



**UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE QUÍMICA**

**Departamento de Físico-Química**

**Estudos de QSAR-2D  
aplicados a  
Diterpenóides Clerodanos e Dibenzoidrazinas**

**TESE DE DOUTORADO**

**Bruno Cramer**

**Orientador: Prof. Dr. Yuji Takahata  
Co-Orientador: Prof. Dr. Munir Salomão Skaf  
Área: Físico-química  
Campinas  
2011**

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

C848e Cramer, Bruno.  
Estudos de QSAR-2D aplicados a diterpenoides clerodanos e dibenzoilidrazinas / Bruno Cramer. --  
Campinas, SP: [s.n], 2011.

Orientador: Prof. Dr. Yuji Takahata.  
Coorientador: Prof. Dr. Munir Salomão Skaf.

Doutorado - Universidade Estadual de Campinas,  
Instituto de Química.

1. QSAR. 2. Diterpenoides. 3. Dibenzoilidrazinas.  
4. Docking. I. Takahata, Yuji. II. Skaf, Munir Salomão.  
III. Universidade Estadual de Campinas. Instituto de  
Química. IV. Título.

**Título em inglês:** Studies of QSAR-2D applied to clerodane diterpenoids and dibenzoylhydrazines

**Palavras-chaves em inglês:** QSAR, Diterpenoids, Dibenzoylhydrazines, Docking

**Área de concentração:** Físico-Química

**Titulação:** Doutor em Ciências

**Banca examinadora:** Prof. Dr. Yuji Takahata (orientador), Profa. Dra. Antonia Tavares do Amaral (IQ-USP-SP), Prof. Dr. Anderson Coser Gáudio (IF-UFES-ES), Prof. Dr. Adalberto Bono Maurizio Bassi (IQ-UNICAMP), Prof. Dr. Roy Edward Bruns (IQ-UNICAMP)

**Data de defesa:** 16/08/2011



*Dedico esta tese à Krystyna Rudzki Cramer, minha espôsa,  
dedicada companheira em todos os momentos.*



*A ciência existe, antes de tudo, para aliviar a nossa perplexidade natural e desenvolver novas faculdades, em nossa vida cada vez mais intensa, para evitar o nocivo e facilitar o útil.*

*Johann Wolfgang von Goethe, 1749-1832*



## **AGRADECIMENTOS**

Ao Prof. Dr. Yuji Takahata, que além de meu orientador no doutorado, concedeu-me o prazer de sua amizade;

Ao Prof. Dr. Munir Salomão Skaf, pelo apoio como co-orientador.

Aos Profs. Drs. Nelson Morgon, Pedro Antonio Muniz Vazquez, Rogerio Custódio, pelo suporte técnico, e por me concederem o prazer de suas amizades.

Aos colegas de doutoramento Euzébio Guimarães Barbosa, Paulo Cesar Telles de Souza, Rodrigo Leandro Silveira e Thiago Gomes, que me auxiliaram na solução de problemas computacionais específicos.

Aos Profs. Drs. integrantes da banca examinadora, pelas suas contribuições, enriquecendo o debate e a tese.

A Izabel Aquino Calasso (Bel) da CPG pela atenção, orientação e amizade.

A VeraChemLLC que gentilmente concedeu a licença dos aplicativos Vconf e Vcharge.

À minha esposa, Krystyna, que durante todos estes anos me apoiou, estimulou, e muitas vezes sacrificou suas próprias aspirações para que eu pudesse manter e alcançar o meu objetivo.



# CURRÍCULO

## Dados Pessoais

Nome: Bruno Cramer

Nascimento: Rio de Janeiro, 11-05-1948

Nacionalidades: Brasileira, Alemã.

## 1. Formação Acadêmica

Mestrado em Ciências Farmacêuticas – USF-Bragança Paulista (2003-2005) – Dissertação: Caracterização Farmacognóstica Comparativa das Espécies *Casearia sylvestris* Swartz e *Casearia decandra* Jacquin Flacourtiaceae

Especialização *Lato sensu*: Validação de Plantas Medicinais / Obtenção de Fitofármacos/Fitofarmacotécnica. Fitoterapia - Instituto Brasileiro de Estudos Homeopáticos – IBEHE-UNAERP (1996-1997 – 620 h/a)  
Monografia: Identificação de Drogas Através de Fotomicrografias.

Bacharelado em Física – PUCSP - (1971-1977)

## 2. Atividade Profissional

Rhodia Indústrias Químicas e Têxteis – Divisão Têxtil – 1966 a 1971-  
Desenho Técnico (4,5 anos)

Indústrias Químicas Eletrocloro (Solvay S.A) -1971 a 1972  
Controle de Projetos – PERT/CPM (1,5 anos)

Volkswagen do Brasil S.A (VWB)– 1972 a 2002 em áreas com função técnica e administrativa (30 anos)

- Engenharia de Manufatura – Montagem Final
- Planejamento de Produção de Veículos no Exterior
- Pesquisa e Desenvolvimento – Energias Alternativas
- Eng<sup>a</sup> de Processos – Desenvolvimento de Novos Veículos na Volkswagen AG – Alemanha - 1998 a 2001 (4 anos)

## 3. Idiomas com proficiência nas três formas de expressão: Alemão, Inglês, Francês e Espanhol.

#### 4. Participação em Eventos Científicos e Publicação

- ❑ 5º Simpósio Brasileiro de Química Medicinal – Ouro Preto – MG 6-9 de Novembro de 2010. Poster: “Conformational sampling and 2D-QSAR: A new approach applied to Anthraquinones”. B. Cramer; Y. Takahata.
- ❑ Cramer, B.; Takahata, Y. QSAR of Clerodane Diterpenoids with Cytotoxic and Antitumoral Activity Against V-79 Cells. *QSAR & Comb. Sci.*, 28, No. 10, 1176 – 1186, 2009.
- ❑ 4º. Simpósio Brasileiro de Química Medicinal – Porto de Galinhas – PE – 9 a 13 de Novembro de 2008. Poster: “QSAR-2D Study of Clerodane Diterpenoids: Compounds with Cytotoxicity and Antitumoral Activity. B. Cramer; Y. Takahata.
- ❑ VII Encontro da SMPMat 2008 – Guarujá –SP – 28 de Setembro a 2 de Outubro de 2008.
- ❑ Sociedade Brasileira de Química - 31ª Reunião Anual – Águas de Lindóia – SP – 26-29 de Maio de 2008. Poster: “Estudo QSAR Comparativo de Dibenzoilidrazinas (DBH) Substituídas”. B. Cramer, Y. Takahata.
- ❑ XIV Simpósio Brasileiro de Química Teórica – Poços de Caldas – MG – de 18 a 21 de Novembro de 2007.
- ❑ 3º. Simpósio Brasileiro de Química Medicinal – São Pedro – SP – 12 a 15 de Novembro de 2006. Poster: Core-Electron Binding Energy as Electronic Molecular Descriptor in QSAR (Quantitative Structure-Activity Relationships) Analysis. B. Cramer, Y. Takahata.
- ❑ XII Simpósio Brasileiro de Química Teórica. São Pedro – SP -20-23 Novembro de 2005.
- ❑ IV Jornada Paulista de Plantas Mediciniais – UNAERP – Ribeirão Preto – SP – 12 a 15 de Outubro de 1999.
- ❑ III Jornada Paulista de Plantas Mediciniais – CPQBA-UNICAMP - Hotel Solar das Andorinhas - Campinas – SP – 11 a 15 de Outubro de 1997. Poster: “Identificação Histológica de Drogas Vegetais”. B. Cramer.
- ❑ II WorkShop de Plantas Mediciniais de Botucatu – UNESP – Botucatu – 14 a 15 de Junho de 1996.

## RESUMO

A tese foi organizada em forma de cinco estudos de caso de QSAR-2D, usando compostos da literatura. Foram utilizados diterpenóides clerodanos e N,N'-dibenzoil-N-t-butilidrazinas (DBHs) para desenvolver modelos de QSAR, propondo o uso de uma metodologia que emprega análise conformacional. Os modelos de QSAR são construídos pela escolha do confômero que melhor reproduz a bioatividade. O critério de seleção dos confômeros é orientado pelo resíduo entre valor observado e o predito. O método proposto aumenta a qualidade de predição interna, auxilia a análise de *outliers*, e em casos específicos, pode proporcionar melhor interpretabilidade do comportamento de descritores em relação à atividade biológica. Após validação interna e externa, os modelos de QSAR foram empregados para simular a proposta de novos compostos. Duas metodologias independentes de regressão foram usadas, Regressão Linear Múltipla (MLR) e Máquinas de Vetores de Suporte (SVM), objetivando produzir modelos de QSAR equivalentes, ambos produzindo resultados próximos de predição no conjunto de treinamento, auxiliando na análise e seleção de resultados preditos de novos compostos, aumentando a confiabilidade e robustez dos modelos de QSAR. Foram propostos novos diterpenóides clerodanos derivados de produtos naturais (n = 113) e outros 53 propostos *in silico* contra células V79. Novas dibenzolidrazinas propostas *in silico* tiveram sua atividade inseticida e larvicida predita para novos bioensaios contra *Spodoptera exigua* (n = 13) e *S. frugiperda* (n = 30). Através do estudo de homologia a tese contribui com o domínio ligante do receptor ecdisona de *S. exigua* (SeEcR-LBD). Empregando *docking*, informações foram obtidas sobre as possíveis interações das DBHs com os aminoácidos, uma contribuição na literatura pertinente. DBHs halogenadas foram analisadas identificando a provável interação dos halogênios com os aminoácidos no domínio ligante do receptor ecdisona de *Heliothis virescens* (HvEcR-LBD), análise estendida a *S. exigua* (SeEcR-LBD).



## ABSTRACT

The thesis was organized in the form of five 2D-QSAR case studies using compounds from literature. Clerodane diterpenoids and N, N'-dibenzoyl-N-t-butylhydrazines (DBHs) were used to develop predictive QSAR models proposing a methodology that uses conformational analysis. QSAR models are built by choosing the conformer that best reproduces the experimental bioactivity. The conformer selection is oriented on a residual criterion between observed and predicted values. The proposed method increases the quality of internal prediction, aids the analysis of outliers, and in specific cases, may provide better interpretability of the descriptor behavior in relation to the bioactivity. After internal and external validation, the QSAR models were used to simulate the proposition of new compounds. Two independent regression methodologies, Multiple Linear Regression (MLR) and Support Vector Machine (SVM) were used to produce equivalent QSARs, both producing close predicted results of the training set and aiding in the analysis and selectivity of MLR-SVM QSARs predicted results of new compounds, increasing the reliability and robustness of the QSAR models. New clerodane diterpenoids derived from natural products (n = 113) and another 53 *in silico* proposed, have their bioactivity predicted against V79 cells. New *in silico* dibenzoylhydrazines against *Spodoptera exigua* (n = 13) and *S. frugiperda* (n = 30) have their insecticidal and larvicidal activity predicted and proposed for new bioassays. Through homology study the thesis contributes with the *S. exigua* ecdyson receptor ligand-binding domain (SeEcR-LBD). Using docking, information was obtained about possible interactions of DBHs with amino acids, a contribution to the entailed literature. Halogen substituted DBHs were analyzed identifying their probable interaction with amino acids of *Heliothis virescens* ecdyson receptor ligand binding domain (HvEcR-LBD), analysis extended to *S. exigua* (SeEcR-LBD).



## ÍNDICE

<b>ÍNDICE</b>	
• Lista de Abreviaturas.....	xxi
• Lista de Tabelas.....	xxiii
• Lista de Figuras.....	xxvii
1. Introdução.....	1
2. Objetivos .....	5
3. Revisão Bibliográfica.....	7
3.1 Diterpenóides e QSAR .....	7
3.2 Dibenzoilidrazinas e QSAR .....	11
4. Material e Métodos – Tópicos Gerais .....	15
4.1 Fundamentos Teóricos – QSAR E Regressão Multivariada.....	15
4.2 Estruturas Químicas – Geometria – Descritores.....	19
4.3 Metodologias de Validação – Domínio de Aplicabilidade ( AD – Applicability Domain ).....	22
4.4 Análise Conformacional.....	27
4.5 Proposta de Novos Compostos – Similaridade Molecular.....	30
5. Estudos de QSAR.....	33
5.1 Diterpenóides Clerodanos contra células V79.....	33
5.1.1 Introdução.....	33
5.1.2 Métodos.....	35
5.1.3 Resultados e Discussão.....	36
5.1.3.1 QSAR e Descritores – Estudo <i>sem</i> Análise Conformacional.....	36
5.1.3.2 QSAR - <i>Outliers</i> e Domínio de Aplicabilidade.....	41
5.1.3.3 QSAR – Testes de Robustez e Validação.....	43
5.1.3.4 Proposta de Diterpenóides Clerodanos (conforme equação 5.1.-1).....	48
5.1.3.5 QSAR e Descritores – Estudo <i>com</i> Análise Conformacional .....	50
5.1.3.6 QSAR – <i>Outliers</i> e Domínio de Aplicabilidade.....	53
5.1.3.7 QSAR – Testes de Robustez e Validação.....	54
5.1.3.8 Proposta de Diterpenóides Clerodanos (conforme equação 5.1-7).....	59

**ÍNDICE**

---

5.1.4	Conclusão.....	65
5.2	<b>Dibenzoilidrazinas contra <i>Chilo suppressalis</i> ( Walker)</b> <b>Lepidóptera, Noctuidae.....</b>	68
5.2.1	Introdução.....	68
5.2.2	Métodos.....	68
5.2.3	Resultados e Discussão.....	69
5.2.3.1	QSAR e Descritores.....	69
5.2.3.2	QSAR - <i>Outliers</i> e Domínio de Aplicabilidade.....	73
5.2.3.3	QSAR – Testes de Robustez e Validação.....	74
5.2.3.4	Proposta de Novos Compostos.....	77
5.2.4	Conclusão.....	82
5.3	<b>Dibenzoilidrazinas contra <i>Leptinotarsa decemlineata</i> (Say)</b> <b>Coleóptera, Crysomelidae.....</b>	83
5.3.1	Introdução.....	83
5.3.2	Métodos.....	84
5.3.3	Resultados e Discussão.....	85
5.3.3.1	QSAR e Descritores.....	85
5.3.3.2	QSAR - <i>Outliers</i> e Domínio de Aplicabilidade.....	97
5.3.3.3	QSAR – Testes de Robustez e Validação.....	98
5.3.3.4	Proposta de Novos Compostos.....	101
5.3.4	Conclusão.....	105
5.4	<b>Dibenzoilidrazinas contra <i>Spodoptera exigua</i> (Huebner)</b> <b>Lepidóptera, Noctuidae.....</b>	106
5.4.1	Introdução.....	106
5.4.2	Métodos.....	107
5.4.3	Resultados e Discussão.....	124
5.4.3.1	QSAR e Descritores.....	124
5.4.3.2	QSAR - <i>Outliers</i> e Domínio de Aplicabilidade.....	133
5.4.3.3	QSAR – Testes de Robustez e Validação.....	133
5.4.3.4	Proposta de Novos Compostos.....	139
5.4.3.5	<i>Docking</i> de Compostos Halogenados.....	140
5.4.4	Conclusão.....	160

**ÍNDICE**

---

<b>5.5</b>	<b>Dibenzoilidrazinas contra <i>Spodoptera frugiperda</i> (<i>J.E.Smith</i>) Lepidóptera, Noctuidae.....</b>	<b>162</b>
5.5.1	Introdução.....	162
5.5.2	Métodos.....	163
5.5.3	Resultados e Discussão.....	167
5.5.3.1	QSAR –MLR e Descritores - QSAR - SVM.....	167
5.5.3.2	QSAR - <i>Outliers</i> e Domínio de Aplicabilidade.....	176
5.5.3.3	QSAR – Testes de Robustez e Validação.....	178
5.5.3.4	Proposta de Novos Compostos.....	182
5.5.3.5	Estudo de Reatividade.....	184
5.5.4	Conclusão.....	196
<b>6.</b>	<b>CONCLUSÃO GERAL E RECOMENDAÇÕES.....</b>	<b>199</b>
•	REFERÊNCIAS BIBLIOGRÁFICAS.....	203
•	ANEXOS.....	219
•	APÊNDICES.....	241
1.	QSAR of Clerodane Diterpenoids with Cytotoxic and Antitumoral Activity Against V79 Cells.	
2.	Pesquisas realizadas com N,N'-dibenzoil-N-t-butilidrazinas (DBHs) Período: 1990 – 2009.	
3.	Estruturas dos diterpenóides clerodanos.	



**ADF** (*Amsterdam Density Functional*)

**AM1** (Austin Model 1) Método quântico semiempírico

**Blind Set** Conjunto de compostos cuja bioatividade é desconhecida. **Pseudo blind set** conjunto de compostos com bioatividade conhecida.

**CV** (*Cross validation*) Validação cruzada

**DBH** (N,N'-dibenzoil-N-t-butilidrazinas) Dibenzoilidrazina

**DNA** (*Deoxidribonucleic acid*) Ácido deoxiribonucleico

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

**EcR-USP** (*Ecdyson Receptor-Ultraspiracl*) Receptor ecdisona - Ultraespiráculo

**F** Parâmetro estatístico de Fisher

**HvEcR** (*Heliothis virescens ecdyson receptor*)

**HOMO** (*Highest Occupied Molecular Orbital*) Orbital molecular mais alto ocupado.

**IC<sub>50</sub>** Concentração molar inibitória que mede a efetividade de um composto inibir uma função química ou biológica em 50% da população

**JH** (*Juvenile Hormones*) Hormônios Juvenis

**LBD** (*Ligand Binding Domain*) Domínio de ligação química do ligante

**LD<sub>50</sub>** Dose letal média (concentração letal para 50% da população)

**LOO-CV** (*Leave-One-Out Cross-validated*) Eliminação de um elemento do conjunto de dados e validação cruzada.

**LNO** (*Leave-N-Out*) Validação com a eliminação de N elementos do conjunto de dados

**MDM** (*Molegro Data Modeler*) Aplicativo

**MLR** (*Multiple Linear Regression*) Regressão Linear Múltipla

**MMFF** (*Molecular Merck Force Field*) Campo de força Merck da Mecânica Molecular.

**MSE** (*Mean Square Error*) Erro Médio Quadrático

**N, n** Número de compostos ou elementos

**PLS** (*Partial Least Squares*) Método dos mínimos quadrados parciais.

**PRESS** (*PREdictive Sum of Squares*)

- Q<sub>Loo-cv</sub>** (*Leave-One-Out Cross-Validation*) Representa a validação cruzada eliminando um composto do conjunto de dados.
- QSAR** (*Quantitative Structure-Activity Relationships*) Relações quantitativas entre estrutura química e atividade biológica
- R** Coeficiente de correlação de Pearson
- R<sup>2</sup>** Coeficiente de determinação
- RBF** (*Radial Basis Function*)
- RDF** (*Radial Distribution Function*)
- RMN** Ressonância Magnética Nuclear
- RMSD(E)** (*Root Mean Square Deviation(Error)*) Desvio Quadrático Médio ou Erro Médio Quadrático
- σ** Desvio padrão ou sigma de Hammett
- SeEcR** (*Spodoptera exigua ecdysone receptor*)
- SAR** (*Structure-Activity Relationships*)
- SDEP** (*Standard Error of Prediction*) Desvio padrão sobre a predição
- SLT** (*Statistical Learning Theory*) Teoria de Aprendizado Estatístico
- SPress** Desvio padrão da validação cruzada
- SVM** (*Support Vector Machine*)
- SVR** (*Support Vector Regression*)

**Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
LISTA DE TABELAS**

---

xxiii

Paginação: [A / nn = anexo/pág.respectiva ] [nn = pág. no texto]

1. Casearinas e Descritores.....	<b>39</b>
2. Comparativo de predição de modelos de QSAR-MLR e QSAR-SVM.....	<b>41</b>
3. Predição interna e externa dos modelos de QSAR – Métodos MLR e SVM....	<b>A/219</b>
4. Validação do conjunto de teste n = 5 (Casearinas Aa, K, L, M, N).....	<b>45</b>
5. Validação do modelo de QSAR-MLR (n = 18) pelos parâmetros RQK .....	<b>47</b>
6. Validação do modelo de QSAR-MLR (n = 13) pelos parâmetros RQK.....	<b>48</b>
7. <b>[a,b,c]</b> Valores preditos IC <sub>50</sub> de diterpenóides clerodanos contra células tumorais V79 com modelos de QSAR-MLR Eq.5.1-6 e QSAR-SVM (n = 13)..	<b>A/220</b>
8. Valores preditos IC <sub>50</sub> da Casearina G substituída em R <sup>1</sup> , R <sup>4</sup> e R <sup>5</sup> contra células tumorais V79 com modelos de QSAR-MLR (Eq. 5.1-7) e QSAR-SVM (n = 18).....	<b>A/223</b>
9. Comparativo de predição dos métodos MLR e SVM aplicando análise conformacional.....	<b>52</b>
10. Validação do conjunto de teste n = 5 (Casearinas Aa, K, L, M,N).....	<b>56</b>
11. <b>[a,b,c]</b> Valores preditos IC <sub>50</sub> de diterpenóides clerodanos contra células tumorais V79 com modelos de QSAR-MLR Eq. (30) e QSAR-SVM (n = 18)..	<b>A/224</b>
12. Validação do modelo de QSAR-MLR (n = 18) pelos parâmetros RQK.....	<b>58</b>
13. Validação do modelo de QSAR-MLR (n = 13) pelos parâmetros RQK.....	<b>59</b>
14. Predição de compostos derivados da Casearborina A substituída em R <sup>4</sup> e R <sup>5</sup> .....	<b>A/227</b>
15. Predição de compostos substituídos derivados das Casearvestrinas A, B e C.....	<b>A/228</b>
16. Predição de compostos substituídos derivados das Caseamembrinas A.....	<b>A/229</b>
17. QSAR de DBHs contra <i>Chilo suppressalis</i> antes e após análise conformacional.....	<b>A/230</b>
18. Validação do modelo de QSAR-MLR (n = 30) pelos parâmetros RQK.....	<b>77</b>

**Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
LISTA DE TABELAS**

---

xxiv

19.	Validação do modelo de QSAR (n = 22) pelos parâmetros RQK.....	77
20.	Valores observados pLD <sub>50</sub> , preditos pelo modelo de QSAR e selecionados pelo índice de similaridade euclidiano.....	A/231
21.	Compostos DBHs usados para desenvolvimento do modelo de QSAR de Dibenzoilidrazinas contra <i>Leptinotarsa decemlineata</i> .....	A/232
22.	Comparativo de resultados obtidos entre modelos de QSAR-MLR e QSAR-SVM Dibenzoilidrazinas contra <i>Leptinotarsa decemlineata</i> .....	A/233
23.	Validação do modelo de QSAR-MLR (n = 30) pelos parâmetros RQK .....	101
24.	Validação do modelo de QSAR-MLR ( n = 24) pelos parâmetros RQK.....	101
25.	Conjunto de compostos (pseudo <i>blind set</i> ) simulando a proposta de novos compostos contra <i>Leptinotarsa decemlineata</i> .....	A/234
26.	Interação não-covalente intermolecular de halogênios <sup>118</sup> .....	122
27.	Distâncias intermoleculares X...Y menores que a soma dos respectivos raios de van der Waals <sup>118</sup> .....	122
28.	Interações de halogênios com aminoácidos <sup>119</sup> .....	122
29.	Interação intermolecular associada a benzenos halogenados <sup>118</sup> .....	122
30.	Compostos usados para desenvolvimento dos modelos de QSAR pelos métodos MLR e SVM de dibenzoilidrazinas contra <i>Spodoptera exigua</i> .....	A/235
31.	Valores para o estudo $I = f(\alpha_{xx}, \alpha_{yy}, \alpha_{zz})$ .....	128
32.	Resultados de predição obtidos com os modelos de QSAR com n = 30 compostos.....	137
33.	Validação do modelo de QSAR-MLR (n = 37) pelos parâmetros RQK.....	138
34.	Validação do modelo de QSAR-MLR (n = 30) pelos parâmetros RQK.....	138
35.	Proposta de novos compostos - Dibenzoilidrazinas contra <i>S. exigua</i> .....	139
36.	Possíveis interações não-covalentes dos átomos halogênios com átomos dos aminoácidos de HvEcR-LBD.....	144

**Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzoidrazinas  
LISTA DE TABELAS**

xxv

---

37.	Possíveis interações não-covalentes dos átomos halogênios com os aminoácidos de SeEcR-LBD de <i>Spodoptera exigua</i> e estimativas da energia pelas forças dispersivas de London.....	146
38.	Possíveis interações- $\pi$ de halogênios da dibenzoidrazina com a Tyr 403, e interação S- $\pi$ da Met380 no domínio ligante SeEcR-LBD.....	147
39.	Valores de RMSD e PMF score obtidos no <i>docking</i> de dibenzoidrazinas halogenadas alojadas no receptor da proteína homóloga SeEcR-LBD de <i>Spodoptera exigua</i> .....	153
40.	Valores X-score extraídos do resultado de <i>docking</i> de dibenzoidrazinas halogenadas alojadas no receptor da proteína homóloga SeEcR-LBD de <i>Spodoptera exigua</i> .....	154
41.	Compostos propostos contra <i>S. exigua</i> . Resultados de modelos de QSAR-MLR e QSAR-SVM e correlação com o PMF_score.....	156
42.	Compostos propostos contra <i>S. exigua</i> . Resultados dos modelos de QSAR-MLR, QSAR-SVM e correlação dos valores pLD <sub>50</sub> com X-score.....	157
43.	Valores utilizados no modelo de QSAR de Ogura <i>et al.</i> <sup>26</sup> e no modelo de QSAR desenvolvido por este estudo de dibenzoidrazinas contra <i>Spodoptera frugiperda</i> .....	A/236
44.	Resultados de predição dos modelos de QSAR-MLR e QSAR-SVM.....	175
45.	Validação do modelo de QSAR-MLR (n = 18) pelos parâmetros RQK.....	180
46.	Validação do modelo de QSAR-MLR (n = 13) pelos parâmetros RQK.....	180
47.	Proposta <i>in silico</i> de DBHs contra <i>Spodoptera frugiperda</i> . Compostos selecionados pelo índice de similaridade molecular ShaEP e respectivo intervalo de predição em que o composto está contido.....	A/237
48.	Compostos propostos contra <i>Spodoptera frugiperda</i> e predição dos valores pIC <sub>50</sub> pelos modelos de QSAR-MLR e SVM.....	A/238
49.	Dibenzoidrazinas com introdução de um fenil substituído contra <i>Spodoptera frugiperda</i> . Compostos produzidos <i>in silico</i> .....	A/239
50.	Valores usados para análise de contribuições eletrofílica e nucleofílica de grupo nas dibenzoidrazinas contra <i>S. frugiperda</i> .....	186

**Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzolidrazinas  
LISTA DE TABELAS**

---

xxvi

<b>51.</b>	Valores usados para análise comparativa de contribuições eletrofílica e nucleofílica de grupo com as dibenzolidrazinas contra <i>S. frugiperda</i> e análise do composto D18.....	<b>186</b>
<b>52.</b>	Contribuição nucleofílica e eletrofílica de grupo.....	<b>187</b>
<b>53.</b>	Contribuição nucleofílica e eletrofílica de grupo.....	<b>192</b>
<b>54.</b>	Valores que geram o modelo de QSAR Eq.(5.5-30).....	<b>193</b>
<b>55.</b>	Validação do modelo de QSAR-MLR (n = 10) pelos parâmetros RQK.....	<b>194</b>

Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
LISTA DE FIGURAS

xxvii

---

1.	Biciclofitanos compostos constituídos por quatro unidades isoprênicas e a numeração empregada.....	8
2.	Diterpenóides clerodanos – Estrutura geral das casearinas.....	9
3.	Estrutura química da dibenzoilidrazina RH-5849.....	11
4.	Hormônio de crescimento <i>20-hidroxiecdisona</i> (20E) – Estrutura química (Nr.CAS 3604-87-3).....	12
5.	Methyl(2E,6E)-10R,11S-(oxiranyl)-3,7-diethyl-11-methyl-2,6-tridecadienoate, nominado também como J01. Exemplo de hormônio juvenil de Lepidoptera.....	12
6.	Dibenzoilidrazina ensaiada contra <i>Choristoneura fumiferana</i> .....	13
7.	Hiperplano (linha central) definindo uma superfície linear separando os elementos de treinamento em duas classes, e vetores suporte (linhas pontilhadas).....	19
8.	Construção do modelo de QSAR pela seleção do confôrmero através do descritor Mor25v, que melhor reproduz a bioatividade. LogP e $\sigma$ -Hammett não variam com a conformação.....	29
9.	Estrutura das Casearinas e a atividade citotóxica medida pelo índice inibitório IC <sub>50</sub> contra células pulmonares (fibroblastos) V79 de Hamster Chinês <sup>76</sup> .....	33
10.	(a)Q <sup>2</sup> <sub>LOO-CV</sub> em função da variação do parâmetro $\varepsilon$ mantendo constantes C e $\gamma$ ; (b) Q <sup>2</sup> <sub>LOO-CV</sub> em função da variação do parâmetro $\gamma$ mantendo constantes C e $\varepsilon$ . Para o menor RMSE seleciona-se o valor Q <sup>2</sup> <sub>LOO-CV</sub> correspondente (Q <sup>2</sup> <sub>LOO-CV</sub> =0,824).....	40
11.	Diagrama de Williams – Domínio de aplicabilidade das Casearinas. Hat crítico 3h=0,923. No conjunto de predição externa obtida pela Eq. (5.1-6) a Casearina M foi classificada como <i>outlier</i> .....	42
12.	Valores pIC <sub>50</sub> observados e preditos pelos modelos de QSAR-MLR e QSAR-SVM contra células tumorais V79 .....	43
13.	Gráfico dos valores R <sup>2</sup> e Q <sup>2</sup> em função das randomizações da bioatividade experimental (Y- <i>randomization</i> ) .....	44

14.	Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR das Casearinas pelas equações <b>(a)</b> $y=ax+b$ , $y=kx$ , $R_0^2$ e <b>(b)</b> $y=a'x+b'$ , $y=k'x$ , $R_0'^2$ .....	46
15.	Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM das Casearinas pelas equações <b>(a)</b> $y=ax+b$ , $y=kx$ , $R_0^2$ e <b>(b)</b> $y=a'x+b'$ , $y=k'x$ , $R_0'^2$ .....	46
16.	<b>(a)</b> $Q^2_{LOO-CV}$ em função da variação do parâmetro $\epsilon$ mantendo constantes C e $\gamma$ ; <b>(b)</b> $Q^2_{LOO-CV}$ em função da variação do parâmetro $\gamma$ mantendo constantes C e $\epsilon$ . Para o menor RMSE seleciona-se o valor $Q^2_{LOO-CV}$ correspondente a 0,969 .....	51
17.	Diagrama de Williams – Domínio de aplicabilidade das Casearinas após análise conformacional. <i>Hat</i> crítico $3h = 0,923$ .....	53
18.	Valores $pIC_{50}$ observados e preditos pelos modelos de QSAR-MLR e QSAR- SVM contra células tumorais V79 após emprego de análise conformacional para alguns compostos .....	54
19.	Gráfico dos valores $R^2$ e $Q^2$ em função das randomizações da bioatividade experimental ( <i>Y-randomization</i> ) do modelo de QSAR-MLR construído com a seleção de confôrmeros .....	55
20.	Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR das Casearinas pelas equações <b>(a)</b> $y=ax+b$ , $y=kx$ , $R_0^2$ <b>(b)</b> e $y=a'x+b'$ , $y=k'x$ , $R_0'^2$ .....	57
21.	Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM das Casearinas pelas equações <b>(a)</b> $y=ax+b$ , $y=kx$ , $R_0^2$ e <b>(b)</b> $y=a'x+b'$ , $y=k'x$ , $R_0'^2$ .....	57
22.	Correlação entre os modelos de QSAR-MLR e QSAR-SVM: <b>(a)</b> antes da eliminação dos compostos do item 10, 20, 33 ,53 ,56, 261, 71, 106, 109 da Tabela 11 e <b>(b)</b> após eliminação dos 9 compostos citados.....	60
23.	Correlação entre os modelos de QSAR-MLR e QSAR-SVM da Casearina G substituída (Fig. 9). Gráfico construído com os valores numéricos da Tabela 8 (v. em Anexos).....	61

- 24.** Efeito substituinte na posição R<sup>4</sup> (Carbono C6) nas Casearvestrinas A, B e C e equivalência de predição entre QSAR-MLR (Y<sub>MLR</sub>) e QSAR-SVM (Y<sub>SVM</sub>). **(a)** Os melhores resultados, considerando a proximidade de valores preditos pIC<sub>50</sub> entre métodos, são obtidos por todos os substituintes da Casearvestrina A **(b)**, seguidos pelos da Casearvestrina B e por fim **(c)**, os substituintes OAc, H, e a própria Casearvestrina C geram os piores resultados na equivalência de predição .. ..... **64**
- 25.** Efeito substituinte nas posições R<sup>4</sup> (carbono C6) e R<sup>5</sup> (Carbono C7) na Casearborina A e equivalência de predição entre QSAR-MLR (Y<sub>MLR</sub>) e QSAR-SVM (Y<sub>SVM</sub>). **(a)** Os melhores resultados, considerando a proximidade de valores preditos pIC<sub>50</sub> entre métodos, são obtidos apenas pelos substituintes OAc e OMe, **(b)**, excluídos os demais e a própria Casearborina A. Verifica-se que há boa equivalência de predição para Cl, = O, OAc, Me, excluindo OMe, F e a própria Casearborina A. **(c)** Os substituintes da Caseamembrina A não mostram equivalência de resultados apenas para Cl e a própria Caseamembrina A ..... **65**
- 26.** Dibenzoilidrazina – Estrutura geral ..... **69**
- 27.** Diagrama de Williams – Domínio de aplicabilidade das DBHs contra *Chilo suppressalis*. Os compostos D20 e D43 do conjunto de treinamento e D7 e D10 do conjunto de teste, estão classificados fora da classe química dos demais compostos,denotado pela linha vertical 3h = 0,545 ..... **73**
- 28.** Valores de letalidade pLD<sub>50</sub> observados e preditos de DBHs contra *Chilo suppressalis*. Em vermelho, estruturas otimizadas pelo método AM1 antes da análise conformacional. Em azul estão os compostos de teste (n = 8) na condição AM1 e campo de força MMFF94 ou Dreiding modificado após análise conformacional..... **74**
- 29.** Valores R<sup>2</sup> e Q<sup>2</sup> obtidos em função da randomização dos valores pLD<sub>50</sub> experimentais ..... **75**
- 30.** Representação gráfica dos parâmetros de validação do QSAR Eq.(5.2-4) das DBHs pelas equações **(a)**  $y=ax+b$ ,  $y=kx$  e **(b)**  $y=a'x+b'$  e  $y=k'x$  ..... **76**
- 31.** Representação gráfica dos resultados de predição do modelo de QSAR para as 28 dibenzoilidrazinas ..... **78**

Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzolidrazinas  
LISTA DE FIGURAS

xxx

- 
32. Representação gráfica e comparativa dos resultados de predição do modelo de QSAR para as 28 dibenzolidrazinas em relação às conformações com AM1 (quadrados vermelhos) e após seleção de confôrmeros (círculos pretos e triângulos) ..... 79
33. Representação da variação de valores  $pLD_{50}$  em função do descritor H8u para o composto D28. Valor estimado de H8u (= 0,319) atribuído à conformação bioativa ( $pLD_{50-exp} = 5,32$ )..... 80
34. Estrutura geral das N,N'-dibenzoil-N-t-butilidrazinas ..... 83
35. Representação gráfica dos parâmetros STERIMOL de Verloop, L e B<sub>1</sub> a B<sub>5</sub>. Representam a projeção de um substituinte **(a)** ao longo do eixo x=L; **(b)** perpendicular ao eixo de substituição, eixo x (perpendicular ao plano do papel). Desenho baseado na pg. 48 da ref. [3],..... 87
36. Resultados de predição interna dos modelos de QSAR-MLR e QSAR-SVM e correlação dos modelos: R = 0,999; F=19704; p < 0,0001..... 88
37. **(a)**  $Q^2_{LOO-CV}$  em função da variação do parâmetro  $\epsilon$  mantendo constantes C e  $\gamma$ ; **(b)**  $Q^2_{LOO-CV}$  em função da variação do parâmetro  $\gamma$  mantendo constantes C e  $\epsilon$ . Para o menor RMSE seleciona-se o valor  $Q^2_{LOO-CV}$  correspondente a 0,987..... 89
38. Representação gráfica da influência dos descritores  $V^{orto}$  e Mor23p na avaliação da bioatividade. Região A: maior bioatividade; região B: menor bioatividade ..... 91
39. Variação do descritor Mor23p para alguns compostos em função dos confôrmeros correspondentes. Para  $pLD_{50}$  crescente, Mor23p decresce. Os pontos nas retas são os valores experimentais de  $pLD_{50}$ ..... 92
40. Representação gráfica da influência dos descritores  $\pi^{meta}$  e Mor23p na avaliação da bioatividade. .... 94
41. Representação gráfica da influência dos descritores  $\Delta B_5$  e Mor23p na avaliação da bioatividade..... 96
42. Diagrama de Williams – Domínio de aplicabilidade das dibenzolidrazinas contra *Leptinotarsa decemlineata*. O composto D23 comporta-se como *outlier* por apresentar o maior resíduo entre os compostos ..... 97

Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
LISTA DE FIGURAS

xxxix

43.	Valores $R^2$ e $Q^2$ obtidos em função da randomização dos valores $pLD_{50}$ experimentais para (a) QSAR-MLR com $n = 30$ e (b) QSAR-MLR com $n = 24$ . .....	99
44.	Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR Eq. (5.3-2) das DBHs pelas equações (a) $y = ax+b$ , $y = kx$ (b) $y = a'x+b'$ e $y = k'x$ .....	100
45.	Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM das DBHs pelas equações (a) $y = ax+b$ , $y = kx$ e (b) $y = a'x+b$ e $y = k'x$ .....	100
46.	Predição da atividade larvicida expressa pelos confôrmeros dos compostos (a) D29, (b) D36 e (c) D26), e a influência do descritor Mor23p na atividade larvicida crescente. Valores auto escalados. ....	103
47.	Ilustração da interação por atração de van der Waals entre um átomo de uma dibenzoilidrazina (DBH) com o átomo de oxigênio de um aminoácido representado por uma alanina (Ala). ....	111
48.	Representação da formação de dipolos transientes entre duas moléculas nos lapsos de tempo $t_1$ a $t_4$ .....	114
49.	Dibenzoilidrazina BYI06830 alojada na cavidade da proteína 1R20 e as modificações feitas no cristal (em vermelho: grupos eliminados), acrescentando os substituintes halogenados $X_n$ .....	117
50.	(a) Representação esquemática da região de interação considerando dibenzoilidrazinas halogenadas substituídas na posição <i>orto</i> (representada), <i>meta</i> e <i>para</i> . $d_{Xn-O}$ varia de 2,99 Å para F...O a 3,50 Å para I...O. O – Y pode ser uma carbonila, hidroxila ou carboxilato quando Y é um carbono; um fosfato quando Y é fósforo; ou um sulfato quando Y é enxofre <sup>119</sup> . (b) Representação da interação molecular e do posicionamento do anel benzênico em base à ilustração da ref.[116] e adaptada para dibenzoilidrazinas.....	124
51.	(a) $Q^2_{CV-LOO}$ em função da variação do parâmetro $\varepsilon$ mantendo constantes C e $\gamma$ ; (b) $Q^2_{CV-LOO}$ em função da variação do parâmetro $\gamma$ mantendo constantes C e $\varepsilon$ . Para o menor RMSE seleciona-se o valor $Q^2_{CV-LOO}$ correspondente ( $Q^2_{CV-LOO} = 0,990$ ).....	125

Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzoididrazinas  
LISTA DE FIGURAS

xxxii

52.	Comparativo da predição pLD <sub>50</sub> : geometria otimizada pelo método semi-empírico AM1 e após análise conformacional com os campos de força MMFF94 ou Dreiding modificado, assim como representação das predições pelos modelos de QSAR-MLR e QSAR-SVM. Correlação entre modelos: R = 0,999; F = 25923; s = 0,032; p < 0,0001.....	126
53.	Influência dos descritores RDF095p, $\alpha_{yy}$ e $\mu_x$ na atividade larvicida dos compostos <i>orto</i> substituídos. O composto D1 (H) não tem substituintes e é utilizado como referência .....	127
54.	Dibenzoididrazinas <i>meta</i> substituídas e influência dos parâmetros $\alpha_{yy}$ , $\mu_x$ e RDF na atividade larvicida. O composto D1 (H) não tem substituintes e é utilizado como referência. ....	131
55.	Dibenzoididrazinas <i>para</i> substituídas e influência dos parâmetros $\alpha_{yy}$ , $\mu_x$ e RDF na atividade larvicida. O composto D1 (H) não tem substituintes e é utilizado como referência .....	132
56.	Diagrama de Williams – Domínio de aplicabilidade das dibenzoididrazinas (n = 37) contra <i>Spodoptera exigua</i> . Composto candidato a <i>outlier</i> D1. Valor crítico: $3^*h = 0,324$ , valor $3^*h$ (D1) = 0,337 (Vide texto).....	133
57.	Valores R <sup>2</sup> e Q <sup>2</sup> obtidos em função da randomização dos valores pLD <sub>50</sub> experimentais para (a) QSAR-MLR com n = 37 (b) QSAR-MLR com n = 30.....	134
58.	Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR Eq. (5.4-25) das DBHs pelas equações (a) $y = ax+b$ , $y = kx$ , (b) e $y = a'x+b'$ e $y = k'x$ .....	135
59.	Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM das DBHs pelas equações (a) $y = ax+b$ , $y = kx$ , (b) $y = a'x + b'$ e $y = k'x$ .....	135
60.	Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR Eq. (5.4-25) das DBHs pelas equações (a) $y = ax+b$ , $y = kx$ e (b) $y = a'x+b'$ e $y = k'x$ aplicado a conjunto n = 5. ....	136
61.	Proteína 1R20, (a) cristalizada em conjunto com o ligante identificado como BYI06830; (b) ligante BYI06830, dibenzoididrazina, situado na cavidade da proteína conforme Billas <i>et al.</i> <sup>96</sup> e aqui reproduzido por este estudo <sup>110</sup> .....	140

---

62.	Cavidade com indicação de três regiões favoráveis à proposta de novos compostos substituídos. Os aminoácidos identificados e os respectivos átomos de interação foram originalmente identificados no trabalho de <i>Billas et al.</i> <sup>96</sup> . Regiões indicadas como 1, 2 e 3 vide texto.....	141
63.	Ilustração da sobreposição, de estruturas, obtidas por <i>docking</i> entre ligante BYI06830 e sua cópia, o ligante originalmente alojado na cavidade indicada pelo número 1. As estruturas em magenta correspondem aos aminoácidos que formam a cavidade receptora .....	142
64.	Dibenzolidrazina D4 e resultado do <i>docking</i> na cavidade da proteína 1R20, e o átomo de Br na posição <i>orto</i> e sua distância de interação de 3,03 Å em relação ao nitrogênio do aminoácido Asn504. ....	143
65.	(a) Alojamento da dibenzolidrazina BYI06830 na cavidade do domínio ligante SeEcR-LBD obtido após a sobreposição <sup>117</sup> do complexo ligante-proteína 1R20 com a proteína obtida por homologia a partir dos aminoácidos de <i>Spodoptera exigua</i> .....	148
	(b) Sobreposição obtida por <i>docking</i> da cópia do composto BYI06830 (coloração por átomos) com o composto BYI06830 (cor magenta, indicado pelo número 1). Obteve-se um valor da distância de sobreposição RMSD = 0,512 Å.....	149
66.	Complexo SeEcR-LBD de <i>Spodoptera exigua</i> com o ligante BYI06830 alojado na cavidade. Linhas pontilhadas verdes indicam as interações com os aminoácidos Asn 504, Thr 343 e Tyr 408. ....	150
67.	Segmentos alinhados do domínio de ligação (LBD) dos aminoácidos formadores das cavidades em SeEcR e HVEcR selecionados pela distância máxima de 3Å de seus átomos, aos átomos do composto D6 através do programa BioMedCache <sup>110</sup> .....	151
68.	Cavidade formada na proteína SeEcR na região do ligante (dibenzolidrazina D6 usada de exemplo). As regiões I e II oferecem oportunidades para desenvolvimento de novos compostos.....	158
69.	Dibenzolidrazina substituída conforme estudo de <i>Smaghe et al.</i> <sup>130</sup> .....	159
70.	(a) Ligante PS4 construído <i>in silico</i> . Estrutura de <i>docking</i> ( em cor cinza) sobreposta à sua similar, mostrando bom preenchimento da cavidade relativa à região II da Figura 68. (b) As setas indicam possíveis interações dos substituintes halogênicos do composto Yn = 4-(Ph-2Cl) no anel B (Figura 69) com os aminoácidos: Cl...O = C-, 3,89Å (Asn 504); Cl...S, 3,24Å (Met 381).....	160

71. Resultado comparativo de valores preditos conforme metodologia de otimização de geometria pelo método AM1 e MMFF94 ou Dreiding modificado. O método da ref. [26] não está especificado na literatura..... 169
72. Ilustração da influência de cada um dos parâmetros na descrição da bioafinidade de ligação das dibenzoilidrazinas com as células Sf-9 ..... 170
73. Comportamento linear de valores de Mor25v dos confômeros do composto não substituído D1, do composto D13 de maior valor  $pIC_{50}$  experimental, e D18 composto de menor valor  $pIC_{50}$  experimental. Pontos vermelhos correspondem aos valores experimentais de  $pIC_{50}$  associado ao valor Mor25v..... 171
74. Correlação do descritor Mor25v com a bioafinidade, com os compostos separados em três grupos, verificando-se correlação linear. Os compostos CN e NO<sub>2</sub> podem eventualmente formar um grupo à parte. Os valores  $pIC_{50}$  são os experimentais..... 172
75. Correlação do descritor LogP com a bioafinidade experimental, com os compostos separados em três conjuntos. C<sub>6</sub>H<sub>5</sub> representa um conjunto de um único elemento..... 173
76. Correlação da constante eletrônica  $\sigma$  de Hammett com a bioafinidade experimental, com os compostos separados em três conjuntos, dois com elementos correlacionados, e outro sem correlação dos elementos (em vermelho). ..... 174
77. Gráfico dos valores preditos pelos modelos de QSAR MLR e SVM comparativamente aos valores experimentais..... 176
78. Diagrama de Williams – Domínio de aplicabilidade das dibenzoilidrazinas (n = 18) contra *S. frugiperda*. O composto D18 comporta-se como *outlier*..... 177
79. Valores R<sup>2</sup> e Q<sup>2</sup> obtidos em função da randomização dos valores  $pLD_{50}$  experimentais para (a) modelos de QSAR-MLR e QSAR-SVM com n = 18 e (b) modelos de QSAR- MLR e QSAR-SVM com n = 13..... 178
80. Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR Eq. (5.5-17) das dibenzoilidrazinas pelas equações (a)  $y = ax+b$ ,  $y = kx$  e (b)  $y = a'x+b'$  e  $y = k'x$ ..... 179

Estudos de QSAR-2D aplicados a  
Diterpenóides Clerodanos e Dibenzolidrazinas  
LISTA DE FIGURAS

---

xxxv

- 81.** Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM (n=13) das dibenzolidrazinas pelas equações **(a)**  $y = ax+b$ ,  $y = kx$  e **(b)**  $y = a'x+b'$  e  $y = k'x$ ..... **181**
- 82.** Predição da bioafinidade  $pIC_{50}$  de compostos propostos (n = 30) e compostos em que os modelos de QSAR-MLR e QSAR-SVM produzem valores próximos (critério de diferença máxima  $Y_{MLR}-Y_{SVM} = \pm 0,10$ )..... **183**
- 83.** Bioafinidades comparativas das dibenzolidrazinas série D (substituídas no anel B) e série T\_D (inclusão de Ph-Y substituído, anel C). Notar que D18 está contido na série T\_D definido pela baixa bioafinidade..... **184**
- 84.** Eletrofilicidade ( $\omega$ ) comparativa das dibenzolidrazinas da série D e série T\_D..... **185**
- 85.** Comparativo de distância e posicionamento do composto D18 de pior bioafinidade ( $d_{Ph} = 3,90 \text{ \AA}$ ), substituinte “rígido”, com o composto D13 com a maior bioafinidade ( $d_{n-C4H9} = 4,70 \text{ \AA}$ ), substituinte flexível. .... **190**
- 86.** *Y-randomization* dos valores  $pIC_{50}$  em relação à variável  $\omega_{gk}^-$  ..... **194**
- 87.** Volume de van der Waals e o potencial eletrostático de superfície dos compostos D1 e D18. Distâncias em  $\text{\AA}$  para informação dimensional..... **196**



## 1.Introdução

A fase da “descoberta” de novas drogas leva em média cinco anos<sup>1</sup>, até chegar aos ensaios pré-clínicos. O desenvolvimento de novos fármacos<sup>1</sup> e de pesticidas agrícolas<sup>2</sup> leva em média, dez anos, e os investimentos são consideravelmente altos. O emprego de métodos de química teórica e modelagem molecular em geral são utilizados na fase inicial das pesquisas, na busca de novos compostos onde QSAR-2D (*Quantitative Structure-Activity Relationships* – em duas dimensões) é uma das metodologias empregadas.

A tese desenvolvida em forma de estudos de caso traz a proposta de construir modelos de QSAR-2D empregando análise conformacional para alguns dos compostos que são selecionados por um critério de resíduo (detalhes em Material e Métodos, seção 4). Pelo critério, são selecionados os confôrmeros que melhor representam o valor da bioatividade. O estudo demonstra que o procedimento aumenta a acurácia da predição interna, auxilia na avaliação de *outliers*, e aumenta a robustez do modelos de QSAR. Através de modelos de QSAR validados, sugerem-se novos compostos propostos *in silico*, e compostos de produtos naturais. QSAR-2D é o enfoque central deste estudo, mas, associado a ele, decorrem estudos de similaridade molecular e *docking*.

Este estudo contempla duas classes de substâncias. Uma natural, com diterpenóides clerodanos extraídos da planta *Casearia sylvestris* Swartz Flacourtiaceae/Salicaceae à qual se atribuem propriedades antitumorais. E outras sintetizadas, conhecidas como N,N'-dibenzoil-N-t-butilidrazinas (ou diacilidrazinas), empregadas em estudos como inseticidas agrícolas, e conhecidas por serem ecologicamente apropriadas, não causando danos à humanos ou insetos benignos, assim como à microorganismos aquáticos.

Partindo-se de dados de literatura, e empregando metodologias da química computacional, este estudo faz diferentes contribuições de natureza teórico-computacional, (no melhor de nosso conhecimento de resultados não contidos na literatura), e que estão relacionadas a seguir: **(a)** Obtenção de modelos de QSAR

por duas metodologias independentes, empregando análise conformacional, onde cada confômero é escolhido de forma a reproduzir o mais próximo possível o valor da conformação bioativa. Os modelos de QSAR são validados e utilizados como modelos preditivos, respectivamente para diterpenóides e dibenzolidrazinas. **(b)** Testam-se métodos de similaridade molecular para verificar o quanto auxiliam na identificação de confômeros na sua suposta conformação bioativa. Os confômeros são utilizados em três estudos de caso de dibenzolidrazinas para avaliar a qualidade de predição da bioatividade com pseudo *blind sets*. **(c)** Partindo de dados cristalográficos de uma estrutura ligante-receptor (resolução 3Å) relativa ao inseto *Heliothis virescens* (Lepidoptera), e resolvida por espectroscopia de raios-X, empregou-se *docking* para obter o possível posicionamento de dibenzolidrazinas substituídas na proteína, e identificar através das distâncias interatômicas a interação dos átomos halogênios das dibenzolidrazinas com os átomos dos aminoácidos integrantes da cavidade receptora. A literatura cita apenas os pontos de ancoragem do ligante, não fazendo nenhuma alusão às demais prováveis interações. **(d)** Utilizando o complexo cristalografado relativo ao inseto *Heliothis virescens* (Lepidoptera), obteve-se a proteína homóloga da lagarta *Spodoptera exigua* (Lepidoptera) e através de *docking*, fêz-se análise similar ao item (c). **(e)** Empregando métodos de avaliação da qualidade de *docking* independentes e adicionais ao RMSD (erro médio quadrático – *Root Mean Square Deviation*) que se utilizam de uma pontuação (*scoring*), extraíram-se informações complementares sobre a qualidade da interação ligante-proteína. **(f)** No estudo de QSAR da lagarta *Spodoptera frugiperda*, através de simulações da química computacional, construiu-se uma hipótese plausível sobre o motivo da baixa bioatividade de dibenzolidrazinas *para*-substituídas, tendo por substituinte um fenil (Ph) e um anel benzênico *para*-substituído (Ph-X). Este aspecto foi deixado por esclarecer na literatura. **(g)** Partindo de estruturas químicas de diterpenóides da literatura de produtos naturais propor novos compostos diterpenóides clerodanos, naturais e modificados *in silico* com bioatividade predita pelos modelos de QSAR. **(h)** Propor dibenzolidrazinas

com novos substituintes com atividade larvicida predita pelos modelos de QSAR. Um caso específico (DBHs contra *S. exigua*), é assistido por estudo de *docking* dos compostos propostos. (i) Decorrente dos estudos de *docking* com dibenzolidrazinas contra *Spodoptera exigua* feitos na cavidade da proteína obtida por homologia, são dadas algumas perspectivas de novos estudos teóricos e complementares aos já realizados.

QSAR-2D, metodologia desenvolvida por Hansch-Fujita<sup>3</sup>, já fez várias contribuições efetivas no desenvolvimento de fármacos, herbicidas, fungicidas e inseticidas, todos ainda presentemente comercializados.<sup>4</sup> Destacam-se: Flobufen (antiinflamatório), Lomerizina (antiemético, antiglaucoma), Norfloxacin (bactericida, foi substituída e é precursora da Ciprofloxacina<sup>5</sup>), Metamitron (herbicida contra o escaravelho da beterraba), Bromobutideo (herbicida para agricultura pantanosa), Bifentrin (inseticida foliar), Metconazol (fungicida, similar ao congêneres Fluconazol).

Na Europa, para novos produtos químicos em geral e em particular de pesticidas, existe a preocupação de se gerarem modelos de QSAR validados com qualidades preditivas que possam ser incorporados às normas europeias de avaliação de agressão ao meio ambiente.<sup>6</sup>

Novos fármacos são também desenvolvidos a partir de produtos naturais, como o Taxol A, diterpeno extraído da *Taxus brevifolia* NUTT, comercialmente semi-sintetizado, um agente anti-mitótico, usado contra tumor de ovário, pulmão e mamas.<sup>5</sup> Na verdade, a síntese total de substâncias naturais candidatas à fármacos, exemplificados pela ref.[7], é a única forma de preservação da espécie animal ou vegetal em função das grandes quantidades do princípio ativo necessárias para a produção em escala industrial.

Produtos naturais produzidos pelo próprio inseto, são também utilizados no controle de pragas como por exemplo, feromônios<sup>8</sup> da própria *Spodoptera frugiperda* (Lepidoptera) aqui estudada, conhecida como lagarta-do-cartucho, causadora de grandes prejuízos na agricultura do milho.<sup>8</sup>



## 2. OBJETIVOS

### Estudos de QSAR

Os estudos estão associados às classes das substâncias enumeradas a seguir com os seguintes objetivos:

**1. Diterpenóides clerodanos:** empregando análise conformacional, desenvolver estudos de QSAR-2D com maior acurácia na predição interna, validado para predição externa, aplicando-o à diterpenóides clerodanos extraídos da literatura na predição de atividade antitumoral em células pulmonares de hamster chinês (células V79). Estender a proposta de novos compostos para diterpenóides desenvolvidos *in silico*.

**2. Dibenzoilidrazinas:** empregando análise conformacional, desenvolver estudos de QSAR-2D com maior acurácia na predição interna, validado para predição externa, aplicando-o na proposição de novos compostos inseticidas desenvolvidos *in silico*.

### Estudo de *docking*

Através de estudos de *docking*, contribuir com informações de possíveis interações de dibenzoilidrazinas halogenadas com os aminoácidos da cavidade da proteína, respectivamente, de *Heliothis virescens* e *Spodoptera exigua*.



### 3. REVISÃO BIBLIOGRÁFICA

---

## 3. REVISÃO BIBLIOGRÁFICA

### 3.1 – Diterpenóides e QSAR

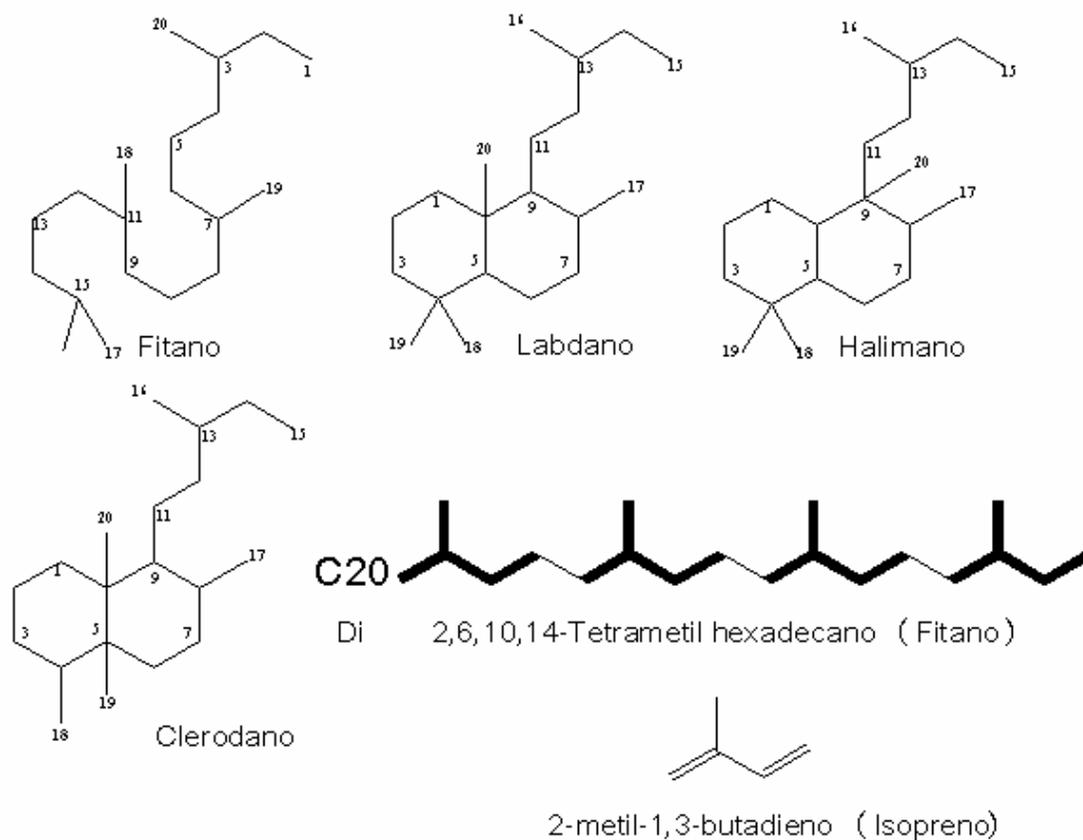
A pesquisa de compostos anticancerígenos originários de plantas remonta à década de 1950, com a descoberta de mais de 90 alcalóides indólicos monoterpénóides na planta semi-arbustiva *Catharanthus roseus* (L.) G. Don (Apocynaceae). Os mais importantes são: o alcalóide indólico catarantina e a unidade deidroindólica vindolina. A maior característica destas drogas é cessar a divisão celular na metáfase. A vincristina e a vinblastina são os representantes terapêuticos mais conhecidos.<sup>9a</sup>

Estudos sistemáticos resultaram no isolamento de muitos produtos naturais que exibem atividade antitumoral, e alguns foram considerados em estudos clínicos, entre eles, alguns terpenóides.<sup>10</sup>

Os terpenóides possuem uma estrutura comum, e são representados pela união de duas, três, quatro, seis, de oito ou de mais unidades de isopreno (2-metil-1,3-butadieno, C<sub>5</sub>). Os diterpenóides, um dos alvos deste estudo, compreendem um grupo de substâncias não voláteis C<sub>20</sub>, formadas por quatro unidades isoprênicas e têm como precursor representativo de diterpenos farmacêuticos o difosfato de *trans*-geranilgeranil<sup>11</sup>, também encontrado na literatura como geranilgeranil pirofosfato.<sup>9b</sup> Conhecem-se aproximadamente 2700 diterpenóides<sup>12</sup> entre acíclicos e cíclicos de origem natural. No presente estudo é de especial interesse os biciclofitanos ilustrados na Figura 1.

### 3. REVISÃO BIBLIOGRÁFICA

---

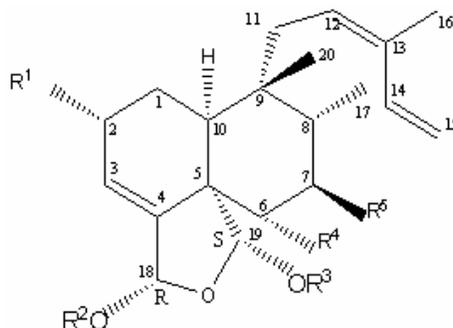


**Figura 1.** Biclofitanos constituídos por quatro unidades isoprênicas e a numeração empregada.<sup>11,12</sup>

O presente estudo emprega diterpenóides do tipo *cis*-clorodanos com anel diacetálico (Figura 2), denominados casearinas.

### 3. REVISÃO BIBLIOGRÁFICA

---



**Figura 2.** Diterpenóides clerodanos – Estrutura geral das casearinas.

As casearinas foram obtidas do extrato etanólico de folhas da *Casearia sylvestris* Swartz família botânica Flacourtiaceae/Salicaceae. Esta família tem aproximadamente 1300 espécies tropicais e subtropicais subordinadas a 86 gêneros. Segundo Barroso<sup>13</sup>, o Brasil possui 19 gêneros e 92 espécies.

A *C. sylvestris*, sinonímia popular Guaçatonga entre muitos outros, é encontrada em quase toda a América do Sul e em especial na região sul do Brasil. Uma síntese sobre esta espécie e suas características farmacognósticas e farmacológicas pode ser encontrada na ref. [14]. Considerando que diterpenóides são reconhecidamente antitumorais, o arsenal botânico brasileiro do gênero *Casearia* – citando apenas um dos gêneros entre muitos outros ricos em diterpenóides – pode conter um número expressivo destas substâncias e contribuir com novos estudos de natureza experimental e teórica.

Além da atividade citotóxica inibindo a atividade tumoral, outro aspecto está relacionado à característica antialimentar causada por diterpenóides à insetos, exemplificado pelo estudo com corimbotinas contidas nas folhas de *Casearia corymbosa* Flacourtiaceae.<sup>15</sup>

Uma fonte bibliográfica importante, são os artigos de revisão<sup>16</sup> sobre diterpenóides que abrangem o período 1983 a 2007. Os artigos reúnem algumas centenas de estruturas destes compostos, incluindo diterpenóides clerodanos,

### 3. REVISÃO BIBLIOGRÁFICA

---

labdanos, kauranos, abietanos e outros, como referência informativa para a realização de estudos complementares, experimentais ou teóricos.

Muitos produtos naturais foram alvos apenas na identificação de novas substâncias e de estudos físico-químicos a elas atrelados, e não se tem conhecimento do seu emprego em bioensaios para avaliar seu potencial de ação biológica. Nestes artigos<sup>16</sup> o autor teve o cuidado de selecionar, da literatura, apenas compostos quimicamente certificados através de espectroscopia de raios-X e ressonância magnética nuclear.

Nada foi encontrado na literatura sobre hipóteses de mecanismos de ação biológica e/ou físico-químicos contemplando a ação do ligante no seu receptor em bioensaios de diterpenóides clerodanos com células tumorais de hamster chinês V79.

Não foram igualmente encontrados artigos com modelos de QSAR, baseados em bioensaios com terpenóides em células V79. O emprego de um reduzido número de compostos nos bioensaios é uma das causas. Quer por extração quer através de síntese, é sempre difícil obter-se substâncias puras de produtos naturais. No processo extrativo muitas vezes as quantidades obtidas são apenas suficientes para elucidação da estrutura química. Por outro lado, em um artigo de revisão especificamente voltado a diterpenos<sup>17</sup>, discorre-se sobre a dificuldade de síntese desta classe de substância.

As dificuldades citadas corroboram o conceito de que um modelo de QSAR validado com qualidades preditivas pode ser de grande utilidade na seleção de prioridades entre os muitos compostos, quando aplicado às substâncias congêneres, otimizando recursos de síntese e de certificação físico-química.

