,75	
30	5,44	5,33	5,33	5,33	5,31	5,31	
31	5,28	5,33	5,33	5,33	5,34	5,34	
32	5,17	5,07	5,06	5,06	5,09	5,09	
33	3,96	3,96	3,97	3,97	3,96	3,96	
34	5,30	5,38	5,38	5,38	5,34	5,34	
35	5,25	5,30	5,30	5,30	5,26	5,26	
36	5,24	5,17	5,16	5,16	5,13	5,13	
37	4,89	4,88	4,87	4,87	4,92	4,92	
38	5,92	5,88	5,93	5,93	5,98	5,98	
39	6,40	6,49	6,50	6,50	6,61	6,61	
ΔpIC_{50}	2,70	2,63	0,74	2,6	0,81	2,59	0,79

¹R_A - alinhamento rígido utilizando apenas o anel da piridina;

²R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_F=1,0$ e $W_H=0,0$).

Tabela XXVI. Valores de pIC_{50} observado e predito através dos modelos CoMSIA, modelo 2 (grupo de treinamento IID), para a série de 23 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}				SEAL	
		R _A ¹		R _B ²		S _C ³	
		Treino	Teste	Treino	Teste	Treino	Teste
01	6,55	6,46	6,47	6,47	6,46	6,46	6,46
13	6,51	6,43	6,41	6,37	6,41	6,41	6,37
14	6,66	6,01	6,01	6,01	6,01	6,01	6,01
20	5,92	6,22	6,22	6,22	6,22	6,22	6,22
21	6,25	6,30	6,30	6,29	6,29	6,29	6,29
22	6,30	6,34	6,32	6,32	6,32	6,32	6,32
23	6,39	5,84	5,86	5,86	5,86	5,86	5,84
24	5,62	6,01	5,99	5,99	5,99	5,99	5,99
25	5,92	5,73	5,72	5,72	5,75	5,75	5,75
26	5,60	6,11	6,12	6,12	6,08	6,08	6,08
27	6,07	4,55	4,57	4,57	4,57	4,57	4,57
28	4,62	5,66	5,69	5,69	5,66	5,66	5,66
29	5,69	5,46	5,48	5,48	5,46	5,46	5,46
30	5,44	5,46	5,46	5,46	5,46	5,46	5,46
31	5,28	5,10	5,10	5,10	5,11	5,11	5,11
32	5,17	3,97	3,94	3,94	3,95	3,95	3,95
33	3,96	5,34	5,35	5,35	5,38	5,38	5,38
34	5,30	5,27	5,28	5,28	5,32	5,32	5,32
35	5,25	5,26	5,24	5,24	5,22	5,22	5,22
36	5,24	4,88	4,89	4,89	4,88	4,88	4,88
37	4,89	5,94	5,97	5,97	5,94	5,94	5,94
38	5,92	6,43	6,44	6,44	6,45	6,45	6,45
39	6,40	1,07	1,02	1,02	1,01	1,01	1,01
ΔpIC_{50}	2,70	2,49	2,53	2,53	2,51	2,51	2,51

¹R_A - alinhamento rígido utilizando apenas o anel da piridina;

²R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiocsemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_F=1,0$ e $W_H=0,0$).

Tabella XXVII. Valores de pIC_{50} observado e predito através dos modelos CoMSIA, modelo 3 (grupo de treinamento IIID), para a série de 23 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}						SEAL	
		Rígido			R_A^2			S_C^2	
		R_A^1	Teste	Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,39		6,39				6,48	
13	6,51	6,47		6,47				6,52	
14	6,66	6,62		6,57				6,61	
20	5,92	5,92		5,93				5,89	
21	6,25		6,19		6,20				6,18
22	6,30	6,31		6,32				6,29	
23	6,39	6,41		6,42				6,35	
24	5,62	5,75		5,75				5,70	
25	5,92	5,90		5,89				5,96	
26	5,60		6,08		6,10				5,99
27	6,07	6,18		6,19				6,09	
28	4,62	4,54		4,53				4,61	
29	5,69	5,67		5,67				5,65	
30	5,44		5,65		5,67				5,70
31	5,28	5,36		5,33				5,31	
32	5,17	5,15		5,17				5,13	
33	3,96	4,02		4,02				3,93	
34	5,30	5,30		5,29				5,32	
35	5,25	5,23		5,22				5,24	
36	5,24		5,41		5,31				5,25
37	4,89	4,82		4,84				4,94	
38	5,92		5,61		5,60				5,77
39	6,40	6,46		6,49				6,50	
ΔpIC_{50}	2,70	2,60	0,78	2,55	0,89			2,68	0,93

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina;

² R_A - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiossemicarbazona e

³ S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_F=1,0$, $W_E=1,0$ e $W_H=0,0$).