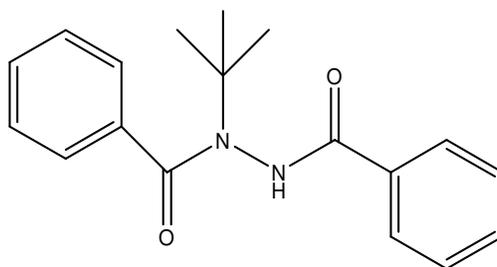
A primeira contribuição de modelos de QSAR com diterpenóides clerodanos está sendo feita com o presente estudo, publicado<sup>18</sup> e integrado a este trabalho como Apêndice 1.

### 3. REVISÃO BIBLIOGRÁFICA

---

#### 3.2 – Dibenzoilidrazinas e QSAR

A primeira bisacilidrazina que mostrou atividade como agonista ecdisteróide (*ecdysteroid*) foi descoberta casualmente pelos cientistas da empresa Haas & Rohm em 1983<sup>19</sup>. Wing *et al.*<sup>20</sup> estudaram um composto análogo ligeiramente mais potente, denominado RH-5849 e ilustrado na Figura 3, com a denominação N,N'-dibenzoil-N-t-butilidrazina (ou 1,2-dibenzoil-1-*tert*-butilidrazina) ensaiado com larvas da lagarta-do-tabaco *Manduca sexta*, Lepidoptera.

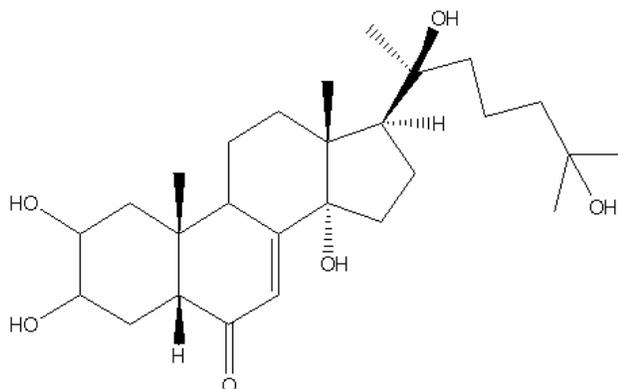


**Figura 3.** Estrutura química da dibenzoilidrazina RH-5849

As dibenzoilidrazinas, substâncias sintetizadas, são agonistas não-esteroidais do hormônio de crescimento conhecido por 20-hidroxiecdisona (20E) (Figura 4), também chamado de hormônio da muda ou ecdisona, que regula o crescimento das larvas e a metamorfose dos insetos. As dibenzoilidrazinas mimetizam o hormônio 20E porém, causando anomalias no desenvolvimento, mais especificamente mudas prematuras, provocando a letalidade.

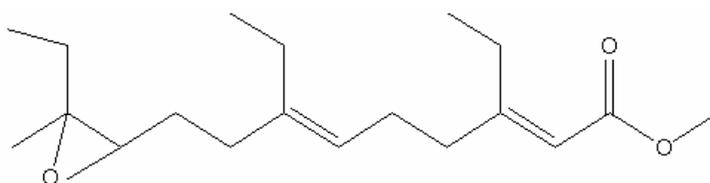
### 3. REVISÃO BIBLIOGRÁFICA

---



**Figura 4.** Hormônio de crescimento *20-hidroxiecdisona (20E)*  
Estrutura química (Nr. CAS 3604-87-3)

Os hormônios juvenís (*Juvenile hormones - JHs*), são sesquiterpenos igualmente ligados à regulação do desenvolvimento das larvas. Este hormônio assegura que cada muda chegue ao próximo estágio de modo imaturo. JHs atuam também de maneira tal que o inseto possa manter sua forma atual além de promover o desenvolvimento de ovos. A Figura 5 exemplifica um tipo de JH encontrado em Lepidoptera.<sup>21</sup>



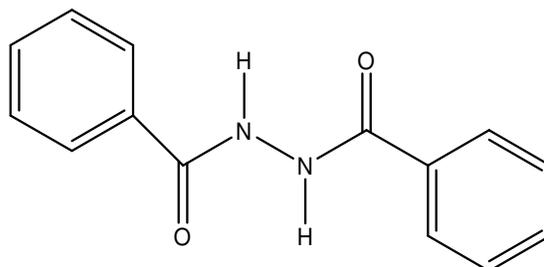
**Figura.5.** Methyl(2E,6E)-10R,11S-(oxiranyl)-3,7-diethyl-11-methyl-2,6tridecadienoate, nominado também como J01. Exemplo de hormônio juvenil de Lepidoptera.

Chan *et al.*<sup>22</sup> sintetizaram a dibenzolidrazina RH 5849 e o composto análogo ilustrado pela estrutura da Figura 6, testando-os em larvas de *Choristoneura fumiferana (Clemens)* Lepidoptera. Segundo os autores, comparativamente aos

### 3. REVISÃO BIBLIOGRÁFICA

---

valores dos resultados larvicidas alcançados com o composto RH 5849, o composto da Figura 6 não apresentou atividade larvicida significativa.



**Figura 6.** Dibenzoilidrazina ensaiada contra *Choristoneura fumiferana*

Se compararmos as estruturas das figuras 3 e 6 com a 4, constatamos que, não há nenhuma similaridade, assim como as respostas larvicidas dos compostos 3 e 6 são marcadamente distintas. Chan *et al.*<sup>22</sup> empregaram espectroscopia de raios-X para estudar os cristais por eles obtidos da RH 5849 e da estrutura da figura 6, mostrando que existem similaridades e pequenas diferenças, respectivamente em alguns comprimentos e ângulos de ligação. A hipótese levantada pelos autores é de que o grupo *tert*-butil presente apenas em RH 5849, poderá ter sido o responsável pela marcante diferença nas atividades larvicidas.

A literatura reúne mais de duas décadas de estudos de natureza teórica e experimental das N,N'-dibenzoil-N-t-butilidrazinas. Na pesquisa do modo de ação destas substâncias encontram-se agonistas da ecdisona e um número expressivo de contribuições continuam sendo feitas por um grupo de estudos no Japão.

Os principais trabalhos feitos entre 1990 e 2009 estão reunidos no Apêndice 2, proporcionando uma visão global. Nesta síntese é possível visualizar a maioria dos estudos no citado período, de natureza experimental (bioensaios) e teóricos com QSAR 2D, QSAR 3D, e *docking* envolvendo dibenzoilidrazinas.

Um estudo de revisão elaborado por Fujita e Nakagawa<sup>23</sup> reúne alguns aspectos importantes de estudos SAR (*Structure-Activity Relationships*) e de QSAR (*Quantitative Structure-Activity Relationships*) com dibenzoilidrazinas contra *Chilo*

### 3. REVISÃO BIBLIOGRÁFICA

---

*suppressalis* Lepidoptera popularmente chamada de broca-do-colmo, conhecida praga do arroz.

As dibenzolidrazinas são inseticidas seguros para os chamados insetos benéficos e têm características ecotoxicológicas positivas: não contaminam a água, o solo, e são fracamente tóxicas para mamíferos. Com base a estes aspectos, e considerando a sua importância, foram selecionados da literatura quatro trabalhos para um estudo complementar de QSAR<sup>24,25,26,27</sup>, porém, dentro do escopo deste estudo. A seleção foi feita em função de produtos agrícolas com importância econômica para o país. Detalhes e citações específicas de tópicos da literatura correlacionados com este estudo serão dadas nas seções 5.2 até 5.5.

O levantamento bibliográfico do emprego de dibenzolidrazinas no Brasil resultou em algumas referências de natureza experimental agrícola, assim como teórica.<sup>28</sup>

## 4. MATERIAL E MÉTODOS – Tópicos Gerais

### 4.1 – Fundamentos Teóricos – QSAR e Regressão Multivariada

QSAR (*Quantitative Structure-Activity Relationships*) descreve a relação entre a atividade biológica e a estrutura química através de modelos matemáticos. A história do desenvolvimento e aplicações do QSAR clássico ou QSAR 2D, também conhecido como método ou análise de Hansch-Fujita, pode ser encontrada nas ref. [3,29]. A análise de regressão linear como método estatístico é um dos mais empregados, correlacionando os dados biológicos com os parâmetros físico-químicos e estruturais.

Duas metodologias de regressão foram empregadas neste estudo para obter os modelos de QSAR. O método MLR (*Multiple Linear Regression* – Regressão Linear Múltipla) foi usado em todo o estudo envolvendo todas as classes de substâncias, diterpenóides e dibenzoilidrazinas.

No caso dos diterpenóides e de três estudos relativos às dibenzoilidrazinas foi também empregado o método SVM (*Support Vector Machine* – Máquina de Vetores de Suporte) com o propósito de utilizar uma metodologia independente para se chegar à resultados que ficassem próximos aos obtidos por MLR. Esta escolha foi motivada por dois aspectos. O primeiro está relacionado à literatura que tem produzido artigos que indicam que em alguns estudos a metodologia SVM gerou bons resultados, comparáveis ou superiores aos métodos MLR, PLS e Redes Neurais (*ANN-Artificial Neural Network*).<sup>30,31,32,33</sup> O segundo aspecto está relacionado ao conjunto de moléculas, que é pequeno. SVM tem características que tornam seu uso apropriado para este estudo: não há sobreajuste (*overfitting*) no conjunto de treinamento e conseqüentemente produz boa predição. Além disto é robusto em relação às variáveis ruidosas (*noisy variables*). Para haver uma segunda certificação dos resultados obtidos por MLR, empregaram-se os mesmos descritores na obtenção do modelo de QSAR por SVM. Desta forma, dois métodos, independentes, gerando resultados próximos, podem ser assumidos como indicadores da qualidade preditiva interna dos modelos de QSAR-MLR e QSAR-SVM.

O método de Regressão Linear Múltipla, que doravante será chamado pela sigla inglesa MLR, pode ser expresso pela equação geral:

$$Y = ax_1 + bx_2 + cx_3 + dx_4 + \dots + \text{constante} \quad (4.1-1)$$

onde  $Y$ , é a variável dependente e representa a atividade biológica. Pode ser representado por  $pIC_{50}$  (ou  $\log 1/IC_{50}$ ), onde  $IC_{50}$  é concentração molar da droga e gera 50% de regressão na proliferação de células tumorais a exemplo do estudo de QSAR com diterpenóides.

Em estudos de QSAR com dibenzoilidrazinas,  $IC_{50}$  é um indicador de bioafinidade dos compostos com as células Sf-9 das ovas de lagarta *Spodoptera exigua* correspondendo a uma concentração inibitória do desenvolvimento em 50% das células.  $Y$  é também expresso por  $pLD_{50}$  (ou  $\log 1/LD_{50}$ ), representando 50% de atividade letal das dibenzoilidrazinas nas ovas das lagartas *Chilo suppressalis*, *Spodoptera exigua* e do besouro *Leptinotarsa decemlineata*. As variáveis independentes (também chamadas de *explicativas*)  $x_1, x_2, \dots$  assumem o papel de descritores físico-químicos e estruturais de cada composto, e  $a, b, c, \dots$  são coeficientes de ajuste. Detalhes técnicos do método MLR podem ser obtidos nas ref. [34,35].

O método de Máquina de Vetores de Suporte, que passa a se chamar apenas de SVM, foi originalmente desenvolvido como um classificador binário por Vapnik.<sup>36a</sup> Posteriormente teve seu uso estendido a problemas de regressão. Pode ser considerada uma metodologia nova em aplicação a estudos de QSAR e de modo geral à tecnologia química.<sup>36b</sup> Esta metodologia modela os dados onde tanto o erro do conjunto de treinamento como a complexidade do modelo é incorporada numa função de perda que é minimizada durante o desenvolvimento do modelo. SVM tem algumas vantagens<sup>36b</sup>: pode ser usado tanto para classificação (SVC, *Support Vector Classification*) como para regressão (SVR, *Support Vector Regression*); é aplicável à dados lineares como não lineares; adequado para conjuntos com poucas amostras o que favorece seu uso em QSAR.

O estudo utiliza o método SVR (*Support Vector Regression* – Regressão de Vetores Suporte) do aplicativo MDM (Molegro Data Modeller).<sup>37</sup> Seguem-se alguns conceitos que compõem o método para ilustrar o seu mecanismo, sem aprofundamento no formalismo matemático que pode ser encontrado na ref. [38] e que corresponde ao que foi implementado no aplicativo MDM. A teoria matemática descrita a seguir foi adaptada das ref. [32,36]. O problema básico é gerar um hiperplano - relativo ao espaço (n+1) dimensional descritores-atividade - que relacione os valores dos descritores às atividades biológicas. Dito de outra forma, trata-se de encontrar um hiperplano que nas suas margens forme classes combinando descritores com a atividade. A margem de cada lado do hiperplano é definida por vetores suporte.

Assim, a atividade biológica de qualquer composto pode ser predita através de seus descritores ao marcar um ponto correspondente a este composto no hiperplano. Utilizando a mesma notação anterior,  $x_i$  para descritores e  $y_i$  para atividade biológica, pode-se escrever a função linear de correlação entre as variáveis como:

$$f(x_i) = \langle \omega_i, x_i \rangle + b \quad (4.1-2)$$

onde  $\omega_i$  é o vetor coeficiente da função linear e  $b$  é o desvio estatístico ('bias'). A minimização dos erros entre o valor observado e o valor calculado é definido por  $\xi_i$

$$|y_i - (\langle \omega_i, x_i \rangle + b)| = \xi_i \quad (4.1-3)$$

SVM se utiliza de uma restrição para resolver o problema de otimização

$$|y_i - (\omega \phi(x_i)) + b_i| = \xi_i \quad (4.1-4)$$

onde os vetores de treinamento  $x_i$  são mapeados em um espaço dimensional maior por uma função chamada de função de *kernel*  $\phi$ . SVM acha uma correlação linear entre a atividade calculada e este espaço maior  $\phi(\xi_i)$ .

O parâmetro  $C$  (chamado de 'custo') que aparece na Eq. (4.1-5) a seguir, é a penalidade introduzida pelo termo do erro que controla o peso entre os dois termos no problema de otimização SVM. Pode ocorrer, entretanto, que a função de *kernel*  $\phi$

não seja capaz de bem representar os compostos do conjunto de treinamento de forma simplificada. Para evitar que o algoritmo produza correlações complicadas do conjunto de treinamento que acaba por produzir previsões não acuradas no conjunto de teste, foi introduzida uma variável de compensação  $\varepsilon$ .

Esta variável  $\varepsilon$  está contida na chamada ‘função de perda’  $L$  introduzida por Vapnik.<sup>36a</sup>

$$L = \min_{\omega, b, \xi} \frac{\|\omega\|}{2} + C \sum_{i=1}^m \begin{cases} 0, & \text{se } \xi_i \leq \varepsilon \\ \xi_i - \varepsilon & \text{se } \xi_i > \varepsilon \end{cases} \quad (4.1-5)$$

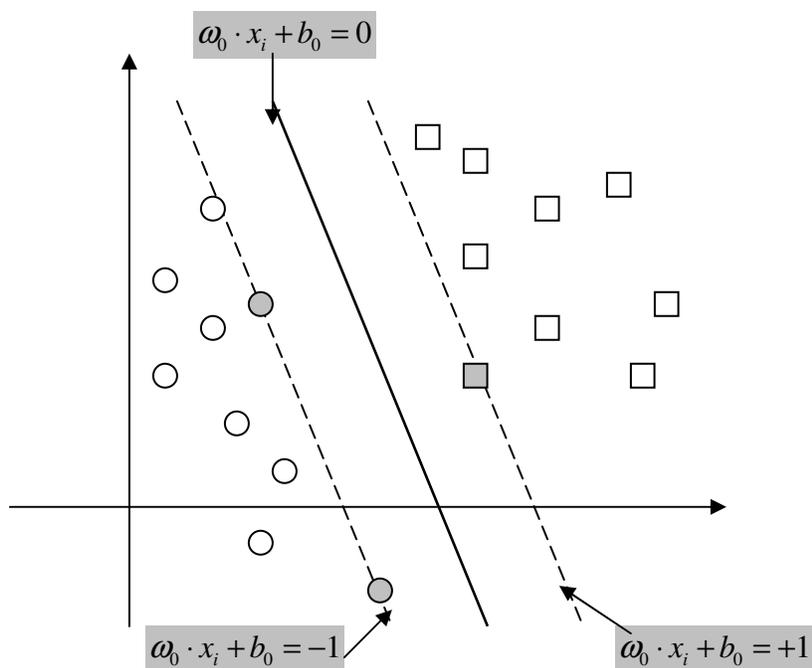
Para se obter o melhor modelo de melhor correlação entre a bioatividade e os descritores é necessário que se faça uma busca por tentativas, variando os valores dos parâmetros  $C$  e  $\varepsilon$ , e um novo parâmetro  $\gamma$  utilizado como um ajuste fino. O programa MDM tem as funções de base radial, polinomial, sigmóide e linear. A função de *kernel* utilizada neste estudo (*default* em SVM do programa MDM<sup>37</sup>) é a função de base RBF (*Radial Basis Function*) definida por

$$\phi(x_i, x) = \exp\left(-\frac{|x_i - x|^2}{2\sigma^2}\right) \quad (4.1-6)$$

onde  $x_i$  e  $x$  são dois pontos do conjunto e  $\sigma$  corresponde ao desvio padrão. A melhor correlação é avaliada utilizando o menor erro médio quadrático (RMSE) como indicador do melhor modelo. Em geral é feita uma representação gráfica de  $(Q^2_{\text{LOO}}/\text{RMSE}, \varepsilon)$  e  $(Q^2_{\text{LOO}}/\text{RMSE}, \gamma)$  para determinado valor de  $C$ , onde o índice LOO (*Leave-One-Out*) representa a validação cruzada  $Q^2_{\text{LOO-cv}}$ .

A Figura 7 ilustra a idéia de vetores suporte de um conjunto de elementos. Verifica-se que o hiperplano ótimo é decidido apenas pelos vetores suporte. As demais amostras (ou componentes), exceto os vetores suporte, não têm influência na posição deste hiperplano. Para finalizar, o método SVM tal como é aqui

empregado deve ser visto como um método complementar ao estudo MLR, sem conotação comparativa.



**Figura 7.** Hiperplano (linha central) definindo uma superfície linear separando os elementos de treinamento, e vetores suporte (linhas pontilhadas).

## 4.2 – Estruturas Químicas - Geometria – Descritores

As estruturas químicas, dos compostos utilizados, baseadas nos artigos foram importadas em 3D, via SciFinder<sup>39</sup> empregando o aplicativo DSViewerLite.<sup>40</sup> Estruturas referentes a compostos propostos por este estudo foram desenhadas em 2D e convertidas para 3D através dos aplicativos ChemDraw/Chem3D.<sup>41</sup> Em estudo específico de *docking*, os compostos foram construídos a partir de uma estrutura cristalografada.

Todas as estruturas químicas deste estudo tiveram a sua geometria pré-otimizada com o campo de força da mecânica molecular MM2<sup>42</sup>, seguindo-se a otimização com o método semi-empírico AM1 (Austin Model1)<sup>43</sup> empregando os pacotes computacionais Gaussian G03W<sup>44</sup> e ORCA.<sup>45</sup>

Estudos específicos com dibenzoilidrazinas utilizadas no estudo de bioafinidade com células Sf-9 de ovas da lagarta *Spodoptera frugiperda* foram feitos empregando, além do semi-empírico AM1, a Teoria do Funcional de Densidade (DFT) com o pacote computacional ADF.<sup>46</sup> Detalhes são dados na sub-seção 5.5.2 em Métodos.

Os descritores físico-químicos e estruturais foram calculados através de diferentes aplicativos e totalizam aproximadamente 2080. Destes, 1664 foram gerados pelo programa DRAGON<sup>47</sup>, 248 pelo programa RECON<sup>48</sup>, e descritores em menor número pelos programas CaCheFujitsu<sup>49</sup>, MollInspiration<sup>50</sup>, PETRA.<sup>51</sup> Constantes hidrofóbicas, eletrônicas e estéricas foram extraídas da ref. [52]. O detalhamento dos descritores ocorre na seção de cada um dos modelos de QSAR.

O procedimento de seleção dos descritores consiste na montagem de uma matriz composta de aproximadamente 2000 descritores. Pelo programa MDM<sup>37</sup> faz-se uma seleção de descritores com correlação de Pearson atividade/descritor preferivelmente com valor mínimo  $R = \pm 0,3$ . A seleção se diferencia de caso para caso, aplicando toda a matriz no estudo de caso dos diterpenóides, ou parcialmente, no caso das dibenzoilidrazinas, com matrizes variando entre 50 a 300 descritores onde se procura selecionar descritores dos modelos de QSAR da literatura<sup>25,26,27</sup>, buscando um descritor sensível à variação conformacional.

Para a seleção dos descritores e construção de modelos de QSAR foram adotados três aplicativos, que operam de forma similar, empregam MLR, mas têm características próprias de seleção de descritores e proposição de modelos de QSAR. Além disto, cada programa apresenta valores de parâmetros estatísticos de forma diferente mas que se completam. Descrição dos Aplicativos:

**(a) MobyDigs<sup>53</sup>:** se utiliza de um algoritmo genético (GA-VSS – *Genetic Algorithm – Variables Subset Selection*) para a seleção de variáveis e permite ao usuário pré-fixar valores para parâmetros específicos de funções de ajuste RQK implementadas como *QUICK Rule*.<sup>54</sup> São basicamente cinco critérios que precisam ser atendidos simultaneamente para que o modelo selecionado seja considerado válido:

1.  $Q^2_{LOO} > Q^0$  , caso contrário rejeitar o modelo.  $Q^0$  é um valor limítrofe entre 0 e 0,5 associado à capacidade de predição.
2.  $K_{xy} - K_x > \delta K$  , caso contrário há alta colinearidade.  $K_x$  corresponde ao índice de correlação calculado no conjunto de dados de  $p_x$  variáveis, e  $K_{xy}$  o índice de correlação obtido no mesmo conjunto acrescido da resposta  $y$ , isto é, da matriz com  $n$  elementos e  $p_x+1$  variáveis.  $K_{xy} < K_x$  caracteriza um modelo com multicolinearidade sem poder de predição.
3.  $Q^2_{LOO} - Q^2_{ASYM} > \delta Q$ , caso contrário a capacidade de predição do modelo é duvidosa. Segundo a referência<sup>54</sup> existe um comportamento assintótico de  $Q^2_{LOO}$  relacionado a  $R^2$  dado por:

$$Q^2_{ASYM} = 1 - (1 - R^2) \left( \frac{n}{n - p'} \right)^2 \text{ onde } n \text{ é o número de compostos e } p' \text{ o}$$

número de variáveis do modelo.

4.  $R^P > t^P$  , caso contrário há redundância nas variáveis explanatórias; o valor de  $t^P$  é definido pelo usuário no programa variando de 0,01 a 0,1. Avalia-se o quanto duas (ou mais) variáveis explicam a mesma coisa. Se nenhum dos valores  $t^P$  satisfizer a condição, o modelo contém redundância nas variáveis.
5.  $R^N > t^N$  , caso contrário há sobreajuste devido às variáveis com ruído.  $t^N$  é modulado pelo usuário por um parâmetro  $\epsilon$  do programa que varia de 0 a 0,1 em intervalos de 0,01. Como no caso anterior, se nenhum valor satisfizer a desigualdade, o modelo deve ser rejeitado. Os detalhes dos critérios 4 e 5 são extensos e podem ser encontrados na ref.[54].

**(b) Molegro Data Modeler (MDM)**<sup>37</sup>: faz uma seleção dos modelos com um até  $n$ -descritores, aos quais estão associados os coeficientes de  $R$  ( correlação de Pearson) e o erro médio quadrático (RMSE) dos modelos.

**(c) BuildQSAR**<sup>55</sup>: utiliza-se de seleção de variáveis através de um algoritmo genético e busca sistemática com até cinco variáveis, apresentando uma listagem de todos os modelos que pode ser ordenada pela correlação de Pearson  $R$  ou  $Q^2$ .

Nesta etapa do estudo, e dependendo do número de compostos, faz-se uma separação do conjunto de treinamento e de compostos que serão utilizados no

conjunto de validação externa, denominado na literatura também por pseudo *blind set*, termo usado ao longo deste estudo que caracteriza conjuntos de compostos com atividade conhecida, não participantes do treinamento. Os conjuntos de treinamento e de validação externa são obtidos por processo randômico do aplicativo MDM.

A seleção do modelo adota a relação de 5 compostos para um descritor.<sup>56</sup>

### 4.3 – Metodologias de Validação – Domínio de Aplicabilidade ( AD – *Applicability Domain*)

De certa forma as funções RQK descritas anteriormente são também critérios de validação e robustez de modelos de QSAR. Além destes são utilizados critérios adicionais cuja definição segue no texto.

#### *Bootstrap*

Uma vez escolhido o modelo, segue-se uma validação pelo método *bootstrap*<sup>57</sup> onde o conjunto de dados é dividido randomicamente em conjunto de treinamento e teste, donde se obtém o valor  $Q^2_{boot}$  que é então comparado ao valor real  $Q^2_{LOO}$ , esperando-se proximidade entre os dois valores. Esta avaliação é feita com o programa MobyDigs que realiza 5000 ciclos pelo processo randômico.

#### *Y-randomization*

Outro teste realizado é chamado de *Y-randomization* onde a variável dependente Y (atividade biológica) é randomicamente redistribuída no conjunto mantendo inalterada a posição dos descritores. Este tipo de validação avalia se o modelo foi obtido casualmente, ou seja, se existem outros modelos cujos valores dos coeficientes de determinação  $R^2$  e capacidade preditiva  $Q^2_{LOO}$  estão próximos aos valores do modelo original. Se este for o caso, pode-se afirmar que o modelo original é casual. Para esta avaliação o programa MobyDigs<sup>37</sup> se utiliza dos valores do *intercept*  $a(R^2)$  e  $a(Q^2)$ , respectivamente das equações de regressão  $R^2_{yrand} = a_R + b_{R^2}$  e  $Q^2_{yrand} = a_Q + b_{Q^2}$ . Segundo a literatura<sup>56</sup>, se forem atendidos os valores  $a(R^2) < 0,3$  e  $a(Q^2) < 0,05$  o modelo QSAR

pode ser considerado robusto. Este estudo se utiliza de dez ciclos de 300 randomizações por ciclo para a validação, perfazendo 3000 randomizações.

Outra forma utilizada para medição da robustez do modelo pode ser feita utilizando os parâmetros  $R^2_{yrand}$  e  $Q^2_{yrand}$  sugeridos nas ref. [58,59] onde se considera que 25 randomizações de Y são suficientes e os valores médios de  $R^2$  e  $Q^2$  são comparados com os valores do modelo inicial  $R^2$  e  $Q^2$ . Os estudos das ref. [58,59] frequentemente citados pela literatura, não sugerem qual deve ser o valor numérico limite como critério de validação do modelo, apenas consideram os valores médios obtidos como 'baixos' (maior valor médio obtido  $R^2_{yrand} = 0,510$  e o maior  $Q^2_{yrand} = 0,205$ ).

Um critério sugerido por Kiralj e Ferreira<sup>60</sup> fixa os seguintes valores para  $R^2_{yrand}$  e  $Q^2_{yrand}$ :

$Q^2_{yrand} < 0,2$  e  $R^2_{yrand} < 0,2$ , sem chances de correlação.

Qualquer  $Q^2_{yrand}$  e  $0,2 < R^2_{yrand} < 0,3$ , chance de correlação negligível.

Qualquer  $Q^2_{yrand}$  e  $0,3 < R^2_{yrand} < 0,4$ , chance de correlação tolerável.

Qualquer  $Q^2_{yrand}$  e  $R^2_{yrand} > 0,4$ , chance de correlação iminente.

Os valores  $R^2_{yrand}$  e  $Q^2_{yrand}$  foram obtidos neste trabalho empregando o aplicativo QSARModeling<sup>61</sup>. A aprovação do teste é feita pelo resultado do *intercept* [ $a(R^2) < 0,3$  e  $a(Q^2) < 0,05$ ] e utilizando o valor médio de  $R^2_{Yrand} \leq 0,40$  a  $Q^2_{Yrand, qualquer}$ . Como a literatura em geral sugere que um modelo QSAR deva ter como valores estatísticos mínimos  $R^2 > 0,6$  e  $Q^2 > 0,5$ , o valor médio de  $R^2 \leq 0,4$  no teste de randomização, concede ao estudo uma margem de segurança razoável de que o modelo de QSAR não tenha sido obtido casualmente.

#### *Leave-one-out* (LOO-CV) e *Leave-N-out* (LNO-CV) - Validação Cruzada

A validação cruzada é um método para se avaliar a acurácia de um modelo de regressão ou classificação. Todo o conjunto de dados será utilizado em teste e treinamento, mas não ao mesmo tempo. Um conjunto de dados é dividido em uma série de conjuntos de teste e treinamento. Um modelo é construído com cada um dos conjuntos de treinamento e validado com o conjunto de teste. O método LOO-CV elimina uma amostra de um conjunto com M amostras, gerando n conjuntos de

treinamento com  $M-1$  amostras e realiza a predição para cada uma das amostras excluídas. A prática tem demonstrado que este processo sozinho não assegura a robustez do modelo na validação interna, o que é obtido pelo método LNO-CV, onde se eliminam  $n$  amostras por vez.<sup>62</sup> Num conjunto com poucas amostras, recomenda-se fixar  $n$  entre 20% - 30% do total. A robustez do modelo é atestada se ocorrer uma diferença entre  $Q^2 - Q^2_{LNO} < 0,1$ .

Em estudos de QSAR o valor mínimo aceitável de  $Q^2$  é maior que 0,5 sendo 0,9 considerado excelente.  $R^2$  em geral deve ser maior que 0,6. A diferença entre  $R^2 - Q^2$  maior que 0,2 - 0,3 indica termos irrelevantes no modelo, sobre ajuste (*overfitting*) ou presença de alguns *outliers*.<sup>56</sup>

### Domínio de Aplicabilidade – *Outliers - Leverage*

O domínio de aplicabilidade (*Applicability Domain – AD*) de um modelo de QSAR é definido como a informação, o conhecimento ou espaço físico-químico, estrutural, ou biológico, utilizado para desenvolver o conjunto de treinamento, que deve ser aplicável em predições para novos compostos. A forma ideal é utilizar o modelo de QSAR somente para predições dentro do domínio por interpolação e não por extrapolação.<sup>63</sup> Neste estudo obtiveram-se algumas predições por extrapolação o que é uma consequência natural do cálculo.

A representação gráfica conhecida por *Williams plot* implementada no programa MobyDigs<sup>37</sup> é utilizada para avaliar *outliers* e *leverage*, ambos parâmetros associados ao domínio de aplicabilidade. *Outliers* e *leverage* são indicadores de compostos situados fora do domínio e estão acima do limite crítico dado por três vezes o desvio padrão ( $\pm 3\sigma$ ). Um composto assim classificado precisa ter sua manutenção ou eliminação do conjunto de treinamento, justificada.

### Validação Externa

A validação externa é feita neste estudo pelo método sugerido por Golbraikh e Tropsha.<sup>64</sup> O método se aplica apenas ao conjunto de compostos selecionados para validação externa (conjunto de teste), com a recomendação que o conjunto tenha pelo menos cinco integrantes.

As equações utilizadas para se fazer esta avaliação e os principais parâmetros propostos pelo método são dados a seguir:

1.)  $Q^2_{LOO}$  com validação cruzada de todos os compostos.

$$Q^2_{LOO} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (4.3-1)$$

onde  $y_i$  é o valor experimental,  $\hat{y}_i$  o valor predito, e  $\bar{y}$  o valor médio experimental.

2.) Cálculo dos coeficientes  $a$ ,  $b$  e  $a'$ ,  $b'$  das retas de regressão do conjunto de teste com  $n$  integrantes. A linha de regressão será bissetriz do ângulo formado pelas direções positivas dos eixos ortogonais  $\tilde{y}$  e  $y$  que são as atividades preditas e calculadas, respectivamente, e podem ser descritas pelas respectivas retas  $y' = a\tilde{y} + b$  e  $\tilde{y}' = a'y + b'$  onde

$$a = \frac{\sum (y_i - \hat{y}_i)(\tilde{y}_i - \bar{\tilde{y}})}{\sum (\tilde{y}_i - \bar{\tilde{y}})^2}, \quad b = \bar{\tilde{y}} - a\bar{\tilde{y}} \quad (4.3-2a)$$

$$a' = \frac{\sum (y_i - \hat{y}_i)(\tilde{y}_i - \bar{\tilde{y}})}{\sum (y_i - \bar{y})^2}, \quad b' = \bar{y} - a'\bar{\tilde{y}} \quad (4.3-2b)$$

onde  $\tilde{y}_i$  e  $y_i$  correspondem ao valor predito e observado, respectivamente.  $\bar{\tilde{y}}$  e  $\bar{y}$  são os valores médios preditos e observados, respectivamente.

3.) Coeficiente de correlação do conjunto de teste

$$R = \frac{\sum (y_i - \bar{y})(\tilde{y}_i - \bar{\tilde{y}})}{\sqrt{\sum (y_i - \bar{y})^2 (\tilde{y}_i - \bar{\tilde{y}})^2}} \quad (4.3-3a)$$

Em um modelo ideal,  $a$ ,  $a'$  são iguais a 1, o *intercept*  $b$ ,  $b'$  igual a zero, e o coeficiente de correlação para a regressão de  $\tilde{y}$  versus  $y$  é igual a 1. Um modelo de QSAR é considerado altamente preditivo se tiver valores próximos aos valores ideais. Isto pode implicar em que o coeficiente de correlação  $R$  entre o valor observado  $y$  e a atividade predita  $\tilde{y}$  devem estar próximos a 1 e as regressões de  $y$  contra  $\tilde{y}$  ou de  $\tilde{y}$  contra  $y$  através da origem deveriam ser caracterizadas pelo

menos por  $k$  ou  $k'$  próximo a 1, isto é,  $y^{r0} = k\tilde{y}$  e  $\tilde{y}^{r0} = k'y$ . As inclinações  $k$  e  $k'$  das retas são dadas por

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2}, \quad (4.3-3b)$$

$$k' = \frac{\sum y_i \tilde{y}_i}{\sum y_i^2}, \quad (4.3-3c)$$

#### 4.) Coeficiente de correlação $R_0^2$ e $R_0'^2$

$$R_0^2 = 1 - \frac{\sum (\tilde{y}_i - y_i^{r0})^2}{\sum (\tilde{y}_i - \bar{\tilde{y}})^2}, \quad (4.3-4a)$$

$$R_0'^2 = 1 - \frac{\sum (y_i - \tilde{y}_i^{r0})^2}{\sum (y_i - \bar{y})^2}, \quad (4.3-4b)$$

onde

$$y_i^{r0} = k\tilde{y}_i \text{ e } \tilde{y}_i^{r0} = k'y_i$$

Para que o modelo de QSAR seja qualificado como modelo preditivo é necessário que sejam atendidas as seguintes condições:

- (a)** Pelo menos um dos coeficientes de correlação  $R_0^2$  ou  $R_0'^2$  (melhor seria ambos) deverá estar próximo a  $R^2$ . O coeficiente  $R_0^2$  mede a correlação  $Y_{\text{Observado}}$  vs.  $Y_{\text{Predito}}$ . O coeficiente  $R_0'^2$  mede a correlação  $Y_{\text{Predito}}$  vs.  $Y_{\text{Observado}}$ . Apesar de parecer redundante, Golbraikh e Tropsha<sup>64</sup> mostram que os respectivos coeficientes de inclinação das retas  $k$  e  $k'$  e os coeficientes das retas  $(a, a')$  e  $(b, b')$  podem ser muito diferentes, como será visto também no presente estudo.
- (b)** Pelo menos um dos valores de  $k$  ou  $k'$  deve estar próximo a 1. O intervalo de referência para aprovação segue:  $0,85 \leq k \leq 1,15$  ou  $0,85 \leq k' \leq 1,15$ .
- (c)**  $[(R^2 - R_0^2) / R^2] < 0,3$  ou  $[(R^2 - R_0'^2) / R^2] < 0,3$

Estes critérios foram empregados em todos os estudos de caso deste estudo QSAR e os resultados apresentados graficamente.

No procedimento Golbraikh-Tropsha<sup>64</sup> é sugerido que o conjunto de teste não participe do conjunto de treinamento. Este aspecto não é obedecido no presente estudo. A experiência acumulada com diversos modelos de QSAR mostrou que

mesmo treinando os compostos do conjunto de teste a validação externa pode não ser obtida.

#### 4.4 – Análise Conformacional

Em geral os estudos de QSAR clássico são desenvolvidos seguindo as seguintes premissas: **(a)** O efeito produzido é causado pelo composto modelado e não por seus metabólitos. **(b)** A atividade biológica é primariamente explicada por processos entálpicos. **(c)** Os termos entrópicos são similares para todos os compostos. **(d)** O sistema é considerado em equilíbrio, e aspectos cinéticos geralmente não são considerados. **(e)** Efeitos farmacocinéticos, como efeito solvente, difusão e transporte, não estão inclusos. **(f)** O sítio ativo é o mesmo para todos os compostos modelados. **(g)** A conformação proposta é também a bioativa. O presente trabalho segue estas premissas, exceto por este último tópico, que está sujeito a um critério: cada uma das estruturas dos compostos tem a sua geometria otimizada, e é considerada a estrutura bioativa se for atendido um valor de resíduo máximo, arbitrado neste estudo em  $Y_{\text{Observado}} - Y_{\text{Predito}} = \pm 0,10$ , o que confere um erro máximo de predição para um determinado composto compreendido no intervalo  $0,80Y_{\text{Obs}} \leq Y_{\text{Pred}} \leq 1,25Y_{\text{Obs}}$ .

O que se objetiva é encontrar uma conformação, caso o composto admita existência de conformações, que melhor reproduza o valor da bioatividade experimental.

Quando os compostos possuem variação conformacional, existe a oportunidade de se construir um modelo de QSAR com predição interna mais acurada. Para isto, é necessário que pelo menos um descritor do modelo de QSAR seja sensível às variações conformacionais e apresente uma variação significativa nos seus valores. Em geral este resultado é alcançado na maioria dos compostos integrantes deste estudo. Para compostos que não tenham uma conformação que esteja contida no valor residual arbitrado ( $Y_{\text{Observado}} - Y_{\text{Predito}} = \pm 0,10$ ), emprega-se a conformação da estrutura com geometria otimizada com o método semiempírico AM1. Este método não se aplica se os compostos não apresentarem variação conformacional.

Um estudo se utilizou de 150 estruturas cristalografadas de complexos ligante-proteína com relevância farmacêutica constatando que 60% dos ligantes não interagem na sua conformação de mínimo global. Uma análise do grau de dobramento dos ligantes confirmou a tendência geral de que moléculas pequenas tendem a se ligar com conformação estendida.<sup>65a</sup> Verifica-se em pesquisas desta natureza<sup>65b</sup> que é recomendável se utilizar mais de um campo de força na avaliação de confôrmeros. O campo MMFF94 é exemplificado como especialmente adequado para ser empregado em pesquisas de fármacos.

Na mesma linha de conclusão do estudo anterior, com o cristal de estruturas do complexo proteína-ligante, mostra-se que conformações bioativas tendem a ter geometrias mais estendidas (mais volumosas) do que as randômicas e estão muitas kcal/mol acima em energia do que os seus respectivos mínimos globais.<sup>66,67</sup> As investigações computacionais com conformações, nos estudos citados, foram feitas com diferentes campos de força.

O presente estudo empregou o campo Merck MMFF94<sup>68,69</sup> e Dreiding modificado<sup>70a,b</sup> utilizando o *default* dos aplicativos, gerando de 20 a 100 conformações por composto. Ambos geram espaços conformacionais que se complementam. O emprego de um único campo, pode, por características de parametrização, gerar uma população de confôrmeros favorável em um caso e desfavorável em outro, onde o termo 'favorável' significa o espaço conformacional gerado conter ou não um confôrmero que atenda à reprodução da bioatividade.

### **Construção do modelo de QSAR – Seleção de Confôrmeros**

Como já citado anteriormente, após calcular o valor dos descritores para todos os confôrmeros de um mesmo composto, e utilizando como referência o valor residual  $Y_{\text{Observado}} - Y_{\text{Predito}} = \pm 0,10$ , faz-se a busca da conformação que melhor reproduza a bioatividade e se substitue o valor do descritor que varia com a conformação no modelo de QSAR. Se o valor calculado (predito) gera o resíduo dentro do intervalo  $-0,10 \leq \text{resíduo} \leq +0,10$  passa-se ao composto seguinte. Caso contrário, procura-se um novo confôrmero até que o valor do intervalo residual seja alcançado. O esquema da Figura 8 a seguir ilustra o procedimento.

**Passo 1.** Escolha do composto que tem resíduo fora do intervalo proposto, p. ex.: composto D6;  $Y_{\text{obs}} - Y_{\text{pred}} = 6,96 - 6,72 = 0,22$ . (Valor correspondente de  $\text{Mor25v} = 0,600$ )

Fração da planilha de QSAR do aplicativo MDM							
	MolID	Yobs	LogP	sigma	Mor25v	MLR-Train [3D]	res
1	D1	7.08	2.59	0	0.552	7.129	-0.05
2	D2	7.29	2.87	0.06	0.519	7.358	-0.07
3	D3	6.89	3.51	0.23	0.586	7.021	-0.13
4	D4	7.69	3.73	0.23	0.51	7.532	0.16
5	D5	7.3	3.96	0.18	0.546	7.43	-0.13
6	D6	6.96	3.68	0.54	0.6	6.742	0.22

**Passo 2.** Busca na planilha de confôrmers do composto D6, um confômero que o aproxime do valor  $\text{pIC}_{50}$  experimental

Fração da planilha de confôrmers (n=30)	
K	L
MolID	Mor25v
DBH6_20	0.612
DBH6_19	0.611
DBH6_21	0.609
DBH6_22	0.609
DBH6_30	0.607
DBH6_29	0.606
DBH6-AM1	0.600
DBH6_27	0.589
DBH6_28	0.588
DBH6_31	0.547
DBH6_32	0.547
DBH6_33	0.538
DBH6_34	0.538
DBH6_3	0.528
DBH6_4	0.526

Valor do descritor Mor25v no QSAR inicial.  
Resíduo  $Y_{\text{Obs}} - Y_{\text{Pred}} = 0,22$

Valor do descritor Mor25v (a ser substituído na planilha de QSAR) que melhor aproxima o valor calculado do experimental  
Resíduo  $Y_{\text{Obs}} - Y_{\text{Pred}} = 0,02$

**Figura 8.** Construção do modelo de QSAR pela seleção do confômero através do descritor Mor25v, que melhor reproduz a bioatividade. LogP e  $\sigma$ -Hammett não variam com a conformação.

O modelo de QSAR final assim obtido é submetido a todo o processo de validação citado anteriormente. Na construção dos modelos de QSAR os descritores têm seus valores autoescalados.

#### 4.5 – Proposta de Novos Compostos – Similaridade Molecular

Os modelos de QSAR deste estudo têm a finalidade principal de serem utilizados como meio para propor novos compostos da mesma classe química em que foram desenvolvidos. No caso dos novos compostos, utilizou-se um pseudo *blind set*, ou seja, um conjunto do qual se conhecem os resultados experimentais obtidos com o mesmo protocolo mas que não integra o conjunto de compostos utilizados no desenvolvimento dos modelos de QSAR. Aplicando-se análise conformacional aos compostos do pseudo *blind set* e através de métodos de similaridade molecular, identifica-se qual o confômero que melhor representa o composto bioativo pelos critérios do método. O método aplicado a um conjunto de compostos e respectivos confômeros, deve conseguir identificar o confômero que mais se aproxima da bioatividade calculada pelo modelo de QSAR à bioatividade observada. O critério de acerto arbitrado como aceitável é de 60%.

Assim, trabalha-se com a hipótese de que compostos análogos propostos, possam ter os confômeros escolhidos como os mais prováveis para uma avaliação de predição da bioatividade pelo modelo de QSAR.

A escolha de um método de similaridade molecular pode dar bons resultados em um caso e falhar em outros. Existem várias abordagens para se medir similaridade química<sup>71</sup> e este estudo optou por três métodos a saber: o método de similaridade molecular por distância euclidiana implementada no aplicativo MDM<sup>37</sup>, o método ShaEP<sup>72</sup> e o terceiro método usa o índice de Carbó exclusivamente aplicado e explicado no trabalho desenvolvido na seção 5.3.

O primeiro método se utiliza da distância euclidiana entre dois pontos  $x$  e  $y$  num espaço  $n$ -dimensional, onde  $n$  é o número de descritores escolhidos para se fazer a avaliação. Os compostos do modelo de QSAR estão dispostos em conjunto com o grupo de confômeros do composto proposto em uma planilha do MDM<sup>37</sup>. A seleção do confômero se faz pela classificação numérica dada entre 0 (alta similaridade) até 1 (nenhuma similaridade).

O segundo método se baseia na sobreposição molecular utilizando a forma molecular e o potencial eletrostático como elementos de avaliação por similaridade. Escolhe-se em geral o composto que tem o maior valor de bioatividade como referência e os confômeros do primeiro composto a ser analisado. São gerados três valores de similaridade que variam de 0 (nenhuma similaridade) até 1 (alta similaridade) a saber: similaridade composta (forma combinada com potencial eletrostático), forma e potencial eletrostático. Escolheu-se para o estudo a similaridade composta como indicador para a seleção do melhor confômero.

Por recomendação dos autores do programa ShaEP, as cargas de Mulliken devem ser substituídas. O estudo presente utiliza cargas AM1-BCC obtidas empregando o programa Chimera<sup>73</sup> e Antechamber.<sup>74a,b</sup> AM1-BCC (*bond charge corrections*) foi parametrizado para reproduzir cargas comparáveis às produzidas pelo método Hartree-Fock HF/6-31G\*. Seu uso é recomendado para qualquer molécula orgânica ou biológica.

O propósito de se utilizar três métodos é verificar se pelo menos um deles é capaz de gerar resultados aceitáveis nos estudos de caso das seções 5.2, 5.3 e 5.5. Não há finalidade comparativa entre métodos.

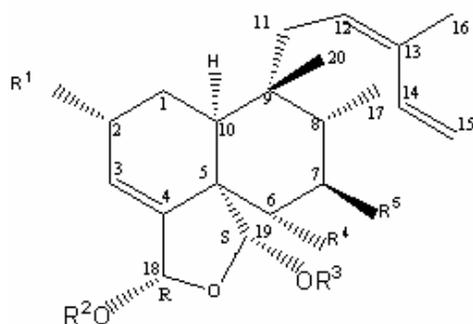


## 5. ESTUDOS de QSAR

### 5.1 – Diterpenóides Clerodanos contra Células Tumorais V79

#### 5.1.1 – Introdução

A partir do extrato sêco de folhas de *Casearia sylvestris* Sw., Itokawa *et al.*<sup>75</sup> e Morita *et al.*<sup>76</sup>, isolaram e elucidaram 22 diterpenóides oxigenados tricíclicos do tipo *cis*-clerodano ensaiados contra células tumorais V79. Empregaram o mesmo protocolo nos bioensaios, o que possibilitou a aplicação do modelo de QSAR proposto. A Figura 9 ilustra as estruturas dos diterpenóides, nominados Casearinas seguidas da letra A até R. As Casearinas Aa, Ab, Ac e Da foram sintetizadas a partir das Casearinas A e D, respectivamente.



Casearin	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	R <sup>5</sup>	IC <sub>50</sub> (μmol/D)
A	OMe	Ac <sup>a</sup>	Ac	OH	OBu <sup>c</sup>	1
B	OMe	Ac	Ac	OAc	OBu	8.5
C	OH	Ac	Ac	OAc	ODc	0.77
D	OH	Bu	Ac	OH	OBu	1.8
E	OH	Et <sup>b</sup>	Ac	OH	ODc <sup>d</sup>	4.7
F	OH	Et	Ac	OH	OBu	29
G	OMe	Ac	Ac	H	OBu	0.17
H	OH	Ac	Ac	H	OBu	0.37
I	OH	Ac	Bu	H	OBu	0.51
J	OMe	Bu	Ac	OH	OBu	1.1
K	OAc	Ac	Ac	OH	OBu	0.52
L	OMe	Bu	Ac	OAc	OH	1.6
M	OH	Bu	Bu	OAc	OH	1.8
N	OMe	Ac	Bu	OAc	OBu	5.9
O	OMe	Bu	Ac	OAc	OBu	6
P	OMe	Ac	Ac	OAc	OAc	7.8
Q	OH	Ac	Ac	OAc	OBu	4.3
R	=O	Ac	Ac	OH	OBu	5.4
Aa	OMe	Ac	Ac	=O	OBu	0.55
Ab	OMe	Ac	Ac	OPr	OBu	17
Ac	OMe	Ac	Ac	OBu	OBu	38
Da	=O	Bu	Ac	=O	OBu	19

(a) OAc = acetato; (b) Et = etil; (c) OBU = butanoato;  
(d) ODc = decanoato

**Figura 9.** Estrutura das Casearinas e a atividade citotóxica medida pela concentração molar IC<sub>50</sub> contra células pulmonares (fibroblastos) V79 de Hamster Chinês<sup>76</sup>.

A literatura<sup>75,76</sup> nada informa sobre réplicas nos bioensaios, consequentemente sem informes sobre erro experimental.

O estudo contribui com uma proposta de QSAR validado, que é usado para prever a bioatividade de 113 diterpenóides clerodanos escolhidos da literatura (vide ref. 46 até 61 do artigo, Apêndice 1), a grande maioria é proveniente de órgãos de plantas (folhas, frutos, sementes, raízes, caule, flores) do gênero *Casearia*, *Laetia* e *Zuelania*, da família Flacourtiaceae/Salicaceae e diterpenóides pertencentes a família Chrisobalanaceae. Alguns destes compostos foram ensaiados com diferentes protocolos e outros, nunca bioensaiados. Todas as estruturas constantes da literatura citada tiveram suas estruturas elucidadas empregando ressonância magnética nuclear (RMN) e espectroscopia de raios-X. Este foi um critério de certificação adotado pelo autor dos artigos de revisão<sup>16</sup>, aspecto este que confere maior certeza da exatidão da estrutura a ser empregada no estudo de QSAR. As 113 estruturas estão ilustradas no Apêndice 3 e estão associadas ao nome da planta que lhes deu origem. Em alguns casos, para evitar ambigüidade, arbitrou-se o nome do composto associando-o ao nome da planta, em substituição a apenas letras ou números atribuídos pelos pesquisadores, ao composto no artigo correspondente. Exemplificando: compostos de *Casearia corymbosa* identificados pelos números 1 até 13, receberam os nomes Ccorymbosa 1, Ccorymbosa 2, etc.

Uma contribuição adicional é dada pela proposta de diterpenóides clerodanos desenvolvidos *in silico*. A Química Medicinal oferece exemplos de produtos naturais que podem ser sintetizados com a incorporação de átomos de cloro e flúor. Neste sentido, existem produtos farmacêuticos comercializados na categoria de agentes anti-mitótico e anti-tumoral, usados no tratamento de câncer.<sup>77</sup> A literatura exemplifica o uso diversificado de compostos fluorados vinculados às propriedades biológicas.<sup>78</sup>

Análogos sintéticos fluorados de produtos naturais são utilizados na quimioterapia, p.ex., a Camptotecina, um alcalóide extraído da árvore *Camptotheca acuminata* Decne., Nyssaceae; a toxina lignana Podofilo extraída de *Podophyllum peltatum* L., é um agente citotóxico contra várias linhas de células tumorais, porém, também apresenta numerosos efeitos colaterais.<sup>79</sup>

### 5.1.2 – Métodos

Os modelos de QSAR para os diterpenóides clerodanos foram inicialmente desenvolvidos com duas metodologias independentes: Regressão Linear Múltipla (MLR-Multiple Linear Regression) e Máquinas de Vetores de Suporte (SVM-Support Vector Machine), empregando os mesmos descritores. O estudo foi desenvolvido em duas fases. A primeira fase, descrita em detalhes no artigo (Apêndice 1) utiliza as estruturas com geometria otimizada pelo método semi-empírico AM1, e que são consideradas as estruturas bioativas, como convencionado por estudos de QSAR. A segunda fase, complementar ao estudo com diterpenóides, é feita empregando análise conformacional, tal como descrito em Material e Métodos, Seção 4.4. Por se tratar de um conjunto pequeno de compostos ( $n = 18$ ) procurou-se empregar dois métodos independentes que pudessem fornecer valores da bioatividade calculada próximos entre si, em outras palavras, os métodos MLR e SVM foram utilizados para obter um mesmo desempenho preditivo dos modelos de QSAR.

Em geral o emprego apenas do método SVM gera modelos de QSAR de difícil interpretabilidade pela própria natureza do método, pois ao contrário do método MLR que gera uma equação linear, no método SVM são geradas várias equações representadas por vetores suporte. Para podermos fazer uma associação dos descritores à bioatividade de forma linear, foram construídos modelos de QSAR com os dois métodos, na tentativa de se buscar resultados próximos ou similares, por metodologias independentes, ficando a interpretabilidade relegada ao modelo de QSAR obtido pelo método MLR.

Para a proposta de novos compostos, foram utilizados dois critérios. O primeiro critério utilizou como base o diterpenóide clerodano Casearina G que apresentou a maior bioatividade (v. fig. 9), e que foi modificado *in silico* pela introdução de átomos de cloro e flúor nas posições específicas  $R^1$ ,  $R^4$  e  $R^5$ . O segundo critério seleciona a Casearborina A, que foi a melhor classificada na predição dos 113 diterpenóides e, com procedimento similar ao da Casearina G, empregando substituintes em  $R^4$  e  $R^5$  (fig. 9). A análise foi estendida também a

Casearvestrina A, B e C, e à Caseamembrina A. Em todos os casos, atenção especial é dada na posição do carbono-sp<sup>3</sup> C-6 mencionado na literatura<sup>76</sup> como sendo um centro sensível às mudanças na bioatividade. A bioatividade (IC<sub>50</sub>) dada em μmol/L foi convertida em  $\log(1/IC_{50}) = pIC_{50}$ .

### 5.1.3 – Resultados e Discussão

#### 5.1.3.1 – QSAR e Descritores - Estudo *sem* Análise Conformacional

A equação de QSAR obtida com os 18 compostos usados como conjunto de treinamento contém dois descritores calculados pelo programa DRAGON (RDF060m e R8p) e um descritor calculado pelo programa RECON (Fuk5), e é representada pela Eq. :

$$pIC_{50} = 0,083 (\pm 0,021) \text{ RDF060m} - 7,637 (\pm 1,486) \text{ R8p} - 0,043 (\pm 0,010) \text{ Fuk5} + 13,219 (\pm 1,664)$$

$$n = 18 \quad R^2 = 0,931 \quad s = 0,148 \quad F = 64 \quad p < 0,0001 \quad Q^2 = 0,889 \quad \text{SPress} = 0,189 \quad \text{SDEP} = 0,172 \quad (5.1-1)$$

RDF060m é definido como uma função de distribuição radial (*Radial Distribution Function-6.0/weighted by atomic masses*) codificando características geométricas e atômicas da estrutura química em 3D de acordo com as propriedades  $A_i$  e  $A_j$ , que no presente estudo corresponde às massas atômicas. A equação (5.1-2), utiliza um termo gaussiano como função da distância

$$RDF(r) = f \sum_{i=1}^{N-1} \sum_{j=i+1}^N A_i A_j e^{-B(r-r_{ij})^2} \quad (5.1-2)$$

onde  $f$  é uma fator de escalamento,  $N$  o número de átomos  $i$  e  $j$ , e  $r_{ij}$  são as distâncias interatômicas. O parâmetro de alisamento  $B$  define a probabilidade de distribuição de distâncias individuais dos átomos que podem ser vistos como sendo as suas vibrações na molécula. Segundo Hemmer *et al.*<sup>80</sup> o descritor RDF pode ser interpretado como a probabilidade de distribuição de se achar um átomo num volume esférico de raio  $r$ . Assim, é plausível se presumir que o descritor RDF060m

codifica uma contribuição parcial da bioatividade das Casearinas em regiões onde os átomos estão localizados a uma distância de até 6,0 Å do centro geométrico da molécula. O programa calculou trinta descritores codificados pela massa atômica, com passo de  $r = 0,5 \text{ \AA}$  e apenas o descritor RDF060m apresentou relevância no modelo de QSAR.

O descritor tridimensional R8p pertence à categoria GETAWAY (*Geometry, Topology, and Atom-Weights Assembly*) e codifica informações estruturais e estereo químicas.<sup>81</sup> A representação estrutural da molécula é definida pela matriz de influência molecular (MIM) denotada por H na equação (5.1-3):

$$\mathbf{H} = \mathbf{M} \cdot (\mathbf{M}^T \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T \quad (5.1-3)$$

onde  $\mathbf{M}$  é a matriz molecular e o sobrescrito  $\mathbf{T}$  se refere a matriz transposta. O descritor R8p pertence à classe R-GETAWAY e é calculado segundo a equação:

$$Rkw = \sum_{i=1}^{N-1} \sum_{j>i}^N \frac{\sqrt{h_{ii} h_{jj}}}{r_{ij}} w_i w_j \delta(k; d_{ij}) \quad (5.1-4)$$

$$k=1, 2, \dots, 8$$

onde R8p representa uma autocorrelação codificado no oitavo intervalo pela polarizabilidade  $p = w_i w_j$  da molécula (*autocorrelation of lag 8/weighted by atomic polarizabilities*).  $N$  é o número de átomos na molécula,  $d_{ij}$  é a distância topológica entre átomos  $i$  e  $j$ ,  $r_{ij}$  é a distância geométrica entre átomos  $i$  e  $j$ , e  $\delta(k; d_{ij})$  é a função delta de Dirac.  $h_{ii}$  e  $h_{jj}$  são os elementos diagonais da matriz de influência (MIM), codificando informações atômicas e representando a influência de cada átomo na molécula na determinação de sua forma geométrica.

O descritor TAE<sup>82,83</sup> (*Transferable Atom Equivalent*) Fuk5 está relacionado com valores escalares da função de Fukui  $F^+$  dado por

$$F^+(r) = \rho_{HOMO}(r) \quad (5.1-5)$$

onde  $\rho_{HOMO}$  é a densidade eletrônica do orbital ocupado mais alto (HOMO) em um ponto  $r$  do espaço ao redor da molécula. Caracteriza ataque nucleofílico e descreve

atividade radical vinculada a uma área de superfície e o número 5 se refere a um valor médio desta função conforme calculado pelo programa RECON.

Em resumo, o modelo de QSAR contém informações físico-químicas dadas por RDF060m dos átomos da molécula em posições favoráveis a uma interação com o receptor (no presente caso desconhecido) da célula V79. Captura informações da forma geométrica e das polarizabilidades atômicas (R8p), e caracteriza a reatividade na superfície da molécula. Contudo, não é possível fazer uma clara interpretação dos descritores individualmente com a bioatividade. A literatura<sup>75,76</sup> não fornece maiores informações sobre eventuais mecanismos de ação entre diterpenóides e as células V79, exceto a constatação da regressão tumoral em base ao grupo controle.

Seguem os coeficientes extraídos da matriz de correlação de Pearson entre descritores: RDF060m/R8p (0,431), RDF060m/Fuk5 (0,096), R8p/Fuk5 (0,230). Correlação entre descritor e bioatividade: RDF060m/pIC<sub>50</sub> (0,349), R8p/pIC<sub>50</sub> (0,437), Fuk5/pIC<sub>50</sub> (0,502). A premissa geral fixada neste estudo é uma correlação entre descritores, menor possível e limitada a 0,500, ao passo que a relação para com a bioatividade é desejável que seja no mínimo igual a  $R = 0,3$ . Um aspecto também a considerar é que um modelo de QSAR de 2 variáveis com o requerimento  $R^2 > 0,60$  não foi encontrado.

### Modelos de QSAR

Os valores numéricos dos descritores e o registro CAS dos diterpenóides estão listados na Tabela 1. O conjunto de dados foi dividido em conjunto de treinamento (n = 13) e conjunto de teste (n = 5), empregando-se para esta seleção o método HCA (Hierarchical Cluster Analysis), do *software* Pirouette (ref. [34] do Apêndice 1).

**Tabela 1.** Casearinas e Descritores

Item	MolID	CAS no.	pIC <sub>50</sub> <sup>a</sup>	Descritores <sup>b</sup>		
				RDF060m	R8p	Fuk5
1	A	134174-73-5	6,00	29,874	0,617	119,23
2	Aa <sup>T</sup>	134955-74-1	6,26	29,562	0,607	111,38
3	B	133577-59-0	5,07	28,128	0,702	119,33
4	C	133577-60-3	6,11	35,602	0,576	132,50
5	D	133557-53-6	5,74	28,914	0,579	126,07
6	E	133535-78-1	5,33	19,374	0,506	129,19
7	G	134955-62-7	6,77	31,047	0,610	105,67
8	H	134955-63-8	6,41	27,228	0,580	108,22
9	I	134955-64-9	6,29	31,794	0,583	114,28
10	J	134955-65-0	5,96	31,524	0,568	125,30
11	K <sup>T</sup>	134955-66-1	6,28	37,110	0,634	121,87
12	L <sup>T</sup>	134955-67-2	5,80	27,584	0,596	119,23
13	M <sup>T</sup>	134955-68-3	5,75	24,939	0,573	127,84
14	N <sup>T</sup>	134955-69-4	5,23	33,333	0,703	125,39
15	O	134955-70-7	5,22	31,519	0,670	125,39
16	P	134955-71-8	5,11	29,160	0,717	113,26
17	Q	134955-72-9	5,37	30,480	0,693	121,87
18	R	134955-73-0	5,27	24,239	0,569	129,27

(a) pIC<sub>50</sub> = Log(1/IC<sub>50</sub>) e IC<sub>50</sub>=μmol/L; (b) valores para estruturas otimizadas com AM1; (T) compostos do conjunto de teste.

A partir do dendograma foram escolhidos dez conjuntos com 5 compostos avaliados pelos métodos MLR e SVM. O melhor modelo foi escolhido utilizando Q<sup>2</sup><sub>ext</sub> como referência, e simultaneamente avaliando a robustez e validação destes conjuntos, resultados que serão apresentados mais adiante.

A equação (5.1-6) abaixo, representa o melhor modelo de QSAR obtido com o conjunto de treinamento n = 13.

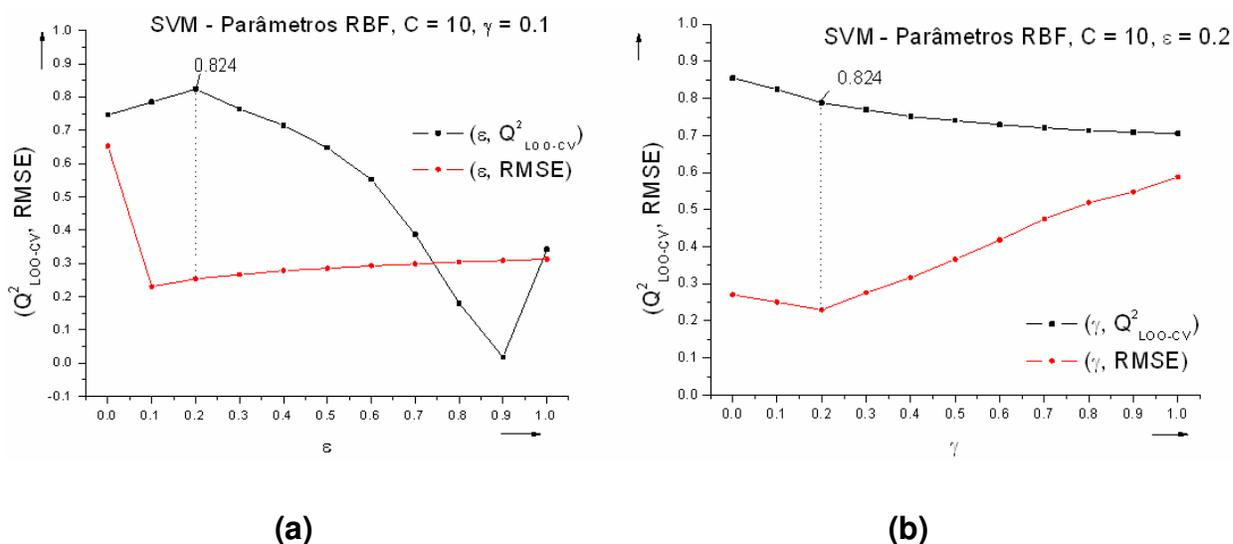
$$pIC_{50} = 0,086 (\pm 0,025) \text{ RDF060m} - 7,581 (\pm 1,616) \text{ R8p} - 0,045 (\pm 0,012) \text{ Fuk5} + 13,363 (\pm 1,954)$$

$$n = 13 \quad R^2 = 0,952 \quad s = 0,141 \quad F = 59,52 \quad p < 0,0001 \quad Q^2 = 0,904 \quad Q^2_{\text{ext}} = 0,816$$

$$SPress = 0,199 \quad SDEP = 0,172) \quad (5.1-6)$$

O modelo de QSAR construído com SVM (n = 18) utilizou o *kernel Radial Basis Function* (RBF) (detalhes em Material e Métodos). Através de um procedimento de ajuste fino, obteve-se o menor valor RMSE (*Root Mean Square Error*) relacionado aos melhores parâmetros de regressão empregando LOO-CV

(*Leave-One-Out Cross-Validation* ( $Q^2_{\text{LOO-CV}}$ )). O melhor valor obtido foi com o ajuste dos seguintes parâmetros: custo (ou capacidade)  $C = 100$ ; extensão da função radial RBF,  $\gamma = 0,1$  e a função de perda  $\varepsilon = 0,2$ . O gráfico da Figura 10 mostra a variação do RMSE quando se varia um dos parâmetros e se mantém os outros dois fixos.



**Figura 10.** (a)  $Q^2_{\text{LOO-CV}}$  em função da variação do parâmetro  $\varepsilon$  mantendo constantes  $C$  e  $\gamma$ ; (b)  $Q^2_{\text{LOO-CV}}$  em função da variação do parâmetro  $\gamma$  mantendo constantes  $C$  e  $\varepsilon$ . Para o menor RMSE seleciona-se o valor  $Q^2_{\text{LOO-CV}}$  correspondente ( $Q^2_{\text{LOO-CV}} = 0,824$ ).

O método SVM se utiliza de vetores suporte ao passo que o método MLR expressa os resultados por uma equação de regressão linear. Assim, para fins comparativos entre métodos, foram utilizados dois parâmetros estatísticos, o coeficiente de determinação ( $R^2$ ) e o erro médio quadrático (MSE-*Mean Square Error*) (Tabela 2). Os resíduos dão uma idéia na diferença de desempenho dos respectivos métodos. O parâmetro do erro médio quadrático é recomendado para ser utilizado no comparativo entre resultados produzidos por equações diferentes. No caso  $\text{MSE-MLR} = 0,0171$  e  $\text{MSE-SVM} = 0,0181$ , o que indica uma relativa equivalência entre as metodologias utilizando  $n = 18$  compostos.

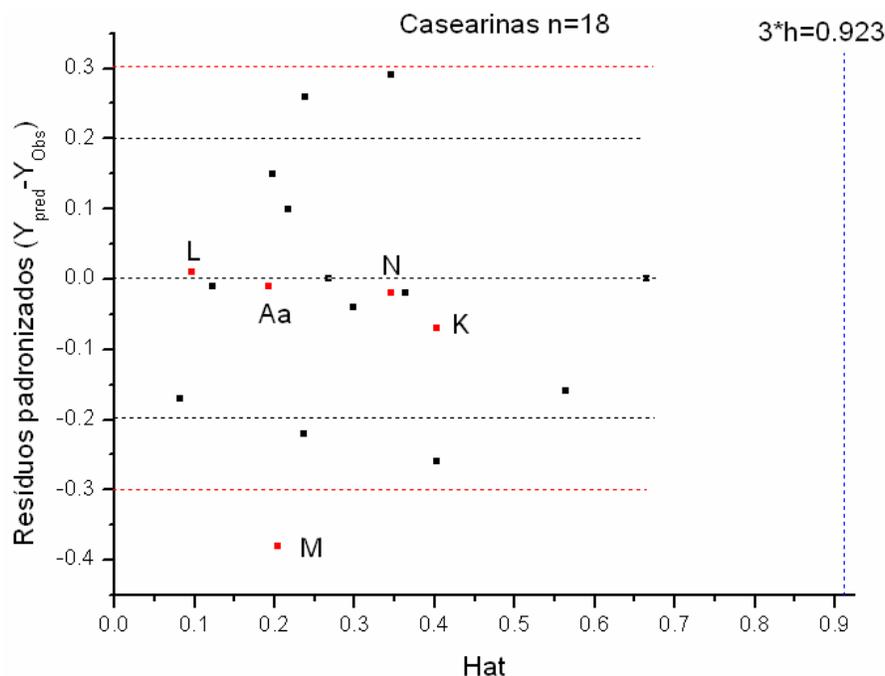
**Tabela 2.** Comparativo de predição de modelos de QSAR-MLR e QSAR-SVM.

Casearina	pIC <sub>50</sub> Obs	Y=pIC <sub>50</sub> -calculado		
		Y <sub>-MLR</sub> <sup>a</sup>	Y <sub>-SVM</sub> <sup>b</sup>	res <sup>c</sup>
A	6,00	5,86	5,88	-0,02
Aa	6,26	6,25	6,29	-0,04
B	5,07	5,07	5,11	-0,04
C	6,11	6,08	5,98	0,11
D	5,74	5,78	5,78	0,00
E	5,33	5,41	5,50	-0,08
G	6,77	6,60	6,60	0,00
H	6,41	6,40	6,47	-0,07
I	6,29	6,50	6,46	0,03
J	5,96	6,11	6,04	0,07
K	6,28	6,22	6,11	0,11
L	5,80	5,84	5,88	-0,04
M	5,75	5,42	5,50	-0,08
N	5,23	5,23	5,27	-0,04
O	5,22	5,33	5,37	-0,04
P	5,11	5,30	5,28	0,02
Q	5,37	5,22	5,24	-0,02
R	5,27	5,33	5,44	-0,10

(a)  $R^2 = 0,932$ ; MSE = 0,0171; (b)  $R^2 = 0,931$ , MSE = 0,0181; (c)  $Y_{MLR} - Y_{SVM}$ .

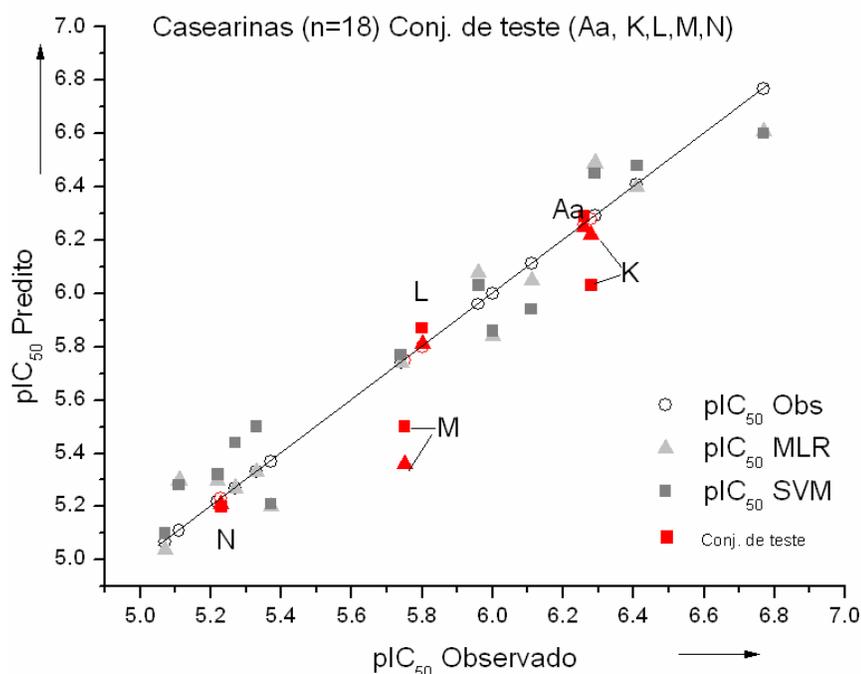
### 5.1.3.2 – QSAR – *Outliers* e Domínio de Aplicabilidade

Partindo-se dos resultados obtidos por MLR aplicando a Eq. (5.1-6), construiu-se o diagrama de Williams através do programa MobyDigs<sup>53</sup> para avaliar o domínio de aplicabilidade do conjunto de treinamento e de teste.



**Figura 11.** Diagrama de Williams – Domínio de aplicabilidade das Casearinas. Hat crítico  $3h = 0,923$ . No conjunto de predição externa obtida pela Eq. (5.1-6) a Casearina M foi classificada como *outlier*.

No diagrama de Williams (Figura 11) estão representadas as predições do conjunto de treinamento e do conjunto de teste (Casearinas Aa, K, L, M, N) onde a Casearina M foi classificada como *outlier* (maior que três desvios padrões,  $> 3\sigma$ ) empregando o método MLR, o que pode ter sido causado pelo grande resíduo que apresenta, razão pela qual o composto não foi eliminado do conjunto. Como será visto mais adiante, esta anomalia desaparecerá no estudo que emprega análise conformacional (v. sub-seção 5.1.3.6, Figura 17). Empregando SVM a Casearina M está contida no diagrama de Williams dentro do intervalo  $\pm 3\sigma$  (não mostrado). O gráfico da Figura 12, e Tabela 3 (v. em Anexos, pg. 219) mostra os resultados de predição interna ( $n = 13$ ) e externa ( $n = 5$ ) obtidos pelos modelos de QSAR com os métodos MLR e SVM. Os valores de predição interna serão bastante melhorados por ambos os modelos de QSAR-MLR e QSAR-SVM após a análise conformacional, como está demonstrado pela proposta deste estudo na Tabela 9 na sub-seção 5.1.3.5.

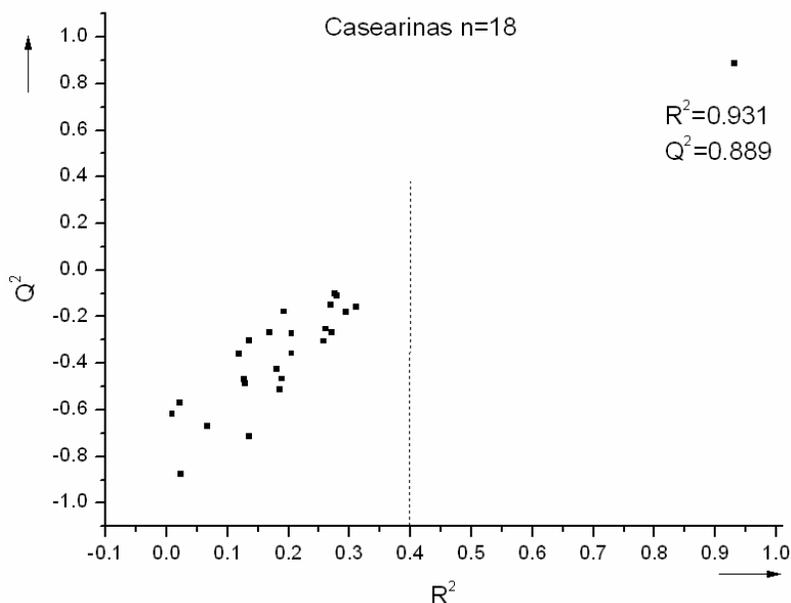


**Figura 12.** Valores pIC<sub>50</sub> observados e preditos pelos modelos de QSAR-MLR e de QSAR-SVM contra células tumorais V79.

### 5.1.3.3 – QSAR – Testes de Robustez e Validação

O conjunto de compostos (n = 18) foi avaliado na sua robustez empregando os métodos *Y-randomization*<sup>53,61</sup> e *Leave-N-out*.<sup>61</sup> O teste de *Y-randomization* foi realizado com 300 iterações (*default* do programa MobyDigs) e 10 ciclos, resultando em 3000 iterações. Os seguintes valores do *intercept* são usados como referência para aprovação [ $a(R^2) < 0,3$  e  $a(Q^2) < 0,05$ ]. Os valores médios obtidos nos dez ciclos estão abaixo destes limites e aprovam o modelo. Segue o valor final do último ciclo:  $a(R^2) = 0,095$  e  $a(Q^2) = -0,554$ .

Utilizando o critério de Eriksson e Wold<sup>58</sup>, que consideram até 25 iterações como suficientes para a determinação dos  $R^2$  e  $Q^2$ , e empregando o programa da ref. [61], obtém-se o gráfico da Figura 13.



**Figura 13.** Gráfico dos valores  $R^2$  e  $Q^2$  em função das randomizações da bioatividade experimental (*Y-randomization*)

A randomização produziu valor médio  $R^2 < 0,4$ . Considerando o critério anterior adicionado ao resultado presentemente obtido, pode-se considerar a Eq. (5.1-1) consistente, gerando assim um modelo de QSAR aceitável neste critério.

A validação interna do conjunto é representada pelos resultados do método *bootstrap*, em que foram realizadas 5000 iterações, resultando um valor final de  $Q^2_{boot} = 0,876$ , que pode ser considerado próximo ao valor  $Q^2_{LOO} = 0,889$  ( $Q^2_{LOO} - Q^2_{boot} = 0,013 < 0,1$ ).

O conjunto de compostos ( $n = 13$ ) foi igualmente avaliado seguindo os mesmos critérios anteriores e os valores que foram obtidos estão abaixo dos limites fixados, o que aprova o modelo do QSAR Eq.(4.1-2):  $a(R^2) = 0,161$  e  $a(Q^2) = -0,900$ .

O critério *Leave-N-Out* foi empregado eliminando até 4 constituintes (30% das amostras) do conjunto obtendo um resultado  $Q^2 - Q^2_{LNO} < 0,1$  para 10 repetições.

Para a validação externa foi adotado o critério proposto por Golbraikh e Tropsha.<sup>64</sup> Respectivamente aos métodos MLR e SVM, os parâmetros de validação estão relacionados na Tabela 4.

**Tabela 4.** Validação do conjunto de teste n = 5 (Casearinas Aa, K, L, M, N)

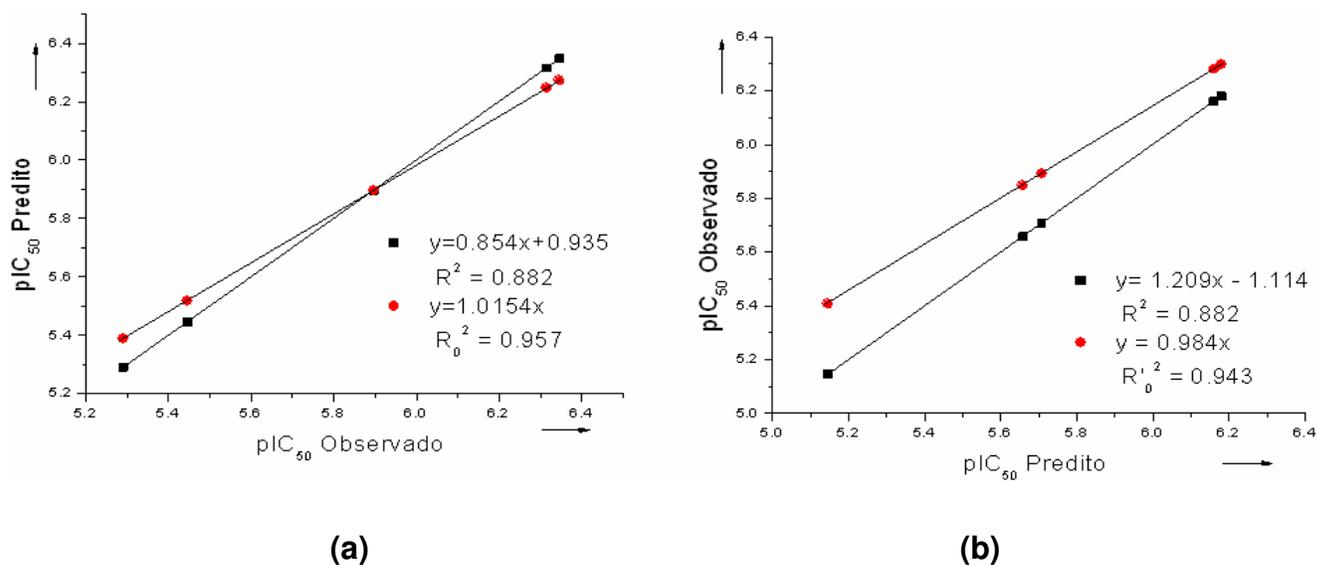
Metodo	$R^2_{\text{test}}$	$R_0^2$	$R'_0{}^2$	$d^a < 0,3$	$0,85 \leq k \leq 0,15$	$0,85 \leq k' \leq 0,15$	MSE
MLR	0,882	0,957	0,943	0,014	1,015	0,984	0,030
SVM	0,879	0,954	0,951	0,003	1,014	0,985	0,026

$$(a) d = [R^2 - R_0^2] / R^2 \text{ ou } d = [R^2 - R'_0{}^2] / R^2$$

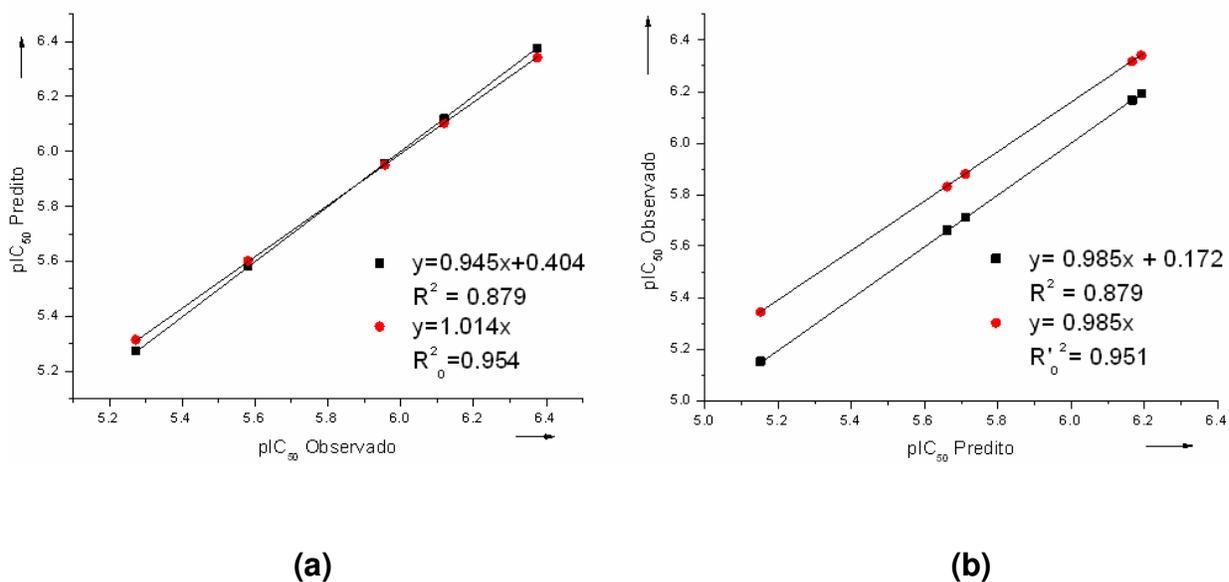
Ambos os modelos passaram nos testes, cujos parâmetros utilizados para aprovação foram apresentados em Material e Métodos, Seção 4. No caso destaca-se o erro médio quadrático (MSE), que em ambos os métodos tem valores próximos (0,03), o que atende ao objetivo de se ter modelos de QSAR que produzam valores de predição aproximadamente equivalentes, ainda que o modelo de QSAR obtido por SVM tenha parâmetros numéricos ligeiramente melhores que o QSAR obtido por MLR.

A seguir encontram-se os gráficos com os resultados obtidos pelo método de validação Golbraikh-Tropsha<sup>64</sup> para os compostos do conjunto de teste (Aa, K, L, M, N) cuja predição se deu pelo método MLR (Figura 14) e SVM (Figura 15).

Pelas regras usadas nas avaliações descritas em Material e Métodos, ambos os modelos de QSAR estão aprovados. Seguem-se os resultados numéricos ou gráficos e comentários correspondentes.



**Figura 14.** Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR das Casearinas pelas equações (a)  $y = ax+b$ ,  $y = kx$ ,  $R_0^2$ , e (b)  $y = a'x+b'$ ,  $y = k'x$ ,  $R'_0{}^2$ .



**Figura 15.** Representação gráfica dos parâmetros de validação do mdelo de QSAR-SVM das Casearinas pelas equações (a)  $y = ax+b$ ,  $y = kx$ ,  $R_0^2$ , e (b)  $y = a'x+b'$ ,  $y = k'x$ ,  $R'_0{}^2$ .

Como visto na seção 4 (Validação externa) não ocorre a proximidade recomendada dos valores  $R^2$ ,  $R_0^2$  e  $R'_0^2$  e  $k$ ,  $k'$ . Ainda assim, os modelos QSAR-MLR e SVM atendem aos critérios,  $0,85 \leq k \leq 1,15$  ou  $0,85 \leq k' \leq 1,15$ , e  $[(R^2 - R_0^2)/R^2] < 0,3$  ou  $[(R^2 - R_0'^2)/R^2] < 0,3$ .

Os coeficientes  $a$  e  $b$  (Figuras 14a, 15a),  $a'$  e  $b'$  (Figuras 14b, 15b), respectivamente nas equações  $y = ax+b$  e  $y = a'x+b'$ , têm os melhores valores no QSAR-SVM. Na Figuras 14a e 15a, ' $x$ ' assume o valor de  $\text{pIC}_{50}$  predito e nas Figuras 14b e 15b o valor de  $\text{pIC}_{50}$  observado.

Como será visto mais adiante (sub-seção 5.1.3.5), quando se aplica seleção conformacional na construção dos modelos de QSAR, ver-se-á que os parâmetros de validação assumirão valores bem mais favoráveis no QSAR-MLR, ou seja,  $R^2$ , próximo a  $R_0^2$  ou a  $R_0'^2$  e próximos a 1, assim como o coeficiente  $a$  mais próximo a 1 e  $b$  mais próximo a zero.

A validação dos modelos de QSAR-MLR das Equações (5.1-1) e (5.1-6) é expressa, respectivamente, através das funções RQK de ajuste (RQK *Fitness functions*) do programa MobyDigs<sup>53,54</sup> (Tabela 5 e 6).

**Tabela 5.** Validação do modelo de QSAR-MLR (n = 18) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis	$K_{xy}$	$K_x$	$K_{xy}-K_x$	DK
$K_{xy} - K_x > DK$	0,375	0,266	0,109	0,000
2.Capacidade de predição				
$Q^2_{\text{LOO}} > Q^0$ (Referência $Q^0 > 0,5$ )	0,889	>0,5		
3.Capacidade de predição do modelo				
$Q^2_{\text{LOO}} - Q^2_{\text{ASYM}} > DQ$	$Q^2_{\text{LOO}}$	$Q^2_{\text{ASYM}}$	$Q^2_{\text{LOO}} - Q^2_{\text{ASYM}}$	DQ
	0,889	0,811	0,078	-0,010
4.Redundância nas variáveis	$R^P$	$t^P$		
$R^P > t^P$	0,566	0,1		
5.Sobre-ajuste (overfitting)	$R^N$	$t^N$		
$R^N > t^N$	0	-0,323		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

**Tabela 6.** Validação do modelo de QSAR-MLR (n = 13) pelos parâmetros RQK

Critérios de aprovação (*)				
1. Colinearidade entre variáveis	$K_{xy}$	$K_x$	$K_{xy}-K_x$	DK
$K_{xy} - K_x > DK$	0,369	0,266	0,103	0,000
2. Capacidade de predição				
$Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,904	> 0,5		
3. Capacidade de predição do modelo				
$Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$	$Q^2_{ASYM}$	$Q^2_{LOO}-Q^2_{ASYM}$	DQ
	0,904	0,842	0,062	-0,010
4. Redundância nas variáveis	$R^P$	$t^P$		
$R^P > t^P$	0,558	0,1		
5. Sobre-ajuste (overfitting)	$R^N$	$t^N$		
$R^N > t^N$	0	-0,323		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

Uma vez que as etapas de robustez e validação dos modelos de QSAR foram atendidas produzindo resultados satisfatórios, o próximo tópico está relacionado à aplicação dos modelos de QSAR obtidos pelas metodologias MLR e SVM na predição de novos compostos.

#### 5.1.3.4 – Proposta de Diterpenóides Clerodanos (conforme Eq. 5.1-1)

Os modelos de QSAR obtidos por MLR e SVM foram aplicados a 113 diterpenóides clerodanos extraídos da literatura referenciada no artigo cujos valores calculados  $Y = pIC_{50}$  e  $IC_{50}$  estão relacionados na Tabela 7 (v. em Anexos, pg. 220). Foram escolhidos como referência principal os resultados obtidos pelo QSAR-SVM, que apresentou resultados globais ligeiramente superiores em relação ao QSAR-MLR. Os resultados estão dispostos em ordem decrescente de  $pIC_{50}$  (coluna tarjada de cinza).

As Casearborinas A e B são as melhores classificadas, com  $pIC_{50} = 7,07$  (referência experimental: Casearina G,  $pIC_{50} = 6,77$ ). Além destas, é interessante citar os grupos das Casearborinas A-E e a Casearina S. Estas substâncias são

encontradas, respectivamente, na *Casearia arborea* (L.C.Rich.) Urban Flacourtiaceae/Salicaceae, e *Casearia sylvestris* Swartz Flacourtiaceae/Salicaceae, árvores encontradas em diversas regiões do Brasil.

Esta condição proporciona a possibilidade de continuidade de estudos de natureza experimental no Brasil, o que já não ocorre com alguns dos demais diterpenóides classificados entre os 20 mais bioativos, que até o presente, só foram encontrados em plantas da espécie *Casearia* no continente asiático (*Casearia grewiifolia* Vent. Flacourtiaceae/Salicaceae), Austrália (*Casearia grayi* L. W. Jessup Flacourtiaceae/Salicaceae), Suriname e Madagascar (*Casearia lucida* Hils. & Bojer ex Tul. Flacourtiaceae/Salicaceae),

Como mencionado anteriormente a Casearina G foi modificada *in silico* introduzindo flúor e cloro como substituintes nas posições R<sup>1</sup> (carbono C2), R<sup>4</sup> (carbono C6 experimentalmente identificado como uma posição sensível às alterações na bioatividade) e R<sup>5</sup> (carbono C7) (v. Figura 9). A Tabela 4 do Apêndice 1 reporta os valores obtidos preditos pelos modelos de QSAR MLR e SVM. Os resultados do modelo de QSAR-SVM estão tarjados em cinza em ordem decrescente dos valores pIC<sub>50</sub>.

Verifica-se que existem oito compostos derivados da Casearina G, especificamente Gv5, Gv11, Gv13, Gv12, Gv9, Gv3, Gv7 e Gv1, com predição de valores iguais ou maiores a ela, causados pelos substituintes nas posições comentadas. Dá-se especial destaque às substituições de flúor em R<sup>1</sup> e R<sup>5</sup>, e cloro em R<sup>1</sup> que geraram os maiores valores. Flúor em R<sup>4</sup> (carbono C6) igualmente causou um aumento da bioatividade predita, porém, o átomo de cloro nesta posição provocou uma redução da bioatividade. Este tópico será revisto e reavaliado com modelos de QSAR obtidos com análise conformacional na sub-seção 5.1.3.5, onde, devido a maior acurácia dos modelos de QSAR-MLR e SVM, o átomo de cloro passa a ter maior importância entre os substituintes.

Não foram encontrados na literatura diterpenóides sintetizados com a incorporação de átomos de flúor ou cloro e que tenham sido ensaiados testando a sua citotoxicidade. A Tabela 4 do Apêndice 1 reúne assim alguns exemplos que ficam como primeira proposta, em analogia a produtos naturais como a

Camptotecina fluorada citada anteriormente e em linha com os objetivos deste estudo, na proposta de novos agentes anticancerígenos.

### 5.1.3.5 – QSAR e Descritores – Estudo com Análise Conformacional

Nesta etapa do estudo, evidencia-se que compostos possuidores de resíduos altos mostram tendência de serem classificados como *outliers* (Casearina M no caso anterior). Com a aplicação da análise conformacional um problema desta natureza tem grande chance de ser superado. O segundo aspecto da análise conformacional está fortemente ligado a um aumento na acurácia da preditividade interna do modelo de QSAR, aproximando o valor predito do valor observado no bioensaio.

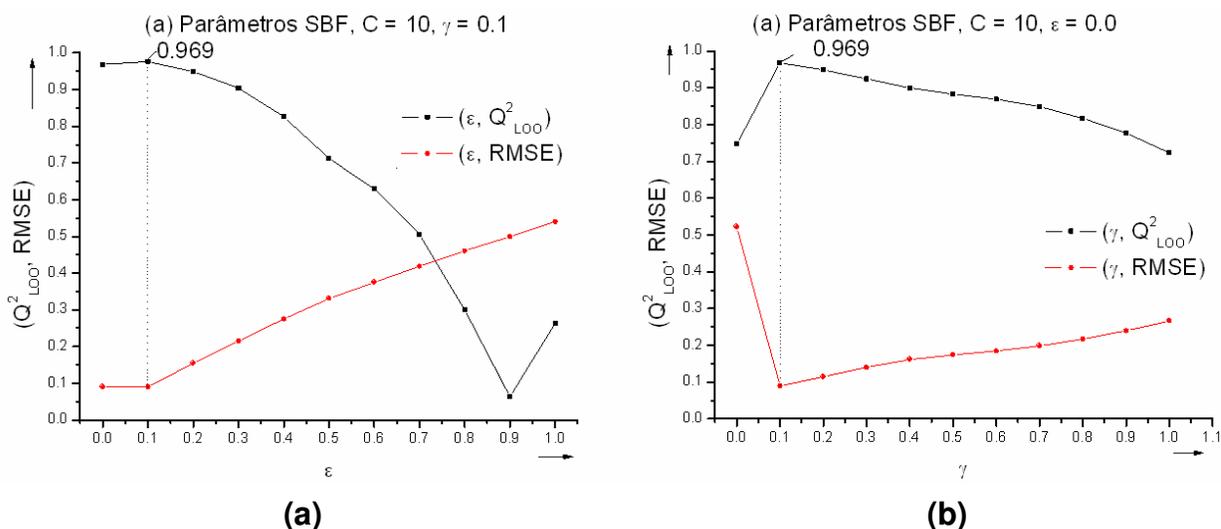
O estudo se utilizou dos mesmos descritores. Os detalhes de como o novo modelo QSAR é obtido já foram mencionados em Material e Métodos.

A nova equação QSAR-MLR obtida para  $n = 18$  está representada a seguir:

$$\begin{aligned} \text{pIC}_{50} = & + 0,088 (\pm 0,007) \text{RDF060m} - 7,202 (\pm 0,520) \text{R8p} - 0,047 (\pm 0,004) \text{Fuk5} + \\ & 13,335 (\pm 0,613) \\ n = 18 \quad R^2 = & 0,991 \quad s = 0,054 \quad F = 502 \quad p < 0,0001 \quad Q^2 = 0,985 \quad \text{SPress} = 0,070 \quad \text{SDEP} \\ = & 0,064 \quad \text{MSE} = 0,0023 \end{aligned} \quad (5.1-7)$$

O modelo de QSAR construído com SVM ( $n = 18$ ) manteve a utilização do *kernel Radial Basis Function* (RBF). Um novo ajuste fino é requerido para se obter o menor valor RMSE (*Root Mean Square Error*), relacionado aos melhores parâmetros de regressão empregando LOO (*Leave-One-Out* ( $Q^2_{\text{LOO}}$ )). O melhor valor obtido foi com o ajuste dos seguintes parâmetros: custo (ou capacidade)  $C = 10$ ; extensão da função radial RBF,  $\gamma = 0,1$  e a função de perda  $\epsilon = 0,0$ .

O gráfico da Figura 16 mostra a variação do RMSE em função dos demais parâmetros  $\epsilon$  e  $\gamma$ , e que define o melhor  $Q^2_{\text{LOO-CV}} = 0,969$ .



**Figura 16.** (a)  $Q^2_{\text{LOO-CV}}$  em função da variação do parâmetro  $\varepsilon$  mantendo constantes  $C$  e  $\gamma$ ; (b)  $Q^2_{\text{LOO-CV}}$  em função da variação do parâmetro  $\gamma$  mantendo constantes  $C$  e  $\varepsilon$ . Para o menor RMSE seleciona-se o valor  $Q^2_{\text{LOO-CV}}$  correspondente a 0,969.

Comparativamente ao valor  $Q^2_{\text{LOO-CV}} = 0,824$  obtido para o modelo de QSAR-SVM anterior (vide Figura 10), nota-se uma forte influência causada pela análise conformacional no novo valor  $Q^2_{\text{LOO-CV}} = 0,969$  e menor RMSE.

Obtidas as equações de QSAR-MLR e QSAR-SVM, foram gerados os valores de predição interna listados na Tabela 9. Usando como referência a diferença de predição entre modelos de QSAR do modelo anterior à análise conformacional, verifica-se que houve manutenção na equivalência (dada pela proximidade dos valores MSE) e aumento na acurácia de predição entre ambos os modelos de QSAR, com MSE igual a 0,0023 (QSAR-MLR) e 0,0021 (QSAR-SVM).

**Tabela 9.** Comparativo de predição dos métodos MLR e SVM aplicando análise conformacional

Casearina	pIC <sub>50</sub>	Y=pIC <sub>50</sub> -		res <sup>c</sup>	res <sup>c,d</sup>
	Observado	calculado			
		Y-MLR <sup>a</sup>	Y-SVM <sup>b</sup>		
A	6,00	6,04	6,04	0,01	-0,02
Aa	6,26	6,31	6,30	0,02	-0,04
B	5,07	5,13	5,12	0,00	-0,04
C	6,11	6,07	6,11	-0,04	0,11
D	5,74	5,77	5,74	0,03	0
E	5,33	5,31	5,33	-0,02	-0,08
G	6,77	6,69	6,68	0,01	0
H	6,41	6,45	6,41	0,04	-0,07
I	6,29	6,23	6,21	0,02	0,03
J	5,96	6,00	5,99	0,01	0,07
K	6,28	6,29	6,28	0,01	0,11
L	5,80	5,85	5,80	0,05	-0,04
M	5,75	5,72	5,71	0,01	-0,08
N	5,23	5,29	5,35	-0,06	-0,04
O	5,22	5,18	5,22	-0,04	-0,04
P	5,11	5,09	5,11	-0,02	0,02
Q	5,37	5,28	5,29	-0,01	-0,02
R	5,27	5,28	5,30	-0,02	-0,10

(a)  $R^2 = 0,991$ ; MSE = 0,0023; (b)  $R^2 = 0,993$ , MSE = 0,0021; (c)  $Y_{MLR} - Y_{SVM}$ ; (d) resíduo antes da análise conformacional.

A equação correspondente ao conjunto de treinamento  $n = 13$  utilizada para a predição do conjunto de teste é dada a seguir:

$$pIC_{50} = + 0,089 (\pm 0,009) RDF060m - 7,156 (\pm 0,628) R8p - 0,047 (\pm 0,005) Fuk5 + 13,337 (\pm 0,790)$$

$$n = 13 \quad R^2 = 0,992 \quad s = 0,057 \quad F = 376 \quad p < 0,0001 \quad Q^2 = 0,983 \quad SPress = 0,084$$

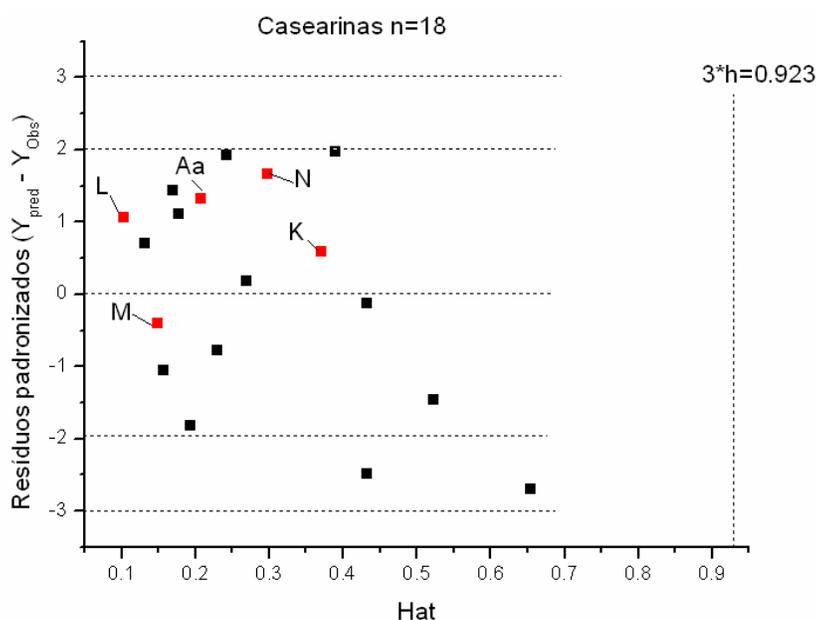
$$SDEP = 0,073 \quad (5.1-8)$$

Verifica-se nesta equação uma sensível melhoria dos parâmetros estatísticos em relação à Eq.(5.1-6) obtida sem análise conformacional.

Os testes de robustez e validação dos modelos de QSAR-MLR e QSAR-SVM sobre o conjunto externo serão discutidos na sub-seção 5.1.3.7.

### 5.1.3.6 – QSAR – *Outliers* e Domínio de Aplicabilidade

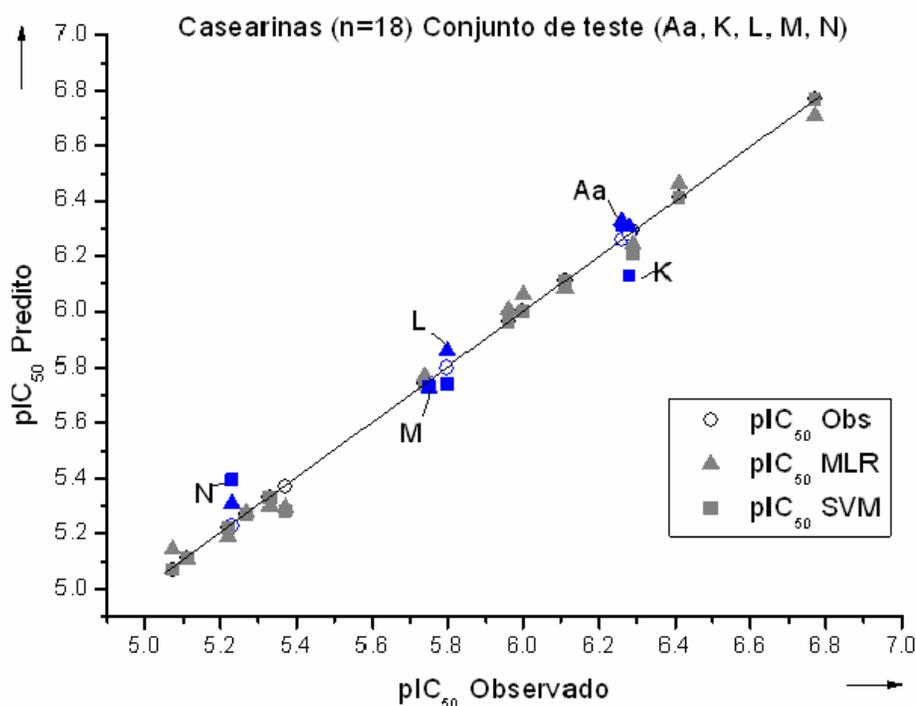
O diagrama de Williams relativo aos resultados obtidos pelo modelo de QSAR-MLR da Eq. (5.1-7) estão reproduzidos na Figura 17. Estão representados os resíduos padronizados no intervalo  $\pm 3\sigma$  das predições das 18 Casearinas do conjunto de treinamento e de teste. Como pode ser verificado em relação à Figura 11 (pg.39), a Casearina M que era classificada como *outlier*, após empregar um confôrmero que melhor representa a bioatividade experimental, passa a estar contida no intervalo  $\pm 3\sigma$ .



**Figura 17.** Diagrama de Williams – Domínio de aplicabilidade das Casearinas após análise conformacional. Hat crítico  $3h = 0,923$ .

A Figura 18 mostra os resultados obtidos na predição interna e externa aplicando os modelos de QSAR obtidos pelo método MLR e SVM após análise

conformacional. Verifica-se que, de forma geral, os modelos de QSAR tornaram-se mais acurados tanto na predição interna como na predição externa (vide Tabela 4 do Apêndice 1 e Tabela 8 em Anexos, pg 223). Ambas as metodologias produziram modelos de QSAR com valores preditos similares.

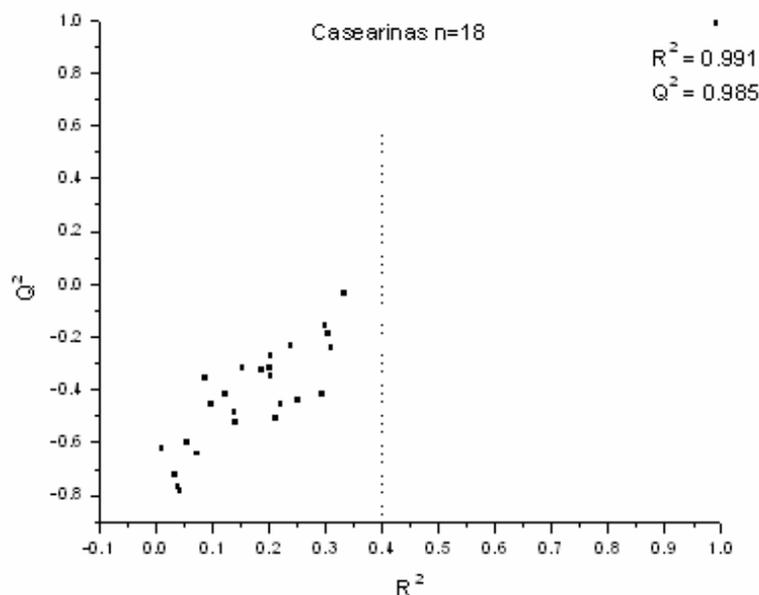


**Figura 18.** Valores pIC<sub>50</sub> observados e preditos pelos modelos de QSAR-MLR e QSAR- SVM contra células tumorais V79 após emprego de análise conformacional para alguns compostos.

### 5.1.3.7 – QSAR – Testes de Robustez e Validação

Repetiram-se os procedimentos *Y-randomization*<sup>53,61</sup> e *Leave-N-out*<sup>61</sup> assim como o critério de Eriksson-Wold<sup>58,61</sup>, avaliando a robustez do modelo de QSAR. O teste de *Y-randomization* foi realizado com 300 iterações (*default* do programa MobyDigs) e 10 ciclos, resultando em 3000 iterações. Os seguintes valores do *intercept* são usados como referência para aprovação [ $\alpha(R^2) < 0,3$  e  $\alpha(Q^2) < 0,05$ ]. Os valores médios obtidos nos dez ciclos estão abaixo destes limites e aprovam o modelo. Segue-se o valor do último ciclo:  $\alpha(R^2) = 0,096$  e  $\alpha(Q^2) = -0,991$ . O critério

de Eriksson-Wold<sup>58</sup> aplicado com 25 iterações aprova o modelo QSAR, ilustrado na Figura 19.



**Figura 19.** Gráfico dos valores  $R^2$  e  $Q^2$  em função das randomizações da bioatividade experimental (*Y-randomization*) do modelo de QSAR-MLR construído com a seleção de confôrmeros.

Pelo critério, se a randomização produziu valor médio  $R^2_{Yrand} \leq 0,4$  pode-se considerar o resultado representado na Figura 19 como aceitável e o modelo de QSAR representado pela Eq. (5.1-7) suficientemente consistente neste critério.

O conjunto de compostos ( $n = 13$ ) foi igualmente avaliado seguindo os mesmos critérios anteriores e os valores que foram obtidos estão abaixo dos limites fixados, o que aprova o modelo de QSAR Eq. (5.1-8). Seguem-se os valores do último ciclo:  $a(R^2) = 0,096$  e  $a(Q^2) = -0,991$ .

O teste *Leave-One-Out* (LNO) foi igualmente repetido com exclusão de 4 amostras (30%) resultando  $Q^2 - Q^2_{LNO} < 0,1$  com 10 repetições.

A validação interna do conjunto é representada pelos resultados do método *bootstrap*, em que foram realizados 5000 ciclos, resultando um valor final de  $Q^2_{boot} = 0,984$ , que pode ser considerado próximo ao valor  $Q^2_{LOO} = 0,985$  ( $Q^2_{LOO} - Q^2_{boot} = 0,010 < 0,1$ ).

Para a validação externa repetiu-se a aplicação do critério Golbraikh-Tropsha<sup>64</sup> ao conjunto de teste (Aa, K, L, M, N) com valores preditos pelos modelos de QSAR- MLR e QSAR-SVM. Os parâmetros de validação estão relacionados na Tabela 10.

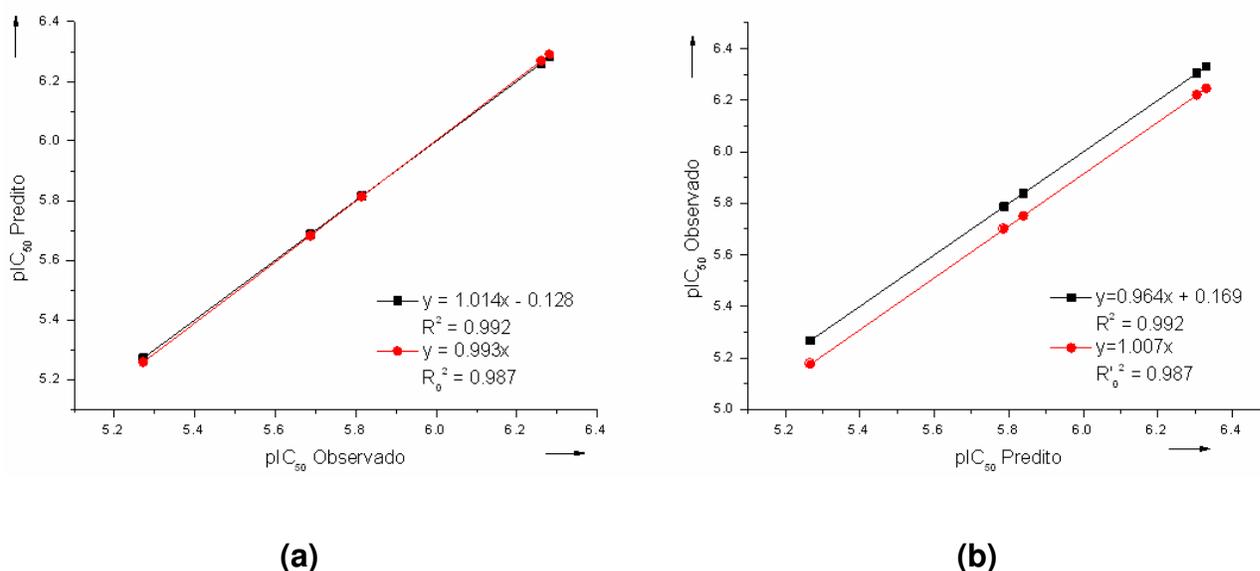
**Tabela 10.** Validação do conjunto de teste n = 5 (Casearinas Aa, K, L, M, N)

Método	$R^2_{test}$	$R_0^2$	$R'_0^2$	$d^a < 0,3$	$0,85 \leq k \leq 0,15$	$0,85 \leq k' \leq 0,15$	MSE
MLR	0,992	0,987	0,988	0,0006	0,993	1,007	0,0023
SVM	0,944	0,999	0,999	0,0001	1,000	0,999	0,0014

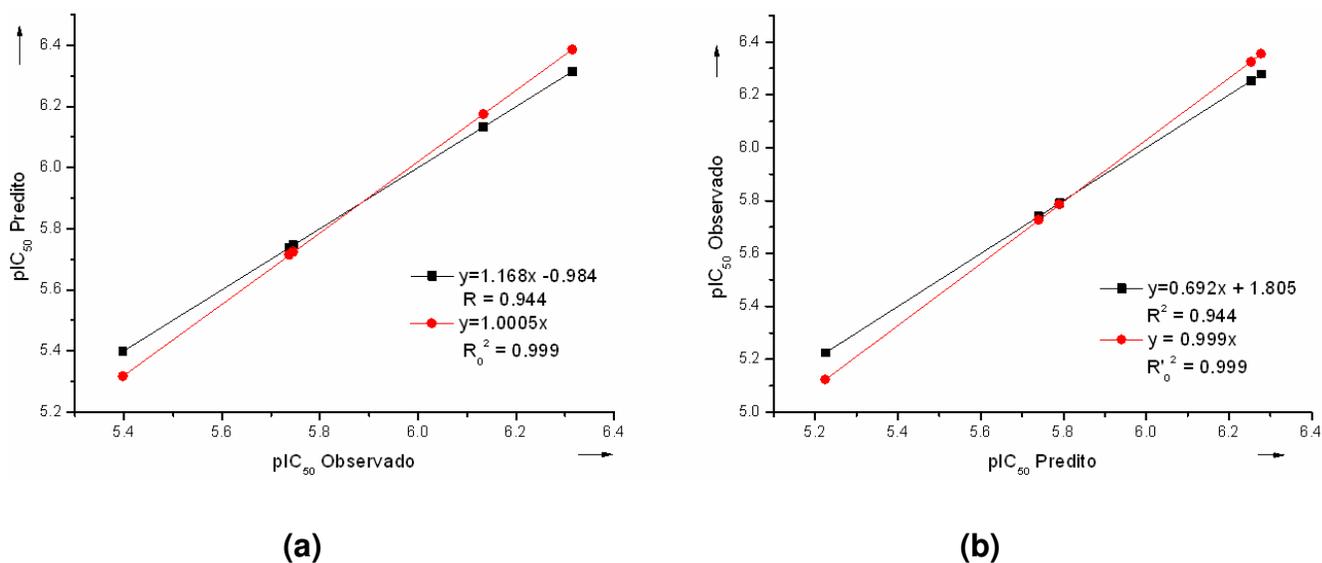
(a)  $d = [R^2 - R_0^2] / R^2$  ou  $d = [R^2 - R_0'^2] / R^2$

Neste caso os valores MSE indicados na Tabela 10 são bem menores do que no caso anterior (Tabela 4, sub-seção 5.1.3.3), e indicam que o modelo SVM gera resultados melhores (MSE menor) que o modelo MLR. Aqui, entretanto, faz-se necessária uma ressalva. Como o objetivo de se aplicar duas metodologias está centrado fundamentalmente em se obter modelos de QSAR equivalentes entre si, por razões de consistência na predição interna de valores próximos, sugere-se adotar os modelos de QSAR que geram predição utilizando todo o conjunto de compostos (n = 18), com MSE = 0,0023 para modelo de QSAR-MLR e MSE = 0,0021 para modelo de QSAR-SVM. O modelo de QSAR-MLR (n = 18) apresenta o mesmo valor MSE = 0,0023 para o modelo de QSAR-MLR (n = 13). Não foram encontrados valores de ajuste dos parâmetros SVM, respectivamente (C,  $\epsilon$ ) e (C,  $\gamma$ ) para n = 13 compostos, tornando possível uma equivalência dos modelos de QSAR em questão.

Encontram-se a seguir os gráficos decorrentes da aplicação do método de validação Golbraikh-Tropsha<sup>64</sup>, validando os compostos do conjunto de teste (Aa, K, L, M, N) cuja predição se deu pelo método MLR (Figura 20) e SVM (Figura 21).



**Figura 20.** Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR das Casearinas pelas equações (a)  $y = ax+b$ ,  $y = kx$ ,  $R_0^2$ , (b) e  $y = a'x+b'$ ,  $y = k'x$ ,  $R'_0^2$



**Figura 21.** Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM das Casearinas pelas equações (a)  $y = ax+b$ ,  $y = kx$ ,  $R_0^2$ , (b) e  $y = a'x+b'$ ,  $y = k'x$ ,  $R'_0^2$

Em relação à validação obtida anteriormente (Figuras 14 e 15) verifica-se substancial melhoria mostrada pela maior proximidade de sobreposição das retas, o que foi gerado por valores melhores dos parâmetros  $R^2$ ,  $R_0^2$ ,  $R'_0^2$ ,  $k$  e  $k'$  obtidos para

o conjunto de teste. Os coeficientes das retas, respectivamente,  $a$  e  $b$ ,  $a'$  e  $b'$  no modelo de QSAR-MLR se aproximam mais dos valores ótimos ( $a = 1$ ,  $b = 0$ ), porém são piores no caso do modelo de QSAR-SVM, compensados, contudo com a proximidade dos valores dos demais parâmetros e atendendo aos critérios de validação.

Seguem-se os parâmetros e valores do modelo de QSAR-MLR avaliado pelas funções RQK de ajuste (RQK *Fitness functions*)<sup>54</sup> do programa MobyDigs<sup>53</sup> que igualmente validam o modelo. Vide Tabelas 12 e 13.

**Tabela 12.** Validação do modelo de QSAR-MLR ( $n = 18$ ) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis	$K_{xy}$	$K_x$	$K_{xy}-K_x$	DK
$K_{xy} - K_x > DK$	0,346	0,223	0,123	0,005
2.Capacidade de predição				
$Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,985	> 0,5		
3.Capacidade de predição do modelo	$Q^2_{LOO}$	$Q^2_{ASYM}$	$Q^2_{LOO}-Q^2_{ASYM}$	DQ
$Q^2_{LOO} - Q^2_{ASYM} > DQ$	0,985	0,974	0,011	0,000
4.Redundância nas variáveis	$R^P$	$t^P$		
$R^P > t^P$	0,574	0,05		
5.Sobre-ajuste (overfitting)	$R^N$	$t^N$		
$R^N > t^N$	0	-0,233		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente.

**Tabela 13.** Validação do do modelo de QSAR-MLR (n = 13) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis	$K_{xy}$	$K_x$	$K_{xy}-K_x$	DK
$K_{xy} - K_x > DK$	0,360	0,253	0,107	0,005
2.Capacidade de predição				
$Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,983	> 0,5		
3.Capacidade de predição do modelo	$Q^2_{LOO}$	$Q^2_{ASYM}$	$Q^2_{LOO}-Q^2_{ASYM}$	DQ
$Q^2_{LOO} - Q^2_{ASYM} > DQ$	0,983	0,973	0,010	-0,001
4.Redundância nas variáveis	$R^P$	$t^P$		
$R^P > t^P$	0,563	0,05		
5.Sobre-ajuste (overfitting)	$R^N$	$t^N$		
$R^N > t^N$	0	-0,233		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente.

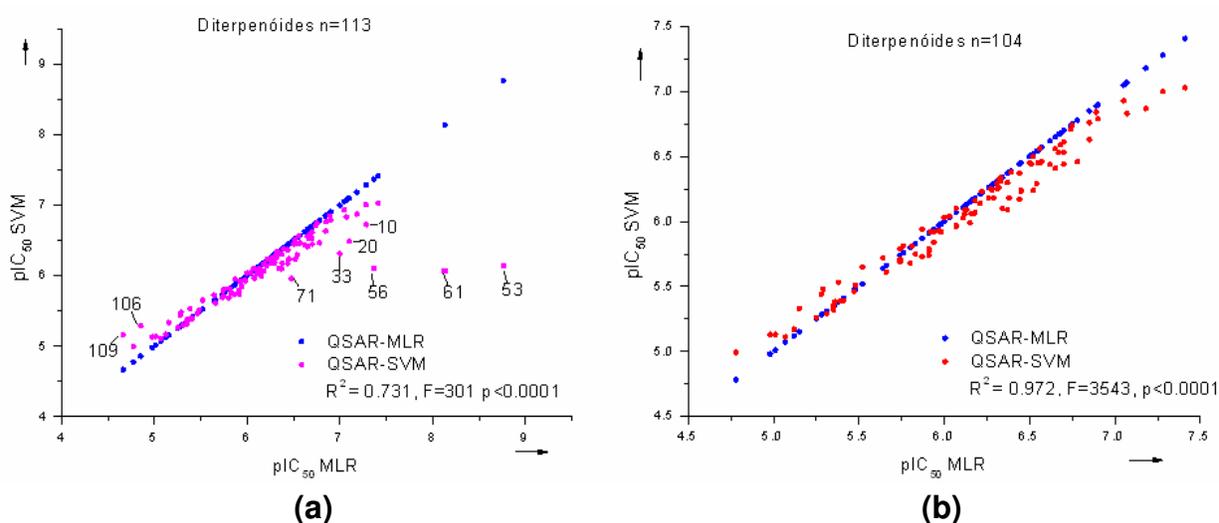
Atendidas as etapas de robustez e validação dos modelos de QSAR produzindo resultados satisfatórios, pode-se aplicar os modelos de QSAR obtidos pelas metodologias MLR e SVM à predição dos 113 compostos nesta nova condição.

### 5.1.3.8 – Proposta de Diterpenóides Clerodanos (conforme Eq. 5.1-7)

A predição dos 113 diterpenóides foi reavaliada com os modelos de QSAR-MLR (Eq. 5.1-7) e QSAR-SVM com 18 compostos, Tabela 11a,b,c, (v. em Anexos, pg. 224). Como já mencionado anteriormente, os modelos de QSAR MLR (n = 18 e n = 13) e o QSAR-SVM (n = 18) se equivalem nas predições internas. A aplicação dos modelos de QSAR revela valores próximos em alguns casos, e distantes em outros quando se comparam os valores  $IC_{50}$  preditos pelas duas metodologias. Em geral os valores  $IC_{50}$  preditos pelo modelo de QSAR-MLR são menores do que os valores preditos pelo modelo de QSAR-SVM.

Vários podem ser os motivos contribuintes para as diferenças: **(a)** A estrutura otimizada com o método semi-empírico AM1 não representa a melhor conformação bioativa. **(b)** Um ou mais descritores não descrevem bem todas as estruturas. **(c)** Diferenças introduzidas pelas peculiaridades dos próprios métodos, causadas por um dos motivos anteriores ou ambos.

A diferença de valores na predição entre o modelo de QSAR-MLR e QSAR-SVM foi usada para avaliar a correlação entre os modelos, representada nos gráficos da Figura 22. Nove compostos que apresentaram uma diferença grande de valores preditos entre os métodos, receberam o tratamento de *outliers*. Quando eliminados, obtém-se:  $n = 104$ ,  $R^2 = 0,972$ ,  $F = 3543$ ,  $p < 0,0001$ .

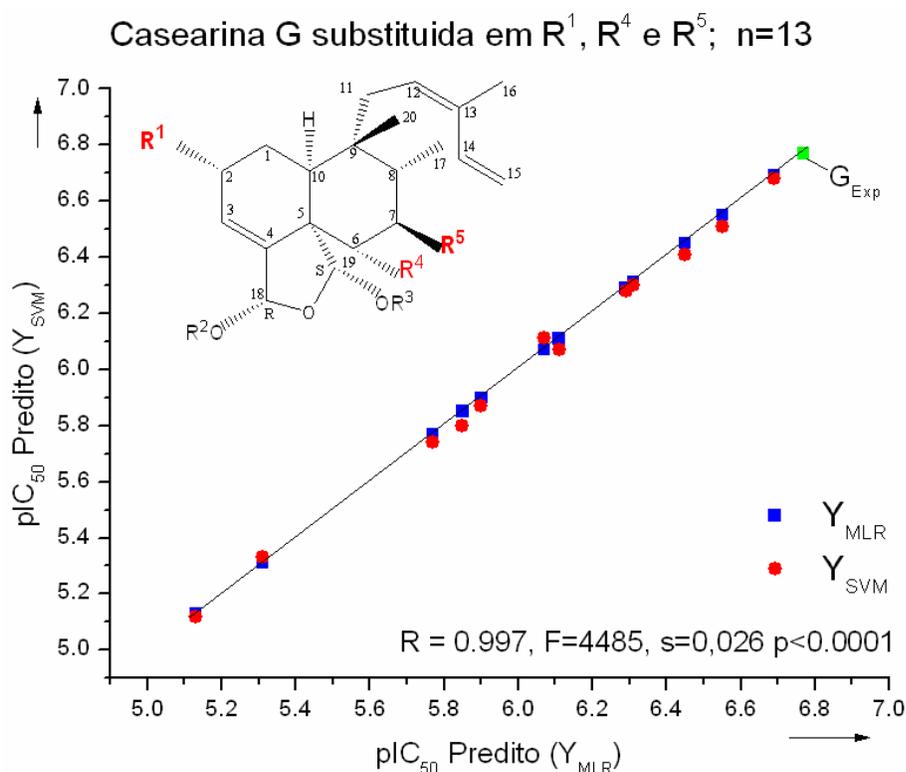


**Figura 22.** Correlação entre os modelos de QSAR-MLR e QSAR-SVM: **(a)** antes da eliminação dos compostos do item 10, 20, 33, 53, 56, 261, 71, 106, 109 da Tabela 11 (v. em Anexos p. 224) e **(b)** após eliminação dos 9 compostos citados.

A discrepância de valores na predição pode estar associada à hipótese da incapacidade de um dos métodos ou de ambos, em descreverem estes compostos com os descritores a eles associados classificando-os fora do Domínio de Aplicabilidade. A inclusão de um número maior de conformêros na análise conformacional dos compostos considerados como *outliers* pode, eventualmente, mudar o enfoque.

Como mencionado anteriormente, selecionaram-se alguns compostos para fazer uma análise *in silico* com os substituintes flúor e cloro nas posições R<sup>4</sup> ou R<sup>5</sup> (v. fig. 9), de forma análoga ao que foi proposto com a Casearina G (v. Tabela 8 em Anexos, pg. 223). Os resultados mostram que cloro na posição R<sup>4</sup> (composto Gv8) gera valores preditos de bioatividade próxima ao valor experimental da Casearina G. Notar que os resultados da Tabela 8 foram obtidos com modelos de QSAR MLR e SVM construídos com análise conformacional. Isto gerou um modelo de QSAR mais acurado, numa faixa de valores sem extrapolações, e com predições da Casearina mais próximas ao valor experimental comparativamente ao resultado mostrado na Tabela 4 contida no apêndice 1.

A correlação obtida entre modelos de QSAR, para o estudo da Casearina G substituída (Tabela 8, v. em Anexos, pg. 223) é dada no gráfico da Figura 23.



**Figura 23.** Correlação entre os modelos de QSAR-MLR e QSAR-SVM da Casearina G substituída (v. Figura 9). Gráfico construído com os valores numéricos da Tabela 8 (v. em Anexos, pg. 223).

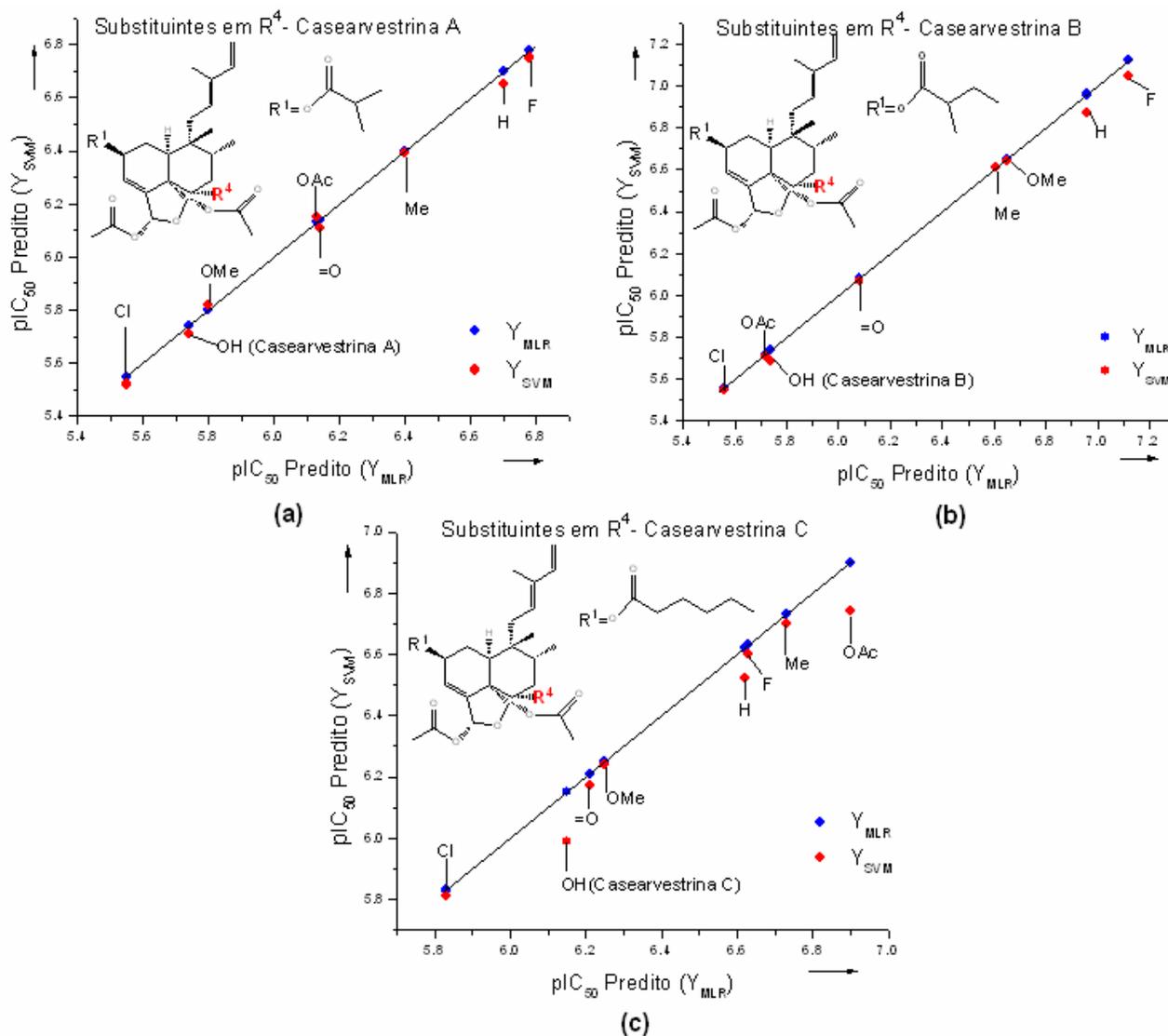
Foram também exploradas algumas alternativas de forma similar à Casearina G. Os compostos selecionados são a Casearborina A, Casearvestrinas A, B e C e Caseamembrina A, que são diterpenóides originários de órgãos de árvores encontradas no Brasil. São substâncias que podem ser encontradas e extraídas, respectivamente de *Casearia arborea* (L.C.Rich.) Urban e *Casearia sylvestris* Sw. A Caseamembrina-A originalmente obtida das folhas de *Casearia membranacea* Hance<sup>84</sup> (encontrada ao norte de Taiwan) foi também extraída de cascas de raízes de *Casearia sylvestris* Sw. var. *lingua*, coletadas no serrado próximo à Brasília, que nos bioensaios mostrou atividade contra *Trypanosoma cruzi*, agente causador da doença de Chagas.<sup>85</sup> Assim, foram desenvolvidos *in silico* 40 compostos derivados dos compostos de origem cuja ação antitumoral ou citotóxica em células V79 foi predita pelos modelos de QSAR propostos, e que serão discutidos a seguir.