Tabella XXVIII. Valores de pIC_{50} observado e predito através dos modelos CoMSIA, modelo 1 (grupo de treinamento IE), para a série de 22 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}						SEAL	
		Rígido			R_A^2			S_C^2	
		R_A^1	Teste	Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,55		6,55		6,54		6,54	6,53
13	6,51		6,55		6,54		6,54		
14	6,66	6,68		6,67				6,69	
20	5,92	5,97		5,95				5,95	
21	6,25	6,19		6,20				6,20	
22	6,30	6,27		6,29		6,31		6,27	
23	6,39		6,29		6,31			6,28	
24	5,62	5,60		5,60		5,63		5,63	
25	5,92		6,20		6,17			6,21	
26	5,60	5,62		5,63				5,62	
27	6,07	6,11		6,12		6,10		6,10	
28	4,62		5,74		5,69			5,76	
29	5,69	5,68		5,68		5,69		5,69	
30	5,44	5,45		5,44		5,42		5,42	
31	5,28	5,26		5,28		5,29		5,29	
32	5,17	5,17		5,17		5,17		5,17	
34	5,30	5,31		5,31		5,31		5,31	
35	5,25	5,22		5,23		5,24		5,24	
36	5,24	5,24		5,24		5,25		5,25	
37	4,89	4,90		4,90		4,88		4,88	
38	5,92	5,92		5,90		5,90		5,90	
39	6,40		6,52		6,55			6,65	
ΔpIC_{50}	2,04	1,78	0,81	1,77	0,86			1,81	0,89

¹ R_A - alinhamento rígido utilizando apenas o anel da piridina;

² R_A - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiossemicarbazona e

³ S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_F=1,0$, $W_E=1,0$ e $W_H=0,0$).

Tabela XXIX. Valores de pIC_{50} observado e predito através dos modelos *CoMSIA*, modelo 2 (grupo de treinamento IIE), para a série de 22 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}						SEAL	
		R _A ¹		R _B ²		R _C ³		S _C ³	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,54		6,50				6,51	
13	6,51	6,49		6,43				6,44	
14	6,66		6,45		6,37				6,39
20	5,92	5,97		6,02				6,01	
21	6,25	6,22		6,23				6,23	
22	6,30	6,31		6,30				6,29	
23	6,39	6,36		6,33				6,31	
24	5,62		5,80		5,86				5,82
25	5,92	5,96		5,98				6,00	
26	5,60	5,62		5,68				5,70	
27	6,07		6,12		6,13				6,11
28	4,62	4,60		4,61				4,62	
29	5,69	5,70		5,67				5,66	
30	5,44	5,41		5,50				5,48	
31	5,28		5,39		5,36				5,28
32	5,17	5,20		5,10				5,10	
34	5,30		5,32		5,33				5,33
35	5,25	5,25		5,26				5,27	
36	5,24	5,25		5,24				5,22	
37	4,89	4,88		4,85				4,86	
38	5,92	5,86		5,92				5,91	
39	6,40	6,42		6,44				6,47	
ΔpIC_{50}	2,04	1,94	1,13	1,89	1,04			1,89	1,11

¹R_A - alinhamento rígido utilizando apenas o anel da piridina;

²R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_E=1,0$ e $W_H=0,0$).

Tabela XXX. Valores de pIC_{50} observado e predito através dos modelos *CoMSIA*, modelo 3 (grupo de treinamento IIE), para a série de 22 compostos, utilizando diferentes procedimentos de alinhamento.

Número do Composto	Observado	pIC_{50}						SEAL	
		R _A ¹		R _B ²		R _C ³		S _C ³	
		Treinamento	Teste	Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
01	6,55	6,51		6,51				6,52	
13	6,51	6,54		6,55				6,54	
14	6,66	6,65		6,64				6,63	
20	5,92	5,89		5,87				5,87	
21	6,25		6,18		6,18				6,18
22	6,30	6,29		6,31				6,30	
23	6,39	6,35		6,38				6,35	
24	5,62	5,67		5,67				5,69	
25	5,92	5,97		5,95				5,96	
26	5,60		6,04		6,08				5,99
27	6,07	6,10		6,06				6,05	
28	4,62	4,63		4,62				4,63	
29	5,69	5,66		5,67				5,67	
30	5,44		5,61		5,63				5,65
31	5,28	5,27		5,29				5,27	
32	5,17	5,15		5,15				5,15	
34	5,30	5,31		5,33				5,31	
35	5,25	5,23		5,25				5,24	
36	5,24		5,42		5,32				5,32
37	4,89	4,89		4,88				4,89	
38	5,92		5,79		5,76				5,82
39	6,40	6,43		6,42				6,48	
ΔpIC_{50}	2,04	2,02	0,76	2,02	0,86			2,00	0,86

¹R_A - alinhamento rígido utilizando apenas o anel da piridina;

²R_B - alinhamento rígido utilizando N(1) e C(2) do anel da piridina e os átomos C=NNC=SN do fragmento da tiossemicarbazona) e

³S_C - alinhamento utilizando o programa SEAL ($\alpha=0,30$, $W_E=1,0$, $W_E=1,0$ e $W_H=0,0$).