Os 12 compostos derivados da Casearborina A, único caso em que foram feitas substituições em R<sup>4</sup> (carbono C-6) ou R<sup>5</sup> (carbono C-7), apresentaram valores pIC<sub>50</sub> preditos inferiores à própria Casearborina A, contudo, as substituições com flúor fazem com que sejam selecionados os derivados como candidatos à síntese tanto em R<sup>4</sup> como em R<sup>5</sup>. As substituições feitas com cloro classificam os compostos derivados com a menor bioatividade predita (Tabela 14, v. em Anexos, pg. 227). Correlação obtida entre modelos de QSAR-MLR e QSAR-SVM: n = 13; R = 0,973; F = 194; p < 0,0001.

Na Tabela 15 (v. em Anexos, pg. 228) encontram-se todos os derivados das Casearvestrinas A, B e C com substituições simuladas apenas em R<sup>4</sup> (Carbono C-6), que tiveram valores pIC<sub>50</sub> preditos melhores que os compostos de origem, destacando-se novamente o flúor como o substituinte de maior influência na bioatividade no caso das Casearvestrinas A e B. Na Casearvestrina C, o óxido acetílico foi o substituinte que influenciou no melhor valor. Exceto nos casos em que comparece cloro como substituinte, que apresentou o pior desempenho, todas as três Casearvestrinas tiveram seus valores de bioatividade predita abaixo dos demais substituintes. Correlação obtida entre modelos de QSAR-MLR e QSAR-SVM: n = 24; R = 0,995; F = 2425; p < 0,0001.

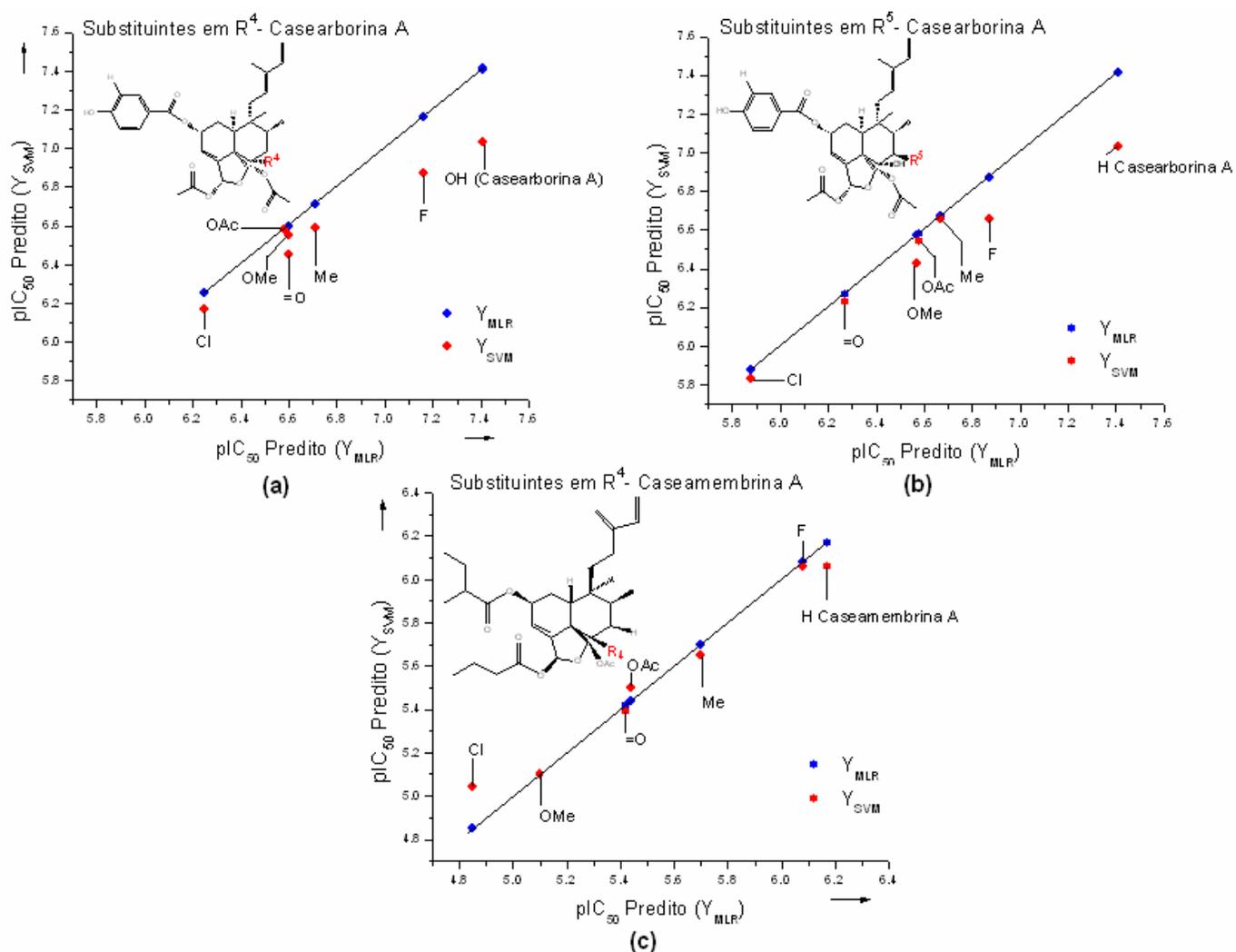
No estudo dos derivados da Caseamembrina A, as substituições em R<sup>4</sup> mostraram que o flúor teve na predição um valor semelhante ao pIC<sub>50</sub> da Caseamembrina A e o cloro apresentou o pior desempenho (Tabela 16, v. em Anexos, pg. 229). Correlação obtida entre QSAR-MLR e QSAR-SVM: n = 8; R = 0,991; F = 327; p < 0,0001.

Os gráficos das figuras a seguir expressam os resultados das simulações com os diterpenóides anteriormente comentados pela atividade antitumoral, onde aqui se chama a atenção para a equivalência de predição. A Figura 24 mostra o efeito do substituinte na posição R<sup>4</sup> em cada uma das Casearvestrinas A, B e C.



**Figura 24.** Efeito substituinte na posição  $R^4$  (Carbono C6) nas Casearvestrinas A,B, e C e equivalência de predição entre modelos de QSAR-MLR ( $Y_{MLR}$ ) e QSAR-SVM ( $Y_{SVM}$ ). (a) Os melhores resultados, considerando a proximidade de valores preditos  $pIC_{50}$  entre métodos, são obtidos por todos os substituintes da Casearvestrina A, (b) seguidos pelos da Casearvestrina B, e por fim (c) os substituintes OAc, H, e a própria Casearvestrina C geram os piores resultados na equivalência de predição.

A Figura 25 ilustra os resultados da Casearborina A substituída em R<sup>4</sup> e R<sup>5</sup>, e Caseamembrina A substituída apenas em R<sup>4</sup>.



**Figura 25.** Efeito substituinte nas posições R<sup>4</sup> (carbono C6) e R<sup>5</sup> (Carbono C7) na Casearborina A e equivalência de predição entre modelos de QSAR-MLR (Y<sub>MLR</sub>) e QSAR-SVM (Y<sub>SVM</sub>). **(a)** Os melhores resultados, considerando a proximidade de valores preditos pIC<sub>50</sub> entre métodos, são obtidos apenas pelos substituintes OAc e OMe, excluídos os demais e a própria Casearborina A. **(b)** Verifica-se que há boa equivalência de predição para Cl, =O, OAc, Me, excluindo OMe, F e a própria Casearborina A. **(c)** Os substituintes da Caseamembrina A não mostram equivalência de resultados apenas para Cl e a própria Caseamembrina A.

### 5.1.4 – Conclusão

O emprego de duas metodologias independentes, nominalmente, modelos de QSAR construídos com Regressão Linear Múltipla (QSAR-MLR) e com Máquinas de Vetores de Suporte (QSAR-SVM), expressam resultados satisfatoriamente próximos. Os resultados obtidos podem ser considerados como os primeiros indicadores para realização de novos bioensaios em células tumorais V79.

Mantendo os mesmos descritores, e aplicando análise conformacional, foi possível produzir modelos de QSAR com predição interna mais acurada, o que se reflete na predição externa. A influência maior ao se usar análise conformacional se deu na reclassificação entre alguns dos 113 diterpenóides, justificando o emprego desta metodologia quanto à acurácia que teoricamente produziu nos modelos de QSAR.

Apesar de não poder ser dada uma interpretabilidade do modelo de QSAR-MLR no sentido de propor uma hipótese de mecanismo de interação ligante-receptor, considerando inclusive que o receptor é desconhecido, os descritores utilizados capturam informações físico-químicas ligadas a conceitos conhecidos, como forma molecular e polarizabilidade (R8p), reatividade da molécula na região da superfície (Fuk5) e fornece um indicador dos átomos em posição favorável à interação com átomos de um receptor pela função de distribuição radial RDF060m.

Os valores preditos pelos modelos de QSAR que são extrapolados ao maior valor experimental (Casearina G:  $pIC_{50} = 6,77$ ;  $IC_{50} = 0,17 \mu\text{mol/L}$ ), não devem ser descartados, mas, considerados com parcimônia, não tanto pelo valor em si, mas porque alguns dos 113 compostos foram testados em bioensaios avaliando a sua atividade com diferentes protocolos, mostrando propriedades antitumorais, característica intrínseca a muitos diterpenóides.

Foi elaborada uma proposta de novos compostos desenvolvidos *in silico*, a partir de seis diterpenóides, a saber, Casearina-G, Casearborina-A, Casearvestrina A, B e C, e Caseamembrina A. Foram escolhidos para avaliar a tendência na variação dos valores  $pIC_{50}$  conforme a substituição principalmente em R<sup>4</sup>, constatando-se que, em nível teórico, a posição R<sup>4</sup> mostra ser um centro que pode influenciar no aumento ou redução da bioatividade, confirmando uma tendência,

constatada em ensaios experimentais.<sup>76</sup> O presente estudo *in silico* não esgota as possibilidades, e a análise pode ser estendida a qualquer um dos compostos da Tabela 11 (v. em Anexos, pg. 224) ou a novos diterpenóides clerodanos da literatura, uma vez que esta contém a identificação de aproximadamente 800 compostos extraídos e purificados com estruturas certificadas por análises espectroscópicas.<sup>16</sup>

A finalidade dos modelos de QSAR desenvolvidos está na predição de diterpenóides que foram extraídos de produtos naturais (no caso de órgãos de plantas). A literatura disponibiliza um banco de dados de diterpenóides que cobre um período de aproximadamente 20 anos, onde a metodologia proposta pode ser aplicada, assim como os modelos de QSAR que atenderam a todos os critérios de validação. Os resultados obtidos neste estudo podem ser usados como base de partida e como indicadores para novas avaliações tanto teóricas como experimentais.

## 5.2 – Dibenzolidrazinas – Inseticida contra *Chilo suppressalis* (Walker) Lepidoptera, Noctuidae

### 5.2.1 – Introdução

A lagarta *Chilo suppressalis*, pertencente à ordem Lepidoptera, conhecida na Europa e no Brasil por lagarta do arroz ou broca-do-colmo-do-arroz (*rice stem borer*), é conhecida pelas perdas causadas no cultivo deste produto agrícola.

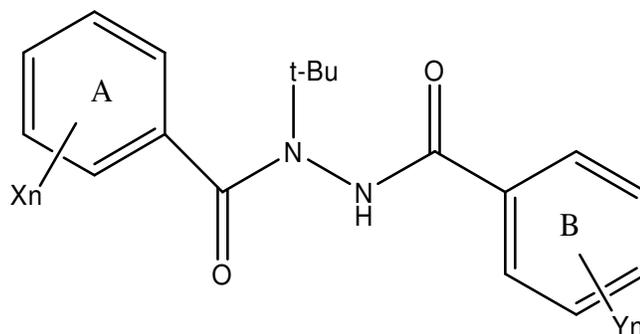
Esta pesquisa se utiliza dos dados experimentais obtidos por Oikawa *et al.*<sup>24</sup> que mediram a atividade larvívica de 1-*tert*-butil-1-(2-clorobenzoil)-2(benzoil substituído) idrazinas na presença do sinergista butóxido de piperonila, um inibidor de metabolismo oxidativo. A utilização de sinergistas é um recurso químico para determinar os mecanismos envolvidos na resistência de insetos pelo mecanismo de seu metabolismo ou detoxificação. O inseto se utiliza do mecanismo para modificar ou detoxificar o inseticida de forma a prevenir a sua ação no sítio alvo, convertendo-o através de processos metabólicos em uma substância não tóxica.<sup>86</sup>

O estudo contribui com o desenvolvimento de modelos de QSAR empregando análise conformacional para obter modelos de QSAR mais acurados na predição interna.

Utilizando o método de similaridade molecular por distância euclidiana, conseguiu-se identificar compostos de um conjunto de confômeros de compostos com valores experimentais conhecidos. Este conjunto é utilizado como pseudo *blind set* para avaliar as possibilidades de acerto do método na escolha do confômero bioativo, quando aplicado na proposição de novos compostos contra *C. suppressalis*.

### 5.2.2 – Métodos

Oikawa *et al.*<sup>24</sup> sintetizaram 67 dibenzolidrazinas (DBHs) e elaboraram alguns estudos de QSAR com o propósito de avaliar o efeito substituinte das dibenzolidrazinas mono substituídas no anel A ( $X_n=2\text{-Cl}$ ) e di-substituídas nos anéis A e B (Figura 26).



**Figura 26.** Dibenzoilidrazina – Estrutura geral.<sup>24</sup>

Em relação ao conjunto original de compostos ( $n = 67$ ) da literatura<sup>24</sup> foram eliminados 9 integrantes em função dos valores  $pLD_{50}$  serem muito baixos. Assim, o estudo de QSAR foi desenvolvido com 58 compostos. Empregou-se análise conformacional, selecionando os confôrmeros que melhor representam o valor experimental obtido. Os detalhes do procedimento estão descritos em Material e Métodos.

Os compostos ( $n = 30$ ) da Tabela 17 (v. em Anexos, pg. 230) correspondem ao conjunto de treinamento ( $n = 22$ ) e o conjunto de teste ( $n = 8$ ) para validação externa. O pseudo *blind set* de 28 compostos consta da Tabela 22 (v. em Anexos, pg. 231).

A atividade larvicida das dibenzoilidrazinas no modelo de QSAR é dada por  $pLD_{50}$ , onde  $LD_{50}$  (mmol/inseto) é a concentração do composto causando letalidade a 50% das larvas. Este estudo não explora a aplicação do método SVM.

## 5.2.3 – Resultados e Discussão

### 5.2.3.1 – QSAR e Descritores

O modelo de QSAR obtido com os 30 compostos usados como conjunto de treinamento contém três descritores calculados pelo programa DRAGON (PCD, RNCG e H8u), e é representado pela Eq.:

$$pIC_{50} = -0,098 (\pm 0,006) PCD - 6,659 (\pm 0,761) RNCG + 6,516 (\pm 0,295) H8u + 7,473 (\pm 0,328)$$

$$n = 30 \quad R^2 = 0,996 \quad s = 0,052 \quad F = 2413 \quad p < 0,0001 \quad Q^2 = 0,995 \quad SPress = 0,059$$

$$SDEP = 0,056 \quad MSE = 0,0024 \quad (5.2-1)$$

O descritor PCD (*difference between multiple path count and path count*) é definido como a diferença entre o número convencional da ordem de ligação (piID) e a contagem total da trajetória TPC (*total path count*). TPC é o número total de trajetos de qualquer comprimento (de 0 até o maior comprimento de ligação) no gráfico da molécula. A ordem de ligação piID é o número total de trajetos aos quais um peso é atribuído. O piID é obtido pela soma dos pesos de todos os trajetos de qualquer comprimento no gráfico. O programa Dragon aplica a transformação logarítmica  $\log(1+x)$  à contagem total das trajetórias TPC e o número de ordem de ligação piID<sup>47</sup>. Nos valores numéricos, portanto, está embutida uma descrição gráfica da molécula considerando as ordens de ligação. Os valores numéricos são muito semelhantes e constantes não variando com a conformação.

RNCG (*Relative Negative Charge*)<sup>47,87</sup> é um descritor associado à carga parcial de superfície, e é dado pela relação do átomo de carga mais negativa dividido pela soma da carga negativa total. Portanto, RNCG representa uma fração da carga negativa da molécula. Este valor varia muito pouco entre confôrmeros de um mesmo composto dos compostos aqui empregados e pode ser considerado constante. RNCG varia apenas de composto para composto.

O descritor H-GETAWAY H8u (GEometry, TOpology and Atom-Weights ASsembly, H autocorrelation of lag 8/unweighted)<sup>88</sup>, está baseado em uma matriz de *leverage* chamada MIM (Matriz de Influência Molecular) semelhante à definida em estatística e usualmente empregada para diagnósticos de regressão.

A matriz MIM denotada por **H** é calculada a partir da matriz **M** de coordenadas cartesianas da molécula como

$$\mathbf{H} = \mathbf{M} \cdot (\mathbf{M}^T \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T \quad (5.2-2)$$

Onde o sobrescrito T representa a matriz transposta de **M**. Trata-se de uma matriz simétrica  $A \times A$ , onde A representa o número de átomos com as seguintes propriedades matemáticas

$$0 \leq h_{ij} \leq 1 \quad \sum_{i=1}^A h_{ii} = D \quad \bar{h} = D/A \quad D=1,2 \text{ ou } 3 \quad \sum_{i=1}^A h_{ij} = 0$$

onde  $h$  denota elementos da matriz de influência,  $D$  é uma classificação da matriz  $\mathbf{M}$  igual a 1, 2 e 3, respectivamente para moléculas lineares, planares e moléculas-3D, e  $\bar{h}$  é um valor médio dos termos da diagonal.

Os elementos diagonais  $h_{ii}$  são chamados de *leverage*, codificam informação atômica e representam a “influência” de cada átomo na determinação da forma de toda a molécula. Cada elemento da diagonal  $h_{ij}$  representa o grau de acessibilidade do  $j$ -ésimo átomo interagir com o  $i$ -ésimo átomo, ou, a atitude dos dois átomos considerados, de interagirem entre si. Sinal negativo dos elementos fora da diagonal representa que os dois átomos ocupam regiões moleculares opostas com respeito ao centro e portanto, o grau de interação mútua deve ser baixo.

Os valores de *leverage* são bastante sensíveis às mudanças conformacionais e a comprimentos de ligações dos tipos atômicos e da multiplicidade de ligações. Estes dois últimos aspectos, comprimento de ligação e multiplicidade de ligações são contemplados pelo descritor PCD anteriormente descrito, mas na sua forma de cálculo pelo programa Dragon, mantém-se constante para diferentes confôrmeros como já comentado.

O descritor H-GETAWAY de auto-correlação H8u é resultante da equação a seguir

$$H_k(\mathbf{w}) = \sum_{i=1}^{A-1} \sum_{j>i} h_{ij} \cdot w_i \cdot w_j \cdot \delta(k; d_{ij}; h_{ij}) \quad k=1, 2, \dots, d \quad (5.2-3)$$

onde  $d_{ij}$  é a distância topológica entre átomos  $i$  e  $j$  e  $d$  é o diâmetro topológico. A função  $\delta(k; d_{ij}; h_{ij})$  é a função delta de Dirac definida por

$$\delta(k; d_{ij}; h_{ij}) = 1 \text{ se } d_{ij} = k \text{ e } h_{ij} > 0 \quad \text{e} \quad \delta(k; d_{ij}; h_{ij}) = 0 \text{ se } d_{ij} \neq k \text{ e } h_{ij} \leq 0$$

e  $w_i$  e  $w_j$  são os pesos (*weights*) dados por propriedades físico-químicas como massa atômica, polarizabilidade, eletro-negatividade, volume atômico de van der Waals, e o peso unitário, caso presente na Eq. (5.2-1).

Tomando por base as características dos descritores, os compostos estão descritos considerando: a distribuição atômica dos átomos na molécula e as ordens de ligação (PCD); a eletro-negatividade da molécula numa determinada região baseada no átomo mais eletro-negativo (RNCG), e a geometria 3D da molécula (H8u) assumida na sua forma bioativa.

Os coeficientes extraídos da matriz de correlação de Pearson entre descritores são os seguintes: RNCG/H8u (0,116), RNCG/PCD (0,011), H8u/PCD (0,189). Correlação entre bioatividade e descritores: pLD<sub>50</sub>/H8u (0,901), pLD<sub>50</sub>/PCD (-0,729) e pLD<sub>50</sub>/RNCG (-0,486).

A Tabela 17 (v. em Anexos, pg. 230) reúne os valores preditos pelo modelo de QSAR da Eq.(5.2-1) antes e após análise conformacional, onde o resíduo gerado pela diferença na predição do valor experimental e predito de cada composto está tarjado em cinza.

Os compostos da Tabela 17 (v. em Anexos, pg. 230) foram divididos em conjunto de treinamento (n = 22) e conjunto de teste (n = 8) empregando processo de seleção randômica do aplicativo MDM. O modelo selecionado está representado pela Eq. (5.2-4).

$$pLD_{50} = - 0,099 (\pm 0,009) PCD - 6,586 (\pm 0,976) RNCG + 6,543 (\pm 0,343) H8u + 7,479 (\pm 0,412)$$

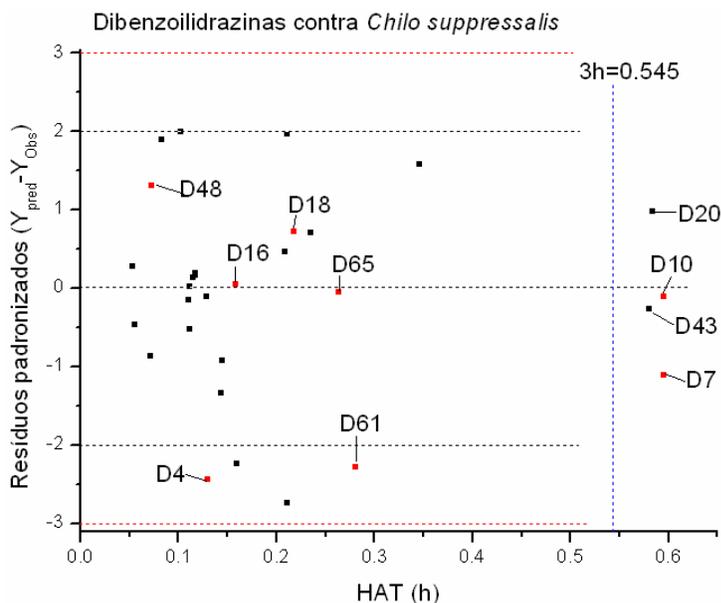
$$n = 22 \quad R^2 = 0,994 \quad s = 0,055 \quad F = 1112 \quad p < 0,0001 \quad Q^2 = 0,992 \quad SPress = 0,069$$

$$SDEP = 0,064 \quad MSE = 0,0025 \quad (5.2-4)$$

A escolha do melhor modelo foi feita pelo maior resultado de predição externa  $Q^2_{ext} = 0,995$  e com MSE = 0,0021 do conjunto de teste.

### 5.2.3.2 – QSAR – *Outliers* e Domínio de Aplicabilidade

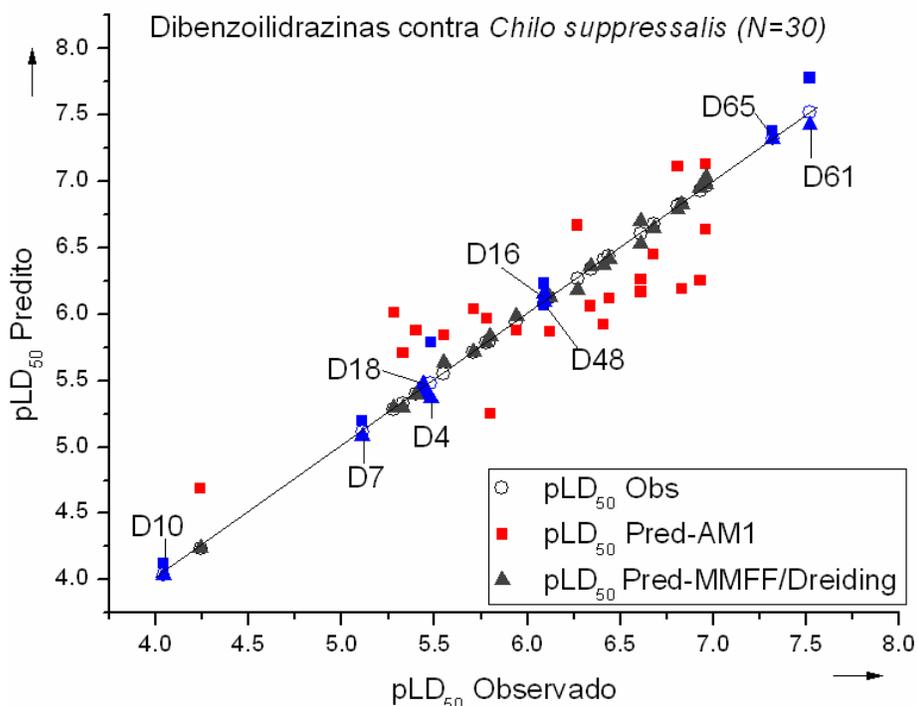
Os resultados de predição obtidos após análise conformacional empregando a Eq. (5.2-4) foram utilizados na construção do diagrama ou gráfico de Williams para avaliar o domínio de aplicabilidade do conjunto de treinamento e teste, ilustrado na Figura 27.



**Figura 27.** Diagrama de Williams – Domínio de aplicabilidade das dibenzolidrazinas contra *Chilo suppressalis*. Os compostos D20 e D43 do conjunto de treinamento e D7 e D10 do conjunto de teste, estão classificados fora da classe química dos demais compostos, denotados pela linha vertical  $3h = 0,545$ .

O valor *leverage* (*hat*) acima de três vezes o valor crítico ( $3h$ ), pode representar uma extrapolação ao modelo e a predição não ser confiável. No presente caso os resíduos são muito baixos e os valores experimentais estão muito próximos aos preditos. Ao eliminar-se todos os quatro compostos em questão, ocorre uma alteração desfavorável nos parâmetros estatísticos significando a piora do modelo de QSAR. Portanto, é mais conveniente preservar a diversidade dos compostos no conjunto. O gráfico da Figura 28 mostra o resultado comparativo das DBHs com estruturas otimizadas obtidas com o método semi-empírico AM1, e a

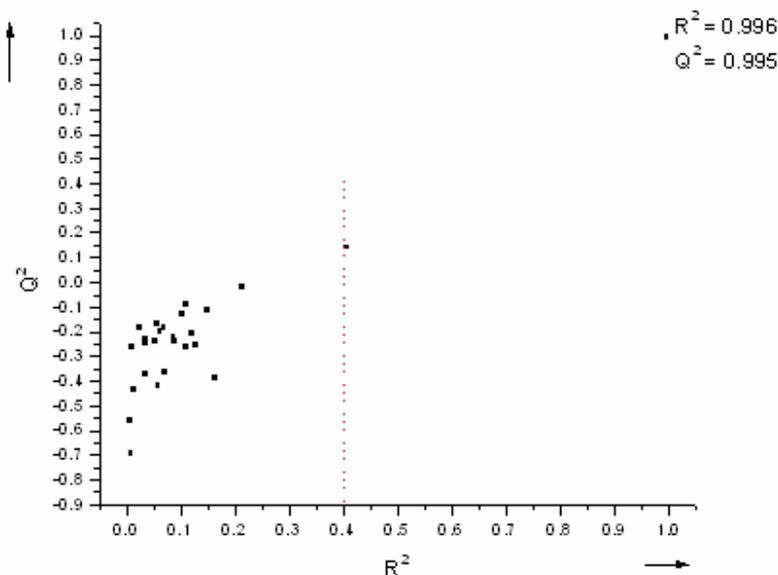
melhoria significativa obtida após uma seleção de confôrmeros que melhor representam os compostos na reprodução do valor experimental.



**Figura 28.** Valores de letalidade pLD<sub>50</sub> observados e preditos de DBHs contra *Chilo suppressalis*. Em vermelho, estruturas otimizadas pelo método AM1 antes da análise conformacional. Em azul estão os compostos de teste (n = 8) na condição AM1 e campo de força MMFF94 ou Dreiding modificado após análise conformacional.

### 5.2.3.3 – QSAR – Testes de Robustez e Validação

A robustez do conjunto de compostos n = 30 foi avaliada com os métodos *Y-randomization*<sup>53,61</sup> e *Leave-N-Out*.<sup>61</sup> O teste *Y-randomization* foi realizado com 300 iterações e 10 ciclos, resultando em 3000 iterações. Os valores do *intercept* usados para a aprovação do modelo sempre ocorreram abaixo dos valores limites  $\alpha(R^2) < 0,3$  e  $\alpha(Q^2) < 0,05$ . Segue o valor final do último ciclo:  $\alpha(R^2) = 0,154$  e  $\alpha(Q^2) = -0,270$ . Empregando o critério Eriksson-Wold<sup>58</sup>, foram feitas 25 iterações para a determinação de R<sup>2</sup> e Q<sup>2</sup> utilizando o programa da ref. [61], obtendo-se o gráfico da Figura 29 que apresenta resultado satisfatório, dando consistência à Eq. (5.2-1).



**Figura 29.** Valores  $R^2$  e  $Q^2$  obtidos em função da randomização dos valores  $pLD_{50}$  experimentais.

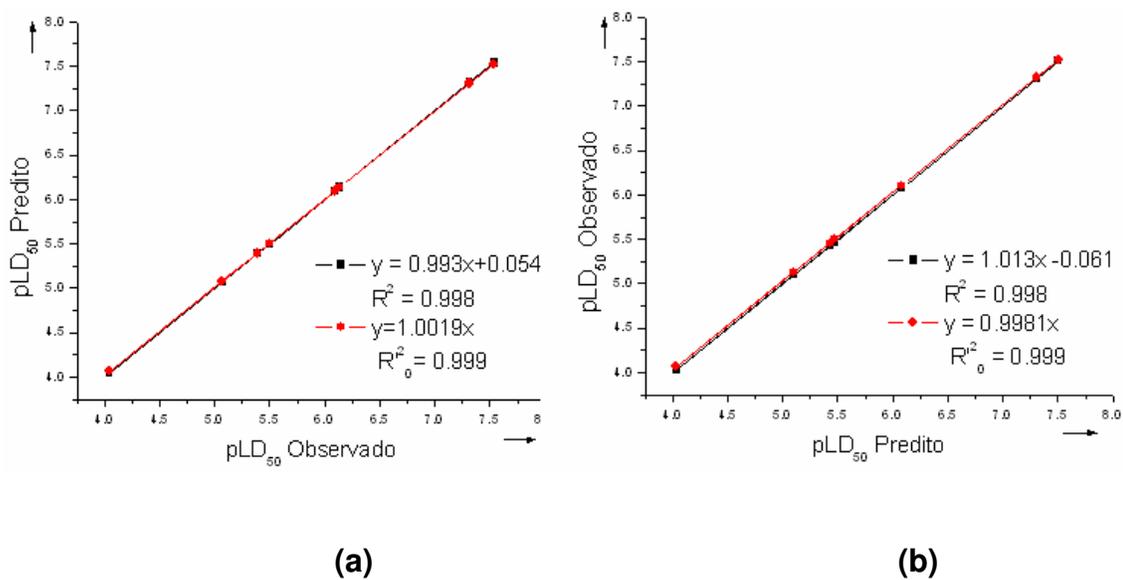
A randomização produziu valor médio  $R^2_{Yrand} \leq 0,4$ . Considerando o critério anterior adicionado ao resultado presentemente obtido, pode-se considerar a Eq. (5.2-1) consistente, gerando assim um modelo de QSAR aceitável neste critério.

A validação interna do conjunto pelo método *bootstrap*, utilizando 5000 iterações, resultando em um valor final de  $Q^2_{boot} = 0,994$  que comparado a  $Q^2_{LOO} = 0,995$  mostra a validação da predição interna da Eq.(5.2-1).

Empregando o critério *Leave-N-out*, foram eliminados 15 compostos (50% do conjunto  $n = 30$ ) obtendo-se um resultado  $Q^2 - Q^2_{LNO} < 0,1$  para 10 repetições.

Aplicando-se os critérios para o conjunto de treinamento  $n = 13$  usado para validação externa do conjunto de teste obtiveram-se no teste de randomização valores  $a(R^2) < 0,3$  e  $a(Q^2) < 0,05$  com os seguintes valores no final do último ciclo:  $a(R^2) = 0,150$  e  $a(Q^2) = -0,290$ .

Para a validação externa foi aplicado o critério de Golbraikh-Tropsha<sup>64</sup> cujo resultado está representado nos gráficos da Figura 30 assim como os parâmetros associados ao critério.



**Figura 30.** Representação gráfica dos parâmetros de validação do modelo de QSAR Eq. (5.2-4) das DBHs pelas equações (a)  $y = ax + b$ ,  $y = kx$ , (b)  $y = a'x + b'$  e  $y = k'x$ .

A sobreposição das retas em ambas as figuras é consequência da proximidade dos valores entre si de  $R^2$ ,  $R_0^2$  e  $R'_0^2$  e  $k$ ,  $k'$  ao valor ideal 1, assim como os coeficientes  $a$  e  $a'$  estão próximos a 1 e  $b$  e  $b'$  mais próximos a zero, respectivamente nas equações  $y = ax + b$  (Fig. 30a) e  $y = a'x + b'$  (Fig. 30b). Esta condição mostra que o modelo de QSAR da Eq. (5.2-4) é apropriado para predições externas segundo o critério empregado.

Empregando-se o critério de validação das funções de ajuste RQK (*RQK Fitness functions*) do programa MobyDigs dos modelos de QSAR Eq.(5.2-1 e 5.2-4), obtiveram resultados que aprovam os modelos (Tabelas 18 e 19).

**Tabela 18.** Validação do modelo de QSAR-MLR (n = 30) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis $K_{xy} - K_x > DK$	$K_{xy}$ 0,531	$K_x$ 0,303	$K_{xy}-K_x$ 0,228	DK 0,000
2.Capacidade de predição $Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,995	> 0,5		
3.Capacidade de predição do modelo $Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$ 0,995	$Q^2_{ASYM}$ 0,993	$Q^2_{LOO}-Q^2_{ASYM}$ 0,002	DQ -0,005
4.Redundância nas variáveis $R^P > t^P$	$R^P$ 0,045	$t^P$ 0,0		
5.Sobre-ajuste (overfitting) $R^N > t^N$	$R^N$ 0	$t^N$ -0,323		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

**Tabela 19.** Validação do modelo de QSAR-MLR (n = 22) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis $K_{xy} - K_x > DK$	$K_{xy}$ 0,448	$K_x$ 0,185	$K_{xy}-K_x$ 0,263	DK 0,000
2.Capacidade de predição $Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,993	> 0,5		
3.Capacidade de predição do modelo $Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$ 0,993	$Q^2_{ASYM}$ 0,981	$Q^2_{LOO}-Q^2_{ASYM}$ 0,012	DQ -0,005
4.Redundância nas variáveis $R^P > t^P$	$R^P$ 0,086	$t^P$ 0,0		
5.Sobre-ajuste (overfitting) $R^N > t^N$	$R^N$ 0	$t^N$ -0,323		

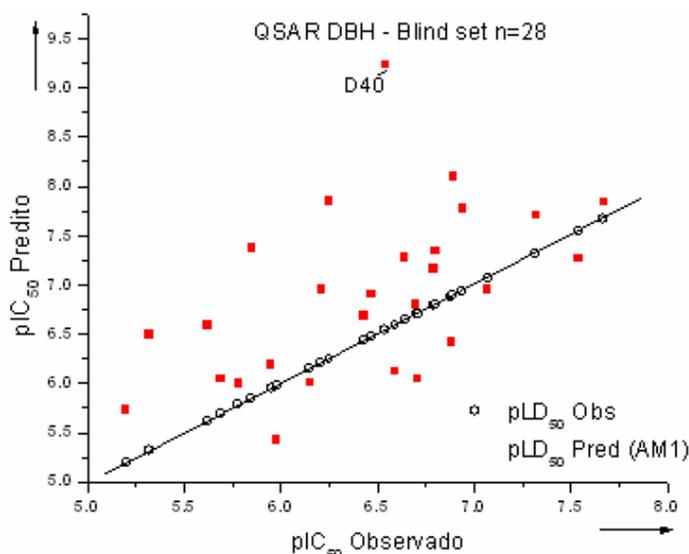
(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

Atendidos os requisitos de robustez e validação dos modelos de QSAR, o próximo passo é a sua aplicação ao conjunto de compostos N = 28 que será usado como um pseudo *blind set*, tratado no próximo tópico.

### 5.2.3.4 – Proposta de Novos Compostos

Considerando que os modelos de QSAR expresso pelas Eq. (5.2-1) e Eq.(5.2-4) são equivalentes utilizando o erro médio quadrático MSE como parâmetro, empregou-se a Eq. (5.2-1) para fazer a predição dos 28 compostos cujas geometrias foram inicialmente otimizadas com o método semi-empírico AM1, e

relembrando que integram o conjunto de 67 compostos que foram testados com o mesmo protocolo. O resultado da predição encontra-se na Tabela 20 (v. em Anexos, pg. 231), onde pode-se verificar através dos resíduos que os valores preditos estão, com poucas exceções, muito distantes dos valores experimentais. A Figura 31 mostra o resultado graficamente. O composto D40 destacado na figura recebeu a pior predição e voltará a ser comentado mais adiante.

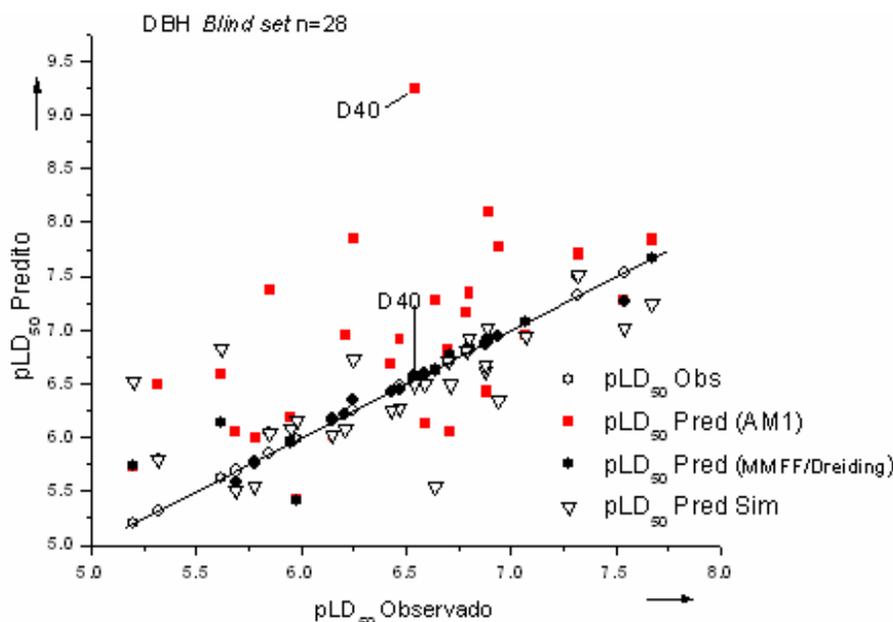


**Figura 31.** Representação gráfica dos resultados de predição do modelo de QSAR para as 28 dibenzolidrazinas.

Considerando que o modelo de QSAR em questão é resultado de um conjunto com poucos compostos, contudo, apresentando consistência dada pela análise de robustez e validação, a causa deste desempenho ruim pode estar contida na hipótese de que o confômero obtido com a otimização da geometria com AM1 não corresponde à conformação bioativa. Precisamente neste aspecto este estudo substitui esta premissa. Pela hipótese, deve haver uma conformação mais apropriada que os descritores possivelmente descreverão com valores que melhor representam na predição de sua bioatividade.

Geradas as conformações dos compostos e empregando o procedimento de similaridade molecular por distância euclidiana descrito em Material e Métodos,

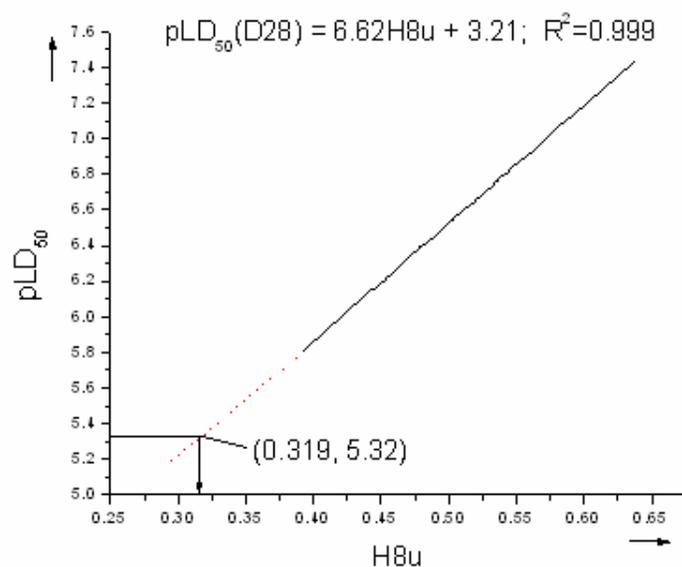
obtiveram-se os valores preditos listados nas quatro últimas colunas na Tabela 20 (v. em Anexos, pg. 231) e graficamente ilustrados na Figura 32.



**Figura 32.** Representação gráfica e comparativa dos resultados de predição do modelo de QSAR para as 28 dibenzimidrazinas em relação às conformações com AM1 (quadrados vermelhos) e após seleção de confôrmeros (círculos pretos e triângulos).

Os valores de predição obtidos para os confôrmeros representados pelos círculos pretos, selecionados por análise conformacional, mostram grande proximidade com os valores experimentais exceto para os compostos D27, D28, D42 e D51. Estes quatro compostos mostram resíduos altos, que se pode atribuir a duas possíveis causas: **(a)** os descritores não conseguem descrever estas estruturas adequadamente; **(b)** apesar de cada composto ter 50 conformações no conjunto, a conformação que melhor representa o valor  $pLD_{50}$  não está contida.

A Figura 33 ilustra a predição do modelo de QSAR para os confôrmeros do composto D28, tomado como exemplo, onde se verifica uma relação linear crescente de  $pLD_{50}$  com H8u, e o valor aproximado que o descritor H8u assumiria caso esta conformação existisse. Análise similar foi estendida aos compostos D27, D42 e D51.



**Figura 33.** Representação da variação de valores pLD<sub>50</sub> em função do descritor H8u para o composto D28. Valor estimado de H8u (= 0,319) atribuído à conformação bioativa (pLD<sub>50-exp</sub> = 5,32)

Foi feita uma tentativa, para cada um dos quatro compostos de se encontrar uma conformação que fosse próxima da bioativa mas, sem sucesso. Aplicando a Eq.(5.2-1 e 5.2-4) ao conjunto de 28 compostos antes da busca conformacional, obtém-se valores estatísticos muito ruins ( $n = 28$ ,  $R^2 = 0,259$ ,  $MSE = 0,866$ ), e os valores preditos correspondentes estão na Tabela 20 (v. em Anexos, pg. 231). Ao executar-se uma análise conformacional, ocorre uma melhoria substancial, como mostram os indicadores estatísticos:  $n = 28$ ,  $R^2 = 0,786$ ,  $MSE = 0,0874$ . Porém, os 4 compostos mostram ter grande influência no resultado, quando eliminados do conjunto, a Eq. (5.2-1) ou Eq.(5.2-4) gera os seguintes valores mostrando melhoria substancial:  $n = 24$ ,  $R^2 = 0,973$ ,  $MSE = 0,0075$ .

Verifica-se para os 24 compostos uma boa aproximação dos valores pLD<sub>50</sub> observados e calculados, resultado só possível de ser obtido pela análise de confôrmeros.

Outro aspecto favorável decorrente da análise conformacional é a análise de *outliers*. Tomando como exemplo D40 (v. Fig. 32), composto que pode ser

considerado pelo seu resíduo inicial um *outlier*, e que pôde ser integrado ao conjunto ao se selecionar o confômero que melhor representa a bioatividade.

Como o *blind set* simboliza a proposta de novos compostos cujo valor experimental na realidade é desconhecido, este estudo utiliza o método de similaridade molecular por distância euclidiana numa tentativa de se identificar o melhor confômero que melhor reproduza o valor experimental. O desempenho do método de similaridade aplicado a estes compostos é medido arbitrando um valor  $pLD_{50}$  de referência como segue.

Baseado em um valor experimental médio  $pLD_{50} = 6,00$  do conjunto de treinamento, adotou-se  $pLD_{50} \geq 6,50$ , ou seja, um pouco acima do valor médio, como valor de referência para avaliar os resultados de desempenho de predição do modelo de QSAR para novos compostos simulados pelo pseudo *blind set*:

- a) Compostos que tiveram predição com resíduo acima do objetivo  $\leq 0,10$ : 7.
- b) Compostos com resíduo no intervalo  $-0,10 \leq 0 \leq +0,10$ : 21.
- c) Confômeros identificados pelo índice de similaridade molecular por distância euclidiana com predições contidas no intervalo  $6,50 \leq pLD_{50} \leq 7,67$ : 13.
- d) Identificações de confômeros falso-positivos ( $pLD_{50}$  incorretamente  $\geq 6,50$ ): 3.
- e) Identificações de confômeros falso-negativos ( $pLD_{50}$  incorretamente  $\leq 6,50$ ): 2.
- f) Compostos corretamente excluídos ( $pLD_{50} < 6,50$ ): 10.

Taxa de acerto obtido considerando os acertos de inclusão e exclusão: 80%. Um resultado considerado favorável neste caso específico de proposição de novos compostos.

### 5.2.4 – Conclusão

Mostrou-se que o emprego de um modelo de QSAR construído com análise conformacional, validado e robusto, é mais favorável na predição interna. Porém, alguns problemas podem surgir, influenciando individualmente ou, em conjunto:

(a) Os descritores têm uma boa correlação com a bioatividade no conjunto dos compostos, mas não descrevem bem todos os integrantes, podendo não ser adequados para a descrição do espaço conformacional de determinado composto (vide Fig. 33 como exemplo).

(b) Mesmo havendo um conjunto de conformações em número razoável, a população de confôrmeros poderá ser insuficiente, exigindo refinamento da análise, o que significa aumentar a quantidade de confôrmeros.

(c) O método de similaridade empregado não consegue identificar todos os melhores confôrmeros corretamente, ou seja, os mais próximos da conformação bioativa.

A similaridade molecular por distância euclidiana neste estudo deu um resultado de acerto de 80% dentro do critério adotado, possibilitando um bom aproveitamento dos compostos propostos, caso não fosse este um exemplo de simulação.

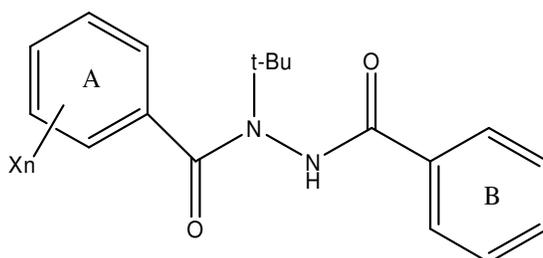
## 5.3 –Dibenzoilidrazinas – contra *Leptinotarsa decemlineata* (Say)

### Coleoptera, Crysomelidae

#### 5.3.1 – Introdução

O escaravelho da batata (*Leptinotarsa decemlineata*) é uma das pragas que afetam a cultura da batata. A perda da área foliar reduz a capacidade da planta em produzir hidratos de carbono para os tubérculos, o que se traduz numa quebra da produção final, bem como na redução da qualidade da batata produzida. Considerando a sua enorme capacidade de reprodução (2 a 3 gerações anuais), associada à sua extrema voracidade, dibenzoilidrazinas podem ser fortes candidatas como agentes larvicidas considerando as características ecológicas que lhes são atribuídas.

A atividade larvicida foi medida por Nakagawa *et al.*<sup>25</sup> numa série de N,N'-dibenzoil-N-t-butilidrazinas sintetizadas, tendo vários substituintes no anel A, perto do grupo *tert*-butila. O anel B não foi substituído (Figura 34) e (v. em Anexos Tabela 21, pg. 232).



**Figura 34.** Estrutura geral das N,N'-dibenzoil-N-t-butilidrazinas

Os efeitos dos substituintes foram analisados pelos autores, empregando vários modelos de QSAR clássicos<sup>25</sup>, com o objetivo de explicar a sua influência em função das posições *orto*, *meta* ou *para* na atividade larvicida. A ação de muitos destes compostos foi testada em ensaio anterior<sup>89</sup> contra *Chilo suppressalis* Lepidoptera, mostrando que as atividades pesticidas governadas pelas propriedades físico-químicas são distintas entre os compostos que agem sobre larvas da espécie Lepidoptera e Coleoptera.

### 5.3.2 – Métodos

Os modelos de QSAR desenvolvidos por Nakagawa *et al.*<sup>25</sup> empregaram os descritores: **(a)**  $V_W$  (parâmetro de Bondi para o volume de van der Waals); **(b)**  $B_1$  e  $B_5$  que correspondem a dois dos parâmetros STERIMOL de Veepl; **(c)**  $\pi^{meta}$  e  $(\pi^{meta})^2$  (parâmetro hidrofóbico para substituintes na posição *meta*); **(d)**  $\text{Log}P$  (coeficiente de partição octanol/água); **(e)**  $I_{para}$  (um parâmetro indicador de substituintes na posição *para*); **(f)**  $I_{multi}$  (parâmetro indicador de compostos multisubstituídos). Desenvolveram cinco modelos de QSAR diferentes, especificamente por grupo de substituintes, avaliando as influências com a atividade larvicida.

Foram utilizados alguns dos descritores acima numa tentativa de preservar parcialmente um dos modelos de QSAR desenvolvidos por Nakagawa *et al.*<sup>25</sup>, porém, introduzindo um novo descritor, cujos valores variam com a conformação, e desenvolver um modelo de QSAR dentro do escopo deste estudo. Como no caso do estudo da seção anterior, através da análise conformacional se busca obter um modelo de QSAR com predição interna mais acurada, mas também utilizando-o como modelo para simular a predição de novos compostos.

O conjunto de dibenzoilidrazinas é integrado por 46 compostos monosubstituídos no anel A (Figura 34). Este conjunto foi dividido em conjunto de treinamento ( $N = 30$ ), (Tabela 21, v. em Anexos, pg. 232) e conjunto de teste simulando novos compostos (*pseudo blind set*,  $n = 16$ ).

Foi utilizado o índice de Carbó<sup>90</sup> para avaliar a similaridade molecular, identificando-se confômeros dos compostos do *pseudo blind set* ( $n = 16$ ) com valores experimentais conhecidos para avaliar a possibilidade de acerto do método.

O método de similaridade molecular de Carbó *et al.*<sup>90</sup> se baseia no valor máximo obtido pela sobreposição das densidades eletrônicas das moléculas:

$$R_{AB} = \frac{\int \rho_A(r) \rho_B(r) dr}{\left( \int \rho_A(r)^2 dr \right)^{1/2} \left( \int \rho_B(r)^2 dr \right)^{1/2}} \quad (5.3-1)$$

Onde  $\rho_A(r)$  e  $\rho_B(r)$  são as densidades eletrônicas das moléculas A e B, respectivamente. A sobreposição entre as densidades eletrônicas das duas moléculas é estimada pelo numerador, e o denominador fornece o fator de normalização cujo índice varia de 0 a 1, onde 1 representa a maior similaridade. O cálculo foi feito pelo programa APBS<sup>91</sup>.

No caso, o pseudo *blind set* simula a proposta de novos compostos. A atividade larvicida das dibenzoilidrazinas no modelo de QSAR é dada por  $pLD_{50}$ , onde  $LD_{50}$  (mmol/inseto) é a concentração molar que provoca a letalidade em 50% das larvas, e  $pLD_{50} = \text{Log}(1/LD_{50})$ . Os compostos do *blind set* que estavam fora do intervalo arbitrado ( $-0,10 \leq pLD_{50\text{-Obs}} - pLD_{50\text{-Pred}} \leq 0,10$ ) foram submetidos à análise conformacional.

Demais pormenores da metodologia aplicada estão descritos nos tópicos da seção 4. O método de regressão SVM como segundo método e independente do método MLR, foi igualmente aplicado.

### 5.3.3 – Resultados e Discussão

#### 5.3.3.1 – QSAR e Descritores

O modelo de QSAR desenvolvido com Regressão Linear Múltipla (MLR) com os 30 compostos como conjunto de treinamento está representado na equação:

$$pLD_{50} = + 0,115 (\pm 0,041) V_w^{\text{orto}} + 0,622 (\pm 0,105) \pi^{\text{meta}} - 1,181 (\pm 0,065) \Delta B_5 - 1,039 (\pm 0,127) I_{2356} - 3,729 (\pm 0,458) \text{Mor23p} + 3,475 (\pm 0,251)$$

$$n = 30 \quad R^2 = 0,992 \quad s = 0,087 \quad F = 607 \quad p < 0,0001 \quad Q^2 = 0,990 \quad \text{SPress} = 0,101$$

$$\text{SDEP} = 0,092 \quad \text{MSE} = 0,006 \quad (5.3-2)$$

O descritor Mor23p pertence à família de descritores 3D-MoRSE (*3D-Molecule Representation of Structures based on Electron diffraction*) é calculado pelo programa DRAGON, e varia significativamente com as conformações de cada um dos compostos. Desenvolvido por Schuur *et al.*<sup>92</sup>, deriva da equação para difração de elétrons. Uma primeira investigação para se avaliar reconhecimento de padrões químicos em estudos de relação estrutura-atividade foi feita por Soltzberg e Wilkins<sup>93</sup>

que partem da equação de difração eletrônica modificando-a para chegar à equação a seguir, transcrita como empregada pelo programa DRAGON:

$$Morsw = I(s) = \sum_{i=1}^{nAT} \sum_{j=i-1}^{nAT} w_i w_j \frac{\text{sen}(s_i r_{ij})}{s_i r_{ij}} \quad (5.3-3)$$

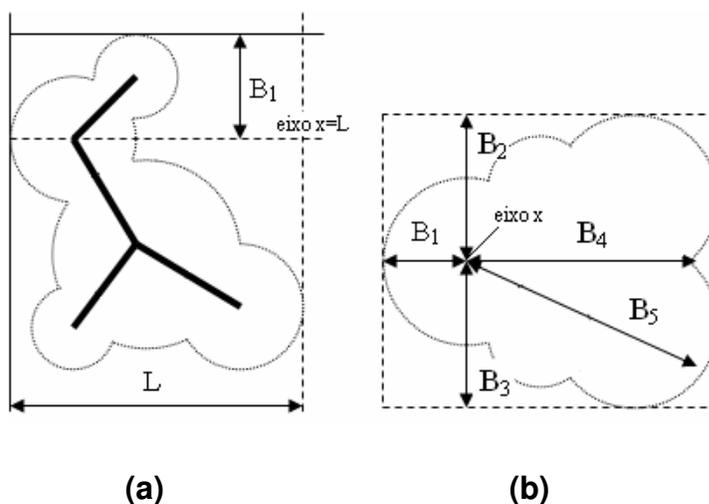
onde  $w$  representa a propriedade,  $r_{ij}$  a distância entre os átomos  $i$  e  $j$ , e  $nAT$  o número de átomos da molécula. O parâmetro  $s$  mede o ângulo de espalhamento dado por  $s = 4\pi \text{sen}(\theta/2)/\lambda$  onde  $\theta$  é o ângulo de espalhamento e  $\lambda$  o comprimento de onda do feixe de elétrons. Em  $Morsw$ ,  $w$  é representado pela polarizabilidade  $p$  e  $s$  é calculado no intervalo que vai de 0 a 31 e para  $s = 0$  a razão de espalhamento é assumida como sendo igual a um. No presente estudo  $s = 23$ .

Pelo parâmetro  $\pi$ , introduzido em estudos QSAR por Hansch-Fujita<sup>29</sup>, calcula-se a lipofilicidade do substituinte dada por

$$\pi_x = \log P_{R-X} - \log P_{R-H} \quad (5.3-4)$$

onde  $P_R$  é o coeficiente de partição octanol/água,  $P_{R-X}$  é o coeficiente de partição do composto para o substituinte  $X$ , e  $P_{R-H}$  corresponde ao coeficiente de partição do composto sem o substituinte, onde  $H$  representa o átomo de hidrogênio.

Os parâmetros estéricos STERIMOL de Verloop estão ilustrados na Figura 35. Seis parâmetros são usados para definir a forma do substituinte na molécula:  $L$ ,  $B_1$  até  $B_5$ ,  $B_1$  definindo uma distância mínima a partir de um átomo  $\alpha$  do substituinte que o conecta ao resto da molécula, e  $B_5$  uma distância máxima e ortogonal a  $L$ .



**Figura 35.** Representação gráfica dos parâmetros STERIMOL de Verloop, L e B<sub>1</sub> a B<sub>5</sub>. Representam a projeção de um substituinte **(a)** ao longo do eixo x=L; **(b)** perpendicular ao eixo de substituição, eixo x (perpendicular ao plano do papel). Desenho baseado na pg. 48 da ref. [3].

Todos os parâmetros de largura B são ortogonais ao comprimento L de um substituinte, compreendendo entre eles um ângulo de 90°. Nakagawa *et al.*<sup>25</sup> definiram os valores dos parâmetros STERIMOL B<sub>1</sub> e B<sub>5</sub> como  $\Delta B_1$  e  $\Delta B_5$  em relação ao hidrogênio, descontando 1 do valor tabelado.<sup>52</sup> Assim, o valor de B<sub>5</sub> do átomo de Cl na posição *para* corresponde a 1,80. Valor usado por Nakagawa *et al.*<sup>25</sup>:

$$1,80 - 1,0 = 0,80.$$

O parâmetro  $I_{2356}$  é um indicador que representa as substituições nas dibenzoilidrazinas, e está correlacionado às posições *orto* e *meta* conjuntas, atribuindo-se o valor 0 ou 1, conforme a posição do substituinte.

A matriz de correlação R apresenta os seguintes valores entre descritores:  $V_w^{\text{orto}}/\pi^{\text{meta}}$  (0,206);  $V_w^{\text{orto}}/\Delta B_5$  (0,412);  $V_w^{\text{orto}}/I_{2356}$  (0,011);  $V_w^{\text{orto}}/\text{Mor23p}$  (0,297);  $\pi^{\text{meta}}/\Delta B_5$  (0,302);  $\pi^{\text{meta}}/I_{2356}$  (0,257);  $\Delta B_5/I_{2356}$  (0,139);  $\Delta B_5/\text{Mor23p}$  (0,107);  $I_{2356}/\text{Mor23p}$  (0,087). Correlação entre descritor e bioatividade:  $pLD_{50}/V_w^{\text{orto}}$  (0,518);  $pLD_{50}/\pi^{\text{meta}}$  (0,267);  $pLD_{50}/\Delta B_5$  (0,899);  $pLD_{50}/I_{2356}$  (0,074);  $pLD_{50}/\text{Mor23p}$  (0,183).

Os compostos da Tabela 21 (v. em Anexos, pg. 232) foram divididos em conjunto de treinamento (n = 24) e conjunto de teste (n = 6) empregando processo de seleção randômica do aplicativo MDM. A escolha do modelo foi feita

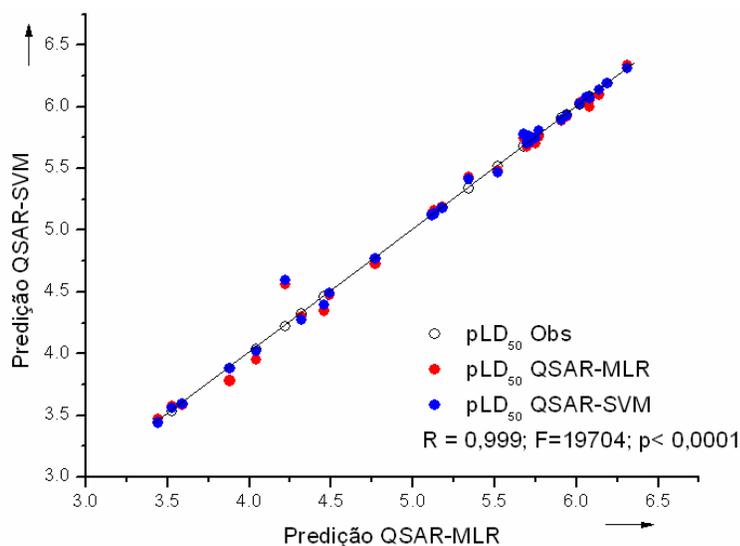
com o aplicativo MDM pelo melhor resultado de predição externa  $Q^2_{\text{ext}} = 0,99$  e com  $\text{MSE} = 0,005$  do conjunto de teste e  $\text{MSE} = 0,006$  com igual valor para os conjuntos de treinamento, respectivamente com  $n = 30$  e  $n = 24$ .

O modelo selecionado está representado pela equação:

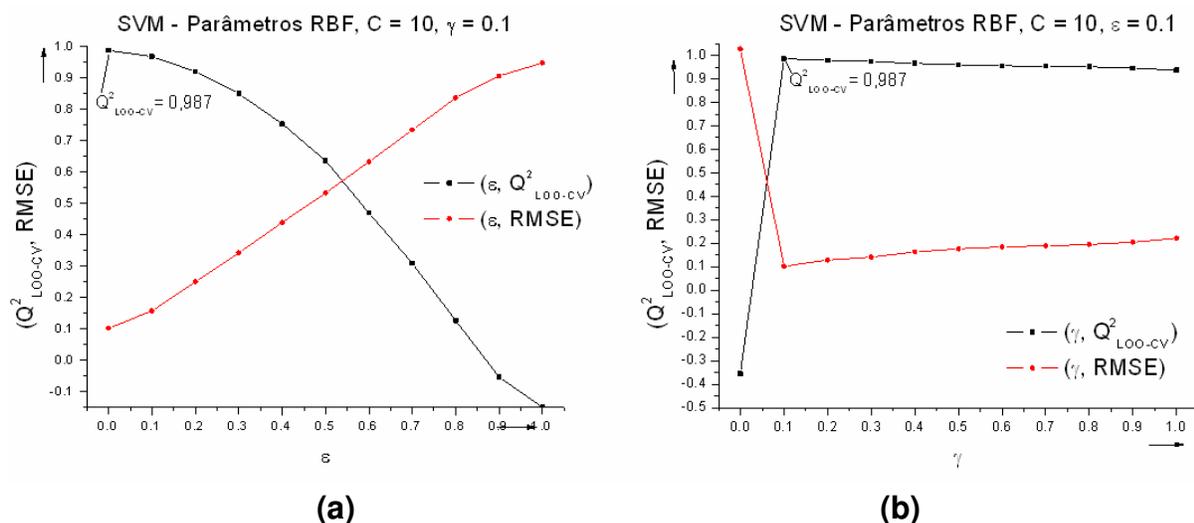
$$\begin{aligned} \text{pLD}_{50} = & + 0,116 (\pm 0,047) V_w^{\text{orto}} + 0,629 (\pm 0,123) \pi^{\text{meta}} - 1,190 (\pm 0,085) \Delta B_5 - 1,044 \\ & (\pm 0,144) I_{2356} - 3,815 (\pm 0,588) \text{Mor23p} + 3,426 (\pm 0,328) \\ n = 24 \quad R^2 = 0,990 \quad s = 0,096 \quad F = 378 \quad p < 0,0001 \quad Q^2 = 0,986 \quad \text{SPress} = 0,117 \\ \text{SDEP} = 0,104 \quad \text{MSE} = 0,006 \end{aligned} \quad (5.3-5)$$

Utilizando o mesmo conjunto de treinamento para MLR ( $n = 24$ ) foi feita uma avaliação do conjunto de dados pelo método SVM que, produziu resultados equivalentes ao MLR, como pode ser verificado pela Figura 36 e pelos dados estatísticos de correlação entre modelos de QSAR. O modelo de QSAR construído com SVM utiliza o *kernel Radial Basis Function* (RBF). O melhor valor obtido foi com o ajuste dos seguintes parâmetros: **(a)** custo (ou capacidade)  $C=10$ ; **(b)** extensão da função radial RBF,  $\gamma = 0,1$  e **(c)** a função de perda  $\varepsilon = 0,0$ .

O gráfico da Figura 37 mostra a variação do RMSE em função dos demais parâmetros  $\varepsilon$  e  $\gamma$ , e que define o melhor  $Q^2_{\text{LOO-CV}} = 0,987$ .



**Figura 36.** Resultados de predição interna dos modelos de QSAR-MLR e QSAR-SVM e correlação dos modelos:  $R = 0,999$ ;  $F = 19704$ ;  $p < 0,0001$ .



**Figura 37.** (a)  $Q^2_{\text{LOO-CV}}$  em função da variação do parâmetro  $\varepsilon$  mantendo constantes C e  $\gamma$ ; (b)  $Q^2_{\text{LOO-CV}}$  em função da variação do parâmetro  $\gamma$  mantendo constantes C e  $\varepsilon$ . Para o menor valor RMSE seleciona-se o valor  $Q^2_{\text{LOO-CV}}$  correspondente a 0,987.

A Tabela 22 (v. em Anexos, pg. 233) mostra os resultados numéricos entre modelos de QSAR-MLR e QSAR-SVM. Os valores expressam certa equivalência entre os modelos de QSAR obtidos por metodologias independentes, o que pode ser verificado pelos valores numericamente próximos. Resta saber se esta condição de equivalência é preservada para predição de substâncias novas, o que será verificado mais adiante com o pseudo *blind set*.

Foi feita uma análise dos descritores e uma tentativa em interpretar a sua influência na bioatividade dividindo os compostos em subconjuntos de substituintes, *orto*, *meta*, e *para*, e combinações, incluindo o composto não substituído D1 para referência em todos os subconjuntos.

### Substituições *orto*

Obteve-se a equação de QSAR:

$$\begin{aligned} \text{pLD}_{50} &= + 0,121 (\pm 0,033) V_w^{\text{orto}} - 3,699 (\pm 0,826) \text{Mor23p} + 3,481 (\pm 0,459) \\ n &= 12 \quad R^2 = 0,972 \quad s = 0,049 \quad F = 154 \quad p < 0,0001 \quad Q^2 = 0,935 \quad \text{SPress} = 0,074 \\ \text{SDEP} &= 0,067 \end{aligned} \quad (5.3-6a)$$

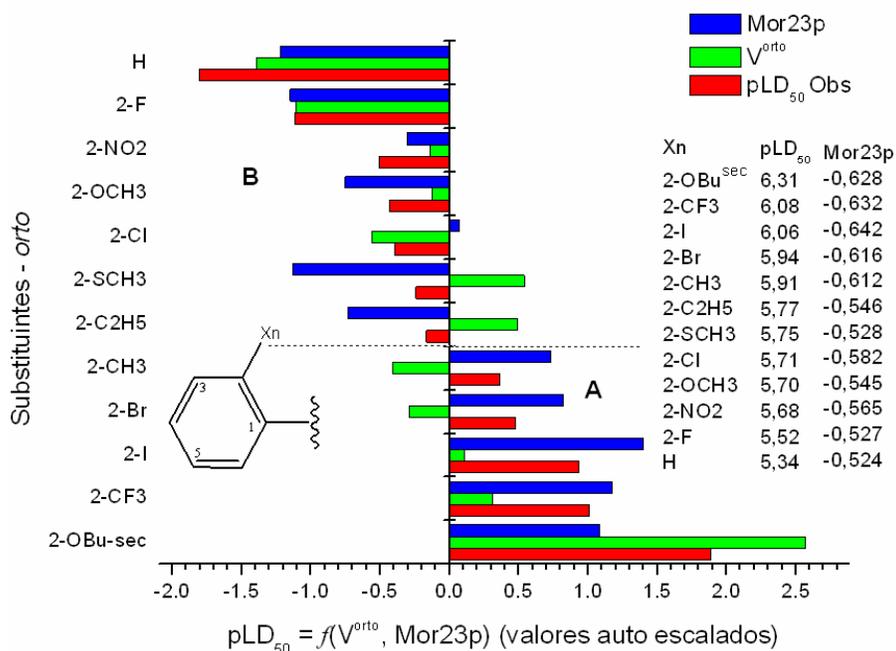
Nakagawa *et al.*<sup>25</sup> se utilizaram dos parâmetros  $V_w^{\text{orto}}$  (parâmetro de Bondi para o volume de van der Waals) e STERIMOL  $\Delta B_1$  de Verloop, que, para substituintes em *orto*, mostram o crescimento da potência das dibenzoilidrazinas o que está relacionado com o aumento de volume das moléculas, expresso pela equação de QSAR

$$\begin{aligned} \text{pLD}_{50} &= + 0,173 (\pm 0,075) V_w^{\text{orto}} + 0,336 (\pm 0,257) \Delta B_1 + 5,334 (\pm 0,200) \\ n &= 12 \quad R = 0,906 \quad s = 0,123 \quad F = 20,69 \quad p = 0,0004 \quad Q^2 = 0,546 \quad \text{SPress} = 0,196 \\ \text{SDEP} &= 0,177 \end{aligned} \quad (5.3-6b)$$

A Eq. (5.3-6a) indica através do parâmetro Mor23p que além da influência do volume, há uma componente relacionada à polarizabilidade da molécula embutida em Mor23p que influi no aumento ou redução da potência na posição *orto*. A matriz de correlação bioatividade/descritor indica esta influência:  $\text{pLD}_{50}/\text{Mor23p}$  (0,867);  $\text{pLD}_{50}/V_w^{\text{orto}}$  (0,805) e  $\text{pLD}_{50}/\Delta B_1$  (0,525) na Eq.(5.3-6b).

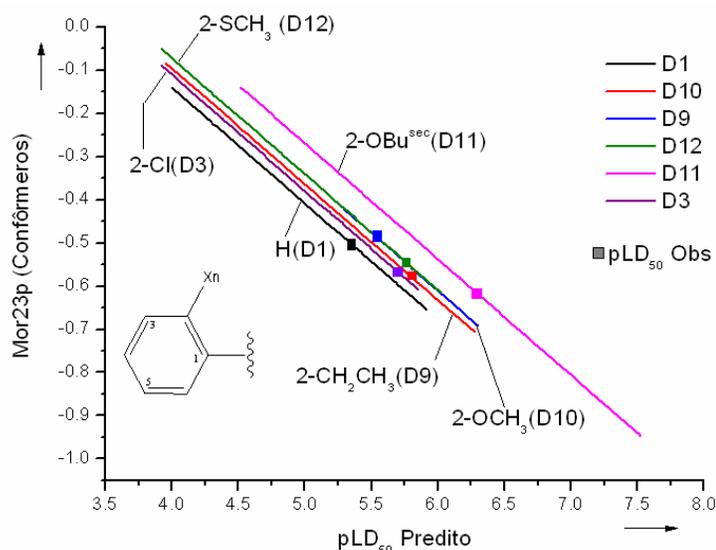
O mérito da Eq.(5.3-6b) está na clara relação da potência e sua variação com o volume, o que na Eq. (5.3-6a) está implícito. A Figura 38 mostra um gráfico construído com valores auto escalados, servindo de auxílio para mostrar os pesos de cada parâmetro  $V_w^{\text{orto}}$  e Mor23p do modelo de QSAR da Eq. (5.3-6a) na bioatividade de cada composto. Por exemplo, o composto com o substituinte 2-OBu<sup>sec</sup> tem a maior potência entre os *orto* substituídos, com grande peso de  $V_w^{\text{orto}}$ , e Mor23p com peso menor no valor da bioatividade. No gráfico são mostrados os valores numéricos de Mor23p associados ao substituinte e à  $\text{pLD}_{50}$ , que sugerem que quanto mais crescentes, ainda que de forma não contínua, não permitindo o reconhecimento de um padrão, o valor de Mor23p implica no decréscimo da potência do composto.

Exceção marcante é dada pelo substituinte 2-Cl. Aqui a potência do composto decai e é preponderantemente influenciada por  $V_w^{\text{orto}}$  desfavorável com uma pequena contribuição favorável da polarizabilidade dada por Mor23p.



**Figura 38.** Representação gráfica da influência dos descritores  $V^{\text{orto}}$  e Mor23p na avaliação da bioatividade. Região **A**: maior bioatividade; região **B**: menor bioatividade.

Para melhor entendimento da variação de  $pLD_{50}$  em função de Mor23p elaborou-se o gráfico da Figura 39 selecionando alguns dos compostos orto-substituídos. A predição de  $pLD_{50}$  dos confôrmeros é feita através da Eq. (5.3-2).



**Figura 39.** Variação do descritor Mor23p para alguns compostos em função dos confômeros correspondentes. Para  $pLD_{50}$  crescente, Mor23p decresce. Os pontos nas retas são os valores experimentais de  $pLD_{50}$ .

Selecionaram-se os compostos com valores  $pLD_{50}$  próximos a 5,70 e que envolvem o grupo  $CH_3$ . O composto D8 (composto substituído por  $CH_3$  e representado na Figura 38) é tomado como referência na análise das variações que os átomos associados ao grupo metila possam causar, conforme segue. Entre parênteses estão os valores experimentais  $pLD_{50}$ .

**D11** 2- $CH_3CH_2CH(CH_3)$  (6,31) maior atividade (2- $OBu^{sec}$ );

**D8** 2- $CH_3$  (5,91);

**D9** 2- $CH_2CH_3$  (5,77);

**D12** 2- $SCH_3$  (5,75);

**D10** 2- $OCH_3$  (5,70).

Pode-se verificar que grupos  $CH_3$  associados a outros átomos que não sejam C e H podem eventualmente ter influência na redução de  $pLD_{50}$ . O composto D8 2- $CH_3$  caracteriza no gráfico da fig. 38 uma região de transição onde os valores  $pLD_{50}$  decrescem. Para o composto D3 2-Cl (5,71) há influência positiva de Mor23p na

bioatividade, porém, com um peso muito pequeno, fato que leva  $V_w^{\text{orto}}$  a determinar o valor influente no valor de  $\text{pLD}_{50}$ .

Pode-se resumir a atuação dos descritores na atividade larvívica pela leitura do gráfico da fig.38 como segue: Mor23p exerce influência positiva no crescimento da atividade até a posição *orto* 2-CH<sub>3</sub>, vindo então a diminuir. Notar também que, em geral, quanto mais decresce o valor de Mor23p, maior a ação larvívica dada por  $\text{pLD}_{50}$ .

### Substituições *meta*

Seis compostos têm substituição *meta*. Incluindo-se o composto não substituído (D1) obtém-se a equação de QSAR a seguir:

$$\text{pLD}_{50} = + 0,663 (\pm 0,125) \pi^{\text{meta}} - 4,057 (\pm 0,730) \text{Mor23p} + 3,277 (\pm 0,425)$$

$$n = 7 \quad R = 0,994 \quad s = 0,055 \quad F = 161 \quad p = 0,0002 \quad Q^2 = 0,961 \quad \text{SPress} = 0,098 \quad \text{SDEP} = 0,080 \quad (5.3-7a)$$

A equação obtida por Nakagawa *et al.*<sup>17</sup> segue abaixo:

$$\text{pLD}_{50} = + 3,345 (\pm 0,811) \pi^{\text{meta}} - 2,358 (\pm 0,630) (\pi^{\text{meta}})^2 + 5,266 (\pm 0,156)$$

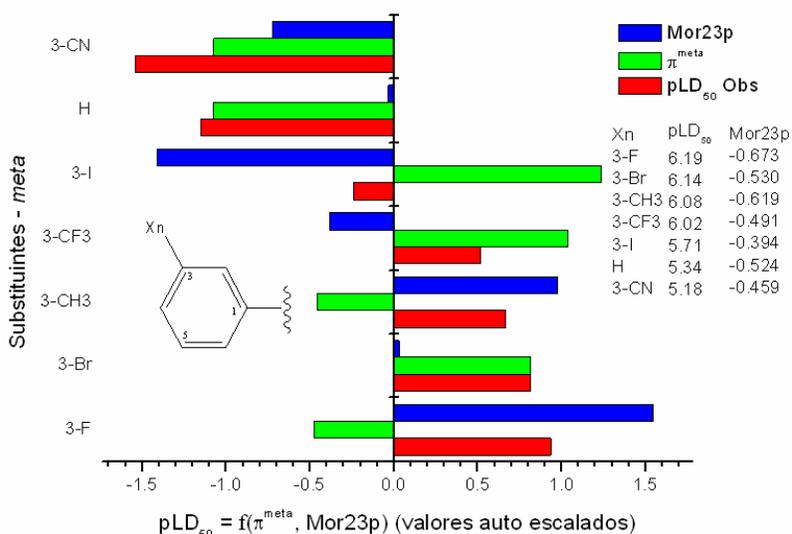
$$n = 7 \quad R = 0,987 \quad s = 0,081 \quad F = 73,4 \quad p = 0,0007 \quad Q^2 = 0,917 \quad \text{SPress} = 0,144 \quad \text{SDEP} = 0,118 \quad (5.3-7b)$$

Na Eq.(5.3-7a) o descritor Mor23p foi incluído no modelo de QSAR formando par com o parâmetro hidrofóbico do substituinte  $\pi$  na posição *meta* ( $\pi^{\text{meta}}$ ). Os valores da matriz de correlação R da bioatividade e descritor seguem:  $\text{pLD}_{50}/\text{Mor23p}$  (0,569);  $\text{pLD}_{50}/\pi^{\text{meta}}$  (0,507) e  $\text{pLD}_{50}/(\pi^{\text{meta}})^2$  (0,321) no caso da Eq. (5.3-7b).

A representação gráfica do modelo de QSAR através da Figura 40 mostra que nos compostos mais potentes 3-F e 3-CH<sub>3</sub> o peso de Mor23p indica maior influência na potência do composto do que o peso do parâmetro hidrofóbico  $\pi^{\text{meta}}$ . No composto com o substituinte 3-Br,  $\pi^{\text{meta}}$  tem marcante prevalência na potência, ao passo que Mor23p tem peso significativamente menor. Nos compostos substituídos 3-CF<sub>3</sub> e 3-I prevalece o peso do parâmetro  $\pi^{\text{meta}}$ , e Mor23p com peso maior em 3-I.

No caso de substituições *meta*, tomando 3-CH<sub>3</sub> a posição de referência no gráfico, verifica-se uma atuação positiva de Mor23p na atividade larvívica

especialmente devido aos halógenos F e Br, vindo então a decrescer até a posição 3-CN.



**Figura 40.** Representação gráfica da influência dos descritores  $\pi^{\text{meta}}$  e Mor23p na avaliação da bioatividade.

CN é menos hidrofóbico do que H (fato experimental) e a potência de 3-CN mostra ação tanto da componente Mor23p como do parâmetro  $\pi^{\text{meta}}$ , H mostra corretamente o peso da componente hidrofóbica e Mor23p com peso muito menor como é de se esperar.

### Substituições *para*

Onze compostos com substituição na posição *para* integram o estudo, adicionando-se o composto não substituído D1 (H). No modelo de QSAR obtido por Nakagawa *et al.*<sup>25</sup> o composto D29 (substituinte 4-O(CH<sub>2</sub>)<sub>3</sub>C<sub>6</sub>H<sub>5</sub>) foi eliminado em razão de apresentar uma potência muito baixa. Além disto no composto D24 substituinte 4-NO<sub>2</sub>, substitui o valor de  $\Delta B_5$  pelo valor de  $\Delta B_1$ , que caso contrário comporta-se como *outlier*. A equação de QSAR resultante é transcrita abaixo:

$$pLD_{50} = -0,694 (\pm 0,283) \Delta B_5^{\text{para}} + 4,925 (\pm 0,389)$$

$$n = 11 \quad s = 0,279 \quad R = 0,880 \quad F_{1,9} = 30,86.$$

(5.3-8a)

Seguindo as indicações do artigo, não foi possível reproduzir o modelo de QSAR da Eq.(5.3-8a) mas obteve-se uma aproximação:

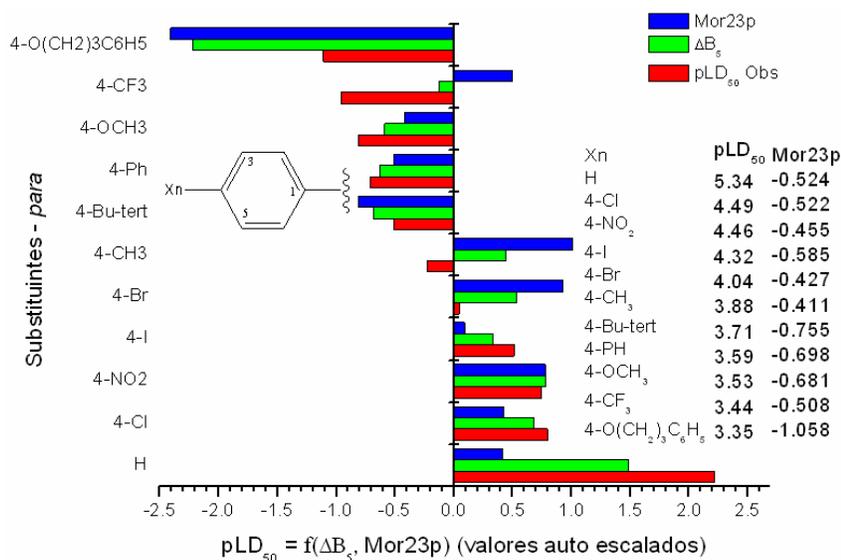
$$\begin{aligned} \text{pLD}_{50} &= - 0,458 (\pm 0,238) \Delta B_5 + 4,541 (\pm 0,334) \\ n &= 10 \quad R = 0,843 \quad s = 0,224 \quad F = 19,69 \quad p = 0,0022 \quad Q^2 = 0,602 \quad \text{SPress} = 0,264 \\ \text{SDEP} &= 0,248 \end{aligned} \quad (5.3-8b)$$

Empregando todos os compostos  $n = 12$ , os dois descritores Mor23p e  $\Delta B_5$  e eliminando o composto D23 como *outlier* (em função do resíduo comparativamente grande em relação aos demais compostos), obtém-se:

$$\begin{aligned} \text{pLD}_{50} &= - 1,109 (\pm 0,160) \Delta B_5 - 3,632 (\pm 0,845) \text{Mor23p} + 3,470 (\pm 0,326) \\ n &= 11 \quad R = 0,989 \quad s = 0,101 \quad F = 173 \quad p < 0,0001 \quad Q^2 = 0,919 \quad \text{SPress} = 0,190 \quad \text{SDEP} \\ &= 0,170 \end{aligned} \quad (5.3-9)$$

A correlação dos descritores com  $\text{pLD}_{50}$  segue:  $\text{pLD}_{50}/\Delta B_5$  (0,830);  $\text{pLD}_{50}/\text{Mor23p}$  (0,509). A Figura 41 utilizada para a análise das contribuições de cada descritor na atividade larvicida mostra que a potência é maior onde o parâmetro estérico  $\Delta B_5$  contribui positivamente com maior peso, invertendo-se a situação a partir do substituinte 4-*tert*-Bu onde a contribuição do parâmetro estérico decresce aumentando a contribuição de Mor23p, exceto para 4-CF<sub>3</sub> onde  $\Delta B_5$  reduz a ação do composto e Mor23p tem peso positivo. O composto com o substituinte 4-O(CH<sub>2</sub>)<sub>3</sub>C<sub>6</sub>H<sub>5</sub> excluído do modelo de QSAR no estudo de Nakagawa *et al.*<sup>25</sup> pela sua baixa potência, tem no gráfico a indicação de ambos parâmetros com peso negativo, o estérico  $\Delta B_5$  e Mor23p.

Vale notar que nas substituições *para*, os valores de Mor23p não se comportam de maneira crescente ou decrescente, indicando menor ou maior potência, respectivamente. Nada pode ser concluído, mas os valores sugerem que as melhores atividades são causadas pelas substituições halogenadas e pelo substituinte NO<sub>2</sub>.



**Figura 41.** Representação gráfica da influência dos descritores  $\Delta B_5$  e Mor23p na avaliação da bioatividade.

### Comentários

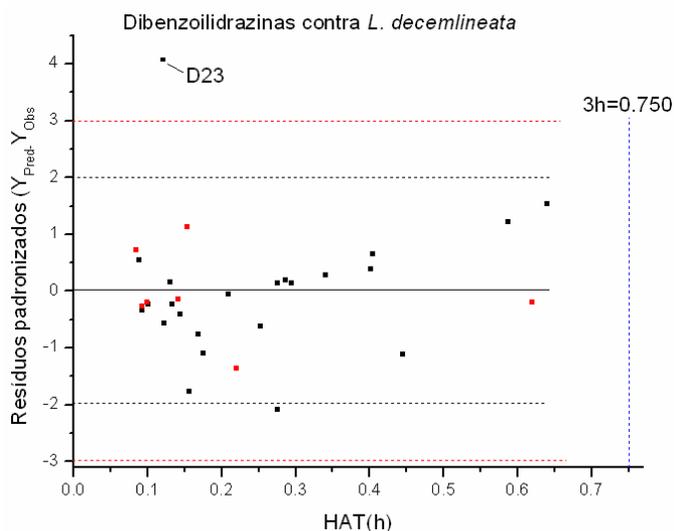
O que se pode observar pelo resultado é que o descritor Mor23p tem uma contribuição complementar quando associado aos descritores originalmente empregados por Nakagawa *et al.*<sup>25</sup>. Assim usando-se os descritores  $V_w^{\text{orto}}$  e  $\Delta B_5^{\text{orto}}$  (Eq. 5.3-2), há uma clara indicação dos valores numéricos crescentes do volume de van der Waals ( $V_w^{\text{orto}}$ ) do substituinte no crescimento favorável da atividade larvicida.

A polarizabilidade implícita no descritor Mor23p mostra que, além do parâmetro geométrico associado ao volume do substituinte, existe uma influência no mecanismo de ação das dibenzoilidrazinas na potência larvicida não associada apenas ao aumento ou diminuição de volume, mas também à polarizabilidade elétrica da molécula. A observação é análoga para Mor23p associado ao descritor hidrofóbico  $\pi^{\text{meta}}$ , Eq.(5.3-7a) e  $\Delta B_5$  Eq.(5.3-9), onde Mor23p complementa a informação em ambos modelos de QSAR.

Não foi possível construir um modelo de QSAR combinando o parâmetro  $I_{2356}$  (ligado às substituições conjuntas em *orto* e *meta*), e o parâmetro Mor23p.

5.3.3.2 – QSAR – *Outliers* e Domínio de Aplicabilidade

Os resultados de predição obtidos empregando a Eq. (5.3-2) foram utilizados na construção do diagrama de Williams para avaliar o domínio de aplicabilidade do conjunto de treinamento e teste, ilustrado na Figura 42.



**Figura 42.** Diagrama de Williams – Domínio de aplicabilidade das dibenzoilidrazinas contra *Leptinotarsa decemlineata*. O composto D23 comporta-se como *outlier* por apresentar o maior resíduo entre os compostos.

O campo de força MMFF e Dreiding Modificado, não gerou nenhuma conformação para o composto D23 (substituinte CN na posição *para*) que aproximasse seu valor calculado do valor experimental, e tem o maior resíduo do conjunto. A eliminação deste composto produz o modelo de QSAR a seguir:

$$\begin{aligned} \text{pLD}_{50} = & + 0,107 (\pm 0,024) V_w^{\text{orto}} + 0,590 (\pm 0,061) \pi^{\text{meta}} - 1,183 (\pm 0,038) \Delta B_5 - \\ & 1,0297 (\pm 0,074) I_{2356} - 3,575 (\pm 0,270) \text{Mor23p} + 3,5851 (\pm 0,149) \\ n = 29 \quad R^2 = 0,998 \quad s = 0,050 \quad F = 1754 \quad p < 0,0001 \quad Q^2 = 0,996 \quad \text{SPress} = 0,062 \\ \text{SDEP} = 0,056 \end{aligned} \quad (5.3-10)$$

Para verificar a influência do composto D23, empregaram-se os modelos de QSAR MLR e SVM, excluindo-o do conjunto. Aplicando-se os modelos de QSAR ao conjunto que será utilizado para predição externa com  $n = 16$ , comparando os modelos de QSAR com e sem o composto D23 obtiveram-se os resultados a seguir.

Composto D23 incluso:

(a) QSAR-MLR - Eq.(5.3-2):  $R^2 = 0,992$  e  $MSE = 0,006$ .

(b) QSAR-SVM:  $R^2 = 0,993$  e  $MSE = 0,006$ .

Composto D23 excluído:

(c) QSAR-MLR – Eq.(5.3-10):  $R^2 = 0,998$  e  $MSE = 0,002$ .

(d) QSAR-SVM:  $R^2 = 0,993$  e  $MSE = 0,009$ .

Os resultados dos modelos de QSAR de uma mesma metodologia, isto é MLR ou SVM, mostram que não existem grandes diferenças entre os erros MSE na predição. Portanto, optou-se por preservar o composto D23 no conjunto.

### 5.3.3.3 – QSAR – Testes de Robustez e Validação

A robustez do conjunto de compostos  $n = 30$  foi avaliada com os métodos *Y-randomization*<sup>53,61</sup> e *Leave-N-Out*.<sup>61</sup> O teste *Y-randomization* foi realizado com 300 iterações e 10 ciclos, resultando em 3000 iterações. Os valores do *intercept* usados para a aprovação do modelo sempre ocorreram abaixo dos valores limites  $a(R^2) < 0,3$  e  $a(Q^2) < 0,05$ . Segue o valor final do último ciclo:  $a(R^2) = 0,117$  e  $a(Q^2) = -0,448$ .

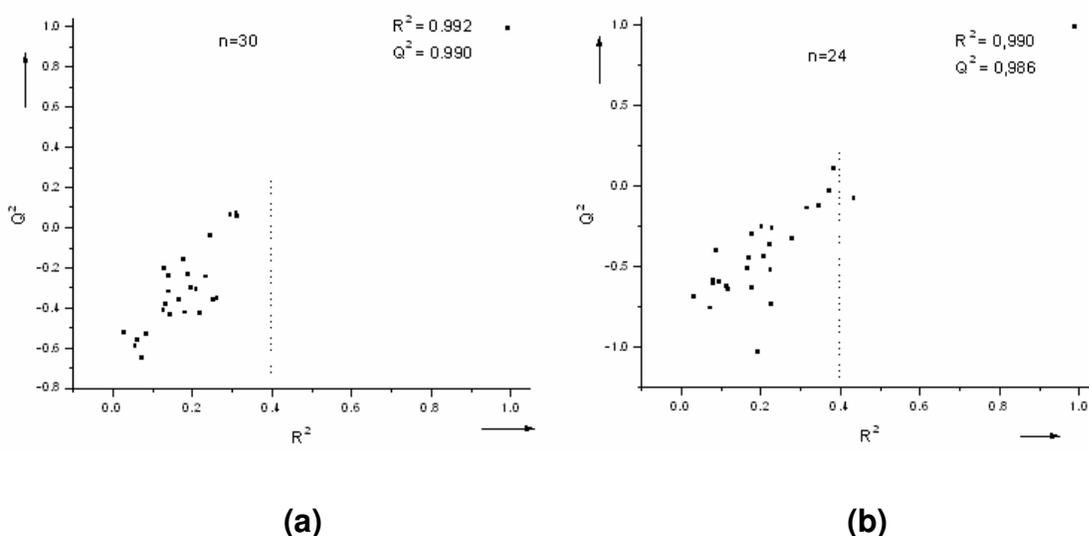
O conjunto de treinamento usado para fazer a predição externa tem o modelo de QSAR-MLR seguinte:

$$\begin{aligned} pLD_{50} = & + 0,116 (\pm 0,047) V_w^{orto} + 0,629 (\pm 0,123) \pi^{meta} - 1,190 (\pm 0,085) \Delta B_5 - 1,043 \\ & (\pm 0,144) I_{2356} - 3,815 (\pm 0,587) Mor23p + 3,426 (\pm 0,328) \\ n = & 24 \quad R^2 = 0,990 \quad s = 0,096 \quad F = 378 \quad p < 0,0001 \quad Q^2 = 0,986 \quad SPress = 0,117 \\ SDEP = & 0,104 \end{aligned} \quad (5.3-11)$$

Aplicando-se os critérios para o conjunto de treinamento  $n = 24$  usado para validação externa do conjunto de teste ( $n = 6$ ), obtiveram-se no teste de

randomização, valores  $a(R^2) < 0,3$  e  $a(Q^2) < 0,05$  com os seguintes resultados no final do último ciclo:  $a(R^2) = 0,151$  e  $a(Q^2) = -0,590$ .

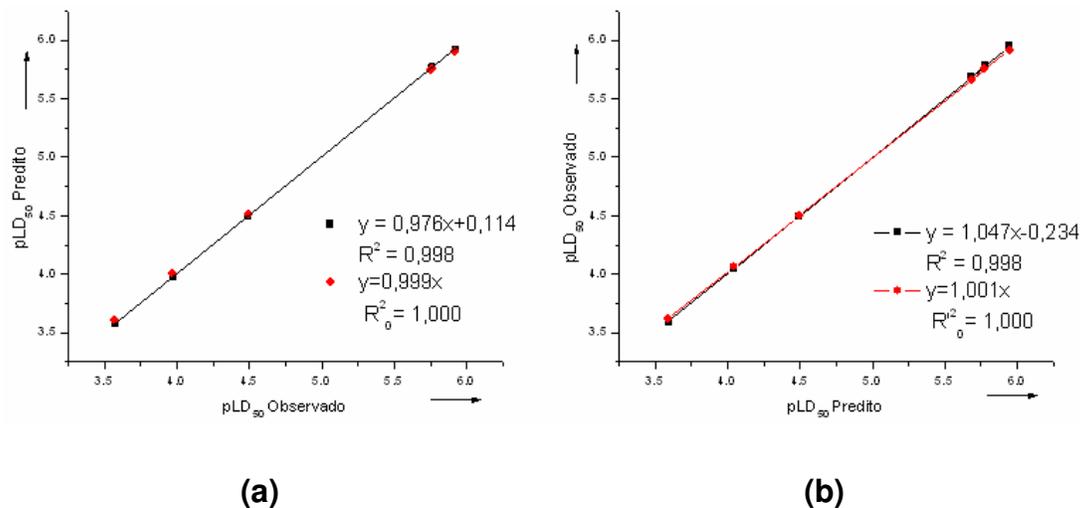
Empregando o critério Eriksson-Wold<sup>58</sup>, foram feitas 25 iterações para a determinação de  $R^2$  e  $Q^2$  utilizando o programa da ref. [61], obtendo-se os gráficos da Figura 43, que apresentam resultados satisfatórios, dando consistência à Eq. (5.3-2) para predições externas.



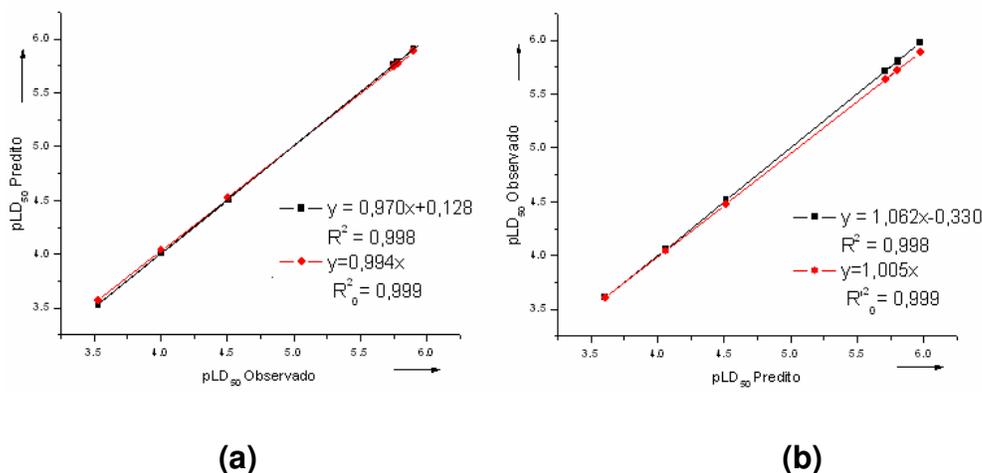
**Figura 43.** Valores  $R^2$  e  $Q^2$  obtidos em função da randomização dos valores  $pLD_{50}$  experimentais para (a) modelos de QSAR-MLR/SVM com  $n = 30$  e (b) modelos de QSAR-MLR/SVM com  $n = 24$ .

Para a validação externa do modelo de QSAR-MLR Eq.(5.3-11) usando o conjunto de teste  $n = 6$ , foi aplicado o critério de Golbraikh-Tropsha<sup>64</sup> cujo resultado está representado nos gráficos da Figura 44 assim como os parâmetros associados ao critério.

A seguir a Figura 45 mostra os resultados de validação utilizando as predições sobre o conjunto de teste pelo modelo de QSAR-SVM, mostrando que ambos os modelos de QSAR se equivalem. Esta condição mostra que o modelo de QSAR da Eq. (5.3-2) é apropriado para predições externas segundo o critério empregado.



**Figura 44.** Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR Eq. (5.3-2) das DBHs pelas equações (a)  $y = ax+b$ ,  $y = kx$ , (b)  $y = a'x+b'$  e  $y = k'x$ .



**Figura 45.** Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM das DBHs pelas equações (a)  $y = ax+b$ ,  $y = kx$ , (b)  $y = a'x+b'$  e  $y = k'x$ .

A validação interna pelo método *bootstrap*, utilizando 5000 iterações, resulta em um valor final  $Q^2_{boot} = 0,979$  que comparado ao  $Q^2_{LOO} = 0,986$  valida a predição interna da Eq.(5.3-2).

Pelo critério *Leave-n-Out*, foram eliminados 15 compostos ( 50% do conjunto  $n = 30$  ), obtendo-se o resultado  $Q^2 - Q^2_{LNO} < 0,1$  para 10 repetições.

Empregando-se o critério de validação das funções de ajuste RQK (*RQK Fitness functions*) do programa MobyDigs dos modelos de QSAR Eq.(5.3-2) e Eq. (5.3-11), obtiveram resultados que aprovam os modelos (Tabelas 23 e 24).

**Tabela 23.** Validação do modelo de QSAR-MLR (n = 30) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis	$K_{xy}$	$K_x$	$K_{xy}-K_x$	DK
$K_{xy} - K_x > DK$	0,415	0,299	0,116	0,000
2.Capacidade de predição				
$Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,990	> 0,5		
3.Capacidade de predição do modelo				
$Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$	$Q^2_{ASYM}$	$Q^2_{LOO}-Q^2_{ASYM}$	DQ
	0,990	0,988	0,002	-0,005
4.Redundância nas variáveis	$R^P$	$t^P$		
$R^P > t^P$	0,067	0,0		
5.Sobre-ajuste (overfitting)	$R^N$	$t^N$		
$R^N > t^N$	-0,142	-0,323		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

**Tabela 24.** Validação do modelo de QSAR-MLR (n = 24) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis	$K_{xy}$	$K_x$	$K_{xy}-K_x$	DK
$K_{xy} - K_x > DK$	0,410	0,289	0,121	0,000
2.Capacidade de predição				
$Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,986	> 0,5		
3.Capacidade de predição do modelo				
$Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$	$Q^2_{ASYM}$	$Q^2_{LOO}-Q^2_{ASYM}$	DQ
	0,986	0,984	0,002	0,000
4.Redundância nas variáveis	$R^P$	$t^P$		
$R^P > t^P$	0,079	0,0		
5.Sobre-ajuste (overfitting)	$R^N$	$t^N$		
$R^N > t^N$	-0,082	-0,190		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

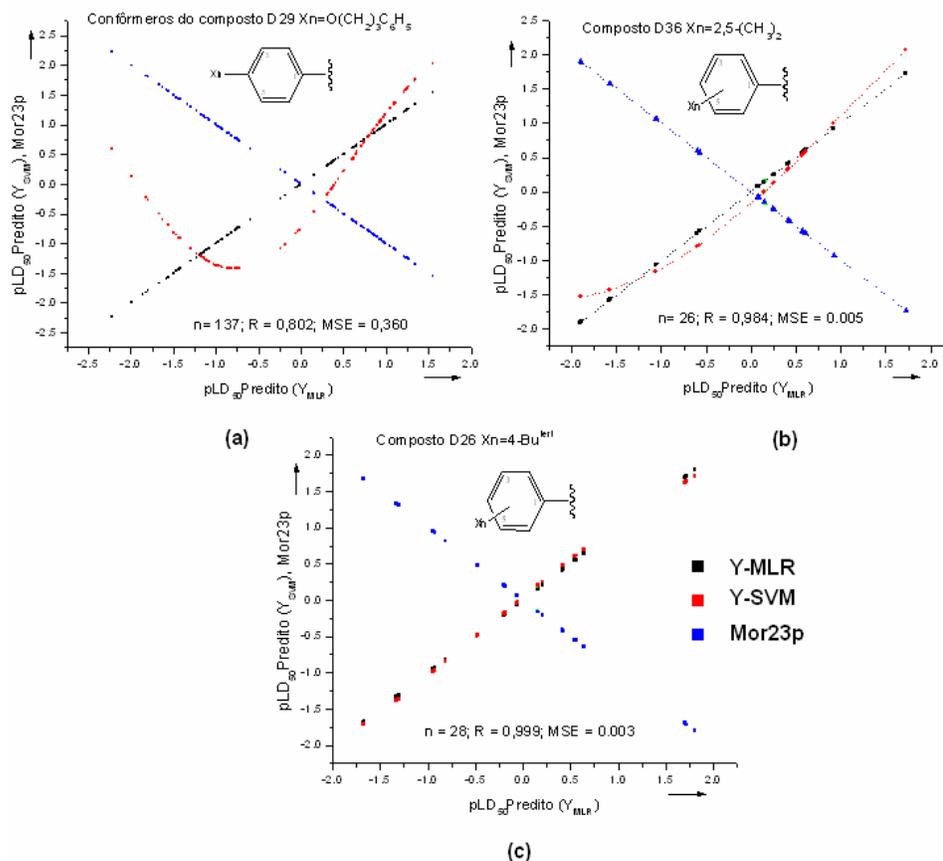
### 5.3.3.4 – Proposta de novos compostos

Considerando que os modelos de QSAR são equivalentes como constatado na predição interna e bem correlacionados (n = 30, R = 0,999, s = 0,034; F = 19603), e utilizando o erro médio quadrático MSE como parâmetro para a avaliação dos

resultados produzidos entre os métodos, empregaram-se os modelos de QSAR respectivamente MLR e SVM para simular a atividade larvicida dos novos compostos. O conjunto (n = 16) foi submetido à análise conformacional, partindo de estruturas otimizadas com o método semi-empírico AM1, para obter o confômero que melhor reproduza o valor pLD<sub>50</sub> experimental. Esta informação servirá para saber qual a margem de erro que os modelos de QSAR irão produzir nos resultados de predição. Obviamente, esta informação não estará disponível no caso de compostos realmente novos. Os parâmetros R de Pearson e MSE de cada modelo de QSAR são dados a seguir:

- (a) pLD<sub>50</sub> Obs vs. pLD<sub>50</sub> Pred QSAR-MLR: n = 16, R = 0,935; MSE = 0,091;  
(Anterior à análise conformacional: n = 16; R = 0,752; MSE = 0,271).
- (b) pLD<sub>50</sub> Obs vs. pLD<sub>50</sub> Pred QSAR-SVM: n = 16, R= 0,973; MSE = 0,027;  
(Anterior à análise conformacional: n = 16; R = 0,809; MSE = 0,112).

Verifica-se pelos valores MSE que cada modelo de QSAR faz a predição dos valores pLD<sub>50</sub> com diferenças entre métodos, mostradas nas cinco primeiras colunas numéricas da Tabela 25 (v. em Anexos, pg. 234). Os valores residuais maiores afetam os compostos D29, D32, D34, D36, D42, D43 e D46. Realizou-se uma análise numérica e gráfica de todos os 16 compostos. A análise gráfica consiste em avaliar o comportamento das predições e do descritor Mor23p associado a cada um dos confômeros dos respectivos compostos. Os valores preditos pLD<sub>50</sub> pelos modelos de QSAR e correspondentes Mor23p são auto escalados, numa tentativa de extrair alguma informação que auxilie no esclarecimento das diferenças de predição. A primeira informação que se obtém é que em alguns casos o modelo de QSAR-SVM mostra um comportamento não linear refletido pelo descritor Mor23p. Outra informação que pode ser extraída do gráfico é a sua característica decrescente associada ao aumento da atividade larvicida, como já foi visto anteriormente. A Figura 46 reflete estas informações para os compostos D29, D36 e D26 escolhidos para exemplificar. Os demais compostos apresentam resultados mais lineares pelo modelo de QSAR-SVM.



**Figura 46.** Predição da atividade larvicida expressa pelos confôrmeros dos compostos **(a)** D29, **(b)** D36 e **(c)** D26, e a influência do descritor Mor23p na atividade larvicida crescente. Valores auto escalados.

Fazendo a análise dos dados numéricos verifica-se, em geral, boa correlação entre os modelos de QSAR-MLR e SVM, porém, isto não assegura uma proximidade na predição externa para todos os compostos. O descritor Mor23p dita a equivalência ou não entre modelos de QSAR. Por exemplo, a diferença nas predições pLD<sub>50</sub> entre modelos de QSAR, com os confôrmeros de D26, estão contidas no intervalo  $-0,05 \leq 0 \leq 0,01$ , e há ótima correlação entre modelos de QSAR, (v. Fig. 46c). O composto D29, apresenta uma variação conformacional que quando medida pelo descritor Mor23p, apresenta diferenças na predição de aproximadamente 2 unidades logarítmicas entre modelos de QSAR-MLR e QSAR-SVM (vide Fig. 46a), indicando um comportamento não-linear na predição. A correlação entre ambas as predições também é menor do que em relação a todos

os outros compostos. Para os confôrmeros do composto D36 (v. Fig 46b), o intervalo das diferenças de predição é muito variado e grande ( $-0,77 \leq n \leq 0,07$ ) ditado pelo descritor Mor23p.

Em resumo, é fundamental proceder a uma análise numérica do conjunto de dados acompanhada por uma análise gráfica para novos compostos, pois, na ausência de valores experimentais, um intervalo com pequenas diferenças de predição entre modelos de QSAR (exemplo dos compostos D26, D31, D35, D37) pode ser um bom indicador na opção de escolha do composto a propor para síntese. Já um intervalo com grandes variações, como no caso dos demais compostos, D29, D32, D34, D36, D38, D39, 40, D41, D42, D43 e D46, indica que os modelos de QSAR têm diferenças de predição pois o descritor regente tem diferentes valores para boas predições em ambos modelos de QSAR. Para estes compostos a busca de um indicador é feita empregando métodos de similaridade molecular.

A similaridade molecular foi feita pelo índice de Carbó<sup>90,91</sup> e o resultado consta nas quatro últimas colunas da Tabela 25 (v. em Anexos, pg. 234). O modelo de QSAR-SVM aplicado aos confôrmeros selecionados pelo índice produziu os resultados mais favoráveis na predição da atividade larvicida.

Utilizando o critério  $-0,10 \leq pLD_{50}Obs - pLD_{50}Pred \leq 0,10$  para dar um peso aos métodos, o modelo de QSAR-SVM obteve o melhor resultado com 9 compostos (56%), pouco abaixo dos 60% arbitrados pelo estudo. O modelo de QSAR-MLR aplicado para a predição dos mesmos confôrmeros gerou apenas 5 compostos (31%) no intervalo.

### 5.3.4 – Conclusão

Foi feita uma contribuição em complementação aos modelos de QSAR de Nakagawa *et al.*<sup>25</sup>, onde a polarização implícita no descritor Mor23p indica exercer influência adicional no aumento ou decréscimo da atividade larvicida, o que pode ser verificado, respectivamente pela diminuição ou aumento do valor do parâmetro Mor23p no estudo conformacional de cada composto. Verifica-se assim, pelos modelos de QSAR propostos, que o parâmetro ligado ao volume de van der Waals ( $V_w$ ) e a hidrofobicidade  $\pi$  não são os únicos contribuintes ligados à atividade larvicida, como sugere a literatura.

Dos 16 compostos, 14 possuem substituições mistas nas posições *orto*, *meta* e *para*, e dois na posição *para*. Comparando, o conjunto de treinamento dos 30 compostos possui apenas 3 substituições mistas, e os demais compostos ou são *orto*, ou *meta* ou *para* substituídos. Apesar desta variabilidade, os parâmetros estatísticos indicam que a calibração ocorreu satisfatoriamente para ambos os modelos de QSAR no conjunto de treinamento ( $n = 30$ ), observando que a equivalência entre modelos de QSAR-MLR e QSAR-SVM é regida pelo descritor Mor23p. A equivalência pôde ser verificada no conjunto de confôrmeros dos compostos onde a diferença entre predições foram próximas para a maioria dos confôrmeros. Valores muito diferentes evidenciaram a falta de equivalência para um determinado composto.

O método de similaridade molecular pelo índice de Carbó<sup>90,91</sup> gerou uma taxa de acêrto próxima ao aceitável, porém, usando o intervalo de resíduos arbitrado por este estudo, excluem-se dois compostos de boa ação larvicida, a saber D31 e D43.

Finalmente, uma boa equivalência entre os modelos de QSAR indicada por diferenças pequenas nas predições entre confôrmeros, pode ser considerada como indicador de inclusão ou exclusão de novos compostos substituídos. Não foram obtidos bons resultados com os métodos de similaridade molecular por distância euclidiana e ShaEP.

## 5.4 –Dibenzoilidrazina – contra *Spodoptera exigua* (Huebner)

### Lepidoptera, Noctuidae

#### 5.4.1 - Introdução

A lagarta da espécie *Spodoptera exigua* ataca uma diversificada população de vegetais, algumas de grande importância econômica, por exemplo, brócolis, beterraba, algodão, feijão, batata, soja, tomate, etc. No verão, as larvas têm mudas 5 vezes a cada 16 dias e seu combate na agricultura deve iniciar-se nesta fase.<sup>94</sup> Existem diversas propostas de combate a esta praga, entre elas a utilização de baculovírus, uma alternativa a inseticidas, que são vírus de DNA dupla fita que infectam invertebrados, em especial a ordem Lepidoptera. Baculovírus conferem total segurança à saúde humana e ao meio ambiente.<sup>95</sup> Como inseticidas considerados ecologicamente aceitáveis, as dibenzoilidrazinas têm sido estudadas contra este tipo de lagarta.

A atividade larvicida foi avaliada em bioensaios com os mesmos 46 compostos de, N,N'-dibenzoil-N-t-butilidrazinas sintetizadas, utilizados contra *L. decemlineata* (Seção 5.3) em novo estudo contra *S. exigua* por Nakagawa *et al.*<sup>27</sup>, tendo vários substituintes no anel A, perto do grupo *tert*-butila e anel B não substituído (v. Figura 34, sub-seção 5.3.1).

A maioria dos compostos empregados nos bioensaios de ação larvicida contra as espécies Lepidoptera são halogenados. Pesquisas recentes de natureza teórica e experimental (citadas na sub-seção 5.4.2 – Métodos-Compostos halogenados), têm reportado características físico-químicas específicas entre os átomos halogênios dos grupos funcionais de ligantes e sua interação com aminoácidos devido à importância dos halogênios no desenvolvimento de novas drogas. Atrelado a este fato, e tomando por referência alguns dados de pesquisas feitos por Billas *et al.*<sup>96</sup>, Kasuya *et al.*<sup>97</sup> e Wurz *et al.*<sup>98</sup>, no presente estudo faz-se uma avaliação das dibenzoilidrazinas halogenadas empregando a técnica de *docking* para obtenção de algumas características da interação não-covalente do ligante receptor em questão. Partiu-se do complexo heterodímero da proteína receptora ecdisona (EcR-USP) da espécie *Heliothis virescens* (Lepidoptera)

cristalizada com uma dibenzoilidrazina código BYI06830<sup>96</sup>, muito semelhante ao cromafenozóideo, uma dibenzoilidrazina comercialmente utilizada como pesticida, identificada pela literatura como ANS-118, e que tem um oxigênio a menos em relação ao anel 1,4-dioxano da BYI06830.

Esta investigação visa contribuir com informações de natureza teórica, no melhor de nosso conhecimento ainda não contempladas pela literatura. Está composta por três partes: **(a)** Desenvolvimento de modelos de QSAR-2D. **(b)** Utilização do complexo ligante-receptor da lagarta *Heliothis virescens* (Lepidoptera) tendo por base as pesquisas de Billas *et al.*<sup>96</sup>, aqui usado para estudar as interações das dibenzoilidrazinas halogenadas e sua interação com os aminoácidos que formam a cavidade protéica. **(c)** Análise por homologia das sequências de *Heliothis virescens* (EcR-LBD) e *Spodoptera exigua* (SeEcR-LBD) (*Spodoptera exigua* Ecdyson Receptor Ligand Binding Domain). A investigação tenta obter as seguintes informações: **(a)** Com quais aminoácidos de SeEcR-LBD as dibenzoilidrazinas halogenadas interagem? **(b)** Qual o provável posicionamento dos grupos funcionais na cavidade da proteína dos compostos halogenados utilizados por Nakagawa *et al.*<sup>27</sup> e, especificamente, **(c)** quais átomos halogênios interagem com os respectivos aminoácidos? **(d)** Utilizar os modelos de QSAR propostos para a predição da atividade larvicida, executar o *docking* destes compostos e avaliar os resultados em relação aos itens (a) e (c). Detalhes adicionais da metodologia empregada estão descritos em Métodos.

### 5.4.2 – Métodos

Nakagawa *et al.*<sup>27</sup> utilizaram 41 compostos dos 46 sintetizados e avaliados contra *L. decemlineata*, realizando algumas análises através dos modelos de QSAR. Desenvolveram vários modelos, sendo que de interesse para este estudo é a equação principal reproduzida a seguir.

$$\text{pLD}_{50} = + 0,781 (\pm 0,227) \log P + 1,271 (\pm 0,494) \sigma^{\text{orto}} - 0,281 (\pm 0,172) \Delta V_w^{\text{orto}} - 0,375 (\pm 0,163) \Delta V_w^{\text{meta}} - 0,897 (\pm 0,140) \Delta V_w^{\text{para}} - 2,186 (\pm 0,522) I_{2,6} - 1,004 (\pm 0,202) I_{2,3,5} + 4,073 (\pm 0,650)$$

$$n = 41 \quad R = 0,959 \quad s = 0,275 \quad F = 54,48 \quad p < 0,0001 \quad Q^2 = 0,883 \quad SPress = 0,334 \\ SDEP = 0,303 \quad (5.4-1)$$

onde  $\log P$  é o coeficiente de partição octanol/água,  $\Delta V_w$  é o parâmetro de Bondi para o volume de van der Waals,  $\sigma^{\text{orto}}$  é a constante de Hammett na posição *orto*,  $I_{2,3,5}$  e  $I_{2,6}$  foi utilizado para caracterizar substituições, respectivamente, *orto-meta* e *orto*.

Foi incorporado um descritor sensível às variações conformacionais em substituição a um dos parâmetros da Eq. (5.4-1), mas com os descritores selecionados não foi possível validar o modelo de QSAR. Partiu-se então para a construção de um modelo de QSAR com novos parâmetros seguindo o procedimento descrito na Seção 4.

Montou-se uma matriz com 40 descritores de natureza estérica e eletrônica. A seleção de descritores foi feita pelo algoritmo genético do programa BuildQSAR. O presente trabalho se utiliza também dos métodos independentes de regressão MLR e SVM na tentativa de obter uma equivalência na predição interna e externa.

Os 46 compostos foram divididos em conjunto de treinamento ( $n = 37$ ) e 9 compostos como *pseudo blind set*.

O modelo de QSAR proposto é composto pela componente do tensor de polarizabilidade estática  $\alpha_{yy}$  da molécula, pela componente  $\mu_x$  do momento de dipolo e por um descritor do programa DRAGON, uma função de distribuição radial balanceada pela polarizabilidade atômica, RDF095p (*Radial Distribution Function weighted on atom polarizability*), cujos valores foram calculados para uma região distante do centro geométrico com  $r = 9,5\text{\AA}$ . As características do descritor RDF foram dadas na seção 5.1 Eq.(5.1-2). A seguir é feita uma breve abordagem de aspectos teóricos dos descritores  $\alpha_{yy}$  e  $\mu_x$ .

A polarizabilidade representa a facilidade com que uma nuvem eletrônica de uma molécula pode ser deformada por uma carga próxima através de seu campo elétrico. Quanto maior a polarizabilidade da molécula maior é o dipolo induzido e a magnitude da atração eletrônica. Em geral, a polarizabilidade aumenta com o aumento da nuvem eletrônica.

Utilizando o enfoque matemático, a polarizabilidade do dipolo elétrico é uma medida da resposta linear da densidade eletrônica na presença de um campo elétrico infinitesimal,  $F$ , e representa uma variação de segunda ordem na energia  $E$ , representada pelo tensor de polarizabilidade conforme a equação<sup>99</sup>

$$\alpha_{a,b} = -\left(\frac{\partial^2 E}{\partial F_a \partial F_b}\right) \quad a,b = x, y, z \quad (5.4-2)$$

e o valor médio dado por

$$\langle a \rangle = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \quad (5.4-3)$$

A polarizabilidade tem a dimensão de volume, sendo da ordem de  $(\hbar^2 / me^2)^3$ , onde  $\hbar$  é a constante de Planck racionalizada ( $h/2\pi$ ),  $e$  a carga do elétron e  $m$  a sua massa. O campo elétrico infinitesimal  $F$  pode ser atribuído, por exemplo, a átomos do receptor, ou produzido pelas cargas parciais e momentos de dipolo induzidos pelo próprio ligante.

O vetor momento de dipolo elétrico permanente de um sistema de estado não perturbado é caracterizado pela ausência de um campo elétrico, na verdade, inteiramente independente deste. A aplicação de um campo dá origem a um momento de dipolo elétrico induzido adicionalmente ao dipolo permanente. O momento de dipolo induzido  $\mu_{ind}$  está relacionado ao campo elétrico  $\mathcal{E}$  por<sup>100</sup>

$$\mu_{ind} = \alpha \mathcal{E} \quad (5.4-4)$$

onde  $\alpha$  corresponde à polarizabilidade do átomo ou da molécula definida anteriormente.

O descritor  $\mu_x$  está relacionado com o momento de dipolo da molécula na direção  $x$ . O momento de dipolo elétrico, medido em debye, é caracterizado por cargas  $q_i$  associadas aos átomos das moléculas e com componentes localizadas numa posição  $\mathbf{r}_i(x,y,z)$  dado por

$$\mu = \sum_i q_i \mathbf{r}_i \quad (5.4-5)$$

As forças atrativas de van der Waals têm três componentes, cada uma com diferentes origens na molécula constituinte: são as forças de Keesom, Debye e

London, responsáveis respectivamente, pelas interações dipolo-dipolo, dipolo-dipolo induzido, e interações instantâneas de dipolo induzido. Todas são consideradas forças interativas fracas. Apesar de cada uma delas poder ser tratada matematicamente independente uma da outra, fisicamente todas atuam de forma simultânea na interação entre átomos ou moléculas idênticas. A energia total envolvida é chamada de “interações de atração de van der Waals”, (podendo ocorrer repulsão) e pode ser escrita como segue<sup>101a</sup>:

$$E_{KDL} = E_K + E_D + E_L = -\frac{2}{3} \frac{1}{r^6} \frac{\mu^4}{kT} - 2 \frac{\alpha\mu^2}{r^6} - \frac{3}{4} \frac{\alpha^2 I}{r^6} \quad (5.4-6)$$

onde  $E_{KDL}$  energia total com contribuições das energias, respectivamente de Keesom (ou de orientação), Debye (ou de indução) e London (ou de dispersão);  $r$  o raio de separação ou distância entre moléculas;  $k$  a constante de Boltzmann;  $T$  a temperatura;  $\mu$  dipolo elétrico (v. Eq. (5.4-5));  $I$  o potencial de ionização da molécula. As interações de Keesom e Debye ocorrem entre moléculas que possuem momento de dipolo permanente. Geralmente a contribuição de cada uma nas interações dadas pela Eq. (5.4-6) são muito fracas. Exemplificando, na interação iodeto de hidrogênio, HI,  $E_K = 0,38$  kcal/mol,  $E_D = 0,35$  kcal/mol e  $E_L = 382$  kcal/mol.<sup>101b</sup>

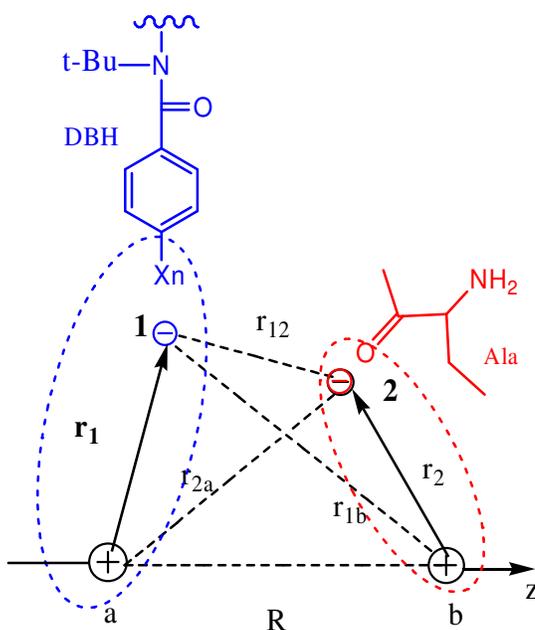
A energia das interações poderá fortalecer ou enfraquecer a união das dibenzoilidrazinas com as cargas dos aminoácidos do receptor, dependendo do valor dos parâmetros acima. Esta avaliação pode ser feita através de simulações quando se conhece a estrutura receptora, procedendo-se a uma análise por *docking* e dinâmica molecular para obter o posicionamento de equilíbrio do ligante em relação ao receptor. Desta forma pode-se obter o valor de  $r$  e conhecer o(s) átomo(s) de interação do receptor com cada uma das dibenzoilidrazinas, sendo então possível calcular cada uma das energias  $E_K$ ,  $E_D$ ,  $E_L$ , e suas contribuições na atividade larvicida. Em geral, as forças de dispersão de London prevalecem de longe sobre as demais (exemplificado acima) quando as moléculas são facilmente polarizadas e mais fracas em caso contrário. Quanto maior a polarizabilidade do átomo ou da molécula, maiores serão as forças de dispersão de London.

Para interações não polares, e que envolvem átomos diferentes, pode-se avaliar a energia de interação ligante-receptor pela relação aproximada<sup>101c</sup> obtida por London como

$$E = -\frac{3}{2} \frac{\alpha_1 \alpha_2}{r^6} \frac{I_1 I_2}{I_1 + I_2} \quad (5.4-7)$$

onde  $\alpha_1$  e  $\alpha_2$  são as polarizabilidades e  $I_1$  e  $I_2$  os potenciais de ionização dos dois átomos envolvidos na interação. Para estimativas,  $r$  pode ser obtido através de dados cristalográficos disponíveis ou estimado pelos resultados de simulações por modelação molecular.

Os detalhes do formalismo matemático contido na dedução da Eq. (5.4-7) pode ser encontrado em Fluege<sup>102a</sup> e Schiff<sup>102b</sup> e aqui resumido para auxiliar na compreensão dos valores que serão estimados pelo uso desta equação após os resultados de *docking*, em especial, para os halogênios. Considere-se a posição  $\mathbf{r}_1$  ( $x_1, y_1, z_1$ ) do elétron **1** relativo ao núcleo **a** e  $\mathbf{r}_2$  ( $x_2, y_2, z_2$ ) a posição do elétron **2** relativo ao núcleo **b** Figura 47.



**Figura 47.** Ilustração da interação por atração de van der Waals entre um átomo de uma dibenzoilidrazina (DBH) com o átomo de oxigênio de um amino ácido representado por uma alanina (Ala).

A direção  $z$  deve coincidir com a direção do eixo nuclear. Utilizando a aproximação de Born-Oppenheimer, tem-se o hamiltoniano

$$H = H^0 + H' \quad (5.4-8.1)$$

com

$$H^0 = -\frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_2^2) - \frac{e^2}{r_1} - \frac{e^2}{r_2} \quad (5.4-8.2)$$

descrevendo dois átomos independentes, e a interação mútua por

$$H' = \frac{e^2}{R} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{1b}} - \frac{e^2}{r_{2a}} \quad (5.4-8.3)$$

$H'$  será utilizado como energia de perturbação. Se  $H'$  for expandido em séries de potências negativas de  $R$  (isto é, para  $r_1 = R$  e  $r_2 = R$ ) (v. Schiff<sup>102b</sup>), o termo de distância grande advém da interação entre os dois dipolos  $a_1$  e  $b_2$  com momentos de dipolo  $p_1 = -er_1$  e  $p_2 = -er_2$ , respectivamente, a saber:

$$H' = \frac{p_1 p_2}{R^3} - 3 \frac{(p_1 R)(p_2 R)}{R^5} \quad (5.4-8.4)$$

Formulando em termos de coordenadas, obtém-se:

$$H' = \frac{e^2}{R^3} (x_1 x_2 + y_1 y_2 - 2z_1 z_2) \quad (5.4-8.5)$$

Expressão que será usada no cálculo seguinte.

Seja  $u_0(r)$  a função de onda do estado atômico fundamental. A função de onda de ordem zero de todo o sistema é então o produto:

$$U(1,2) = u_0(r_1)u_0(r_2) \quad (5.4-8.6)$$

onde a simetrização foi omitida, uma vez que contribuições de troca tendem exponencialmente a zero com distâncias  $R$  muito grandes e podem, portanto, ser desprezadas. Em aproximação de ordem zero a soma das energias dos dois átomos é a energia do sistema. Em primeira ordem temos que adicionar

$$E' = \langle U | H' | U \rangle = 0 \quad (5.4-8.7)$$

Pode ser verificado que este termo se anula. Tomando por exemplo o primeiro termo da Eq. (5.4-8.5), temos:

$$\frac{e^2}{R^3} \langle U | x_1 x_2 | U \rangle = \frac{e^2}{R^3} \langle u_0 | x | u_0 \rangle^2 = \frac{e^2}{R^3} \left\{ \int d\tau u_0^2(r) x \right\}^2, \quad (5.4-8.8)$$

sendo que estas integrais descrevendo momentos de dipolo de estados atômicos com simetria esférica de fato se anulam.

A energia de perturbação de segunda ordem é dada por:

$$E'' = \sum_n \frac{|\langle 0 | H' | n \rangle|^2}{E_0 - E_n}, \quad (5.4-8.9)$$

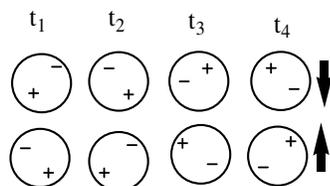
onde a soma se estende sobre todos os estados excitados, e 0 se refere ao estado fundamental. Como  $E_n > E_0$ , todos os denominadores da soma são negativos de modo que  $E'' < 0$  surgindo assim uma atração. Os elementos da matriz dependem apenas de R através do fator constante  $R^{-3}$  na Eq. (5.4-8.5). Assim,  $E''$  deve ser da forma:

$$E'' = -\frac{C}{R^6} \quad (5.4-8.10)$$

onde C é uma constante positiva, chamada de constante de Hamaker, dada por:

$$C = \frac{3I_1 I_2}{2(I_1 + I_2)} \alpha_1 \alpha_2 \text{ o que nos remete à Eq. (5.4-7).}$$

O potencial de ionização mede a energia mínima de remoção do elétron de um átomo isolado ou molécula (no seu estado vibracional fundamental) na fase gasosa, e a polarizabilidade elétrica de dipolo é a resposta de um sistema atômico à ação de um campo elétrico externo. Moléculas que não possuem momento de dipolo permanente, como é caso das apolares, possuem flutuações no posicionamento dos elétrons e com isto variações na densidade eletrônica, de forma que são gerados momentos de dipolo, mas que se modificam rapidamente no valor e direção. Estes dipolos transientes de uma molécula induzem momentos de dipolo momentâneos nas moléculas vizinhas, resultando numa atração temporária entre elas (Figura 48).<sup>103</sup> As interações entre dois dipolos também depende da polarizabilidade da molécula alvo (um amino ácido por ex.) de modo que ela determina o quanto um dipolo pode ser prontamente induzido por outra molécula.



**Figura 48.** Representação da formação de dipolos transitórios entre duas moléculas nos lapsos de tempo  $t_1$  a  $t_4$ .

### **Docking**

*Docking* é uma ferramenta importante em biologia molecular estrutural e em química medicinal no desenvolvimento de novas drogas, (ou, no caso presente, inseticidas) assistido por computador. O objetivo maior do *docking* ligante-proteína é prever formas predominantes de interação do ligante (dibenzoilidrazinas) com a proteína. A rigor, é a formulação de uma hipótese de interação do ligante com o alvo, e é neste sentido que este estudo contribui.

A qualidade do resultado, isto é, o sucesso obtido no *docking*, em geral é apontado pela literatura com a obtenção de um desvio médio quadrático (RMSD – *Root Mean Square Deviation*) em relação à posição das coordenadas cartesianas dos grupos atômicos menor do que 2 Å.<sup>104</sup> Dito de outro modo, é uma medida geométrica entre a estrutura experimental e a estrutura inserida por *docking* considerada como predita. A resolução do complexo ligante-proteína recomendado deve ser inferior a 2,5 Å. É comum encontrar na literatura trabalhos que não conseguem atender à resolução recomendada. Assim é o caso do presente estudo, que se utiliza de um complexo do ligante BY106830 (dibenzoilidrazina) com a proteína (código 1R20.pdb)<sup>105</sup> com resolução de 3,0 Å, obtido por Billas *et al.*<sup>96</sup>, e que tem sido base para vários outros pesquisadores em diferentes tipos de análises tais como: (i) QSAR-3D por Hormann *et al.*<sup>106</sup>. (ii) Wheelock *et al.*<sup>107</sup>, talvez um dos estudos mais abrangentes até o presente envolvendo HTS (*High-throughput screening*), bioensaios e modelação molecular envolvendo por sua vez, sequências do domínio ligante (LBD) de *Bombyx mori*, *Heliothis virescens* e *Bemisia tabaci*; (iii) Carmichael *et al.*<sup>108</sup>, cristalografaram a estrutura com código 1Z5X.pdb<sup>105</sup> com o

esteróide ecdisona Ponasterona A como ligante, (resolução do complexo de 3,0 Å) relacionada à espécie hemíptera *Bemisia tabaci* e fazem uma análise comparativa extensiva a outras espécies no domínio ligante de Coleópteras, Lepidópteras e Dípteras.

Swevers *et al.*<sup>109</sup> fizeram estudo experimental obtendo o domínio ligante SeEcR-LBD de *S. exigua* depositado no GenBank sob código de acesso A59205, informação que possibilitou a realização de parte do presente estudo.

Foi utilizado o programa BioMedCACHe<sup>110</sup> que usa a pontuação (*scoring*) PMF (*potentials of mean force*) cuja teoria é resumida a seguir.<sup>111</sup> O objetivo do *scoring* é identificar o correto modo de ligação com a menor energia, assumindo-se que a conformação é também a da observação experimental, ou seja, a bioativa.

A energia livre de interação proteína-ligante (PMF) entre um átomo da proteína do tipo *i* e átomos do ligante do tipo *j* pode ser dada por:

$$A_{ij}(r) = -k_B T \ln \left[ f_{Vol-corr}^j(r) \frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}} \right] \quad (5.4-9)$$

onde  $k_B$  é a constante de Boltzmann,  $T$  a temperatura absoluta, e  $r$  a distância entre o par de átomos  $ij$ .  $f_{Vol-corr}^j(r)$  é o fator de correção de volume do ligante,  $\rho_{seg}^{ij}(r)$  é a densidade do número de pares do tipo  $ij$  da proteína que ocorre num certo intervalo de raios indicado por 'seg', e  $\rho_{bulk}^{ij}$  representa a distribuição de  $i$  e  $j$  quando não ocorre nenhuma interação entre  $i$  e  $j$ .  $A_{ij}(r)$  é derivado com uma esfera de referência de 12 Å.

O quociente  $\rho_{seg}^{ij}(r)/\rho_{bulk}^{ij}$  na Eq.(5.4-9) designa a função de correlação de pares (distribuição radial) de uma proteína com átomo de tipo  $i$  com um átomo do ligante tipo  $j$  no complexo proteína-ligante.

A função da pontuação (*scoring function*) é definida como a soma sobre todas as interações do complexo ligante-proteína como

$$PMF\_score = \sum_{ij} A_{ij}(r) \quad e \quad r < r_{cut\_off}^{ij} \quad (5.4-10)$$

onde  $r < r_{cut\_off}^{ij}$  é o raio de corte para os átomos do par  $ij$ . A distância ótima de *cut-off* é dada como 6 Å para interações carbono-carbono e 9 Å para as demais.<sup>111</sup>

Os detalhes para a obtenção da Eq. (5.4.10) podem ser encontrados no Apêndice A da ref.[111]. Os parâmetros do programa BioMedCache<sup>110</sup> foram utilizados com os valores de *default*, exceto os seguintes:

- Número de gerações: 40.000 (valor recomendado para dados com qualidade de publicação).<sup>111,112</sup>
- Campo de força de otimização proteína-ligante: MM3.
- Tipo de *docking*: ligante flexível e proteína com cavidade rígida.
- Distância de seleção dos átomos dos aminoácidos vizinhos ao ligante para a formação da cavidade: 3 Å.

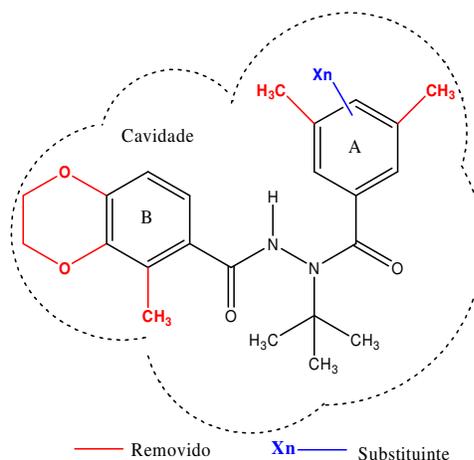
Os aspectos a seguir estão associados ao conceito geral de *docking*, e têm relação com as diferenças apresentadas pelo programa BioMedCache com outros programas do gênero. A afinidade de ligação  $\Delta G_{lig}$  é ditada pela mudança de energia livre, induzida pelo processo de ligação, que tem contribuições entálpicas e entrópicas. Em geral, os fatores mais importantes são os estéricos, eletrostáticos, ligação-hidrogênio, características determinantes na grandeza  $\Delta G_{lig}$ . A dessolvatação favorece entropicamente a ligação do ligante ao receptor indiferentemente da polaridade.<sup>113</sup> O aspecto mais relevante que contrasta com o que acabou de ser citado é que o BioMedCache executa o *docking* com o potencial simplificado pois, como o PMF reflete as energias livres de Helmholtz, as contribuições de solvatação e entrópicas estão implicitamente contidas, incluindo a entropia de dobra da proteína-ligante na conformação do complexo, valores estes, que são derivados dos dados cristalográficos. Também não há necessidade de se conhecer as afinidades de ligação para derivar os potenciais.<sup>114</sup>

O programa se utiliza de um algoritmo genético que codifica uma possível solução de *docking* que corresponde aos chamados cromossomas. Os cromossomas para um ligante flexível contemplam translações, rotações e números de ângulos de diedro que descrevem a posição, orientação e conformação do ligante no sítio ativo. Iniciando com a criação de uma população randômica e através de

operadores de recombinação, o programa preserva as conformações que melhor se ajustam à cavidade, preservando as melhores soluções pelos valores da pontuação (*score*) PMF.

Antes de iniciar o *docking* dos compostos selecionados, foi feita uma validação do complexo da proteína EcR-USP (*Ecdysone Receptor – Ultraspiracle*) complexada com a dibenzoilidrazina BYI06830, código do complexo 1R20.pdb.<sup>105</sup>

As dibenzoilidrazinas foram construídas dentro da cavidade, partindo do cristal da estrutura BYI06830 e modificando-a conforme os grupos funcionais Xn de cada composto, gerando uma população proteína-ligante (v. Figura 49). Dos compostos selecionados foram avaliados os quesitos (a) até (c) citados na Introdução da seção 5.4



**Figura 49** Dibenzoilidrazina BYI06830 alojada na cavidade da proteína 1R20 e as modificações feitas no cristal (em vermelho: grupos eliminados), acrescentando os substituintes halogenados Xn.

Em estudos desta natureza é conveniente utilizar-se de uma metodologia complementar, uma segunda ferramenta para avaliar a qualidade do resultado ligante-receptor ou, dito de outra forma, avaliar a predição da afinidade de ligação. Utilizou-se a metodologia proposta por Wang *et al.*<sup>115</sup> através do programa X-score por eles desenvolvido. O programa é utilizado para extrair resultados pós-*docking* obtidos com qualquer outro aplicativo. Retira-se a estrutura do ligante do complexo ligante-proteína, e as seguintes estruturas devem estar contidas no diretório: a proteína (pdb), o ligante de referência (no caso utilizou-se o cristal BYI06830) e a

estrutura a ser avaliada. O programa foi testado pelos seus autores com 200 complexos ligante-proteína com resolução melhor que 3,0Å e reproduzindo energias de ligação com desvio de 2,2 kcal/mol a 2,0 kcal/mol. Os algoritmos e detalhes técnicos podem ser encontrados na ref.[115], porém, resumem-se a seguir as principais equações que geram os dados de saída do programa X-score, que serão utilizados no estudo. Assume-se neste método que a energia livre global que ocorre no processo de ligação ligante-proteína pode ser obtida pelos termos seguintes:

$$\Delta G_{lig} = \Delta G_{vdw} + \Delta G_{H-lig} + \Delta G_{deform} + \Delta G_{hidrofobico} + \Delta G_0 \quad (5.4-11)$$

onde  $\Delta G_{vdw}$  representa a energia de interação de van der Waals entre o ligante e a proteína;  $\Delta G_{H-lig}$  refere-se à ligação de hidrogênio entre ligante e proteína;  $\Delta G_{deform}$  é o efeito de deformação;  $\Delta G_{hidrofobico}$  é relativo ao efeito hidrofóbico;  $\Delta G_0$  é uma constante de regressão a qual implicitamente inclui a perda de entropia causada pelo efeitos de rotação e translação.

As interações de van der Waals são obtidas empregando a equação de Lennard-Jones que reflete o balanço das repulsões de curto alcance, e as forças de dispersão de longo alcance:

$$VdW = \sum_i^{ligante} \sum_j^{proteina} VdW_{ij} = \sum_i^{ligante} \sum_j^{proteina} \left[ \left( \frac{d_{ij,0}}{d_{ij}} \right)^8 - 2 \left( \frac{d_{ij,0}}{d_{ij}} \right)^4 \right] \quad (5.4-12)$$

onde  $d_{ij,0} = r_i + r_j$  representa a soma dos raios de van der Waals dos átomos  $i$  e  $j$ .

A interação de ligação-hidrogênio entre o ligante e a proteína é calculada somando todas as ligações hidrogênio:

$$HB = \sum_i^{ligante} \sum_j^{proteina} HB_{ij} \quad (5.4-13)$$

Durante a ligação, tanto o ligante como a proteína serão compelidos a uma determinada conformação diferente dos respectivos estados livres em meio solvente, o que provoca mudanças entrópicas, e que traz implicações negativas que precisam ser superadas no processo de ligação. Descrito de forma sucinta, a variação entrópica está associada diretamente ao número de rotações livres que gerarão uma

variação dos estados conformacionais. Por conveniência computacional, as rotações são contadas assumindo a partilha de cada átomo ligante por:

$$RT = \sum_i^{\text{ligante}} RT_i \quad (5.4-14)$$

onde  $RT_i$  assume valores por exemplo de 0 ou 0,5 ou 1 dependendo se, respectivamente, não houver envolvimento rotacional, apenas uma rotação, ou mais de um rotor. Segundo os desenvolvedores do X-score, o fator de deformação expresso por  $RT_i$  faz uma avaliação apenas aproximada, porém, o algoritmo auxilia na qualidade da função de *scoring*.

A interação do ligante e da proteína ocorre sob efeito de um processo de dessolvatação e sob ação de mudanças na entropia como também na entalpia. Um dos resultados é que os grupos não polares tendem a se favorecer entre si, referido como o ‘efeito hidrofóbico’. O programa se utiliza de dois algoritmos a saber: o algoritmo hidrofóbico de superfície (*HS*) e o hidrofóbico de contacto (*HC*).

O algoritmo hidrofóbico de superfície é assumido como sendo proporcional à superfície hidrofóbica enterrada do ligante, descrito pela superfície acessível do solvente (*SAS*). *HS* é dado por:

$$HS = \sum_i^{\text{ligante}} SAS_i \quad (5.4-15)$$

O efeito hidrofóbico por contacto (*HC*) é calculado somando os pares de átomos hidrofóbicos formados entre ligante e proteína, dado por:

$$HC = \sum_i^{\text{ligante}} \sum_j^{\text{proteína}} f(d_{ij}) \quad (5.4-16)$$

onde

$$\begin{aligned} f(d) &= 1,0 & d \leq d_0 + 0,5\text{\AA} \\ &= (1/1,5) \cdot (d_0 + 2,0 - d) & d_0 + 0,5\text{\AA} < d \\ &= 0,0 & d > d_0 + 2,0\text{\AA} \end{aligned}$$

A função de distância reflete intuitivamente que a força das interações hidrofóbicas irão atingir o seu máximo quando dois átomos hidrofóbicos formam contacto por van der Waals, e irão diminuir gradualmente com o aumento das distâncias interatômicas.

O último algoritmo que integra o *X-score* é o de ajuste (*matching*), *HM* fornece o ajuste global entre o ligante e a proteína, dado por:

$$HM = \sum_i^{\text{ligand}} \log P_i \cdot HM_i \quad (5.4-17)$$

$HM_i$  é uma variável de indicação. Assume o valor 1 se um átomo hidrofóbico  $i$  estiver num ambiente hidrofóbico, e em caso contrário será 0.  $\log P_i$  é o coeficiente de partição octanol/água relativo à contribuição do átomo  $i$  ao coeficiente de partição  $\log P$  da molécula.

Finalizando, a função de *scoring* para avaliar a afinidade de ligação ligante-proteína é expressa por unidades  $pK_d$  somando todos os termos dos três algoritmos citados anteriormente, resultante de três funções de *scoring*:

$$X\text{-score} = (pK_{d,1} + pK_{d,2} + pK_{d,3})/3 \quad (5.4-18)$$

### Homologia – Proteína SeEcR complexada com a DBH BYI06830

Partiu-se da sequência dos aminoácidos do complexo ligante-proteína 1R20 de *Heliothis virescens* e da sequência dos aminoácidos da *Spodoptera exigua* referenciada na literatura.<sup>109</sup> Utilizando o servidor UniProt<sup>116a</sup> obteve-se a sequência dos aminoácidos para construir a proteína através do servidor I-TASSER.<sup>116b,c</sup> Empregando o programa LovoAlign<sup>117</sup>, foi feita a sobreposição do complexo 1R20 com a proteína construída chamada de 'SeEcR'. Uma vez verificado o bom ajuste entre proteínas foi feita a transferência manual das coordenadas do ligante BYI06830, obtendo-se finalmente a estrutura do complexo SeEcR-LBD + BYI06830. Este complexo foi submetido a otimização de geometria usando o método MM3 do programa BioMedCache.<sup>110</sup> Após, seguem-se as etapas descritas anteriormente em *Docking* (Sub-seção 5.4.2).

No estudo de *docking* dos compostos halogenados resultou um *scoring* PMF para cada um deles. São feitas duas avaliações de correlação entre a atividade

larvicida e o PMF-score. A primeira se refere ao resultado obtido pelo *docking* apenas das dibenzoilidrazinas halogenadas, empregando o valor pLD<sub>50</sub> obtido experimentalmente por Nakagawa *et al.*<sup>27</sup> e o correspondente valor PMF-score obtido neste estudo. A segunda avaliação é feita com o conjunto das dibenzoilidrazinas propostas, onde se utilizam os valores preditos pelos modelos de QSAR MLR e SVM .

Segundo a literatura<sup>111,114</sup>, a correlação geralmente obtida se situa entre 0 e 0,37. Casos muito especiais geraram R<sup>2</sup> acima de 0,9. Alguns detalhes do método com relação à regressão serão complementados na sub-seção 5.4.3 ‘Resultados e Discussão’.

## Compostos Halogenados

Sob aspecto farmacológico para humanos, a halogenação é utilizada para aumentar a estabilidade metabólica e a lipofilicidade, aumentando a afinidade de ligação ao receptor da proteína alvo. Contudo, algumas vezes causa efeitos indesejáveis por aumentar a toxicidade, aspecto este de interesse para o desenvolvimento de larvicidas ou, usando o termo geral, inseticidas.

Segundo a literatura, estudos mais específicos sobre os halogênios e sua respectiva ação intermolecular não-covalente somente receberam maior atenção a partir do início da década de 1990.<sup>118</sup>

Como já citado anteriormente, empregou-se *docking* neste estudo com a principal finalidade de avaliar alguns dos compostos halogenados. Nos estudos de *docking* com dibenzoilidrazinas citados<sup>97,98</sup> não havia halogênios presentes, assim como também não foram encontradas informações específicas sobre a interação intermolecular de substâncias halogenadas desta classe de substâncias. Neste aspecto este estudo faz também uma contribuição.

Para que os resultados gerados por *docking* possam ser avaliados, foram obtidos valores de outros estudos, e criada uma base de referência usando dados teóricos e experimentais da literatura<sup>118,119</sup> baseada na área biológica. Vale salientar que o estudo de interação intermolecular de halogênios é de grande interesse na área de desenvolvimento de novos materiais para engenharia.<sup>120</sup>

A seguir foram construídas quatro tabelas com os parâmetros de interação intermolecular não-covalente, seus valores e/ou características extraídas da literatura<sup>118,119</sup> e que formarão a base para uma avaliação comparativa dos resultados obtidos por este estudo.

**Tabela 26.** Interação não-covalente intermolecular de halogênios<sup>118</sup>

C-X...Y	PDB's <sup>a</sup>	Contatos <sup>b</sup>	X...Y <sub>c,d</sub>
C-Cl...O	85	139	3,040
C-Br...O	60	90	3,132
C-I...O	50	76	3,269
C-Cl...N	27	37	3,082
C-Br...N	8	8	3,121
C-I...N	5	7	3,193
C-Cl...S	11	17	3,363
C-Br...S	3	3	3,008
C-I...S	5	5	3,415
<b>C-X...π<sup>e</sup></b>			
C-Cl...π	106	124	3,854
C-Br...π	29	53	3,944
C-I...π	11	12	4,112
<b>C-X...H</b>			
C-F...H	31	494	2,577
C-Cl...H	20	584	2,803
C-Br...H	15	300	2,832
C-I...H	8	42	3,123

(a) Número de arquivos PDB; (b) número de contatos; (c) distância média (Å); (d) vide Figura 50; (e) corresponde aos elétrons no orbital π

**Tabela 28.** Interações de halogênios com aminoácidos<sup>119</sup>

X...N,O	Número de Interações	Distância Média (Å)	Ângulo Médio(°)
F...N	6862	3,05	96,09
F...O	7615	2,80	120,55
Cl...N	21394	3,10	92,69
Cl...O	18358	2,90	120,04
Br...N	2827	3,20	101,23
Br...O	2289	3,05	118,91
I...N	2560	3,30	95,88
I...O	2704	3,15	117,96

(a)Vide Figura 50

**Tabela 27.** Distâncias intermoleculares X...Y menores que a soma dos respectivos raios de van der Waals<sup>118</sup>

X...Y	Distância ( Å )
Cl...O	< 3,27
Br...O	< 3,37
I...O	< 3,50
Cl...N	< 3,30
Br...N	< 3,40
I...N	< 3,53
Cl...S	< 3,55
Br...S	< 3,65
I...S	< 3,78

Ângulo C-X...Y >140 °

(a)Vide Figura 50

**Tabela 29.** Interação intermolecular associada a benzenos halogenados<sup>118</sup>

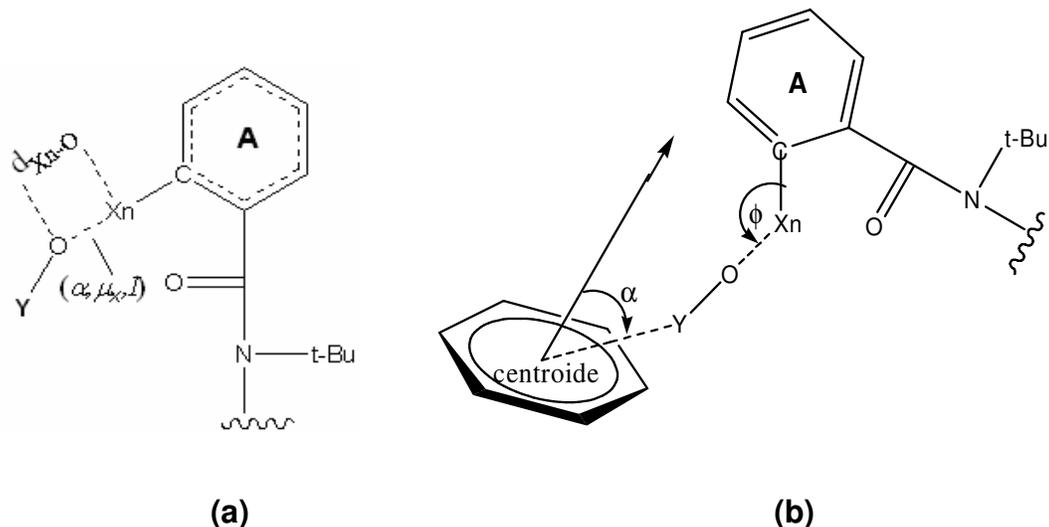
Sistema	Distância d(X...Y) (Å)	Ângulo médio C-X...Y
PhCl...NH <sub>3</sub>	3,193	179,8
PhBr...NH <sub>3</sub>	3,102	179,7
PhI...NH <sub>3</sub>	3,108	179,9
PhI...OH <sub>2</sub>	3,140	179,2
PhI...SH <sub>2</sub>	3,620	177,2
<b>Sistema</b>		
	Distância ( Å )	Ângulo (°)
	X...Centróide <sup>a</sup>	C-X...Centróide <sup>a</sup>
PhCl...Ph	3,233	180,0
PhBr...Ph	3,272	180,0
PhI...Ph	3,423	180,0

(a)Vide Figura 50

Os dados tabelados foram obtidos pelos autores da pesquisa utilizando complexos ligante-proteína halogenados, em que os cristais resolvidos por espectroscopia de raios-X têm resolução de 3,0 Å ou menor.

Ainda fazendo parte da base de referências, considera-se o trabalho de Auffinger *et al.*<sup>121</sup> em estudos com ácidos nucleicos e proteínas, onde se constatou que ligações com halogênios: **(a)** são estabilizantes tanto de intraligações como interligações moleculares, **(b)** afetam a ligação do ligante no dobramento da proteína. Segundo a citada referência, a ligação com halogênios pode ser definida por interações curtas, com distância  $d_{Xn-O}$  menor ou igual à soma dos respectivos raios de van der Waals que variam conforme segue:  $R_{vdW}(F...O) = 2,99\text{Å}$ ,  $R_{vdW}(Cl...O) = 3,27\text{Å}$ ,  $R_{vdW}(Br...O) = 3,37\text{Å}$  e  $R_{vdW}(I...O) = 3,50$ , informações que ratificam as constatadas por Lu *et al.*<sup>118</sup> e listadas na Tabela 27. A Figura 50 ilustra simbolicamente os parâmetros mencionados utilizando uma dibenzoilidrazina parcialmente representada, mostrando a região de interação e que será avaliada por este estudo, onde os substituintes Xn são os halogênios (F, Cl, Br, I) e Y o átomo do aminoácido. Segundo Lu *et al.*<sup>118</sup> os ângulos C-X...Y são maiores do que 140°.

Como última complementação da base de referências, Gonnade *et al.*<sup>122a</sup> em um artigo de revisão, mostram em gráfico uma nuvem de distâncias do cloro com relação a oxigênios hibridizados  $sp^2$ , variando, aproximadamente de 2,8 Å a 4,0 Å. O levantamento foi feito com algumas centenas de compostos cristalografados do banco de dados CSD (*Cambridge Crystall Structure Database*).<sup>122b</sup>



**Figura 50. (a)** Representação esquemática da região de interação considerando dibenzoilidrazinas halogenadas substituídas na posição *orto* (representada), *meta* e *para*.  $d_{Xn-O}$  varia de 2,99 Å para F...O a 3,50 Å para I...O. O – Y pode ser uma carbonila, hidroxila ou carboxilato quando Y é um carbono; um fosfato quando Y é fósforo; ou um sulfato quando Y é enxofre<sup>121</sup>. **(b)** Representação da interação molecular e do posicionamento do anel benzênico em base à ilustração da ref.[118] e adaptada para dibenzoilidrazinas.

A região de interação na figura é representada pela polarizabilidade, momento de dipolo e potencial de ionização ( $\alpha_{yy}$ ,  $\mu_x$ ,  $I$ ), parâmetros que serão comentados novamente mais adiante.

### 5.4.3 – Resultados e Discussão

#### 5.4.3.1 – QSAR e Descritores

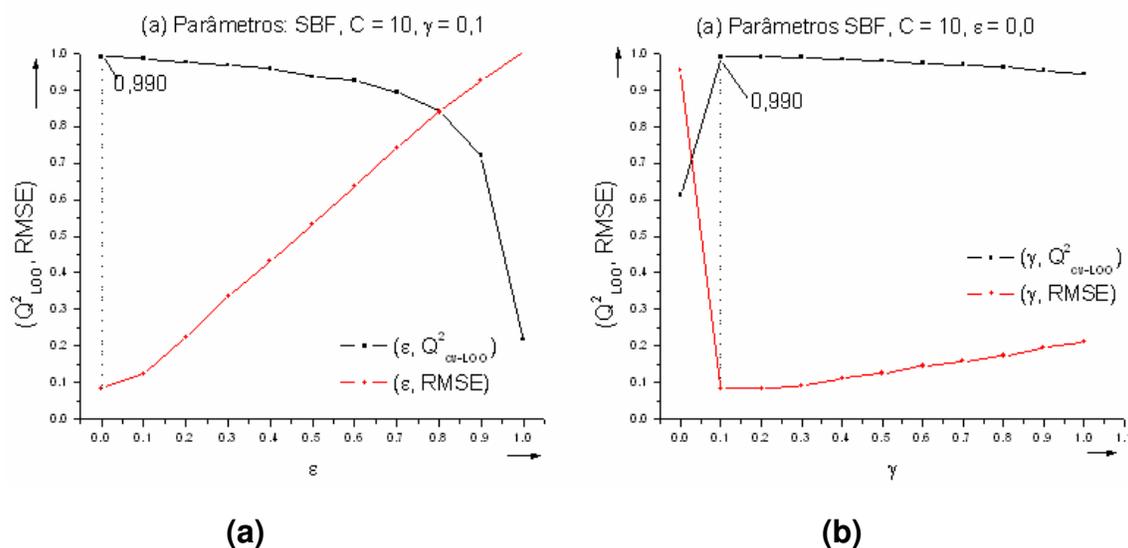
Como já citado, o modelo de QSAR deste estudo é integrado pelo descritor RDF095p, o tensor de polarizabilidade  $\alpha_{yy}$ , e o momento de dipolo na direção x,  $\mu_x$ . Portanto, o modelo de QSAR foi basicamente construído por três parâmetros ligados de alguma forma à polarizabilidade, aspecto importante nas interações químicas, tanto intramolecular como intermolecular. A equação a seguir descreve o modelo de QSAR obtido com o conjunto de treinamento ( $n = 37$ ):

$$pLD_{50} = -0,053 (\pm 0,003) \alpha_{yy} + 0,242 (\pm 0,009) \mu_x - 0,288 (\pm 0,011) RDF095p + 8,461 (\pm 0,121)$$

$$n = 37 \quad R^2 = 0,997 \quad s = 0,052 \quad F = 3270 \quad p < 0,0001 \quad Q^2 = 0,996 \quad SPress = 0,060 \\ SDEP = 0,057 \quad MSE = 0,002 \quad (5.4-19)$$

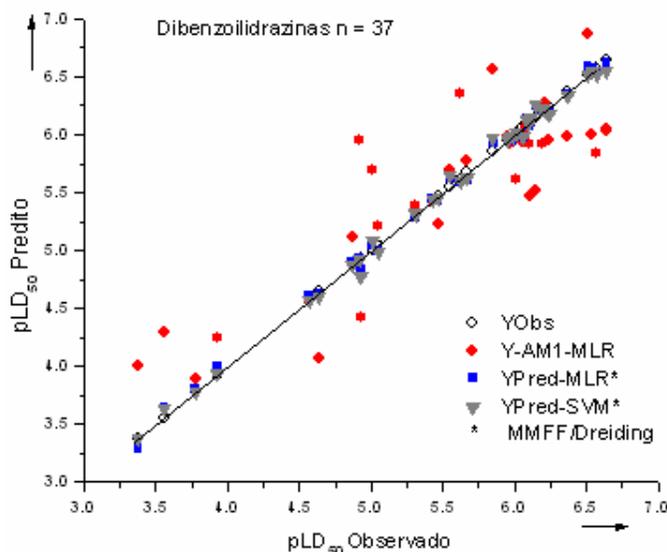
A matriz de correlação entre descritores tem os valores seguintes:  $\mu_x/\alpha_{yy}(0,273)$ ;  $\mu_x/RDF095p(-0,227)$ ;  $\alpha_{yy}/RDF095p(0,251)$ . Correlação entre descritores e atividade larvicida:  $pLD_{50}/\alpha_{yy}(-0,409)$ ;  $pLD_{50}/\mu_x(0,596)$ ;  $pLD_{50}/RDF095p(-0,822)$ .

Empregando os mesmos descritores e o mesmo conjunto de treinamento ( $n = 37$ ) foi feita uma avaliação do conjunto de dados pelo método SVM buscando resultados equivalentes ao modelo de QSAR-MLR, o que foi possível como pode ser verificado pela Tabela 30 (v. em Anexos, pg. 235), e pelos dados estatísticos de correlação entre modelos de QSAR. O modelo de QSAR construído com SVM ( $n = 37$ ) utiliza o *kernel Radial Basis Function* (RBF). O melhor valor obtido foi com o ajuste dos seguintes parâmetros: **(a)** custo (ou capacidade)  $C=10$ ; **(b)** extensão da função radial RBF,  $\gamma = 0,1$  e **(c)** a função de perda  $\varepsilon = 0,0$  (v. Figura 51).



**Figura 51. (a)**  $Q^2_{cv-LOO}$  em função da variação do parâmetro  $\varepsilon$  mantendo constantes  $C$  e  $\gamma$ ; **(b)**  $Q^2_{cv-LOO}$  em função da variação do parâmetro  $\gamma$  mantendo constantes  $C$  e  $\varepsilon$ . Para o menor RMSE seleciona-se o valor  $Q^2_{cv-LOO}$  correspondente ( $Q^2_{cv-LOO} = 0,990$ ).

A representação gráfica da predição interna dos modelos de QSAR da Eq. (5.4-1) está ilustrada na Figura 52.



**Figura 52.** Comparativo da predição  $pLD_{50}$ : geometria otimizada pelo método semi-empírico AM1 e após análise conformacional com os campos de força MMFF94 ou Dreiding modificado, assim como representação das predições pelos modelos de QSAR-MLR e QSAR-SVM. Correlação entre modelos:  $R = 0,999$ ;  $F = 25923$ ;  $s = 0,032$ ;  $p < 0,0001$ .

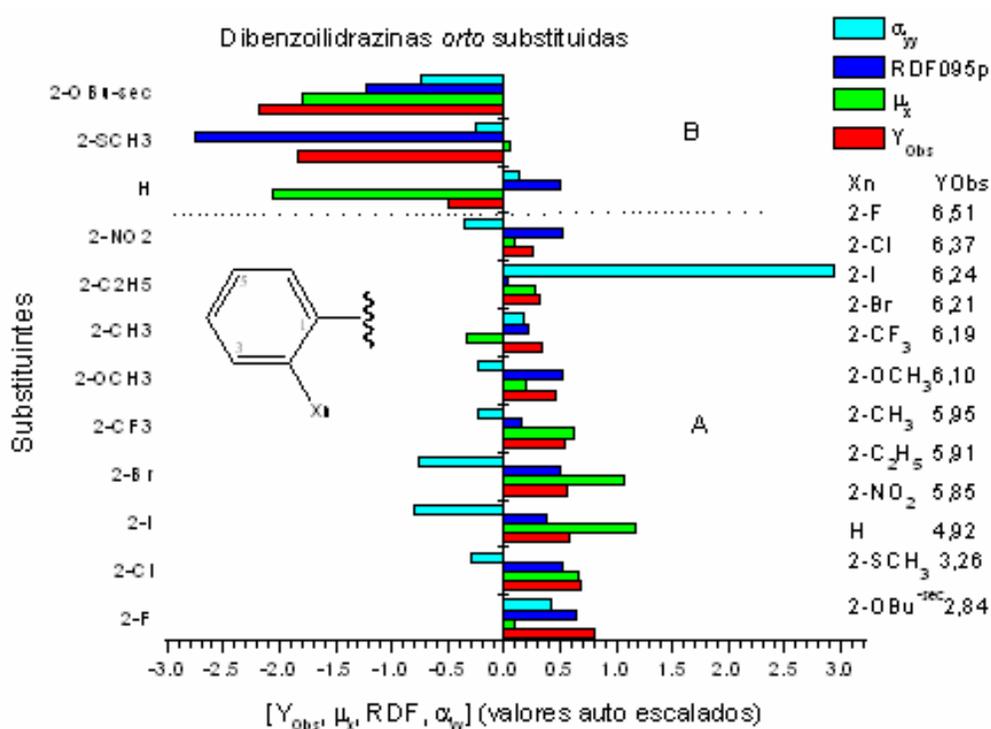
A análise dos descritores e seu peso desempenhado na atividade larvica são dados a seguir conforme as posições *orto*, *meta*, e *para* ocupadas pelos substituintes no anel A da estrutura da figura 34.

### Substituições *orto*

Na posição *orto*, o peso da polarizabilidade molecular é mostrado pelas barras associadas aos três descritores  $\alpha_{yy}$ ,  $\mu_x$  e RDF095p, este último caracterizado por uma polarizabilidade global implícita em cada composto.

Pelo gráfico e pelos valores numéricos da atividade  $pLD_{50}$ , em disposição decrescente incluídos na Figura 53, é possível identificar os substituintes halogênicos (F, Cl, I, Br) na posição *orto* que apresentaram bons resultados na atividade larvica. A grande eletronegatividade do átomo de flúor, a sua fraca polarizabilidade atômica ( $0,557 \text{ \AA}^3$  – é a menor polarizabilidade entre os átomos halogênicos), e as

ligações do carbono  $sp^3$  ( $\delta^+$ ) fortemente polarizadas na direção do flúor ( $\delta^-$ ), são fatores que conferem uma energia de ligação do flúor ao carbono, superior aos outros halogênios.<sup>78</sup> O flúor na posição *orto* apresentou a maior atividade larvica entre as dibenzoilidrazinas *orto*-substituídas. Apesar de não haver um padrão uniforme na distribuição do efeito da polarização envolvendo as três variáveis, pode-se distinguir duas regiões no gráfico: a região **A**, que contempla a atividade larvica mais potente destacando a polarizabilidade global RDF095p e o momento de dipolo  $\mu_x = \sum_i q_i x_i$  como mais atuantes no efeito larvica, e região **B**, onde a atividade decai por uma ação desfavorável destes parâmetros.



**Figura 53.** Influência dos descritores RDF095p,  $\alpha_{yy}$  e  $\mu_x$  na atividade larvica dos compostos *orto* substituídos. O composto D1 (H) não tem substituintes e é utilizado como referência.

A seguir são citados dois trabalhos cujas informações são utilizadas neste estudo para auxiliar no entendimento do modelo de QSAR desenvolvido.

Um trabalho desenvolvido por Fricke<sup>123</sup> em um estudo teórico com átomos da tabela periódica, mostra que há correlação do potencial de ionização  $I$  com a polarizabilidade atômica. Além dos valores por ele próprio obtidos, reproduz os resultados do trabalho por ele citado como T.H.Miller e B. Bederson, Adv. At. Mol. Phys., 13, 1, 1977, mostrando a perfeita correlação dos parâmetros em questão para os halogênios assim como para outros grupos da tabela periódica. No presente estudo esta análise foi estendida para o nível molecular, empregando as dibenzoilidrazinas *orto*, *meta* e *para* substituídas. Os compostos D11 e D12 foram excluídos do estudo de QSAR devido à baixa potência, mas foram preservados para a análise da influência dos parâmetros em questão.

A análise envolvendo 8 dos 11 substituintes na posição *orto*, gerou a equação 5.4-20, obtida com os valores da Tabela 31, onde  $I$  corresponde ao potencial de ionização expresso em função das componentes de polarizabilidade tal como definido pela equação 5.4-3.

**Tabela 31.** Valores para o estudo  $I = f(\alpha_{xx}, \alpha_{yy}, \alpha_{zz})$

MolID <sup>a</sup>	Xn	$I$ (eV)	$\alpha_{xx}$ (Å <sup>3</sup> )	$\alpha_{yy}$ (Å <sup>3</sup> )	$\alpha_{zz}$ (Å <sup>3</sup> )	$\langle\alpha\rangle$ (Å <sup>3</sup> )	$\sigma_{Ind}^d$	$\sigma_R^d$
D1 <sup>c</sup>	H	9,788	45,454	31,226	47,362	41,347	-	-
D2 <sup>b,c</sup>	2-F	9,581	45,506	30,074	46,785	40,788	0,52	-0,46
D3 <sup>b,c</sup>	2-Cl	9,565	46,394	32,895	46,362	41,884	0,47	-0,24
D4 <sup>b,c</sup>	2-Br	9,563	46,073	40,888	39,776	42,245	0,44	-0,22
D5 <sup>c</sup>	2-I	9,582	49,218	47,282	41,597	46,032	0,39	-0,25
D6 <sup>b</sup>	2-CF <sub>3</sub>	9,696	47,555	43,210	33,485	41,417	0,42	0,10
D7 <sup>b</sup>	2-NO <sub>2</sub>	9,719	49,496	42,365	34,634	42,165	0,56	0,22
D8 <sup>b</sup>	2-CH <sub>3</sub>	9,520	48,893	33,297	49,175	43,788	0,04	-0,15
D9 <sup>b</sup>	2-C <sub>2</sub> H <sub>5</sub>	9,454	50,684	36,153	51,564	46,134	0,05	-0,11
D10	2-OCH <sub>3</sub>	9,274	49,985	42,176	38,882	43,681	0,29	-0,43
D11 <sup>b</sup>	2-OBu <sup>sec</sup>	9,329	55,256	47,208	50,097	50,854	-	-
D12	2-SCH <sub>3</sub>	8,260	52,310	40,422	46,794	46,509	0,23	-0,16

(a) Estruturas com geometria otimizada com o método semi-empírico AM1; (b) compostos da Eq.(5.4-20); (c) compostos da Eq. (5.4-21); (d) valores conforme Hansch *et al.*<sup>52</sup> e ref. [78]

$$I = + 0,017 (\pm 0,008) \alpha_{xx} - 0,020 (\pm 0,005) \alpha_{yy} - 0,025 (\pm 0,003) \alpha_{zz} + 10,575 (\pm 0,178)$$

$$n = 8 \quad R = 0,998 \quad s = 0,010 \quad F = 353 \quad p < 0,0001$$

$$(5.4-20)$$

Utilizando-se apenas os 4 substituintes halogênicos com inclusão do composto D1 não substituído pôde-se obter a regressão linear dada a seguir:

$$I = -0,272\alpha_{xx} + 0,099\alpha_{yy} + 0,134\alpha_{zz} + 12,756$$

$$n = 5 \quad R = 0,999 \quad s = 0,003 \quad F = 1827 \quad (5.4.21)$$

Tomando o flúor novamente por base comparativa, quando ligado a um átomo de carbono insaturado ou a um anel aromático, que é o caso presente, verifica-se um efeito de atração eletrônica por efeito indutivo  $\sigma_{\text{Ind}}$  e um efeito doador por ressonância  $\sigma_{\text{R}}$  superior aos outros halogênios<sup>78</sup> verificável pelos valores das constantes de Hammett  $\sigma_{\text{Ind}}$  e  $\sigma_{\text{R}}$  para alguns dos compostos da Tabela 31.

A correlação de  $I$  e  $\alpha$  mostra ser menor quando  $\alpha$  é utilizado no lugar de seus componentes  $\alpha_{xx}$ ,  $\alpha_{yy}$  e  $\alpha_{zz}$ :

$$I = -0,032 (\pm 0,019) \alpha + 10,969 (\pm 0,819)$$

$$n = 8 \quad R = 0,866 \quad s = 0,068 \quad F = 18 \quad (5.4.22)$$

Na verdade isto não constitui nenhuma anormalidade, apesar dos parâmetros aqui empregados para a análise, terem sido calculados com o método semi-empírico AM1, o que certamente gera resultados mais imprecisos. Torna-se oportuno citar uma situação análoga a esta, ocorrida no estudo de QSAR feito por Fujiwara *et al.*<sup>124a</sup> com o parâmetro LogP (coeficiente de partição octanol/água) para álcoois graxos substituídos. Obtiveram o modelo de QSAR:

$$pID_{50} = 0,678\text{LogP} + 1,475$$

$$n = 11 \quad R = 0,639 \quad s = 0,570 \quad (5.4.23a)$$

onde  $pID_{50}$  representa 50% de potência inibitória de tiopurina metiltransferase.

Empregando calorimetria por titulação com os álcoois, partiram da equação  $\Delta G_p^0 = -RT \ln P$ , onde  $R$  é a constante universal dos gases,  $T$  a temperatura,  $\Delta G$  representa a energia de Gibbs, e  $\ln P$  é obtido em função da componente entálpica ( $\Delta H$ ) e entrópica ( $\Delta S$ ), representado por:

$$\log P = -\Delta H_p^0 / (2,303RT) + \Delta S_p^0 / (2,303R) = P_H + P_S \quad (5.4.23b)$$

O novo modelo de QSAR, agora expresso pelas componentes entálpica e entrópica de log P resultou em:

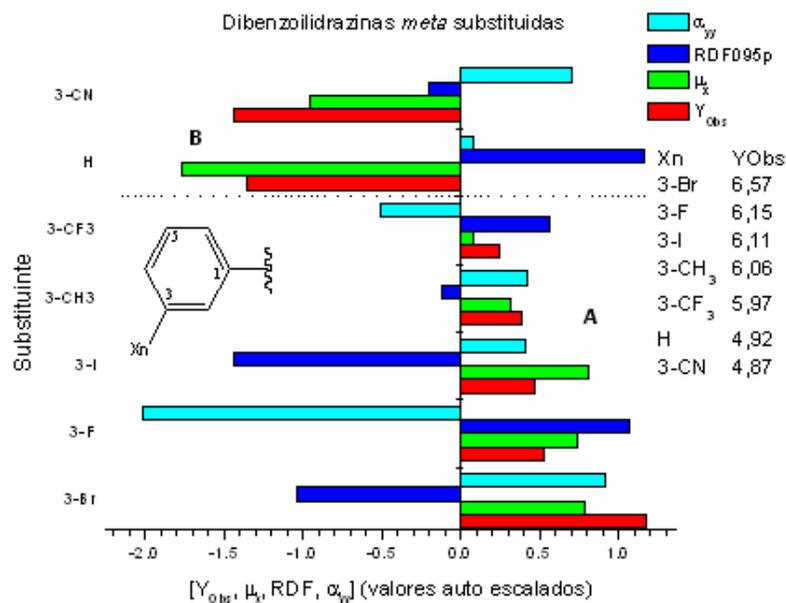
$$pID_{50} = 1,006P_H + 0,335P_s \quad n = 11 \quad R = 0,952 \quad s = 0,241 \quad (5.4-23c)$$

resultado este expressivamente superior ao obtido pela Eq. (5.4-23a), mostrando que a utilização das partes que compõem uma grandeza física podem contribuir com melhores resultados.

Este é também um momento oportuno para chamar a atenção de que o parâmetro LogP pode ser excluído indevidamente por algoritmos de seleção de variáveis de programas de QSAR, quando na realidade, tem importante papel na bioatividade, mas somente constatável se houver o seu desmembramento nas componentes entálpica e entrópica como foi mostrado anteriormente. Além disso, recomenda-se no desenvolvimento de novas drogas, que as componentes entálpica e entrópica de moléculas biológicas sejam estudadas em benefício do desenvolvimento.<sup>124b,c</sup>

### **Substituições *meta***

As dibenzoilidrazinas mono-substituídas na posição *meta* Figura 54, têm nos halogênios os melhores representantes da atividade larvicida, com o bromo superando a ação do flúor tanto na posição *meta* como na posição *orto* do caso anterior.

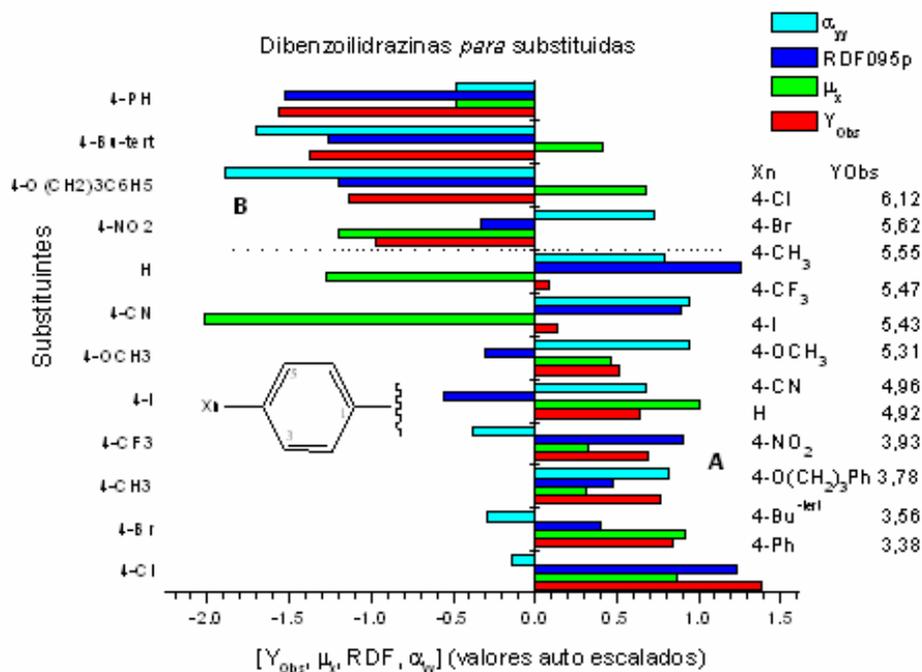


**Figura 54.** Dibenzoilidrazinas *meta* substituídas e influência dos parâmetros  $\alpha_{yy}$ ,  $\mu_x$  e RDF na atividade larvicida. O composto D1 (H) não tem substituintes e é utilizado como referência.

Neste caso verifica-se que na região **A** o momento de dipolo elétrico e a polarizabilidade, respectivamente parâmetros  $\mu_x$  e  $\alpha_{yy}$  contribuem mais na influência da atividade larvicida, e a polarizabilidade global descrita por RDF095p contribui nos compostos substituídos por F e CF<sub>3</sub>. Na região **B** encontra-se o substituinte ciano, com a menor atividade.

### Substituições *para*

Os compostos halogenados na substituição *para*, Figura 55, são os que mostram maior atividade larvicida em relação aos demais substituintes. De maneira similar ao caso anterior, a região **A** mostra o momento de dipolo elétrico  $\mu_x$  e a polarizabilidade  $\alpha_{yy}$  com maior contribuição. Acima da linha de referência pontilhada, a influência combinada dos três parâmetros, ora favoráveis ora desfavoráveis à atividade, demonstra uma queda de potência dos compostos.



**Figura 55.** Dibenzoilidrazinas *para* substituídas e influência dos parâmetros  $\alpha_{yy}$ ,  $\mu_x$  e RDF na atividade larvicida. O composto D1 (H) não tem substituintes e é utilizado como referência.

Das 12 dibenzoilidrazinas substituídas na posição *para*, foi possível obter uma equação relacionando o potencial de ionização  $I$  com a polarizabilidade estática  $\alpha$  para 7 compostos.

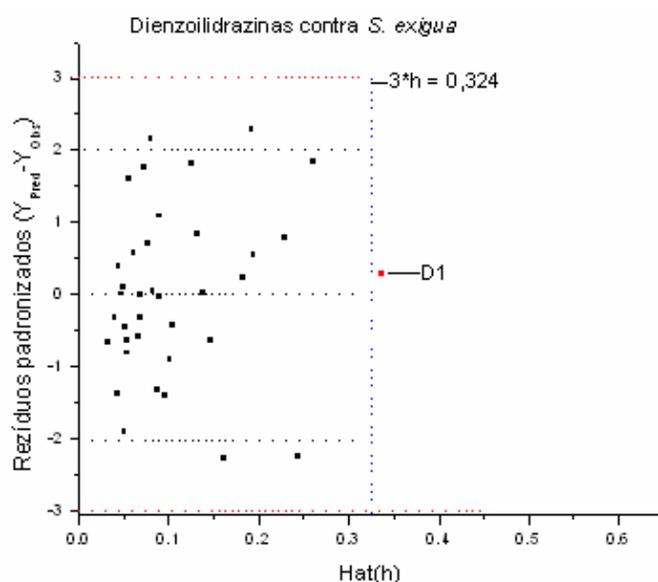
$$I = -0,030 (\pm 0,013) \alpha + 11,014 (\pm 0,589)$$

$$n = 7 \quad R = 0,936 \quad s = 0,087 \quad F = 35 \quad (5.4-24)$$

Os resultados obtidos não possibilitam a formulação de uma hipótese única ou obtenção de um padrão uniforme para todos os compostos e seus substituintes relacionando as propriedades dos descritores à atividade larvicida. Contudo, os resultados são indicadores de que para alguns compostos, destacando-se os halogenados, as interações ocorrem por efeitos de polarização, eletrostáticos, transferência de carga e contribuições dispersivas, concorrendo com constatações de alguns autores<sup>118,119,121</sup> o que confere plausibilidade de uso dos descritores  $\mu_x$ ,  $\alpha_{yy}$  e RDF095p no modelo de QSAR obtido.

### 5.4.3.2 – QSAR – *Outliers* e Domínio de Aplicabilidade

Construindo-se o diagrama de Williams com os resultados de predição obtidos pelo emprego da Eq. (5.4-19) verifica-se que o composto D1 é candidato a *outlier*, Figura 56. Dois fatores são fortemente favoráveis a mantê-lo no conjunto: **(a)** possui baixo resíduo na predição (0,01); **(b)** o composto D1 corresponde à dibenzoilidrazina não substituída, também utilizada como referência em alguns pontos deste estudo sendo que ambos modelos de QSAR, por MLR e SVM, fazem uma predição dos compostos próximo ao valor experimental como pode ser verificado pela Tabela 30 (v. em Anexos, pg. 235). Além disto, excluí-lo do conjunto não traz nenhuma diferença expressiva na equação de QSAR-MLR e nos resultados preditos, apresentando *leverage* alto ( $3h = 0,337$ ), porém, favorável.



**Figura 56.** Diagrama de Williams – Domínio de aplicabilidade das dibenzoilidrazinas ( $n = 37$ ) contra *Spodoptera exigua*. Composto candidato a *outlier*: D1. Valor crítico:  $3 \cdot h = 0,324$ , valor  $3 \cdot h$  (D1) = 0,337 (Vide texto).

### 5.4.3.3 – QSAR – Testes de Robustez e Validação

A robustez do conjunto de compostos  $n = 37$  foi avaliada com os métodos *Y-randomization*<sup>53,61</sup> e *Leave-N-Out*.<sup>61</sup> O teste *Y-randomization* foi realizado com 300 iterações e 10 ciclos, resultando em 3000 iterações. Os valores do *intercept* usados

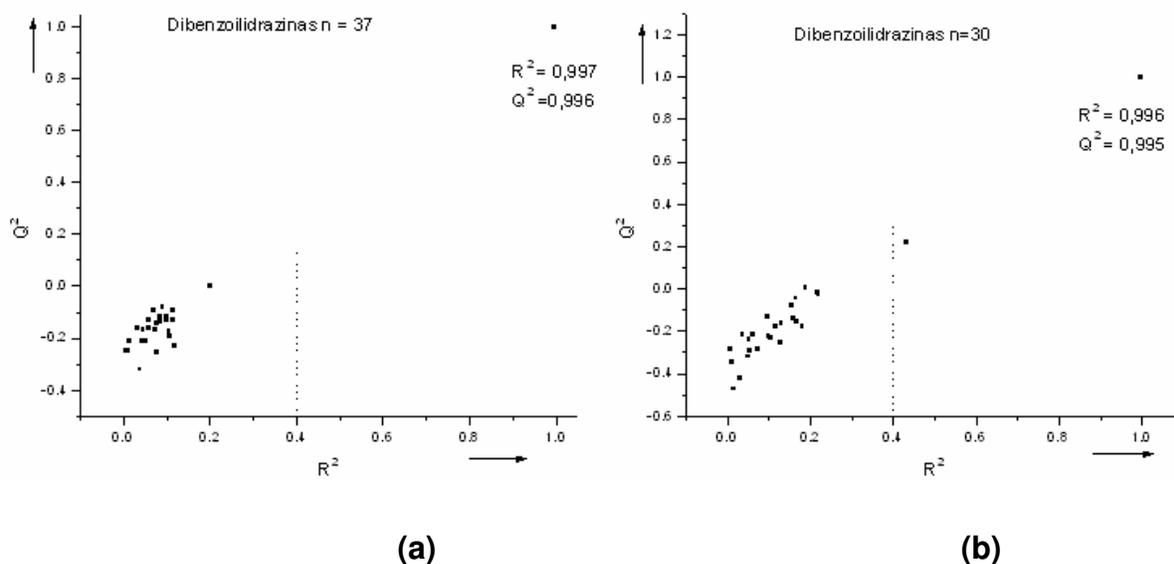
para a aprovação do modelo sempre ocorreram abaixo dos valores limites  $a(R^2) < 0,3$  e  $a(Q^2) < 0,05$ . Segue o valor final do último ciclo:

$a(R^2) = 0,004$  e  $a(Q^2) = -0,273$ .

O método *bootstrap* para a validação interna utilizou 5000 iterações, resultando em um valor final  $Q^2_{boot} = 0,995$  que comparado ao  $Q^2_{LOO} = 0,996$  valida a predição interna da Eq.(5.4-19).

Pelo critério *Leave-n-Out*, foram eliminados 18 compostos (50% do conjunto  $n = 37$ ), obtendo-se o resultado  $Q^2 - Q^2_{LNO} < 0,1$  para 10 repetições.

Empregando o critério de Ericsson-Wold<sup>58</sup>, foram feitas 25 iterações para a determinação de  $R^2$  e  $Q^2$  utilizando o programa da ref.[61], obtendo-se os gráficos da Figura 57, com resultados satisfatórios.



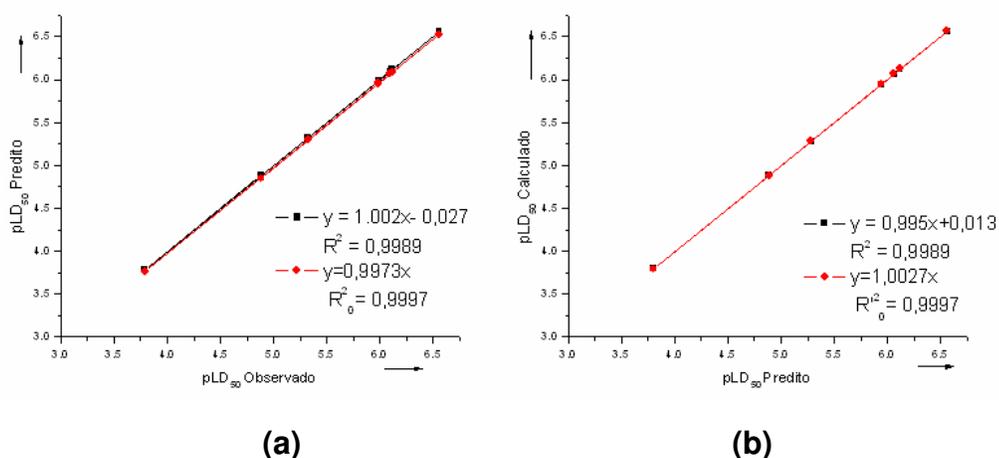
**Figura 57.** Valores  $R^2$  e  $Q^2$  obtidos em função da randomização dos valores  $pLD_{50}$  experimentais para modelos de QSAR (a) MLR com  $n = 37$  e (b) MLR com  $n = 30$ .

Para a validação externa foi desenvolvido o QSAR-MLR usando o conjunto de teste  $n = 7$ :

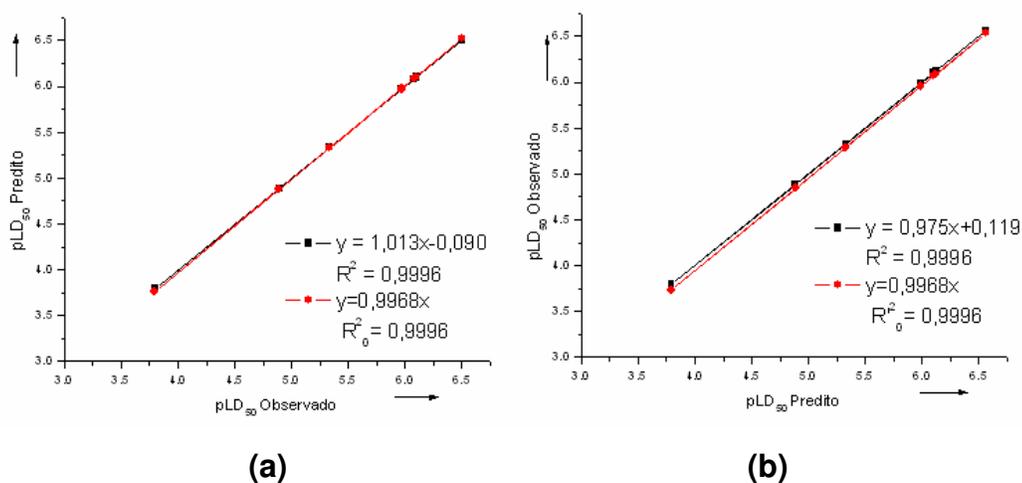
$$pLD_{50} = -0,053 (\pm 0,003) \alpha_{yy} + 0,244 (\pm 0,011) \mu_x - 0,286 (\pm 0,013) RDF095p + 8,460 (\pm 0,142)$$

$$n = 30 \quad R^2 = 0,996 \quad s = 0,057 \quad F = 2218 \quad p < 0,0001 \quad Q^2 = 0,995 \quad SPress = 0,067 \quad SDEP = 0,063 \quad MSE = 0,003 \quad (5.4-25)$$

Aplicando o critério de Golbraikh-Tropsha<sup>64</sup> obtiveram-se os resultados representados nos gráficos da Figura 58 para o modelo de QSAR-MLR, e Figura 59 para o modelo de QSAR-SVM, assim como os parâmetros associados ao critério. Em ambos os casos ocorreram as validações e pode-se verificar graficamente pelos parâmetros de avaliação que há ótima concordância entre os resultados de ambos os modelos de QSAR.



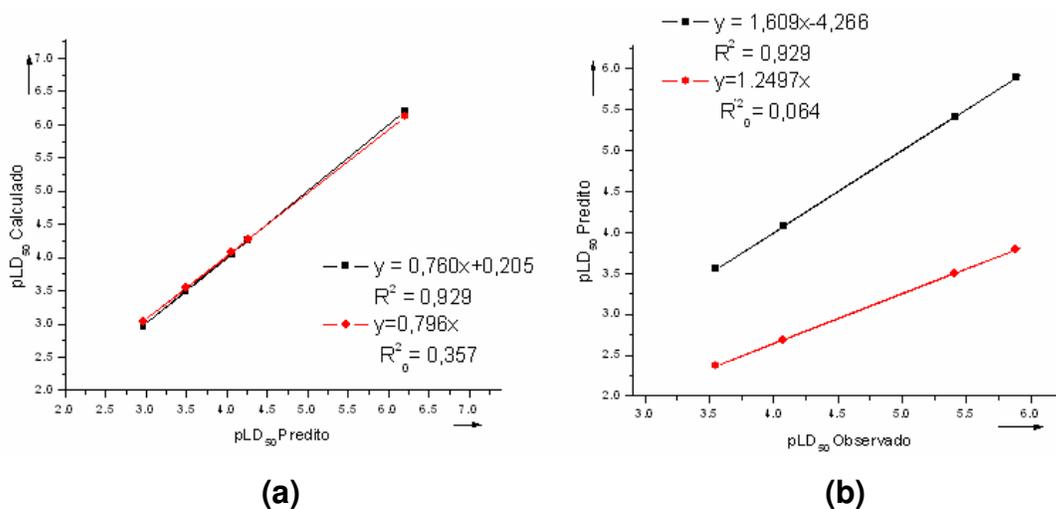
**Figura 58.** Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR Eq. (5.4-25) das DBHs pelas equações (a)  $y = ax+b$ ,  $y = kx$ , (b)  $y = a'x+b'$  e  $y = k'x$ .



**Figura 59.** Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM das DBHs pelas equações (a)  $y = ax+b$ ,  $y = kx$ , (b)  $y = a'x+b'$  e  $y = k'x$ .

Em qualquer uma das duas metodologias, MLR ou SVM, ao se aplicar os modelos de QSAR aos nove compostos (pseudo *blind set*), a correlação de Pearson na predição é baixa. Como exemplo, empregando os modelos de QSAR ( $n = 30$ ) obteve-se pelo método MLR,  $R = 0,693$  e pelo SVM,  $R = 0,706$ . Subdividindo-se o conjunto dos nove compostos em dois conjuntos, com quatro e cinco compostos, respectivamente, este último concentrando os halogenados, obtêm-se coeficientes de determinação mais favoráveis (para  $n = 5$ ): modelo de QSAR-MLR,  $R^2 = 0,929$ ; modelo de QSAR-SVM,  $R^2 = 0,981$ . Apesar dos valores numéricos mostrarem um resultado estatístico aparentemente satisfatório, aplicando o método de Golbraikh-Tropsha<sup>64</sup> verifica-se o contrário pelos parâmetros de validação, mostrando uma grande discordância na predição externa que é ilustrada na Figura 60.

Relembrando o que foi mencionado em Material e Métodos, este estudo submete todos os compostos ao conjunto de treinamento, contrariando o proposto pelo método de validação Golbraikh-Tropsha. Estes nove compostos entretanto, obedeceram ao critério e não participaram do treinamento. Contudo, foram submetidos à análise conformacional.



**Figura 60.** Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR Eq. (5.4-25) das DBHs pelas equações **(a)**  $y = ax + b$ ,  $y = kx$ . **(b)**  $y = a'x + b'$  e  $y = k'x$  aplicado a conjunto  $n = 5$ .

Os dados empregados encontram-se na Tabela 32 a seguir.

**Tabela 32.** Resultados de predição obtidos com os modelos de QSAR com n = 30 compostos

Item	MolID	Xn	pLD <sub>50</sub> <sup>b</sup>	Descritores			Métodos empregados	
				$\alpha_{yy}$	$\mu_x$	RDF095p	QSAR-MLR	QSAR-SVM
1	D9	2-C <sub>2</sub> H <sub>5</sub>	5,91	8,945	0,634	0,144	7,79	6,90
2	D11	2-OBu <sup>-sec</sup>	2,84	47,780	-6,86	2,478	3,73	3,80
3	D12	2-SCH <sub>3</sub>	3,26	51,609	-3,14	1,824	4,39	4,44
4	D23	4-CN	4,96	33,281	-9,45	0,964	4,21	4,30
5	D34 <sup>a</sup>	2,3-(CH <sub>3</sub> ) <sub>2</sub>	4,33	50,095	2,74	0,180	5,09	5,20
6	D40 <sup>a</sup>	2,6-F <sub>2</sub>	4,71	47,156	2,241	0,184	5,35	5,43
7	D41 <sup>a</sup>	2-F,6-Cl	4,71	46,645	4,403	0,925	4,64	4,71
8	D45 <sup>a</sup>	2,3,4,5-F <sub>4</sub>	4,48	39,212	2,889	2,380	4,99	4,96
9	D46 <sup>a</sup>	2,3,4,5,6-F <sub>5</sub>	3,04	41,213	1,388	3,251	5,00	4,97

(a) Compostos usados no gráfico da Figura 60 ; (b) pLD<sub>50</sub> Observado (LD<sub>50</sub> = mmol/inseto).

Uma vez que as predições dos modelos de QSAR-MLR e QSAR-SVM estão razoavelmente próximas entre si, mas distantes dos valores experimentais, o resultado sugere três hipóteses que podem conter o motivo do insucesso no teste de validação externa: **(a)** Os descritores não conseguem descrever todas as estruturas moleculares com a mesma eficácia. **(b)** O espaço conformacional pode estar incompleto. **(c)** Podem ocorrer as duas situações combinadas. Através de simulações, verificou-se que o causador do mau resultado está com grande probabilidade de estar contido no item (c). No presente caso, apenas o composto D41 obteve boa predição da atividade larvicida.

Empregando-se o critério de validação das funções de ajuste RQK (*RQK Fitness functions*) do programa MobyDigs dos modelos de QSAR Eq.(5.4-19) e Eq. (5.4-25), obteve-se resultados que aprovam os modelos gerados (v. Tabela 33 e Tabela 34).

**Tabela 33.** Validação do modelo de QSAR-MLR (n = 37) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis $K_{xy} - K_x > DK$	$K_{xy}$ 0,495	$K_x$ 0,251	$K_{xy}-K_x$ 0,244	DK 0,000
2.Capacidade de predição $Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,995	> 0,5		
3.Capacidade de predição do modelo $Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$ 0,995	$Q^2_{ASYM}$ 0,996	$Q^2_{LOO}-Q^2_{ASYM}$ -0,001	DQ -0,005
4.Redundância nas variáveis $R^P > t^P$	$R^P$ 0,142	$t^P$ 0,0		
5.Sobre ajuste (overfitting) $R^N > t^N$	$R^N$ 0,000	$t^N$ -0,233		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente.

**Tabela 34.** Validação do modelo de QSAR-MLR (n = 30) pelos parâmetros RQK

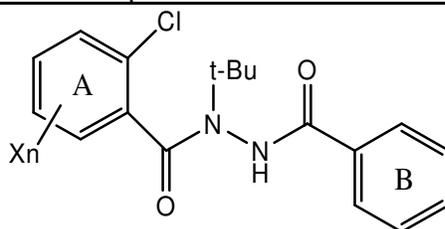
Critérios de aprovação (*)				
1.Colinearidade entre variáveis $K_{xy} - K_x > DK$	$K_{xy}$ 0,484	$K_x$ 0,240	$K_{xy}-K_x$ 0,244	DK 0,000
2.Capacidade de predição $Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,995	> 0,5		
3.Capacidade de predição do modelo $Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$ 0,995	$Q^2_{ASYM}$ 0,994	$Q^2_{LOO}-Q^2_{ASYM}$ 0,001	DQ -0,005
4.Redundância nas variáveis $R^P > t^P$	$R^P$ 0,169	$t^P$ 0,0		
5.Sobre ajuste (overfitting) $R^N > t^N$	$R^N$ 0	$t^N$ -0,233		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente.

## 5.4.3.4 – Proposta de novos compostos

Foram propostos treze compostos *in silico* de dibenzoilidrazinas, na maioria *orto* substituídas, todos halogenados, e pelos resultados de predição, identificam-se candidatos potenciais para novos bioensaios (Tabela 35).

**Tabela 35.** Proposta de novos compostos - Dibenzoilidrazinas contra *S. exigua*



Item	MolID <sup>a</sup>	Xn	Descritores			Métodos e valores preditos		
			$\alpha_{yy}$	$\mu_x$	RDF095p	$Y_{MLR}^b$	$Y_{SVM}^b$	$Y_{MLR}-Y_{SVM}^c$
1	P1	2,6-Cl <sub>2</sub>	50,0458	-1,768	1,187	5,03	5,01	0,02
2	<b>P2</b>	2-CF <sub>3</sub> ,6-OH	47,7173	4,246	3,015	6,08	5,83	<b>0,25</b>
3	P3	2-F,6-CH <sub>3</sub>	44,9731	-0,483	1,350	5,56	5,61	-0,04
4	P4	2-Cl,6-CH <sub>3</sub>	39,6707	1,062	1,199	6,26	6,23	0,03
5	<b>P5</b>	2-Br,6-CH <sub>3</sub>	45,9469	1,489	3,733	5,30	5,18	<b>0,12</b>
6	<b>P6</b>	2-OH,6-SCF <sub>3</sub>	47,2449	2,398	4,938	5,10	4,88	<b>0,22</b>
7	P7	2-CF <sub>3</sub> ,6-CH <sub>3</sub>	46,3575	-0,361	2,564	5,17	5,17	0,00
8	P8	3,5-F <sub>2</sub>	39,8195	1,936	2,408	6,12	6,04	0,08
9	P9	2,4-F <sub>2</sub>	36,0982	0,108	1,839	6,04	6,05	-0,01
10	P10	2,3-F <sub>2</sub>	29,8617	-0,836	0,738	6,46	6,51	-0,06
11	P11	3-Cl	36,2183	-0,154	2,181	5,87	5,89	-0,02
12	P12	4-F	36,3931	-0,413	2,802	5,62	5,67	-0,05
13	<b>P13</b>	2,5-F <sub>2</sub>	32,2238	0,454	1,542	6,41	6,52	<b>-0,11</b>

(a) Compostos com geometria AM1; (b)  $Y_{MLR, SVM} = pLD_{50}$  ( $LD_{50} = \text{mmol/inseto}$ ); (c)  $Y_{MLR}-Y_{SVM} = \text{resíduo entre predição dos modelos de QSAR-MLR e QSAR-SVM}$ .

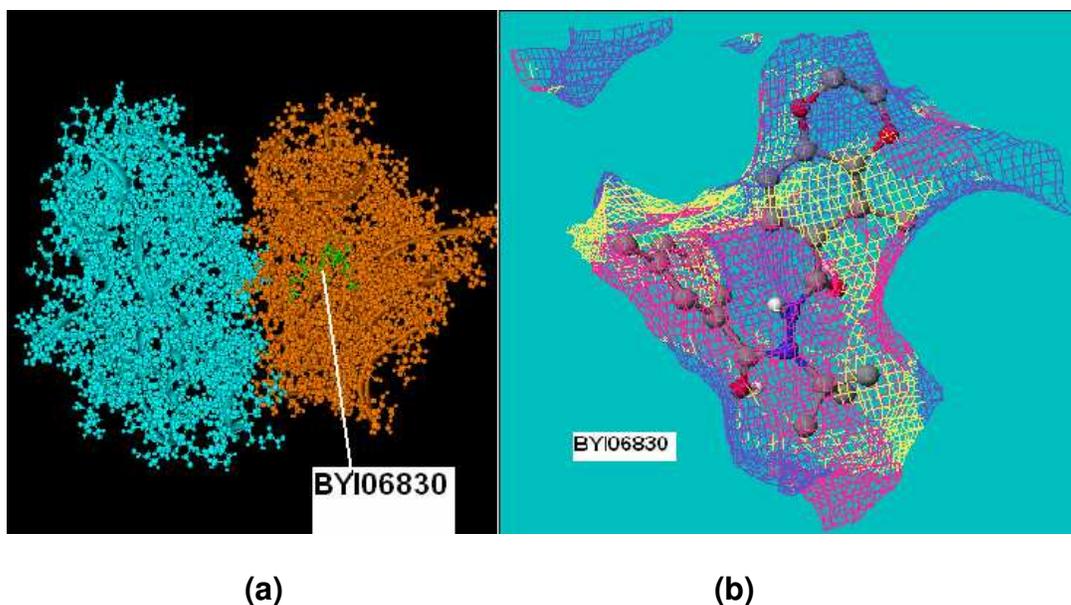
Na Tabela 35 foram assinalados quatro compostos em negrito por apresentarem uma grande diferença entre os demais na predição dos modelos de QSAR-MLR e QSAR-SVM. Os comentários feitos para os compostos relacionados na Tabela 32 valem igualmente para este caso. Pelo estudo de *docking* foi possível agregar um pouco mais de informação para a interpretabilidade dos resultados.

### 5.4.3.5 – Docking de compostos halogenados

Como descrito em Métodos, os resultados que se seguem foram gerados através da técnica de *docking* na seguinte ordem: **(I)** Validação do método empregado pelo programa BioMedCache em relação ao complexo cristalino da proteína 1R20.pdb com a dibenzoilidrazina BY106830. **(II)** *Docking* dos compostos halogenados propostos. **(III)** *Docking* dos compostos propostos pelos modelos de QSAR. **(IV)** Análise dos compostos alojados na cavidade em conjunto com a base de dados constantes nas Tabelas 26, 27, 28 e 29.

#### (I) Validação do método pelo composto BY106830

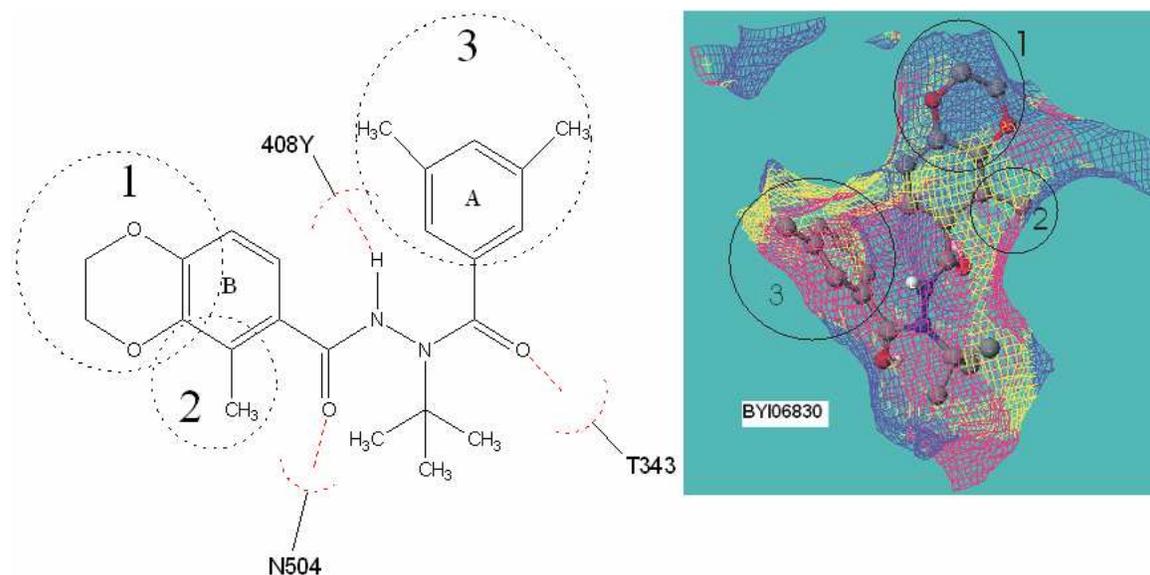
A Figura 61 ilustra a proteína, código PDB 1R20. A estrutura em cor verde se situa no monômero indicado correspondente ao ligante BY106830.



**Figura 61.** Proteína 1R20, **(a)** cristalizada em conjunto com o ligante identificado como BY106830; **(b)** ligante BY106830, dibenzoilidrazina, situado na cavidade da proteína conforme Billas *et al.*<sup>96</sup> e aqui reproduzido por este estudo.<sup>110</sup>

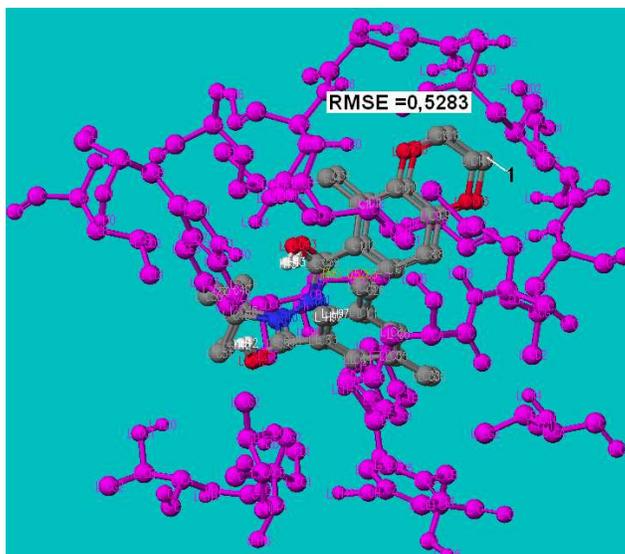
A Figura 62 ilustra a cavidade com a identificação de três regiões que podem ser exploradas na proposição de novos compostos, porém, o presente estudo explora apenas a região 3 à qual está associado o anel A do composto BY106830. Pela convenção das cores, a região azul representa que a proteína necessita de

ligantes doadores de H (p. ex. grupos C=O-) e a região vermelha indica que a proteína necessita de receptores de H; (3) A região 3, que é o alvo deste estudo, indica ser uma região predominantemente receptora de H com sítios hidrofóbicos assinalados na cor creme.



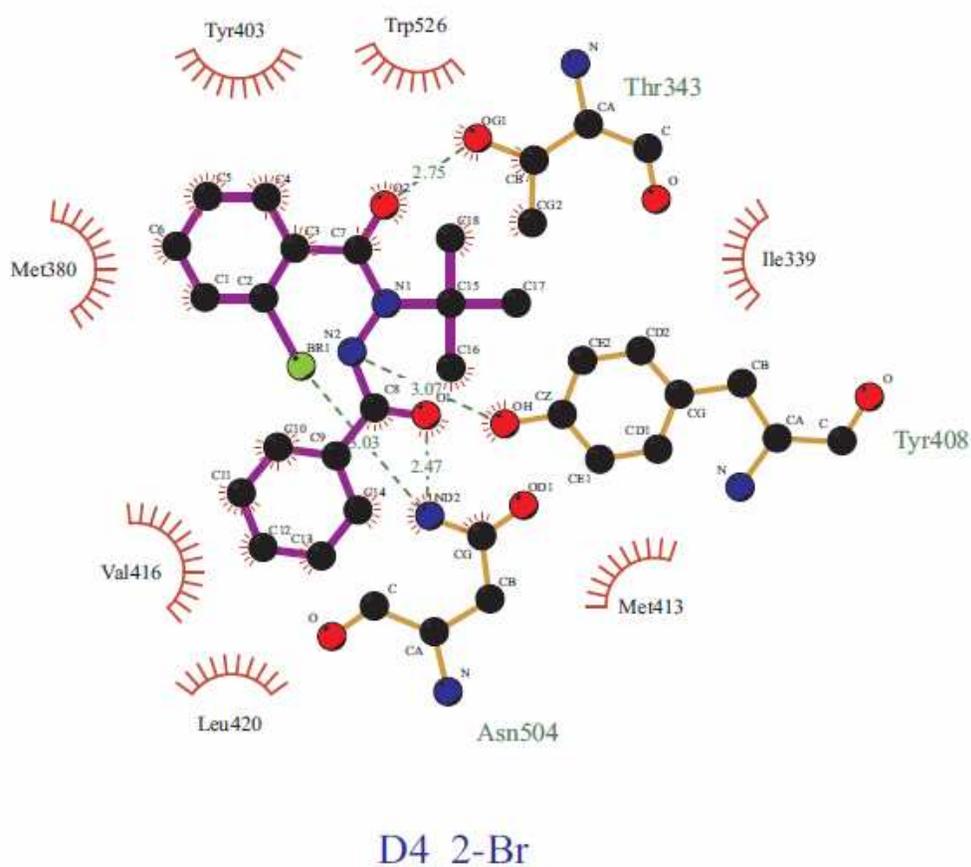
**Figura 62.** Cavidade com indicação de três regiões favoráveis à proposta de novos compostos substituídos. Os aminoácidos identificados e os respectivos átomos de interação foram originalmente identificados no trabalho de *Billas et al.*<sup>96</sup>. Regiões indicadas como 1, 2 e 3 vide texto.

A validação ocorre utilizando uma cópia do ligante que fica desalojada da cavidade, e que na operação de *docking* deverá ocupar a posição o mais próximo possível do ligante. Obteve-se o valor RMSD = 0,5283 Å. A Figura 63 ilustra a posição original do ligante indicada pelo número 1 e a sua cópia sobreposta por *docking*.



**Figura 63.** Ilustração da sobreposição de estruturas, obtidas por *docking* entre ligante BYI06830 e sua cópia, o ligante originalmente alojado na cavidade indicada pelo número 1. As estruturas em magenta correspondem aos aminoácidos que formam a cavidade receptora.

A seguir são apresentados os dados obtidos da interação dos átomos halogenados das dibenzoilidrazinas empregadas contra *S. exigua* e estudadas na cavidade da proteína 1R20. A Figura 64 ilustra um exemplo gráfico da interação. Após *docking*, cada estrutura proteica complexada com o respectivo ligante halogenado foi submetida ao programa LigPlot<sup>125</sup> para extrair as informações relativas à interação ligante-proteína. Os dados estão listados na Tabela 36.



**Figura 64.** Dibenzoilidrazina D4 e resultado do *docking* na cavidade da proteína 1R20, e o átomo de Br na posição *orto* e sua distância de interação de 3,03 Å em relação ao nitrogênio do aminoácido Asn504.

Os aminoácidos Asn504, Thy408 e Thr343 formam uma ancoragem do ligante na proteína conforme revelado no estudo de Billas *et al.*<sup>98</sup>. Tomando por base os estudos de interação não-covalente e buscando apenas distâncias inferiores a 4,0 Å, constatou-se no presente trabalho que pode haver influência de interação dos átomos halogenados das dibenzoilidrazinas com os átomos ou grupo funcional dos aminoácidos Trp 526, Met 380, Met 381 e Ser 377, conforme listado na Tabela 36. Pode-se admitir a hipótese de que as interações com estes aminoácidos são também as que podem fazer alguma diferença na atividade larvicida.

**Tabela 36.** Possíveis interações não-covalentes dos átomos halogênios com átomos dos aminoácidos de HvEcR-LBD

MolID	Interação <sup>a</sup> C-X...Y	Distância d <sub>Xn...Y</sub> (Å)	Aminoácido	Ângulo médio ( <sup>o</sup> )
D2_2-F	PhF...π <sup>b</sup>	2,61	Trp 526	154
D3_2-Cl	Cl...O=C-	2,96	Asn 504	
	PhCl...π	3,05	Trp 526	164
D4_2-Br	Br...O=C-	2,86	Asn 504	
	PhBr...π	3,08	Trp 526	166
	Br...N	3,03	Asn 504	
D5_2-I	I...N	3,07	Asn 504	
	PhI...π	3,16	Trp 526	166
	I...O=C-	2,87	Asn504	
D6_2-CF <sub>3</sub>	F...S	3,78	Met 380	
	F...O=C-	3,25	Asn 504	
	PhF...π	2,89	Trp 526	160
D13_3-F	F...S	3,08	Met 380	
	F...O=C-	3,31	Ser377	
D14_3Br	Br...S	3,89	Met 380	
	Br...O=C-	2,89	Ser377	
D15_3-I	I...S	3,54	Met 380	
	I...O=C-	2,74	Ser377	
D16_3-CF <sub>3</sub>	F...S	3,88	Met 380	
	F...S	3,65	Met381	
	F...O=C-	2,48	Ser 377	
	F...O=C-	3,06	Asn 504	
D19_4-Cl	Cl...S	3,24	Met 381	
D20_4-Br	Br...S	3,18	Met 381	
D21_4-I	I...S	3,12	Met 381	
D22_4-CF <sub>3</sub>	F...S	1,92	Met 381	
	F...S	3,16	Met 380	

(a) Y corresponde ao aminoácido, vide Figura 50; (b) corresponde aos elétrons do orbital π.

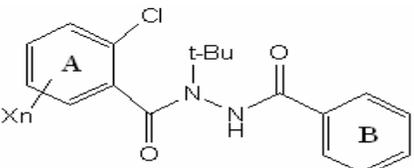
Utilizando os dados das Tabelas 26, 27 e 28 como referência para avaliação dos resultados, verifica-se que todos os valores estão razoavelmente condizentes. Não ocorreram casos da Tabela 29. No caso dos compostos com substituinte CF<sub>3</sub>, escolheu-se o átomo de flúor com a menor distância ao átomo de interação do aminoácido.

Analogamente ao domínio ligante HvEcR-LBD, foi elaborada a Tabela 37 para o domínio ligante de SeEcR-LBD com algumas dibenzoilidrazinas halogenadas e os aminoácidos vizinhos dentro de uma distância menor que 4Å. Foram calculadas as energias de cada interação átomo-ligante/átomo-aminoácido pelo uso da equação 5.4-7, para se obter uma noção aproximada da ordem de grandeza destas interações. Os compostos D5, D6 e D31 aparecem com valores altos dada a curta distância de interação dos respectivos átomos. Qualquer pequena variação (ou erro) de  $r$  influi enormemente no resultado final em função do fator  $1/r^6$ . Pela grande dificuldade de se obter valores da energia de interação átomo-a-átomo em estruturas experimentais, os valores obtidos são ilustrativos e devem ser vistos com ponderação.

Como se utilizou uma proteína desenvolvida por homologia e *docking* como ferramenta para formar o complexo ligante-proteína é necessário que os resultados sejam vistos sob a luz da fundamentação e de algumas consistências e restrições já citadas ao longo do texto. Sob este enfoque, as informações produzidas por este estudo podem ser úteis no desenvolvimento de novos compostos. A primeira informação é dada pela identificação dos aminoácidos com os quais as dibenzoilidrazinas interagem: Tyr 408 já foi identificado anteriormente como um dos aminoácidos de ancoragem da dibenzoilidrazina, mas, adicionalmente, ocorrem também interações dos substituintes halogenados com outros átomos deste aminoácido. Interações novas identificadas são dadas por Met 380, Met 381 e Tyr 403. O composto D33 tem a maior atividade larvicida e mostrou ter interação com a Ser 377. Investigou-se, se há correlação da energia  $E_L$  com a bioatividade, mas isto não ocorre. Qualquer hipótese sobre este aspecto, neste momento, seria demasiadamente especulativa.

O estudo segue produzindo também informações de possíveis interações- $\pi$  do anel benzênico substituído com o átomo de enxôfre (S) com o aminoácido Met380, e dos átomos halogênios da dibenzoilidrazina com o anel benzênico de Tyr403.

**Tabela 37.** Possíveis interações não-covalentes dos átomos halogênicos com os aminoácidos de SeEcR-LBD de *Spodoptera exigua* e estimativas da energia pela forças dispersivas de London ( $E_L$ ).



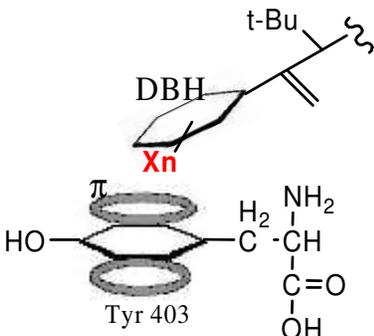
$$E_L = -\frac{3}{2} \frac{\alpha_1 \alpha_2}{r^6} \frac{I_1 I_2}{I_1 + I_2}$$

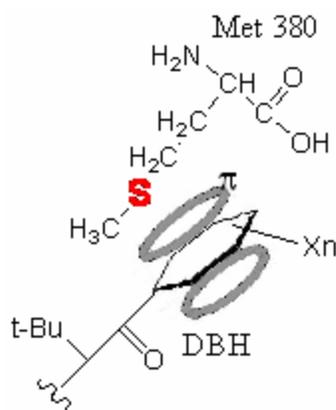
MolID	Interação <sup>a</sup>	Distância	AA Y <sup>b</sup>	E <sub>L</sub> parcial		E <sub>L</sub> Tot (kcal/mol) <sup>(+)</sup>
	C-X...Y	dXn...Y (Å)		(eV)	kcal/mol	
D2_2-F	F...O=C-	2,24	Tyr408	-0,041	-0,935	<b>-0,935</b>
D3_2-Cl	Cl...O-H	2,08	Tyr408	-0,215	-4,961	<b>-4,961</b>
D4_2-Br	Br...O-H	2,00	Tyr408	-0,363	-8,363	<b>-8,363</b>
D5_2-I	I...O-H	1,08	Tyr408	-23,983	-553,053	<b>-553,053</b>
D6_2-CF <sub>3</sub>	F...O-H	1,07	Tyr408	-3,413	-78,702	<b>-79,253</b>
	F...O-H	2,57	Tyr408	-0,018	-0,410	
	F...O-H	3,07	Tyr408	-0,006	-0,141	
D13_3-F	F...O-H	3,67	Tyr408	-0,002	-0,048	<b>-0,048</b>
D14_3-Br	Br...O-H	2,86	Tyr408	-0,042	-0,978	<b>-0,978</b>
D15_3-I	I...S	3,73	Met381	-0,045	-1,037	<b>-1,596</b>
	I...O=C-	3,93	Met380	-0,010	-0,238	
	I...N-H	3,96	Met380	-0,014	-0,321	
	I...O=C-	3,62	Tyr408	-0,002	-0,052	
D16_3-CF <sub>3</sub>	F...H-O	3,84	Tyr408	-0,001	-0,031	<b>-0,190</b>
	F...O-H	3,47	Tyr403	-0,003	-0,068	
	F...H-O	3,69	Tyr403	-0,002	-0,039	
	F...O=C-	3,34	Met380	-0,013	-0,289	
D19_4-Cl	Cl...S	3,91	Met380	-0,015	-0,352	<b>-1,951</b>
	Cl...H-O	3,62	Met380	-0,006	-0,148	
	Cl...N-H	3,15	Met380	-0,025	-0,582	
	Cl...S	3,60	Met381	-0,025	-0,579	
	Cl...O=C-	3,57	Met381	-0,015	-0,335	
D20_4-Br	Br...N-H	3,03	Met381	-0,016	-0,363	<b>-1,160</b>
	Br...O=C-	3,24	Met380	-0,020	-0,463	
	Br...S	3,53	Met381	-0,063	-1,443	
D21_4-I	I...S	3,53	Met381	-0,063	-1,443	<b>-5,065</b>
	I...O=C-	3,10	Met380	-0,043	-0,989	
	I...H-O	3,41	Met380	-0,020	-0,464	
	I...N	2,88	Met380	-0,094	-2,169	
D30_2,3-Cl <sub>2</sub>	Cl...O-H	2,08	Tyr408	-0,215	-4,961	<b>-6,041</b>
	Cl...H-O	2,60	Tyr408	-0,047	-1,080	
D31_2,4-Cl <sub>2</sub>	Cl...O=C-	3,35	Met380	-0,012	-0,284	<b>-61,110</b>
	Cl...N-H	3,15	Met380	-0,025	-0,582	
	Cl...H-N	3,62	Met380	-0,006	-0,148	
	Cl...O-H	2,08	Tyr408	-0,215	-4,961	
D33_3,5-Cl <sub>2</sub>	Cl...H-O	1,35	Tyr408	-2,391	-55,135	<b>-1,598</b>
	Cl...S	3,48	Met380	-0,031	-0,709	
	Cl...O=C-	2,77	Ser377	-0,039	-0,889	

(a) Vide Figura 47 e 50; (b) Y representa o aminoácido (AA), vide Figura 50; (+) 1eV = 23,0605 kcal/mol.

Relativamente às distâncias e aos ângulos encontrados neste estudo, há uma correspondência satisfatória com os dados contidos na literatura.<sup>118</sup>

**Tabela 38.** Possíveis interações- $\pi$  de halogênios da dibenzoilidrazina com a Tyr 403, e interação S- $\pi$  da Met380 no domínio ligante SeEcR-LBD.

MolID	Interação <sup>a</sup> C-X...Y	Distância <sup>b</sup> dXn...cent. (Å)	Aminoácido	Ângulo médio ( <sup>o</sup> ) <sup>b</sup>
				
D2_2-F	PhF... $\pi$	3,10	Tyr403	150
D3_2-Cl	PhCl... $\pi$	2,95	Tyr403	146
D5_2-I	PhI... $\pi$	2,78	Tyr403	140
D6_2-CF <sub>3</sub>	PhF... $\pi$	2,73	Tyr403	179



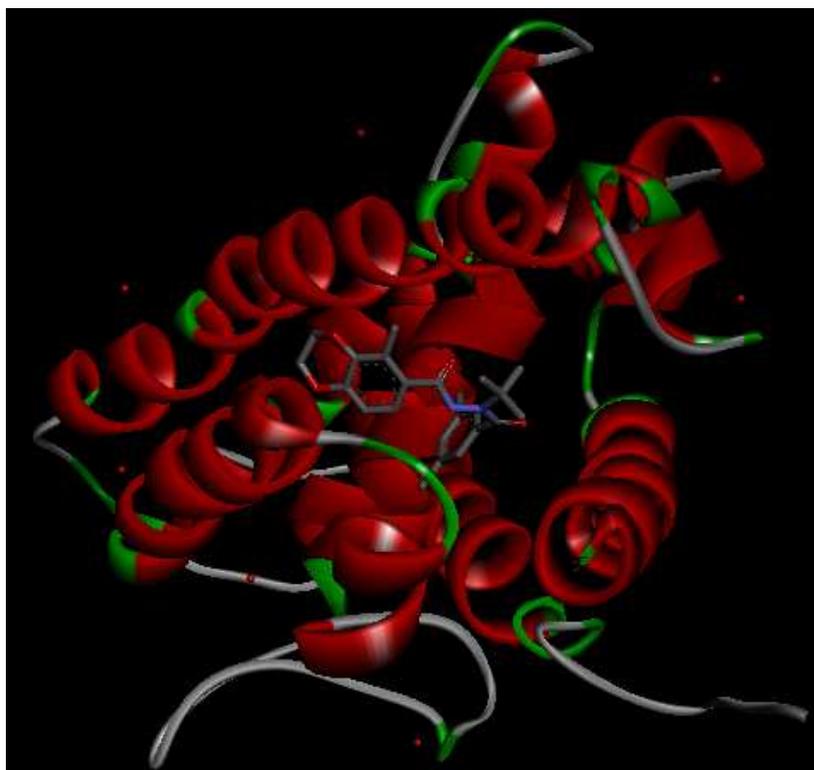
D15_3-I	S... $\pi$	2,87	Met380	118
D16_3-CF <sub>3</sub>	S... $\pi$	3,67	Met380	173
D19_4-Cl	S... $\pi$	3,67	Met380	173
D20_4-Br	S... $\pi$	3,67	Met380	173
D21_4-I	S... $\pi$	3,67	Met380	173
D30_2,3-Cl <sub>2</sub>	S... $\pi$	3,67	Met380	173
D31_2,4-Cl <sub>2</sub>	S... $\pi$	3,67	Met380	173
D33_3,5-Cl <sub>2</sub>	S... $\pi$	3,67	Met380	173

(a) Y representa o amino ácido; (b) distância corresponde a Xn até ao centróide ilustrado na Figura 50.

Tauer<sup>126</sup> em sua dissertação sobre contactos S- $\pi$  baseou-se em 946 proteínas cristalografadas com 1,5 Å de resolução, encontrou distância média de 3,6 Å. Saraogi *et al.*<sup>127</sup> encontraram uma variação maior de valores estudando proteínas com resolução de 2,5Å ou melhor: C-F... $\pi$  = 3,21Å, C-Cl... $\pi$  = 3,53Å, C-Br... $\pi$  = 3,56Å e C-I... $\pi$  = 3,75Å .

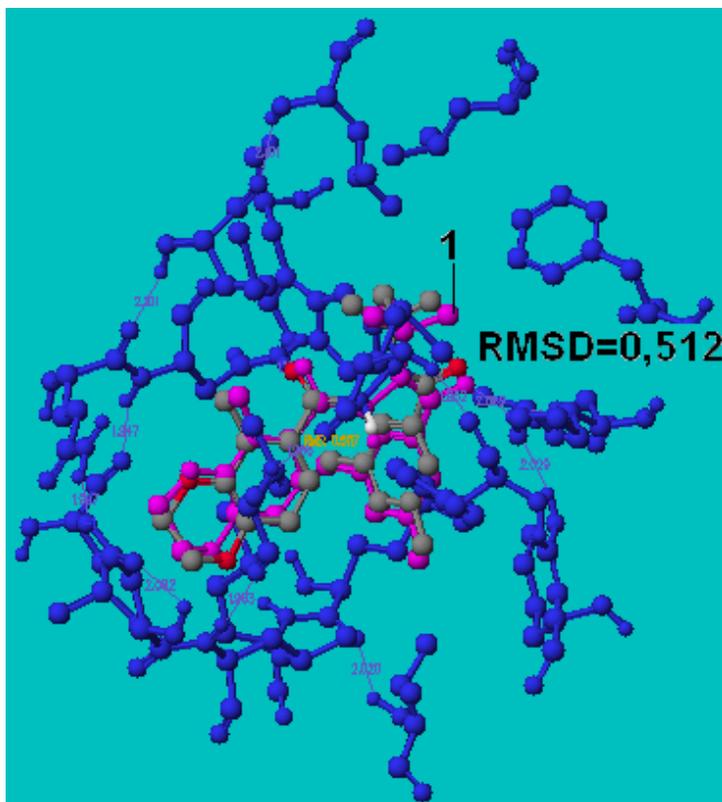
## (II) *Docking* das Dibenzoilidrazinas Halogenadas Propostas

A Figura 65a mostra o alojamento da dibenzoilidrazina BY106830 na cavidade da proteína SeEcR obtida por homologia através da proteína 1R20.



**Figura 65a.** Alojamento da dibenzoilidrazina BY106830 na cavidade do domínio ligante SeEcR-LBD obtido após a sobreposição<sup>117</sup> do complexo ligante-proteína 1R20 com a proteína obtida por homologia a partir dos aminoácidos de *Spodoptera exigua* (MW = 26,5 kDa).

A Figura 65b é o resultado do *docking* da dibenzoilidrazina BYI06830 na cavidade da proteína SeEcR da ecdisona *Spodoptera exigua*, obtida por homologia com a proteína 1R20.



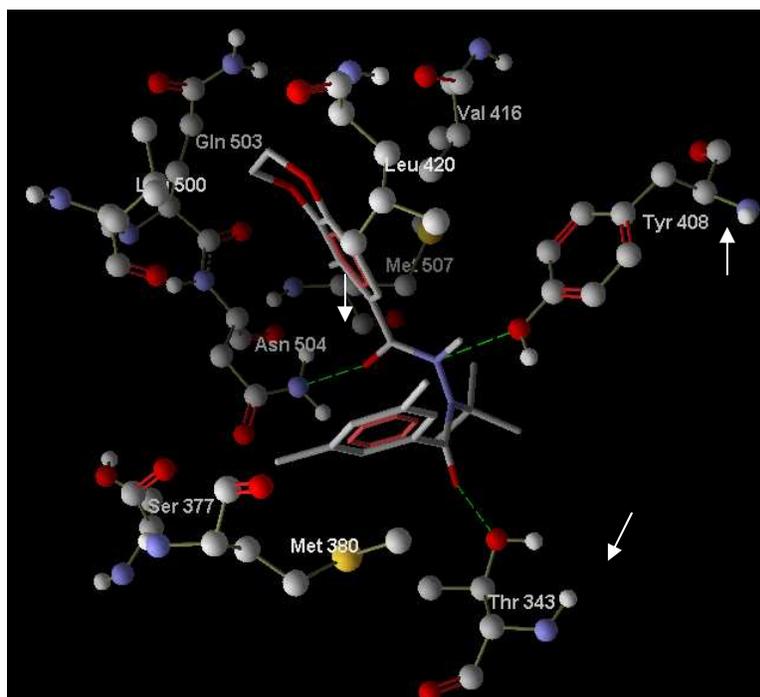
**Figura 65b.** Sobreposição obtida por *docking* da cópia do composto BYI06830 (coloração por átomos) com o composto BYI06830 (cor magenta, indicado pelo número 1). Obteve-se um valor da distância de sobreposição RMSD = 0,512 Å.

O resultado de sobreposição obtido pelo *docking* (RMSD = 0,512 Å) pode ser comparado por utilizar a mesma metodologia, com o resultado obtido anteriormente com HvEcR-LBD + BYI06830 de 0,528 Å e este último, comparado com o obtido por Bordas *et al.*<sup>128</sup> de 0,540 Å para o mesmo complexo.

Na impossibilidade de se fazer uma validação experimental, o RMSD é o único parâmetro escolhido e utilizado neste estudo para avaliar e validar, em teoria, o *docking* do presente estudo.

Como primeira informação obtida pelo *docking*, verifica-se que a ancoragem do ligante BYI06830 ocorre com os mesmos aminoácidos Thr343, Tyr408 e Asn504

(de forma análoga à proteína 1R20), conforme indicado na figura gerada pelo programa MMV (*Molegro Molecular Viewer*)<sup>129</sup> (v. Figura 66).



**Figura 66.** Complexo SeEcR-LBD de *Spodoptera exigua* com o ligante BY106830 alojado na cavidade. Linhas pontilhadas indicam as interações com os aminoácidos Asn 504, Thr 343 e Tyr 408.

Este fato não causa surpresa e ao mesmo tempo fortalece o resultado obtido pelo presente estudo. Hormann *et al.*<sup>106</sup>, em pesquisa com 77 dibenzoilidrazinas e também utilizando CoMFA (*Comparative Molecular Field Analysis*) e CoMSIA (*Comparative Molecular Similarity Indices Analysis*) com *Chilo suppressalis* e *Leptinotarsa decemlineata*, partiram das sequências do domínio de ligação (LBD) de *Heliothis virescens* cristalizado com BY106830 constatando os mesmos aminoácidos como pontos de contacto.

As diferenças do domínio ligante (LBD) entre proteína SeEcR e HvEcR estão ilustradas na Figura 67.

SeEcR-LBD: Ile339, \*\*\*\*\*, Thr343, Ser377, Met380, Met381, Val384, Tyr403, Tyr408, Met413,

HvEcR-LBD: Ile339, Thr340, Thr343, Ser377, Met380, Met381, Val384, Tyr403, Tyr408, Met413,

SeEcR-LBD Val416, Leu420, Leu500, Gln503, Asn504, Met507, \*\*\*\*\*, Trp526.

HvEcR-LBD Val416, Leu420, \*\*\*\*\*, Gln503, Asn504, Met507, Leu511, Trp526.

**Figura 67.** Segmentos alinhados do domínio de ligação (LBD) dos aminoácidos formadores das cavidades em SeEcR e HvEcR selecionados pela distância máxima de 3Å de seus átomos, aos átomos do composto D6 através do programa BioMedCache.<sup>110</sup>

O conjunto destes resultados evidencia que, em nível teórico, existe boa consistência nas metodologias usadas na proposição de novos bioensaios destes compostos ou de parte deles.

Antes de fazer a avaliação dos resultados dos compostos propostos, é necessário obter algumas informações geradas sobre os compostos halogenados com valores de atividade larvicida experimental.

Os resultados obtidos pelo *docking* na cavidade receptora SeEcR-LBD da proteína homóloga à proteína 1R20 são fornecidos em duas tabelas distintas. A Tabela 39 apresenta a avaliação pelo PMF-score, e a Tabela 40 pelo X-score, parâmetros sem propósitos comparativos, mas para verificar o que cada um deles pode fornecer de informações sobre a afinidade de ligação ligante-proteína.

Iniciando a análise pelo PMF-score com os valores da Tabela 39, verifica-se que a correlação R obtida com a atividade larvicida pLD<sub>50</sub> no conjunto dos 25 compostos é muito baixa, mas em concordância com a literatura (0 a 0,37)<sup>112,114</sup> para valores globais: R = 0,100. Se, entretanto, a avaliação for feita por grupos de compostos (assinalados na tabela por GI, GII, etc.), a correlação passa a ser representativa como mostrado a seguir:

Grupo I (n = 11): R = 0,860; Grupo II (n = 7): R = 0,975; Grupo III (n = 5): R = 0,888. Os compostos D41 e D46 funcionam como *outliers* nos três grupos.

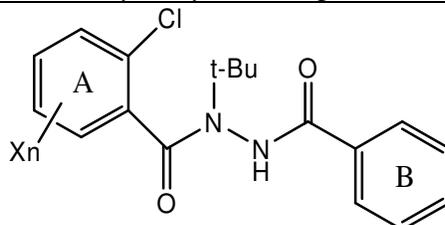
A correlação obtida pela atividade larvicida pLD<sub>50</sub> e os valores X-score de todos os 25 compostos da Tabela 40 foi de R = 0,572. Procedendo a análise de correlação por grupos como anteriormente feito obtém-se:

Grupo I (n = 12): R = 0,963; Grupo II (n = 4): R = 0,995; Grupo III (n = 6): R = 0,997. Grupo IV (n = 3): R = 0,869.

Em ambos os casos as afinidades de ligação ligante-proteína preditas pelo PMF-score e X-score têm boa correlação com a atividade larvicida ( $pLD_{50}$ ), fator que será considerado por este estudo como indicador da aceitabilidade dos resultados de *docking*. Utilizando duas metodologias independentes operando por algoritmos diferentes, para avaliar a qualidade do *docking*, qualificam-se os resultados finais com aceitabilidade teórica suficiente para se atribuir consistência ao estudo.

A questão que se pode formular em vista dos bons resultados de correlação obtidos por grupos de compostos é saber o quanto eles são consistentes, por indicadores independentes, aspecto que será abordado a seguir.

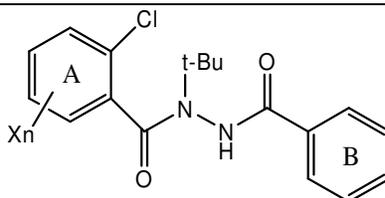
**Tabela 39.** Valores de RMSD e PMF-score obtidos no *docking* de dibenzoilidrazinas halogenadas alojadas no receptor da proteína homóloga SeEcR-LBD de *Spodoptera exigua*.



Item	MolID(+)	Substituintes Xn	RMSD Å <sup>a</sup>	pLD <sub>50</sub> <sup>b</sup>	PMF-score
1	D1 <sup>GII</sup>	H	0,973	4,92	-54,425
2	D2 <sup>GIII</sup>	2-F	0,922	6,51	-50,172
3	D3 <sup>GIII</sup>	2-Cl	1,879	6,37	-57,602
4	D4 <sup>GI</sup>	2-Br	1,870	6,21	-52,411
5	D5 <sup>GIII</sup>	2-I	1,442	6,24	-44,876
6	D6 <sup>GI</sup>	2-CF <sub>3</sub>	0,924	6,19	-44,662
7	D13 <sup>GI</sup>	3-F	0,891	6,15	-55,085
8	D14 <sup>GI</sup>	3-Br	0,977	6,57	-48,623
9	D15 <sup>GII</sup>	3-I	1,530	6,11	-34,361
10	D16 <sup>GI</sup>	3-CF <sub>3</sub>	0,490	5,97	-61,046
11	D19 <sup>GI</sup>	4-Cl	0,461	6,12	-59,557
12	D20 <sup>GI</sup>	4-Br	1,598	5,62	-58,853
13	D21 <sup>GII</sup>	4-I	1,168	5,43	-46,169
14	D22 <sup>GII</sup>	4-CF <sub>3</sub>	1,859	5,47	-48,762
15	D30 <sup>GI</sup>	2,3-Cl <sub>2</sub>	1,585	6,01	-63,404
16	D31 <sup>GIII</sup>	2,4-Cl <sub>2</sub>	0,608	6,54	-60,390
17	D32 <sup>GI</sup>	2,5-Cl <sub>2</sub>	1,738	5,99	-61,705
18	D33 <sup>GIII</sup>	3,5-Cl <sub>2</sub>	1,957	6,64	-61,527
19	D39 <sup>GII</sup>	2-CH <sub>3</sub> ,3-Cl	1,400	5,31	-49,820
20	D40 <sup>GII</sup>	2,6-F <sub>2</sub>	1,437	4,93	-60,407
21	D41 <sup>(*)</sup>	2-F-6-Cl	1,793	4,71	56,804
22	D43 <sup>GI</sup>	2,3,4-Cl <sub>3</sub>	1,765	5,67	-66,692
23	D44 <sup>GI</sup>	2,5-Cl <sub>2</sub> -3-CF <sub>3</sub>	2,014	4,93	-74,290
24	D45 <sup>GII</sup>	2,3,4,5-F <sub>4</sub>	1,642	4,48	-71,750
25	D46 <sup>(*)</sup>	2,3,4,5,6-F <sub>5</sub>	0,973	3,04	-61,744

(a) Valor de referência com o cristal de BYI06830: RMSD = 0,512Å; (b) valor experimental (LD<sub>50</sub> = mmol/inseto); (+) GI, GII e GIII formam grupos para análise de correlação da atividade larvívica pLD<sub>50</sub> com o PMF-score; (\*) D41 e D46 têm comportamento de *outliers* nos três grupos.

**Tabela 40.** Valores X-score extraídos do resultado de *docking* de dibenzoilidrazinas halogenadas alojadas no receptor da proteína homóloga SeEcR-LBD de *Spodoptera exigua*.



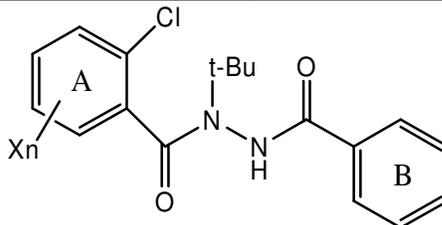
Item	MolID	Substituintes Xn	pLD50 <sup>a</sup>	X-score <sup>b</sup>	Componentes X-score			LBE <sup>d</sup> (kcal/mol)
					HP <sup>c</sup>	HM <sup>c</sup>	HS <sup>c</sup>	
1	D1 <sup>GI</sup>	H	4,92	7,34	6,94	8,00	7,09	-10,01
2	D2 <sup>GIV</sup>	2-F	6,51	7,47	7,07	8,20	7,13	-10,19
3	D3 <sup>GIII</sup>	2-Cl	6,37	7,56	7,18	8,22	7,28	-10,31
4	D4 <sup>GIII</sup>	2-Br	6,21	7,45	6,97	8,19	7,18	-10,16
5	D5 <sup>GI</sup>	2-I	6,24	7,67	7,14	8,45	7,41	-10,46
6	D6 <sup>GI</sup>	2-CF <sub>3</sub>	6,19	7,70	7,33	8,38	7,39	-10,50
7	D13 <sup>GIV</sup>	3-F	6,15	7,31	6,91	7,95	7,07	-9,97
8	D14 <sup>GIII</sup>	3-Br	6,57	7,57	7,07	8,31	7,32	-10,32
9	D15 <sup>GI</sup>	3-I	6,11	7,62	7,08	8,42	7,35	-10,39
10	D16 <sup>GI</sup>	3-CF <sub>3</sub>	5,97	7,58	7,25	8,29	7,21	-10,34
11	D19 <sup>GIII</sup>	4-Cl	6,12	7,53	7,05	8,23	7,29	-10,26
12	D20 <sup>GI</sup>	4-Br	5,62	7,55	7,04	8,29	7,31	-10,30
13	D21 <sup>GI</sup>	4-I	5,43	7,54	6,99	8,34	7,29	10,29
14	D22 <sup>GI</sup>	4-CF <sub>3</sub>	5,47	7,47	7,14	8,19	7,10	-10,20
15	D30 <sup>GI</sup>	2,3-Cl <sub>2</sub>	6,01	7,58	7,07	8,36	7,32	-10,34
16	D31 <sup>GIII</sup>	2,4-Cl <sub>2</sub>	6,54	7,60	7,07	8,36	7,36	-10,36
17	D32 <sup>GII</sup>	2,5-Cl <sub>2</sub>	5,99	7,77	7,30	8,50	7,51	-10,60
18	D33 <sup>GI</sup>	3,5-Cl <sub>2</sub>	6,64	7,70	7,16	8,47	7,49	-10,51
19	D39 <sup>GII</sup>	2-CH <sub>3</sub> ,3-Cl	5,31	7,62	7,18	8,32	7,35	-10,39
20	D40 <sup>GIII</sup>	2,6-F <sub>2</sub>	4,93	7,20	6,85	7,81	7,00	-9,85
21	D41 <sup>GI</sup>	2-F,6-Cl	4,71	7,35	6,90	8,04	7,11	-10,03
22	D43 <sup>GII</sup>	2,3,4-Cl <sub>3</sub>	5,67	7,76	7,18	8,59	7,50	-10,58
23	D44 <sup>GIV</sup>	2,5-Cl <sub>2</sub> ,3-CF <sub>3</sub>	4,93	7,84	7,37	8,65	7,49	-10,69
24	D45 <sup>GI</sup>	2,3,4,5-F <sub>4</sub>	4,48	7,17	6,84	7,73	6,95	-9,79
25	D46 <sup>GII</sup>	2,3,4,5,6-F <sub>5</sub>	3,04	7,16	6,82	7,70	6,96	-9,77

(a) Valor experimental (LD<sub>50</sub> = mmol/inseto); (b) estrutura de referência: cristal de BY106830 (RMSD = 0,512Å); (c) HP = contribuição hidrofóbica e de ajuste dada pelos parâmetros HS, HC, HM descritos em Métodos); (d) LBE (*Ligand Binding Energy*); (+) GI, GII, GIII e GIV formam grupos para análise de correlação da atividade larvicida pLD<sub>50</sub> com X-score.

Podem-se considerar os seguintes indicadores favoráveis: **(a)** partiu-se da proteína 1R20 cristalografada com o ligante, complexo bastante utilizado por outros pesquisadores como referência para trabalhos similares, conforme citado na subseção 5.4.2 em Métodos no parágrafo *Docking*. **(b)** A homologia obtida pelo servidor I-TASSER que produziu o resultado foi dado pela qualificação *C-score*<sup>116b,c</sup> com um RMSD = 0,75 em relação à estrutura nativa, e o *TM-score* = 0,91. Um *TM-score*<sup>116b,c</sup> > 0,5 indica um modelo com topologia correta e < 0,17 uma similaridade randômica. **(c)** Foi possível obter uma validação do *docking* da estrutura do cristal da dibenzoilidrazina BYI06830 com um RMSD = 0,528 Å contra um RMSD = 0,540 Å da literatura<sup>128</sup>. **(d)** O *docking* da estrutura do composto BYI06830 no receptor da proteína homóloga SeEcR-LBD de *S. exigua* resultou um RMSD = 0,521 Å, o que sugere que qualitativamente, a nível teórico, o domínio receptor da proteína SeEcR-LBD é muito similar à de *Heliothis virescens* HvEcR-LBD. **(e)** Constatou-se neste estudo que a ancoragem das dibenzoilidrazinas sempre se dá com os átomos dos aminoácidos comuns aos domínios ligantes das duas proteínas, respectivamente de SeEcR-LBD e HvEcR-LBD: Thr323; Asn 504 e Tyr 408. Como já citado, este resultado foi também obtido para *C. suppressalis* (Lepidoptera) e *L. decemlineata* (Coleoptera) por Hormann *et al.*<sup>106</sup>. **(f)** As afinidades de ligação dos compostos dadas pelos valores *X-score* foram obtidas utilizando o cristal BYI06830 como referência que por sua vez apresentou um bom RMSD de *docking* (v. item(c)).

Considerando todos estes aspectos positivos, torna-se plausível estender o estudo ao conjunto de compostos propostos.

A seguir, estão relacionados na Tabela 41 os resultados obtidos por *docking* das treze dibenzoilidrazinas halogenadas propostas.

**Tabela 41.** Compostos propostos contra *S. exigua*. Resultados de modelos de QSAR-MLR e QSAR-SVM e correlação com o PMF-score

Item	MolID <sup>(+)</sup>	Substituintes Xn	RMSD Å	pLD <sub>50</sub> <sup>a</sup> QSAR_MLR	pLD <sub>50</sub> <sup>a</sup> QSAR-SVM	PMF-score
1	P1 <sup>GII</sup>	2,6-Cl <sub>2</sub>	2,020	5,03	5,01	-52,261
2	P2 <sup>GII</sup>	2-CF <sub>3</sub> ,6-OH	1,656	6,08	5,83	-47,445
3	P3 <sup>GI</sup>	2-F,6-CH <sub>3</sub>	0,722	5,56	5,61	-48,288
4	P4 <sup>GII</sup>	2-Cl,6-CH <sub>3</sub>	0,748	6,26	6,23	-49,971
5	P5 <sup>GI</sup>	2-Br,6-CH <sub>3</sub>	0,608	5,30	5,18	-42,481
6	P6 <sup>GI</sup>	2-OH,6-SCF <sub>3</sub>	0,362	5,10	4,88	-35,349
7	P7 <sup>GI</sup>	2-CF <sub>3</sub> ,6-CH <sub>3</sub>	0,363	5,17	5,17	-44,838
8	P8 <sup>GI</sup>	3,5-F <sub>2</sub>	1,886	6,12	6,04	-62,689
9	P9 <sup>GI</sup>	2,4-F <sub>2</sub>	1,686	6,04	6,05	-68,109
10	P10 <sup>GII</sup>	2,3-F <sub>2</sub>	0,991	6,46	6,51	-50,067
11	P11 <sup>GI</sup>	3-Cl	1,871	5,87	5,89	-59,023
12	P12 <sup>GI</sup>	4-F	1,440	5,62	5,67	-50,621
13	P13 <sup>GI</sup>	2,5-F <sub>2</sub>	0,709	6,41	6,52	-66,626

(+) GI e GII formam grupos para análise de correlação da atividade larvicida pLD<sub>50</sub> com PMF-score; (a) valores pLD<sub>50</sub> preditos (LD<sub>50</sub> = mmol/inseto).

Foi feita uma correlação com as atividades larvicidas preditas tanto pelos modelos de QSAR-MLR como SVM com o índice PMF-score, onde ambas as correlações estão próximas ( $n = 13$ ,  $R_{MLR} = 0,604$  e  $R_{SVM} = 0,663$ ) e satisfatórias se comparadas aos valores citados pela literatura ( $R = 0$  a  $0,37$ ).<sup>111,114</sup> A correlação entre os valores pLD<sub>50</sub> preditos pelos modelos de QSAR MLR e SVM resulta igualmente ser satisfatória: ( $n = 13$ ,  $R = 0,982$ ;  $s = 0,099$ ;  $F = 300$ ;  $p < 0,0001$ ). O pior resultado de *docking* foi obtido pelo composto P1, com RMSD de 2,020Å, portanto, acima do limite preconizado ( $< 2,0$  Å). Para concluir a análise, e conforme praticado antes, fez-se uma separação em dois grupos com os compostos melhor correlacionados, obtendo-se os resultados seguintes para  $pLD_{50} = f(\text{PMF-score})$ :

Grupo I ( $n=9$ ):  $R_{MLR} = R_{SVM} = 0,953$   $s=0,169$   $F = 70$   $p = 0,0001$ .

Grupo II ( $n=4$ ):  $R_{MLR,SVM}$  sem significância estatística.

A mesma análise é feita com os valores da Tabela 42 correlacionando a atividade larvicida  $pLD_{50}$  com os valores *X-score*,  $pLD_{50} = f(X\text{-score})$ , como segue:

Grupo I (n=8) :  $R_{MLR} = 0,956$   $s = 0,189$   $F = 63$   $p = 0,0002$ .

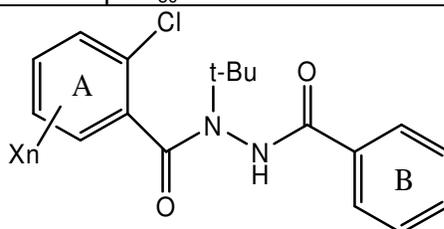
$R_{SVM} = 0,938$   $s = 0,245$   $F=43$   $p = 0,0006$ .

Grupo II (n=4):  $R_{MLR} = 0,998$   $s = 0,024$   $F = 474$   $p = 0,002$ .

$R_{SVM} =$  sem significância estatística.

O composto P5 tem comportamento de *outlier*.

**Tabela 42.** Compostos propostos contra *S. exigua*. Resultados dos modelos de QSAR-MLR, QSAR-SVM e correlação dos valores  $pLD_{50}$  com *X-score*.



Item	MolID	Substituintes Xn	$pLD_{50}^a$		<i>X-score</i> <sup>b</sup>	Componentes <i>X-score</i>			LBE <sup>d</sup> (kcal/mol)
			QSAR_MLR	QSAR_SVM		HP <sup>c</sup>	HM <sup>c</sup>	HS <sup>c</sup>	
1	P1 <sup>GI</sup>	2,6-Cl <sub>2</sub>	5,03	5,01	7,48	7,00	8,23	7,22	-10,21
2	P2 <sup>GII</sup>	2-CF <sub>3</sub> ,6-OH	6,08	5,83	7,65	7,35	8,28	7,31	-10,43
3	P3 <sup>GII</sup>	2-F,6-CH <sub>3</sub>	5,56	5,61	7,38	7,00	8,00	7,15	-10,07
4	P4 <sup>GII</sup>	2-Cl,6-CH <sub>3</sub>	6,26	6,23	7,73	7,37	8,40	7,44	-10,55
5	P5 <sup>(*)</sup>	2-Br,6-CH <sub>3</sub>	5,30	5,18	7,78	7,37	8,47	7,49	-10,61
6	P6 <sup>GI</sup>	2-OH,6-SCF <sub>3</sub>	5,10	4,88	7,45	7,16	8,07	7,12	-10,16
7	P7 <sup>GI</sup>	2-CF <sub>3</sub> ,6-CH <sub>3</sub>	5,17	5,17	7,43	7,08	8,11	7,08	-10,13
8	P8 <sup>GI</sup>	3,5-F <sub>2</sub>	6,12	6,04	7,24	6,88	7,87	6,98	-9,88
9	P9 <sup>GI</sup>	2,4-F <sub>2</sub>	6,04	6,05	7,30	6,91	7,92	7,06	-9,95
10	P10 <sup>GI</sup>	2,3-F <sub>2</sub>	6,46	6,51	7,26	6,88	7,87	7,04	-9,90
11	P11 <sup>GII</sup>	3-Cl	5,87	5,89	7,52	7,06	8,23	7,27	-10,26
12	P12 <sup>GI</sup>	4-F	5,62	5,67	7,33	6,93	7,97	7,08	-9,99
13	P13 <sup>GI</sup>	2,5-F <sub>2</sub>	6,41	6,52	7,27	6,90	7,88	7,02	-9,91

(a) Valor predito ( $LD_{50} = \text{mmol/inseto}$ ); (b) estrutura de referência: cristal de BY106830 (RMSD = 0,512Å); (c) parâmetros de afinidade de ligação:  $(HP+HM+HS)/3 = X\text{-score}$ ; (d) LBE (*Ligand Binding Energy*-Energia de ligação); (+) GI e GII formam grupos para análise de correlação da atividade larvicida  $pLD_{50}$  com *X-score*;(\*) P5 é *outlier* em ambos os grupos.

O parâmetro LBE no programa *X-score* representa o valor predito da energia de ligação ligante-proteína como descrito pela Eq. (5.4-11) e também pode ser usado como um indicador na qualidade dos resultados. Utilizando-se o procedimento por grupos como nos casos anteriores, e empregando o conjunto de compostos propostos obtém-se para os valores  $pLD_{50} = f(LBE)$  do modelo de QSAR-MLR:

Grupo I (n=8):  $R_{MLR} = 0,956$   $s = 0,188$   $F = 63$   $p = 0,0002$ .

Grupo II (n=4):  $R_{MLR} = 0,998$   $s = 0,021$   $F = 631$   $p = 0,0016$ .

Para os valores  $pLD_{50} = f(LBD)$  do modelo de QSAR-SVM:

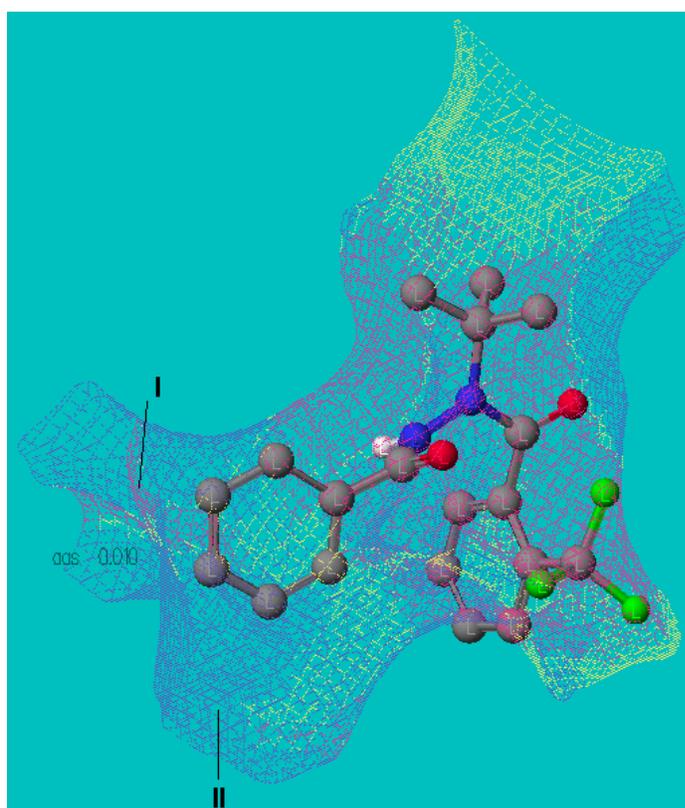
Grupo I (n=8):  $R_{SVM} = 0,940$   $s = 0,241$   $F = 45$   $p = 0,0005$ .

Grupo II (n=4):  $R_{SVM}$  = sem significância estatística.

O composto P5, como anteriormente citado, tem comportamento de *outlier*.

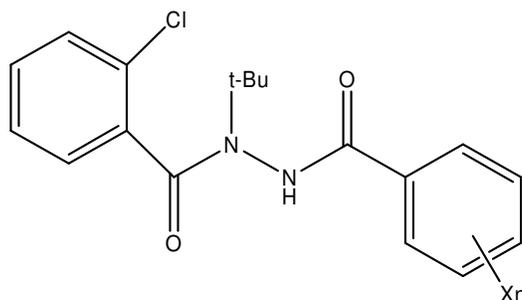
Os valores das correlações obtidas em cada um dos grupos de  $pLD_{50}$  preditos com os parâmetros independentes PMF-score, X-score e LBE, são indicadores de que o *docking* pode ser considerado qualitativamente satisfatório.

Através da malha tridimensional que forma a cavidade no domínio ligante da proteína SeEcR-LBD, pode-se verificar também dois espaços livres ao redor do anel B da dibenzoilidrazina representada na Figura 68 pelas regiões I e II, e que oferecem oportunidades para o desenvolvimento de novas propostas.



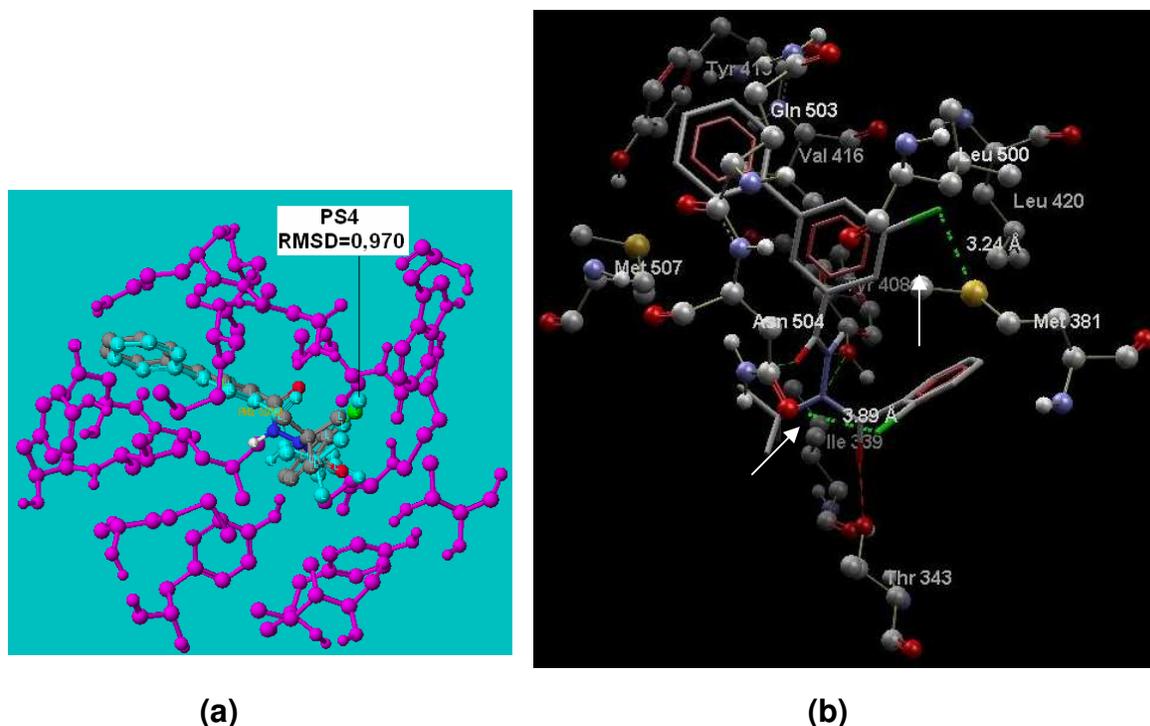
**Figura 68** Cavidade formada na proteína SeEcR na região do ligante (dibenzoilidrazina D6 usada de exemplo). As regiões I e II oferecem oportunidades para desenvolvimento de novos compostos.

Dibenzoilidrazinas substituídas no anel B (Figura 69) foram bioensaiadas por Smaghe *et al.*<sup>130</sup> com quarenta e duas dibenzoilidrazinas contra *S. exigua*. Grande parte dos substituintes são os mesmos usados por Nakagawa *et al.*<sup>27</sup> no qual se baseou este trabalho.



**Figura 69.** Dibenzoilidrazina substituída conforme Smaghe *et al.*<sup>130</sup>.

Não é propósito deste estudo explorar esta região, mas apenas indicá-la como uma oportunidade para novas pesquisas em aditamento ao trabalho de Smaghe *et al.*<sup>130</sup>. Através de *docking* de um composto proposto *in silico*, obteve-se um RMSD = 0,970 Å. Com isto foi possível exemplificar o preenchimento da cavidade (Figura 70a) assim como possíveis interações não-covalentes dos halogênios (Figura 70b).



**Figura 70.** (a) Ligante PS4 construído *in silico*. Estrutura de *docking* (em cor cinza) sobreposta à sua similar, mostrando bom preenchimento da cavidade relativa à região II da Figura 68. (b) As setas indicam possíveis interações dos substituintes halogênicos do composto Yn = 4-(Ph-2Cl) no anel B (Figura 69) com os aminoácidos: Cl...O = C-, 3,89 Å (Asn 504); Cl...S, 3,24 Å (Met 381).

#### 5.4.4 – Conclusão

Através do emprego de dois métodos de regressão, MLR e SVM, construíram-se modelos de QSAR a partir de confôrmeros que melhor representaram o composto na atividade larvicida experimental, e que reproduziram os valores experimentais com boa proximidade, mostrando equivalência. Verificou-se o peso da polarizabilidade estática  $\alpha_{yy}$ , do momento de dipolo  $\mu_x$  e da função de distribuição radial calculada pela polarizabilidade atômica RDF095p na atividade larvicida para cada um dos compostos nas posição *orto*, *meta* ou *para*. Para alguns dos compostos halogenados, é possível associar o valor das forças dispersivas de London e obter uma idéia aproximada da energia de interação não-covalente do composto com o respectivo átomo do aminoácido.

Os modelos de QSAR mostraram-se consistentes, pois os descritores mencionados representam conceitos físicos de polarizabilidade que estão muito presentes na literatura relativa a interações ligante-proteína. Como complemento, foi possível mostrar a existência de uma correlação entre a polarizabilidade estática e o potencial de ionização entre alguns dos compostos.

O estudo de *docking* feito com vinte e cinco dibenzoilidrazinas halogenadas, partindo-se de um cristal de referência complexado com a proteína 1R20, obteve valores de RMSD abaixo de 2,0 Å para vinte e quatro compostos. Este resultado qualitativo permitiu investigar quais as possíveis interações não-covalentes dos substituintes com os átomos de aminoácidos vizinhos, contribuindo com uma informação teórica cujos resultados são compatíveis com a literatura. O bom resultado obtido pelo *docking* com o domínio ligante HvEcR-LBD motivou a pesquisa complementar, empregando metodologias conhecidas para obtenção de uma proteína construída por homologia, baseada na proteína 1R20 e através de novo estudo de *docking*, conhecer as interações das dibenzoilidrazinas na proteína homóloga de *Spodoptera exigua*. O emprego de PMF-score e X-score, duas metodologias independentes empregadas para obter a bioafinidade predita ligante-proteína gerada pelo *docking*, permitiu correlacionar os valores numéricos com a atividade larvicida experimental para o complexo HvEcR-LBD e predita pelos modelos de QSAR MLR e SVM, assim como para o complexo teórico SeEcR-LBD, com resultados satisfatórios e consistentes sob aspecto metodológico.

Foram propostas treze dibenzoilidrazinas halogenadas contra a lagarta *S. exigua*, igualmente submetidas a *docking* e à avaliação da bio-afinidade (estimada) ligante-proteína. Destas, doze tiveram RMSD < 2.0 Å, indicando assim bom ajuste ligante-proteína. Assumiram-se os valores preditos pelos modelos de QSAR como adequados para verificação da correlação bioafinidade ligante-proteína, novamente gerando correlações que qualificam os resultados para sua utilização prática.

Como já citado, os resultados de *docking* das dibenzoilidrazinas partiram da validação do *docking* no complexo cristalizado HvEcR-LBD da proteína 1R20. Estes resultados assim como o *docking* de DBHs no domínio ligante-proteína SeEcR-LBD, requerem validação experimental, não contida no escopo desta tese.

## 5.5–Dibenzoilidrazinas contra *Spodoptera frugiperda* (J.E. Smith) Lepidoptera, Noctuidae

### 5.5.1 – Introdução

A lagarta-do-cartucho *Spodoptera frugiperda* (Lepidoptera) é uma das maiores predadoras da cultura do milho, classificada como a praga mais importante desta cultura, ocasionando perdas na produção<sup>131,132,133</sup> que se aproximam dos 40%. O cartucho da planta é o seu alvo preferencial, mas ataca também o pendão e espigas em formação.<sup>134</sup>

Utilizando células Sf-9 dos ovários de *Spodoptera frugiperda* Ogura *et al.*<sup>26</sup> fizeram estudos experimentais de bioafinidade de ligação com dezoito dibenzoilidrazinas *para*-substituídas. Dibenzoilidrazinas são agonistas da ecdisona, e têm o propósito de inibir a ligação da Ponasterona A, hormônio esteroide usado no ensaio, regulador da metamorfose do inseto. A afinidade de ligação dos compostos ao receptor da célula é dada por  $pIC_{50} = \log(1/IC_{50})$ , onde  $IC_{50}$  (M) representa a concentração para 50% da inibição da atividade da Ponasterona A.

No modelo de QSAR desenvolvido, o composto D18 por apresentar grande valor residual, foi classificado como *outlier*. Tem o menor  $pIC_{50}$ , porém, tem o segundo melhor valor do parâmetro hidrofóbico Log P. Os compostos com Log P mais elevados foram os que apresentaram a melhor bioafinidade, excluído D18. Segue o modelo de QSAR obtido por Ogura *et al.*<sup>26</sup> e transcrito conforme publicado:

$$pIC_{50} = + 0,607 (\pm 0,238) \text{ LogP} - 0,822 (\pm 0,532)\sigma - 0,364 (\pm 0,441) \Delta B_1 + 5,502 (\pm 0,723)$$

$$n = 17 \quad R^2 = 0,914 \quad s = 0,243 \quad F = 46 \quad p < 0,0001 \quad Q^2 = 0,864 \quad SPress = 0,305 \quad SDEP = 0,275 \quad MSE = 0,047 \quad (5.5-1)$$

onde  $\sigma$  é a constante substituinte de Hammett,  $\Delta B_1$  é o descritor estérico STERIMOL de Verloop, e LogP o coeficiente de partição octanol/água que quantifica a hidrofobicidade do composto. O erro ( $\pm 0,441$ ) do coeficiente  $\Delta B_1$  é justificado<sup>26</sup> e corresponde a 90,4% no intervalo de confiança.

Os dados experimentais contidos na literatura<sup>26</sup> serviram de base para o desenvolvimento de um novo modelo de QSAR e uma proposta de novos compostos por este estudo.

### **5.5.2 – Métodos**

Foram utilizados os métodos descritos em Material e Métodos, Seção 4, acrescidos de outros métodos para estudo de problemas específicos conforme segue mais adiante.

Os parâmetros LogP e a constante eletrônica de Hammett  $\sigma$ , além de descritores clássicos de QSAR, são possivelmente os mais adequados para interpretar o mecanismo de ação em função de diversos modelos de QSAR desenvolvidos por outros grupos de estudos com lagartas do gênero *Lepidoptera*. Por esta razão foram preservados. Utilizou-se também neste estudo o descritor Mor25v, escolhido por algoritmo genético pelo aplicativo BuildQSAR, e que substitui o descritor STERIMOL  $\Delta B_1$ . Desta forma, introduziu-se um descritor sensível às variações conformacionais. Mor25v é um descritor do programa DRAGON pertencente ao grupo 3D-MoRSE (*3D-Molecular Representation of Structures based on Electron diffraction*). Detalhes adicionais são dados em Resultados e Discussão na sub-seção 5.5.3.

### **Estudo de Reatividade Química das DBHs contra *S. frugiperda***

O modelo de QSAR de dibenzoilidrazinas contra *S. frugiperda* da literatura<sup>26</sup> deixou em aberto uma questão de interesse prático. A hidrofobicidade expressa pelo descritor de partição octanol/água (LogP) teve marcada importância no estudo de bioafinidade, o que foi verificado experimentalmente. Recordando, o composto D18 apresentou o pior resultado experimental mas tem o segundo melhor valor para LogP em relação aos demais compostos.

Numa tentativa de estudar a possível causa da baixa bioafinidade de D18, foi feita uma análise sob o enfoque da reatividade envolvendo aspectos nucleofílicos e eletrofílicos. A análise se estende a todas as 18 dibenzoilidrazinas. A metodologia

empregada foi a mesma desenvolvida por Parr *et al.*<sup>135</sup> assim como aplicações com complementações introduzidas por Roy *et al.*<sup>136</sup> e Parthasarathi *et al.*<sup>137</sup>.

Relacionam-se a seguir as variáveis que são calculadas e utilizadas na análise. Para o completo formalismo matemático acompanhado de todos os conceitos físico-químicos sugere-se a consulta dos artigos citados e de algumas citações adicionais ao longo do texto.

### Índice de eletrofilicidade

Proposto por Parr *et al.*<sup>135,136</sup>, o índice de eletrofilicidade global é dado por

$$\omega \equiv \mu^2 / 2\eta \quad (5.5-2)$$

A Eq. (5.5-2) mede a eletrofilicidade de um ligante representado pela dibenzoilidrazina ao interagir com o sítio receptor da proteína na célula de ovas Sf-9 da lagarta *S. frugiperda*, onde:

$$\mu = (\partial E / \partial N)_{v(r)} \quad (5.5-3)$$

representa o potencial químico, que é variação da energia em relação à transferência de elétrons sob a ação de um potencial externo  $v(r)$  constante.  $\mu = \mu(N, v)$  mede a tendência dos elétrons de escaparem de um sistema em equilíbrio.<sup>138</sup> O parâmetro  $\eta$  é definido como ‘dureza química’ (*hardness*)<sup>139</sup>:

$$\eta = (\partial \mu / \partial N)_{v(r)} = (\partial^2 E / \partial N^2)_{v(r)} \quad (5.5-4)$$

Empregando o método de aproximação por diferença finita na Eq. (5.5-3) obtém-se:

$$\mu = -\chi = -\frac{I + A}{2} \quad (5.5-5)$$

onde  $I$  representa o Potencial de Ionização vertical e  $A$  a afinidade eletrônica.  $\chi$  representa a eletronegatividade da molécula. O potencial de ionização é definido como a quantidade de energia requerida para remover um elétron da molécula  $\text{PH}_2$ , onde P representa um átomo, computada como a diferença entre a energia do cátion  $E(\text{PH}_2^+)$  e a energia da molécula neutra  $E(\text{PH}_2^0)$ :

$$I = E(\text{PH}_2^+) - E(\text{PH}_2^0) \quad (5.5-6a)$$

A afinidade eletrônica  $A$  é definida como a energia liberada quando um elétron é adicionado à molécula neutra, calculada como a diferença de energias entre a forma neutra e a energia do ânion  $E(\text{PH}_2^-)$ :

$$A = E(\text{PH}_2^0) - E(\text{PH}_2^-) \quad (5.5-6b)$$

Empregando novamente o método de aproximação por diferença finita na Eq. (5.5-3), chega-se à equação:

$$\eta = I - A \quad (5.5-7)$$

Outra grandeza calculada é a maciez (*softness*)  $S$ , que é a variação de transferência de elétrons em função do potencial químico, e expressa pelo inverso da dureza:

$$S = \left( \frac{\partial N}{\partial \mu} \right)_{v(r)} = \frac{1}{I - A} \quad (5.5-8)$$

#### Índice de Reatividade de Fukui (Função de Fukui)

A função de Fukui ou função de fronteira foi definida por Parr e Yang<sup>140</sup> e dado pela Eq. (5.5-9). É utilizada para avaliar a perturbação externa e como o potencial químico de um sistema é afetado. A avaliação é feita pela derivada do funcional do potencial químico com relação à mudança no potencial externo, ou, segundo às relações de Maxwell, idêntica à variação da densidade eletrônica em relação ao número de elétrons:

$$f(\vec{r}) = \left( \frac{\partial \mu}{\partial v(\vec{r})} \right)_N = \left( \frac{\partial \rho(\vec{r})}{\partial N} \right)_{v(\vec{r})} \quad (5.5-9)$$

Os cálculos neste estudo se utilizam da forma simplificada, proposta por Yang e Mortier<sup>141</sup> resumida nas equações abaixo:

$$\text{Para ataque nucleofílico } f_i^+ = q_i(N+1) - q_i(N) \quad (5.5-10a)$$

$$\text{Para ataque eletrofílico } f_i^- = q_i(N) - q_i(N-1) \quad (5.5-10b)$$

$$\text{Para ataque radical } f_i^0 = 0,5[q_i(N+1) + q_i(N-1)] \quad (5.5-10c)$$

onde  $q_i$  é a carga de cada átomo da molécula.

Segundo diversos estudos comparativos<sup>142,143</sup>, esta pesquisa se orientou pelas recomendações da literatura para emprego da análise populacional de Hirshfeld (HPA).

### Índice de Eletrofilicidade Local – Maciez Local (*Local softness*)

Chattaraj *et al.*<sup>144</sup> em sua análise sugerem a adoção de um índice de eletrofilicidade local,  $\omega_k^\alpha$ , que supostamente é mais sensível em medir propensões de ataque eletrofílico, nucleofílico e radical, que varia de átomo para átomo em uma molécula, sendo que o ataque eletrofílico global está propenso a todos os ataques e se conserva.  $\omega_k^\alpha$  portanto, está associado a alguns átomos de uma região da molécula, e pode ser obtido pela equação geral :

$$\omega^\alpha(\vec{r}) = \omega f^\alpha(\vec{r}) \quad (5.5-11)$$

onde  $\alpha = +, -, e 0$  significa, ataque nucleofílico, eletrofílico e radical, respectivamente. A representação condensada das variáveis  $\omega^\alpha$  no sítio atômico  $k$  se escreve como:

$$\omega_k^\alpha(\vec{r}) = \omega f_k^\alpha \quad (5.5-12)$$

A maciez local  $S_k^\alpha$  é um descritor igualmente utilizado na análise, diferenciando-se da maciez global pela equação :

$$S_k^\alpha = S f_k^\alpha \quad (5.5-13)$$

e a maciez global dada por:

$$S_g = \sum_{k=1}^n S_k \quad (5.5-14)$$

### Índice de Eletrofilicidade de Grupo (ou filicidade de grupo)

A filicidade condensada somada sobre o grupo de átomos relevantes ao estudo pode ser expressa pela equação<sup>137</sup>

$$\omega_g^\alpha = \sum_{k=1}^n \omega_k^\alpha \quad (5.5-15)$$

onde  $n$  é o número de átomos coordenados ao átomo reativo,  $\omega_k^\alpha$  a eletrofilicidade local do átomo mais próximo ligado, e  $\alpha = +, -, e 0$  têm o mesmo significado anterior: ataque nucleofílico, eletrofílico e radical, respectivamente.

O cálculo de contribuição eletrofílica e nucleofílica de grupo se utiliza de todos os átomos de uma região de interesse, que no caso específico deste estudo, são todos os átomos do anel benzênico B mais todos os átomos que integram o substituinte (v. estrutura na Tabela 43, em Anexos, pg. 236).

Para obter os valores numéricos de todos os parâmetros descritos, utilizou-se o pacote computacional ADF<sup>46</sup> utilizando Teoria do Funcional de Densidade (DFT) para a otimização da geometria com o conjunto de base TZP (Triplo Zeta com Polarização) e o funcional Becke-Perdew 86. Além disto fez-se o cálculo das frequências para assegurar que a estrutura esteja no seu mínimo, o que foi avaliado pela ausência de frequências imaginárias.

A sequência de cálculo é a seguinte: **(a)** otimização da geometria inicial com o método semi-empírico AM1. **(b)** Segue-se o método DFT da otimização da geometria. **(c)** Cálculo das frequências. **(d)** Cálculo da energia *Single Point* da molécula neutra, do cátion, e do ânion. **(e)** Disposição de dados numa planilha, para cálculo dos parâmetros necessários juntamente com as cargas de Hirshfeld, que são automaticamente geradas na otimização da geometria pelo pacote ADF. Os dados são extraídos através de um *script* em linguagem Perl para este propósito.

### 5.5.3 – Resultados e Discussão

#### 5.5.3.1 – QSAR–MLR, QSAR–SVM e Descritores

##### QSAR-MLR e Descritores

Após análise conformacional, obteve-se o seguinte modelo de QSAR, sem eliminar D18 do conjunto, cuja predição está próxima ao valor observado, contrariamente ao valor obtido pelo estudo da ref.[26], onde o valor residual é grande:

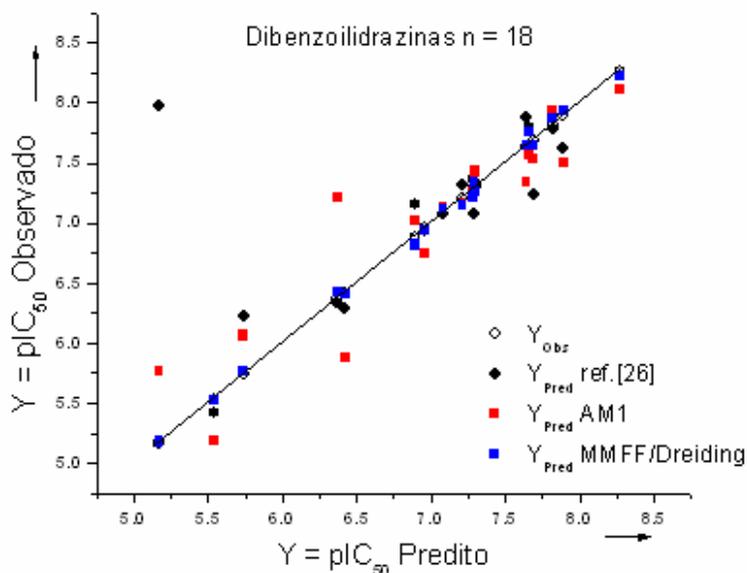
$$pIC_{50} = + 0,215 (\pm 0,041) \text{ LogP} - 1,061 (\pm 0,102) \sigma - 7,106 (\pm 0,316) \text{ Mor25v} + 10,476 (\pm 0,231)$$

$$n = 18 \quad R^2 = 0,996 \quad s = 0,055 \quad F = 1384 \quad p < 0,0001 \quad Q^2 = 0,994 \quad \text{SPress} = 0,070 \\ \text{SDEP} = 0,064 \quad \text{MSE} = 0,0023 \quad (5.5-16)$$

onde Mor25v (3D-Morse – *signal 25/ weighted by atomic van der Waals volumes*), é o descritor sensível às variações conformacionais e é através dele que os confôrmeros são escolhidos como explicado em Material e Métodos, Seção 4.

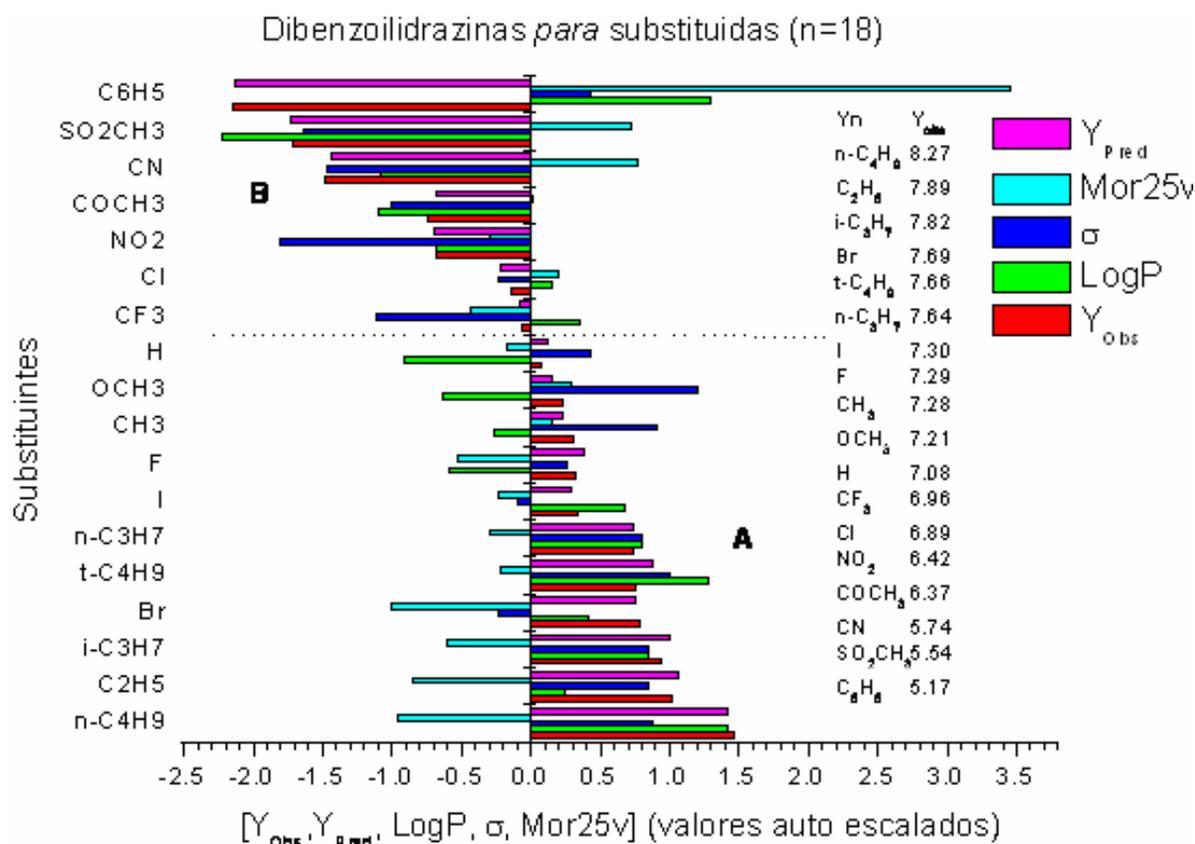
As predições para cada um dos compostos antes e após análise conformacional podem ser verificadas na Tabela 43 (v. em Anexos, pg. 236), incluindo os resultados da literatura.<sup>26</sup> Os valores dos descritores utilizados para construir o modelo de QSAR estão contidos na referida tabela e o gráfico correspondente na Figura 71. A matriz de correlação R apresenta os seguintes valores entre variáveis: LogP/ $\sigma$  (-0,594); LogP/Mor25v (-0,048);  $\sigma$ /Mor25v (0,109). Entre descritores e bioafinidade: pIC<sub>50</sub>/LogP (0,509); pIC<sub>50</sub>/ $\sigma$  (-0,643); pIC<sub>50</sub>/Mor25v (-0,810).

A teoria envolvendo descritores do tipo 3D-MoRSE já foi descrita na seção 5.3, porém, serão dados detalhes de como o descritor se comporta em função da variação conformacional e sua relação com a bioafinidade.



**Figura 71.** Resultado comparativo de valores preditos conforme metodologia de otimização de geometria pelo método AM1 e MMFF94 ou Dreiding modificado. O método da ref. [26] não está especificado na literatura.

A equação (5.5-16) está representada graficamente na Figura 72, onde as barras representam a contribuição de cada um dos parâmetros na bioafinidade de ligação das dibenzoilidrazinas com as células Sf-9 de *S. frugiperda*. Verifica-se que há grande contribuição do  $\sigma$  de Hammett que tem sinal negativo na equação QSAR e que pela interpretação de Ogura *et al.*<sup>26</sup> sugerem um efeito elétron-doador do substituinte favorecendo uma interação do grupo CONH com o receptor, este último até o presente desconhecido na literatura. Independentemente desta interpretação, pode-se verificar pela Figura 72 que a constante eletrônica  $\sigma$  tem prevalência em quase todas as atividades maiores (região **A**). Na região **B** há uma influência redutora da atividade, exceto para o composto substituído por um fenil ( $C_6H_5$ ) que será alvo de uma análise em tópico específico mais adiante, na sub-seção 5.5.3.5.

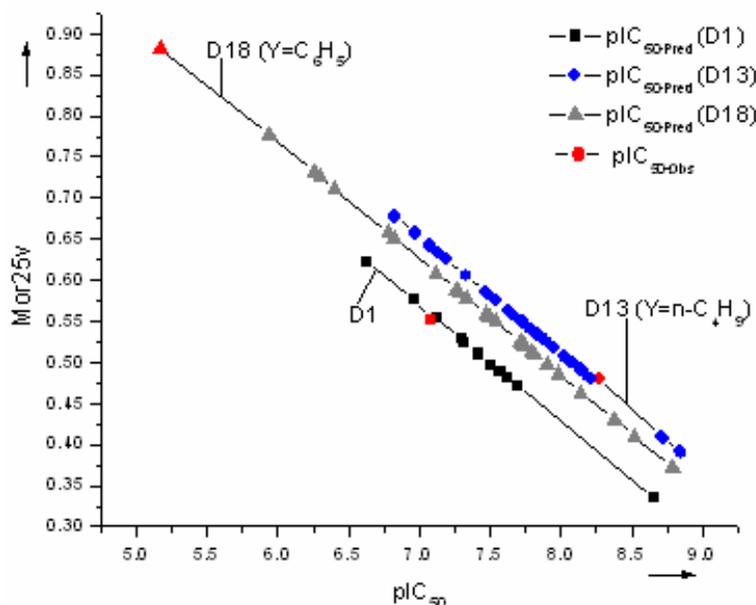


**Figura 72.** Ilustração da influência de cada um dos parâmetros na descrição da bioafinidade de ligação das dibenzoilidrazinas com as células Sf-9.

Mor25v, codificado pelo volume de van der Waals, marca presença maior na região **A**, porém, indicando agir de modo desfavorável à bioafinidade, como é também o caso do parâmetro STERIMOL  $\Delta B_1$ , da equação (5.5-1) desenvolvida por Ogura *et al.*<sup>26</sup>.

A Figura 73 mostra que há resposta linear entre a bioafinidade  $pIC_{50}$  e o valor do parâmetro Mor25v para os confômeros correspondentes de cada composto. Esta situação se repete para todos os compostos além dos três representados no gráfico. Verifica-se que na medida em que Mor25v decresce a bioafinidade é crescente. Para todos os compostos juntos não há uma boa correlação linear, mas quando se formam grupos por tipo de substituinte, ela se verifica, como está caracterizado na

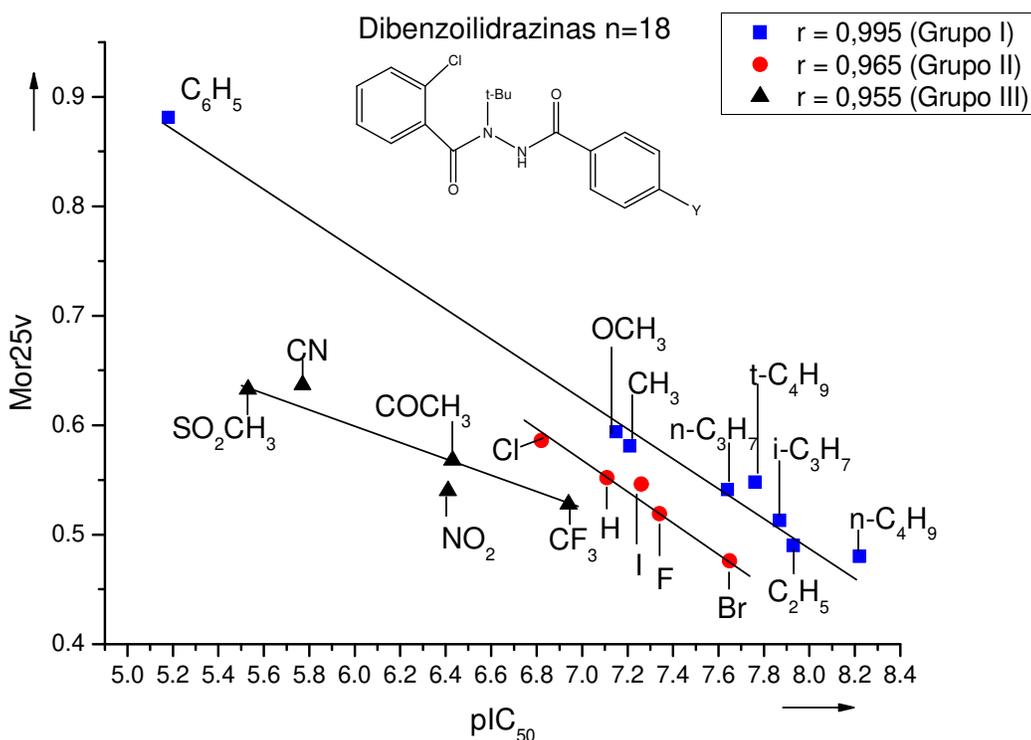
Figura 74. Notar que os halogênios estão relacionados na mesma reta com o composto não substituído H, assim como os grupos  $\text{CH}_3$ ,  $\text{C}_3\text{H}_7$  e  $\text{C}_4\text{H}_9$ .



**Figura 73.** Comportamento linear de valores de Mor25v dos confôrmeros do composto não substituído D1, do composto D13 de maior valor  $\text{pIC}_{50}$  experimental, e D18 composto de menor valor  $\text{pIC}_{50}$  experimental. Pontos vermelhos correspondem aos valores experimentais de  $\text{pIC}_{50}$  associados aos valores Mor25v.

Os compostos substituídos por  $\text{SO}_2\text{CH}_3$ ,  $\text{COCH}_3$  e  $\text{CF}_3$  parecem pertencer à mesma classe descrita linearmente por Mor25v. CN e  $\text{NO}_2$  podem, talvez, formar uma classe própria, apesar de formarem uma correlação ainda aceitável com os três substituintes citados. Esta investigação, contudo, não está no escopo deste trabalho, mas não deixa de apresentar um aspecto cujos resultados poderiam servir de indicadores para seleção de substituintes na proposta de novos compostos.

As correlações obtidas em cada grupo e associadas ao gráfico da Figura 74 são as seguintes: Grupo I ( $n = 8$   $R = 0,995$   $s = 0,101$   $F = 613$ ); Grupo II ( $n = 5$   $R = 0,965$   $s = 0,090$   $F = 41$ ); Grupo III ( $n = 5$   $R = 0,955$   $s = 0,196$   $F = 31$ );

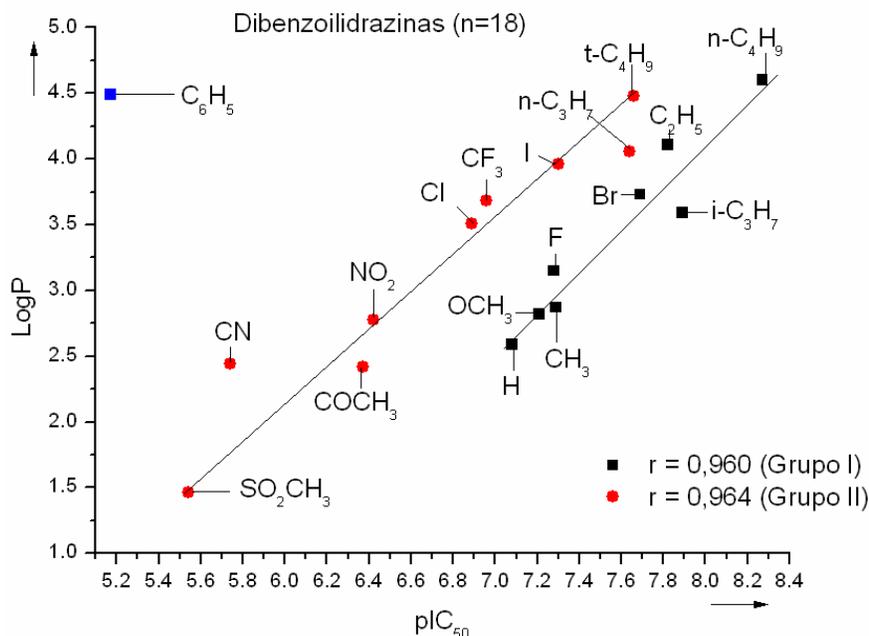


**Figura 74.** Correlação do descritor Mor25v com a bioafinidade, com os compostos separados em três grupos, verificando-se correlação linear. Os compostos CN e NO<sub>2</sub> podem eventualmente formar um grupo à parte. Os valores pIC<sub>50</sub> são os experimentais.

O gráfico sugere que compostos substituídos halogenados, assim como a série integrada por C<sub>x</sub>H<sub>y</sub>, oferece oportunidades de investigação na proposição de novos compostos.

Uma avaliação do parâmetro hidrofóbico revela que existem dois grupos de compostos fortemente correlacionados com LogP (Figura 75), porém, o composto D18, que tem o segundo melhor valor como já foi citado, se comporta como *outlier*. Diferentemente do descritor Mor25v que mostrou alguma especificidade de correlação nas classes de substituintes (halogênios, por exemplo), LogP não mostra esta especificidade, e na separação de grupos as correlações são as seguintes: Grupo I (n = 8 R = 0,960 s = 0,125 F = 71); Grupo II (n = 9 R = 0,964 s = 0,218 F = 92).

Assim, pelo gráfico verifica-se que, em termos gerais, a bioafinidade é crescente com compostos mais hidrofóbicos, fator a considerar na proposta de dibenzoilidrazinas na atividade larvicida de *S. frugiperda*.

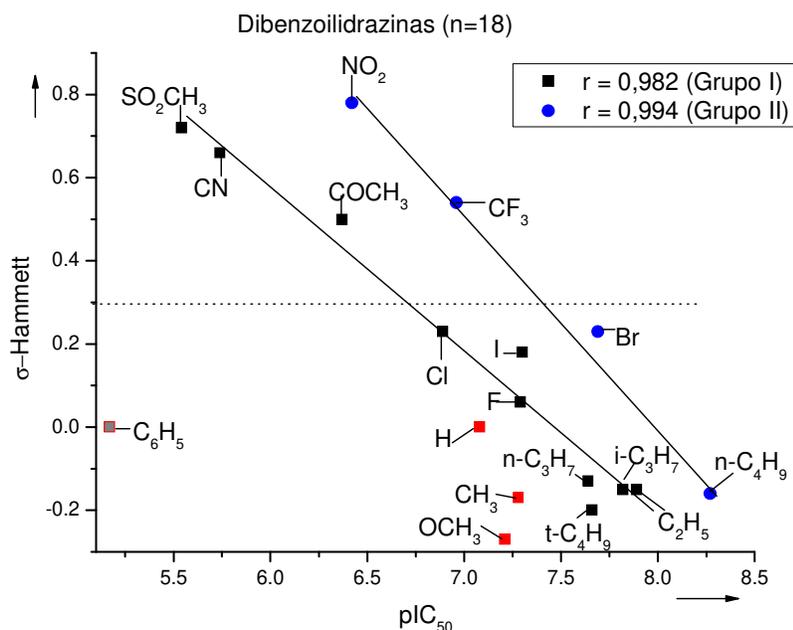


**Figura 75.** Correlação do descritor LogP com a bioafinidade experimental, com os compostos separados em três conjuntos. C<sub>6</sub>H<sub>5</sub> representa um conjunto de um único elemento.

Fazendo a mesma análise com o parâmetro  $\sigma$  de Hammett, destacam-se também três grupos, como ilustrado na Figura 76. Constata-se uma boa correlação em dois conjuntos de compostos: Grupo I (n = 10 R = 0,982 s = 0,170 F = 220); Grupo II (n = 4 R = 0,994 s = 0,108 F = 168). Usando os halogenados ou série C<sub>x</sub>H<sub>y</sub> novamente como exemplo, não há diferenciação de classe de substituintes. Verifica-se que os compostos com valores  $\sigma$  mais baixos (abaixo da linha pontilhada no gráfico) são os que apresentam bioafinidade mais favorável.

Em síntese, e para conclusão da análise, partindo-se das tendências mostradas nos gráficos, novas propostas de compostos que tenham baixo valor

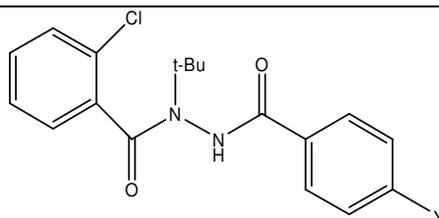
Mor25v, alto valor LogP, ou seja, boa hidrofobicidade, e uma constante eletrônica  $\sigma$  de valores baixos, têm possibilidades de apresentarem boa bioafinidade. O composto D13 (Y=n-C<sub>4</sub>H<sub>9</sub>) se classifica dentro destes critérios.



**Figura 76.** Correlação da constante eletrônica  $\sigma$  de Hammett com a bioafinidade experimental, com os compostos separados em três conjuntos, dois com elementos correlacionados, e outro sem correlação dos elementos (em vermelho).

### QSAR-SVM

Como o conjunto de compostos tem poucos integrantes, é conveniente utilizar um método independente da regressão MLR, na tentativa de reproduzir a predição dos valores experimentais o mais próximo possível dos preditos e verificar a correlação dos resultados de ambos os métodos. Neste sentido foi construído o modelo de QSAR-SVM cujos resultados de predição interna estão reunidos na Tabela 44.

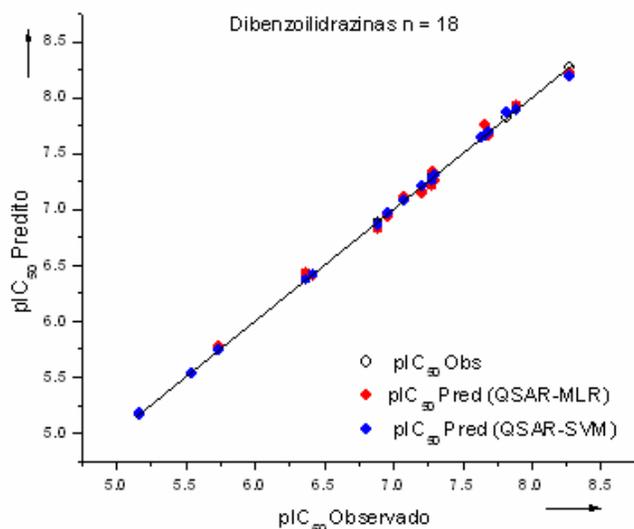
**Tabela 44.** Resultados de predição dos modelos dos modelos de QSAR-MLR e QSAR-SVM

No.	MolID	Substituinte Y	Y <sub>Obs</sub> <sup>a</sup>	Descritores			QSAR e Método			
				LogP	$\sigma$	Mor25v	Y <sub>Pred</sub> <sup>b</sup> MLR	Res <sup>c</sup>	Y <sub>Pred</sub> <sup>d</sup> SVM	Res <sup>e</sup>
1	D1	H	7,08	2,59	0	0,552	7,11	-0,03	7,08	0,00
2	D2	F	7,29	2,87	0,06	0,519	7,34	-0,05	7,29	0,00
3	D3	Cl	6,89	3,51	0,23	0,586	6,82	0,07	6,86	0,03
4	D4	Br	7,69	3,73	0,23	0,476	7,65	0,04	7,69	0,00
5	D5	I	7,30	3,96	0,18	0,546	7,26	0,04	7,30	0,00
6	D6	CF <sub>3</sub>	6,96	3,68	0,54	0,528	6,94	0,02	6,96	0,00
7	D7	NO <sub>2</sub>	6,42	2,78	0,78	0,540	6,41	0,01	6,42	0,00
8	D8	CN	5,74	2,44	0,66	0,637	5,77	-0,03	5,74	0,00
9	D9	CH <sub>3</sub>	7,28	3,15	-0,17	0,581	7,21	0,07	7,26	0,02
10	D10	C <sub>2</sub> H <sub>5</sub>	7,89	3,59	-0,15	0,490	7,93	-0,04	7,89	0,00
11	D11	n-C <sub>3</sub> H <sub>7</sub>	7,64	4,06	-0,13	0,541	7,64	0,00	7,64	0,00
12	D12	i-C <sub>3</sub> H <sub>7</sub>	7,82	4,11	-0,15	0,513	7,87	-0,05	7,86	-0,04
13	D13	n-C <sub>4</sub> H <sub>9</sub>	8,27	4,60	-0,16	0,480	8,22	0,05	8,19	0,08
14	D14	t-C <sub>4</sub> H <sub>9</sub>	7,66	4,48	-0,20	0,548	7,76	-0,10	7,66	0,00
15	D15	OCH <sub>3</sub>	7,21	2,82	-0,27	0,594	7,15	0,06	7,21	0,00
16	D16	SO <sub>2</sub> CH <sub>3</sub>	5,54	1,46	0,72	0,633	5,53	0,01	5,54	0,00
17	D17	COCH <sub>3</sub>	6,37	2,42	0,50	0,568	6,43	-0,06	6,37	0,00
18	D18	C <sub>6</sub> H <sub>5</sub>	5,17	4,49	0	0,881	5,18	-0,01	5,17	0,00

(a)  $Y = \text{pIC}_{50}$  (M); (b)  $\text{MSE} = 0,0023$ ; (c)  $Y_{\text{obs}} - Y_{\text{Pred}}$  (QSAR-MLR); (d) parâmetros do QSAR-SVM:  $C = 10$ ;  $\gamma = 0,10$ ;  $\epsilon = 0,0$ ;  $\text{MSE} = 0,0027$ ; (e)  $Y_{\text{obs}} - Y_{\text{Pred}}$  (QSAR-SVM).

O gráfico correspondente a esta tabela está representado na Figura 77.

A correlação obtida entre os valores de predição dos dois modelos de QSAR é dada a seguir: ( $n = 18$   $R = 0,999$   $s = 0,044$   $F = 6305$   $p < 0,0001$ ). Este resultado mostra ser um bom indicador de que os dois métodos independentes geram resultados próximos na predição interna dos modelos de QSAR. Outro detalhe a notar é que o modelo de QSAR-SVM reproduz o valor experimental do composto D18.



**Figura 77.** Gráfico dos valores preditos pelos modelos de QSAR MLR e SVM comparativamente aos valores experimentais.

### 5.5.3.2 – QSAR - *Outliers* e Domínio de Aplicabilidade

Empregando o gráfico de Williams (Figura 78), D18 é classificado como *outlier* mesmo apresentando resíduo pequeno. Contudo, D18 foi mantido no conjunto, que foi dividido em treinamento (n = 13) e teste (n = 5; D8, D13, D14, D16, D17) utilizado na validação externa do modelo, gerando o modelo de QSAR a seguir,:

$$pIC_{50} = + 0,206 (\pm 0,060) \text{ LogP} - 1,043 (\pm 0,111) \sigma - 7,046 (\pm 0,354) \text{ Mor25v} + 10,484 (\pm 0,2379)$$

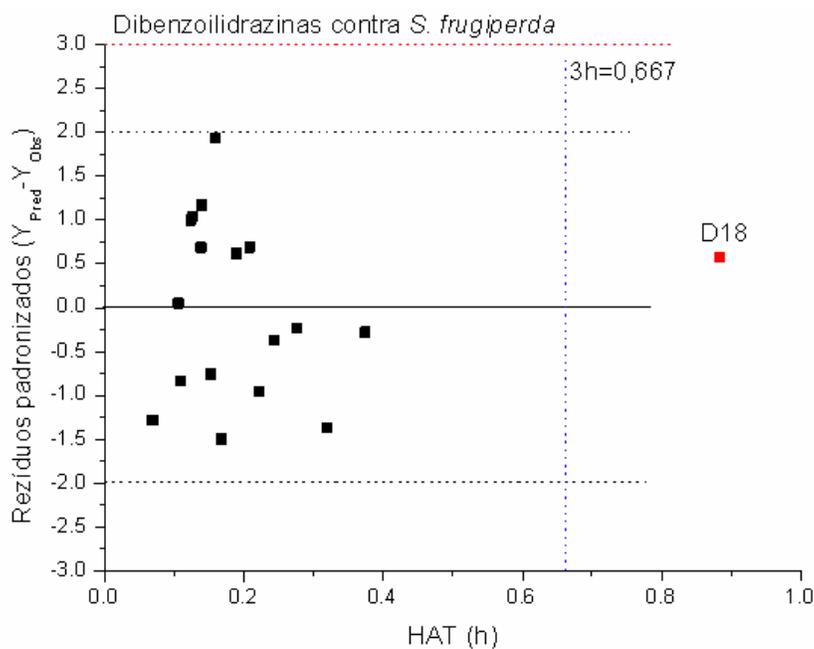
$$n = 13 \quad R^2 = 0,998 \quad s = 0,051 \quad F = 785 \quad p < 0,0001 \quad Q^2 = 0,971 \quad SPress = 0,140$$

$$SDEP = 0,121 \quad MSE = 0,0018 \quad (5.5-17)$$

Um aspecto a notar é a equivalência das equações (5.5-16) e (5.5-17) pela similaridade de coeficientes, característica positiva indicando estabilidade e uma tendência a produzir previsões próximas, conforme indica também o índice MSE.

A eliminação de D18 produz um resultado próximo ao da Eq. (5.5-16), e além disto, os resíduos obtidos encontram-se todos no intervalo  $-0,10 \leq \text{res} \leq 0,10$ .

D18 pode ser de fato um *outlier*, no sentido de apresentar características estéricas e físico-químicas peculiares, e em decorrência disto, ter uma interação desfavorável em relação aos demais compostos.



**Figura 78.** Diagrama de Williams – Domínio de aplicabilidade das dibenzoilidrazinas ( $n = 18$ ) contra *S. frugiperda*. O composto D18 comporta-se como *outlier*.

Verifica-se na Figura 78 que todos os compostos estão contidos dentro do domínio de aplicabilidade exceto o composto D18.

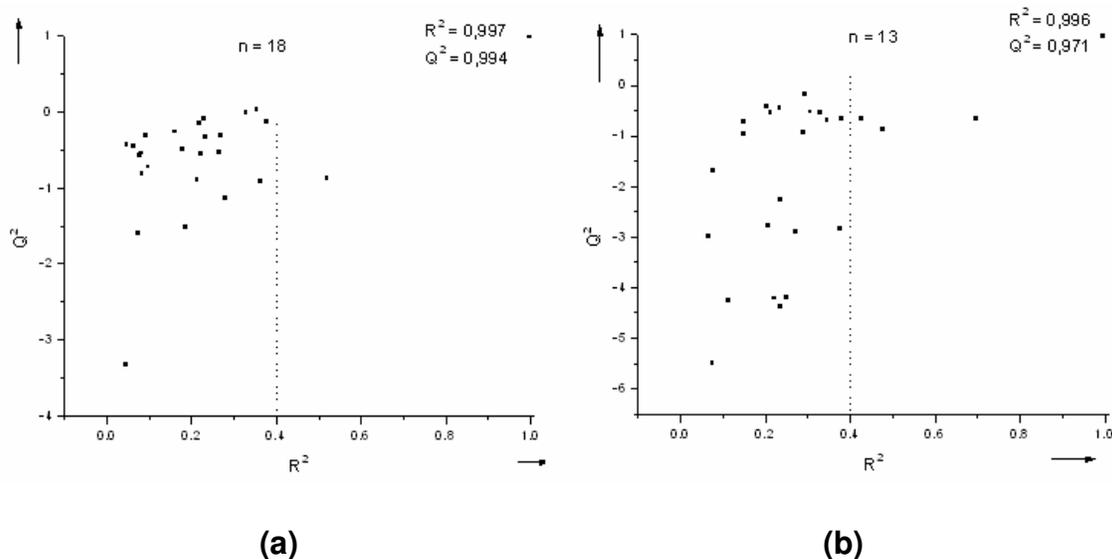
### 5.5.3.3 – QSAR – Testes de Robustez e Validação

#### Modelos de QSAR-MLR e QSAR-SVM

A robustez do conjunto de compostos  $n = 18$  (Eq. (5.5-16)) foi avaliada com os métodos *Y-randomization*<sup>53,61</sup> e *Leave-N-Out*.<sup>61</sup> O teste *Y-randomization* foi realizado com 300 iterações e 10 ciclos, resultando em 3000 iterações. Os valores do *intercept* usados para a aprovação do modelo sempre ocorreram abaixo dos valores limites  $a(R^2) < 0,3$  e  $a(Q^2) < 0,05$ . Segue o valor final do último ciclo:  $a(R^2) = 0,068$  e  $a(Q^2) = -1,004$ .

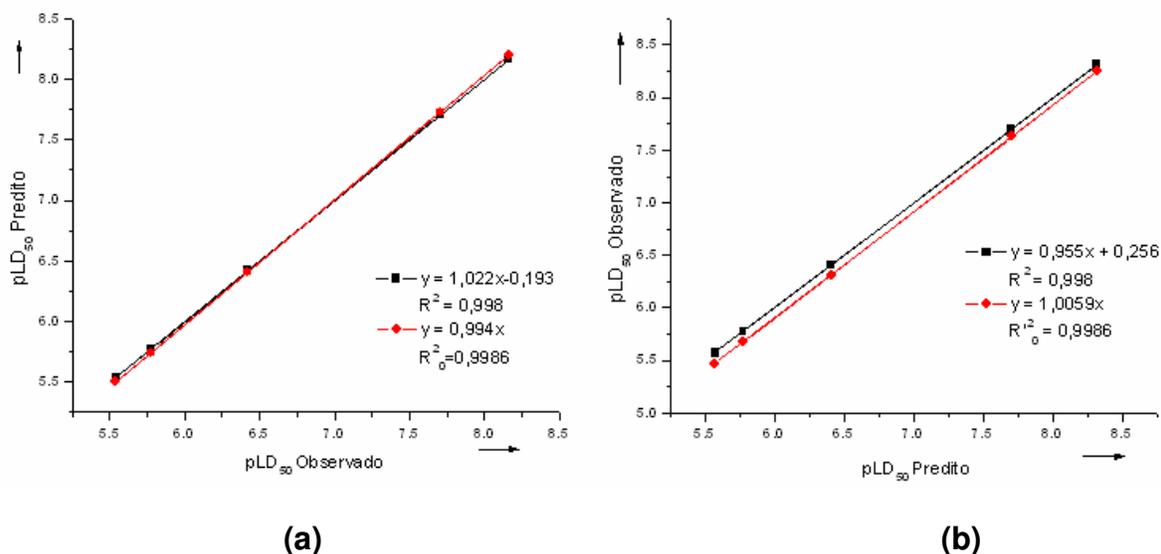
A mesma avaliação foi feita para os compostos  $n = 13$  (Eq. (5.5-17)) obtendo-se os valores  $a(R^2) = 0,081$  e  $a(Q^2) = -0,603$ .

Empregando o critério Eriksson-Wold<sup>58</sup>, foram feitas 25 iterações para a determinação de  $R^2$  e  $Q^2$  utilizando o programa da ref. [61], obtendo-se os gráficos da Figura 79, que apresentam resultados satisfatórios, dando robustez às Eq. (5.5-16 e 5.5-17) para predições externas.



**Figura 79.** Valores  $R^2$  e  $Q^2$  obtidos em função da randomização dos valores  $pLD_{50}$  experimentais para (a) modelos de QSAR-MLR e QSAR-SVM com  $n = 18$  e (b) modelos de QSAR-MLR e QSAR-SVM com  $n = 13$ .

Para a validação externa do modelo de QSAR, Eq.(5.5-17), usando o conjunto de teste  $n = 5$ , foi aplicado o critério de Golbraikh-Tropsha<sup>64</sup> cujo resultado está representado nos gráficos da Figura 80 assim como os parâmetros associados ao critério.



**Figura 80.** Representação gráfica dos parâmetros de validação do modelo de QSAR-MLR Eq. (5.5-17) das dibenzoilidrazinas pelas equações (a)  $y = ax+b$ ,  $y = kx$  e (b)  $y = a'x+b'$  e  $y = k'x$ .

A validação interna pelo método *bootstrap*, utilizando 5000 iterações, resulta em um valor final  $Q^2_{boot} = 0,989$  que comparado ao  $Q^2_{LOO} = 0,995$  valida a predição interna da Eq.(5.5-16).

Pelo critério *Leave-n-Out*, foram eliminados 6 compostos (30% do conjunto  $n = 18$ ), obtendo-se o resultado  $Q^2 - Q^2_{LNO} < 0,1$  para 10 repetições.

Empregando-se o critério de validação das funções de ajuste RQK (*RQK Fitness functions*) do programa MobyDigs dos modelos de QSAR Eq.(5.5-16 e 5.5-17), obteve-se resultados que aprovam os modelos (Tabelas 45 e 46).

**Tabela 45.** Validação do modelo de QSAR-MLR (n = 18) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis $K_{xy} - K_x > DK$	$K_{xy}$ 0,531	$K_x$ 0,307	$K_{xy}-K_x$ 0,224	DK 0,010
2.Capacidade de predição $Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,994	> 0,5		
3.Capacidade de predição do modelo $Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$ 0,994	$Q^2_{ASYM}$ 0,994	$Q^2_{LOO}-Q^2_{ASYM}$ 0,000	DQ -0,015
4.Redundância nas variáveis $R^P > t^P$	$R^P$ 0,111	$t^P$ 0,010		
5.Sobre ajuste (overfitting) $R^N > t^N$	$R^N$ 0,000	$t^N$ -0,323		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

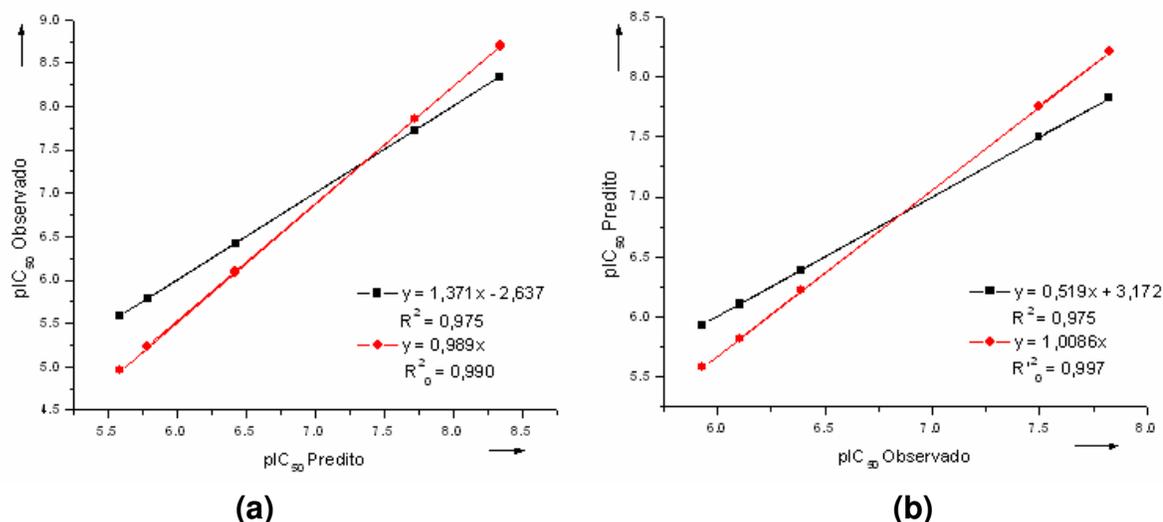
**Tabela 46.** Validação do modelo de QSAR-MLR (n = 13) pelos parâmetros RQK

Critérios de aprovação (*)				
1.Colinearidade entre variáveis $K_{xy} - K_x > DK$	$K_{xy}$ 0,411	$K_x$ 0,219	$K_{xy}-K_x$ 0,192	DK 0,000
2.Capacidade de predição $Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,971	> 0,5		
3.Capacidade de predição do modelo $Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$ 0,971	$Q^2_{ASYM}$ 0,993	$Q^2_{LOO}-Q^2_{ASYM}$ -0,022	DQ -0,025
4.Redundância nas variáveis $R^P > t^P$	$R^P$ 0,185	$t^P$ 0,0		
5.Sobre ajuste (overfitting) $R^N > t^N$	$R^N$ -0,221	$t^N$ -0,233		

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

## QSAR-SVM

Analogamente ao procedimento do modelo de QSAR-MLR, para a validação externa do modelo de QSAR-SVM foi usado o conjunto de teste n = 5, aplicando-se o critério de Golbraikh-Tropsha<sup>64</sup> cujo resultado está representado nos gráficos da Figura 81 assim como os parâmetros associados ao critério.



**Figura 81.** Representação gráfica dos parâmetros de validação do modelo de QSAR-SVM ( $n = 13$ ) das dibenzoilidrazinas pelas equações (a)  $y = ax+b$ ,  $y = kx$ ; (b)  $y = a'x+b'$  e  $y = k'x$ .

Estes gráficos ilustram as diferenças entre os valores dos coeficientes de inclinação ( $k, k'$ ) e os coeficientes das retas ( $a, a'$ ) e ( $b, b'$ ) comentados em Material e Métodos (Seção 4). A rigor, se observado apenas por este aspecto, o modelo de QSAR-SVM está reprovado se comparado aos resultados obtidos para o modelo de QSAR-MLR (v. Figura 80). Contudo, do critério de validação participam também os intervalos de referência

$$0,85 \leq k \leq 1,15 \text{ ou } 0,85 \leq k' \leq 1,15, \text{ e } [(R^2 - R_0^2) / R^2] < 0,3 \text{ ou } [(R^2 - R_0'^2) / R^2] < 0,3.$$

No caso:  $0,85 \leq k=0,989 \leq 1,15$  ou  $0,85 \leq k'=1,008 \leq 1,15$ , e  $[(R^2 - R_0^2) / R^2] = -0,015 < 0,3$  ou  $[(R^2 - R_0'^2) / R^2] = -0,023 < 0,3$ . Assim, o modelo QSAR-SVM será considerado validado.

O modelo de QSAR-SVM será útil para auxiliar parcimoniosamente na avaliação das predições que serão feitas com os compostos propostos, como será visto a seguir.

### 5.5.3.4 – Proposta de Novos Compostos.

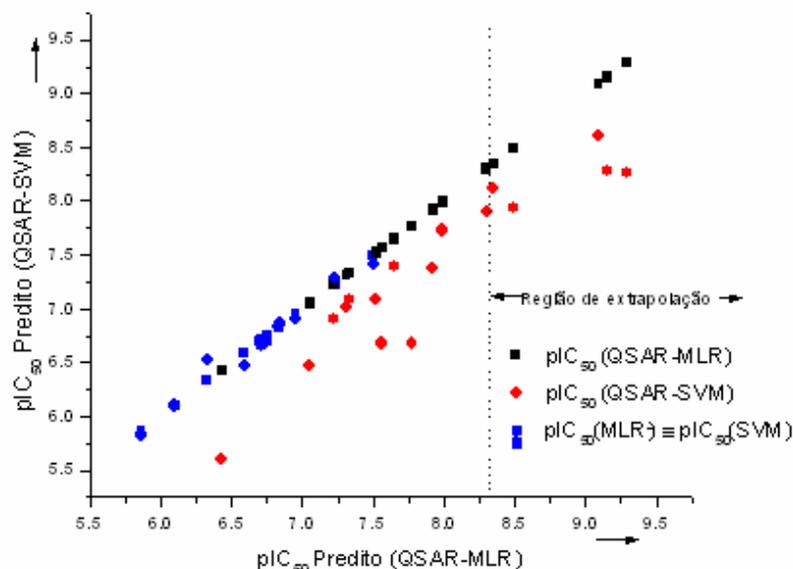
Testou-se o método de similaridade molecular ShaEP (parte teórica descrita no capítulo 4) para avaliar seu desempenho na construção de um modelo de QSAR utilizando confôrmeros por ele selecionados, obtendo-se o seguinte resultado, onde o composto D18 foi excluído como *outlier*:

$$[\text{pIC}_{50}]_{\text{ShaEP}} = + 0,592 (\pm 0,190) \text{ LogP} - 1,056 (\pm 0,405) \sigma + 2,998 (\pm 2,370) \text{ Mor25v} + 3,776 (\pm 1,612)$$

$$n = 17 \quad R^2 = 0,931 \quad s = 0,216 \quad F = 59 \quad p < 0,0001 \quad Q2 = 0,872 \quad \text{SPress} = 0,296$$

$$\text{SDEP} = 0,267 \quad \text{MSE} = 0,036 \quad (5.5-18)$$

Os valores MSE das equações (5.5-18 e 5.5-1) estão muito próximos e tornam as equações praticamente equivalentes. Neste caso, o método ShaEP mostrou-se adequado como seletor de confôrmeros. Em base a este resultado, foram avaliados pela equação (5.5-16), a predição da bioafinidade de 30 compostos propostos *in silico*, cujos confôrmeros foram selecionados pelo índice global ShaEP. A Tabela 47 (v. em Anexos, pg. 237) contém, além do confôrmero selecionado, o intervalo dos resultados da predição. Tomando por referência o valor experimental médio  $[\text{pIC}_{50}]_{\text{exp}} = 7,0$ , tem-se dezessete DBHs selecionadas pelo método ShaEP como propostas para novos ensaios, porém, com a ressalva, que estão contidas cinco extrapolações com valores acima do maior valor experimental ( $\text{pIC}_{50} = 8,27$ ). A causa destas extrapolações está no confôrmero selecionado pelo método, cujos valores calculados para os descritores estão situados dentro da faixa de valores favoráveis, aspecto este comentado na análise de cada um deles na sub-seção 5.5.3.1: Log P com valor alto,  $\sigma$  e Mor25v com valores baixos. Para valores demasiadamente extrapolados, sugere-se optar pelo lado conservador. Neste sentido, se utilizarmos o modelo de QSAR-SVM aplicado ao conjunto dos compostos propostos obtém-se os resultados contidos na Tabela 48 (v. em Anexos, pg. 238). Pode-se, por exemplo, optar apenas pelos compostos em que as predições por ambos os modelos de QSAR tenham sido próximas, alternativa representada graficamente na Figura 82 (pontos em azul), que indica os compostos da referida tabela.

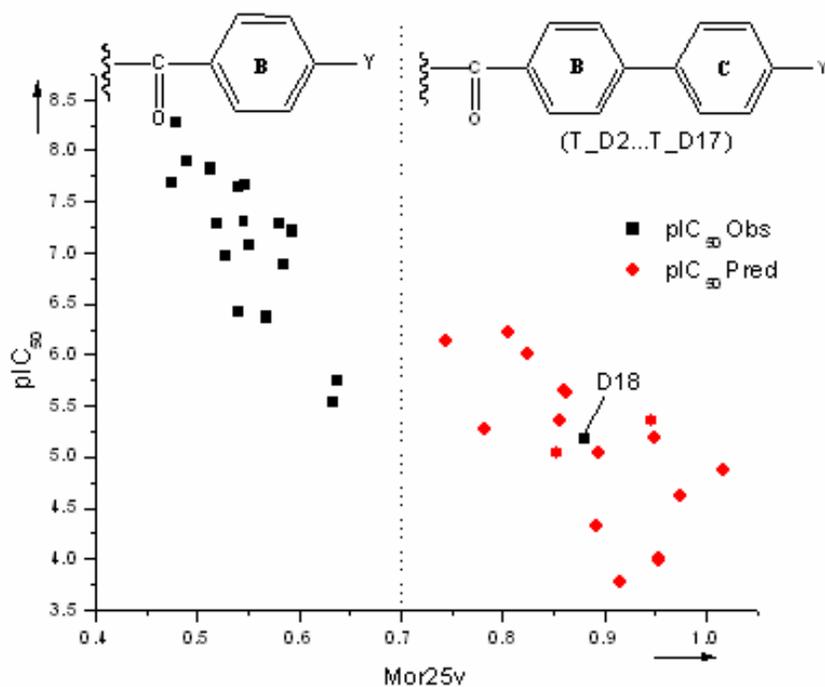


**Figura 82.** Predição da bioafinidade  $pIC_{50}$  de compostos propostos ( $n = 30$ ) e compostos em que os modelos de QSAR-MLR e QSAR-SVM produzem valores próximos (critério de diferença máxima  $Y_{MLR} - Y_{SVM} = \pm 0,10$ ).

Ao se optar apenas pelos compostos cujo valor  $pIC_{50}$  seja igual ou menor do que o maior valor experimental ( $pIC_{50} = 8,27$ ) contido na construção dos modelos de QSAR, ainda restariam onze compostos que poderiam ser de interesse para novos bioensaios.

Como mencionado anteriormente, não se encontrou uma explicação para a baixa bioafinidade de D18 e uma tentativa no presente estudo é feita para o entendimento da causa. A hipótese que foi formulada por Ogura *et al.*<sup>26</sup>, é que a aromaticidade e rigidez do grupo fenil podem ser os agentes responsáveis pela baixa bioafinidade. Partindo-se da hipótese citada, fenilas substituídas devem ter, pressupostamente, comportamento similar, mantendo baixa bioafinidade. Neste sentido foram gerados os compostos *in silico* T\_D2 até T\_D17 que contêm substituintes Ph-Y, onde Y percorre a série de substituintes na posição *para* utilizados nas dibenzoilidrazinas do modelo de QSAR. Após otimização da geometria, como citado no capítulo 4, e obtendo-se as predições pela Eq. (5.5-16), foi utilizado o descritor Mor25v, que por ser sensível às variações conformacionais, é mais adequado para descrever os compostos em função da bioafinidade.

Os resultados obtidos constam da Tabela 49 (v. em Anexos, pg. 239), ilustrados na Figura 83. Pode-se evidenciar que todos os compostos Ph-Y (série T\_D) apresentam na predição, baixa bioafinidade, como D18. Estes compostos acabam por formar uma classe própria, o que explica o comportamento de *outlier* de D18.



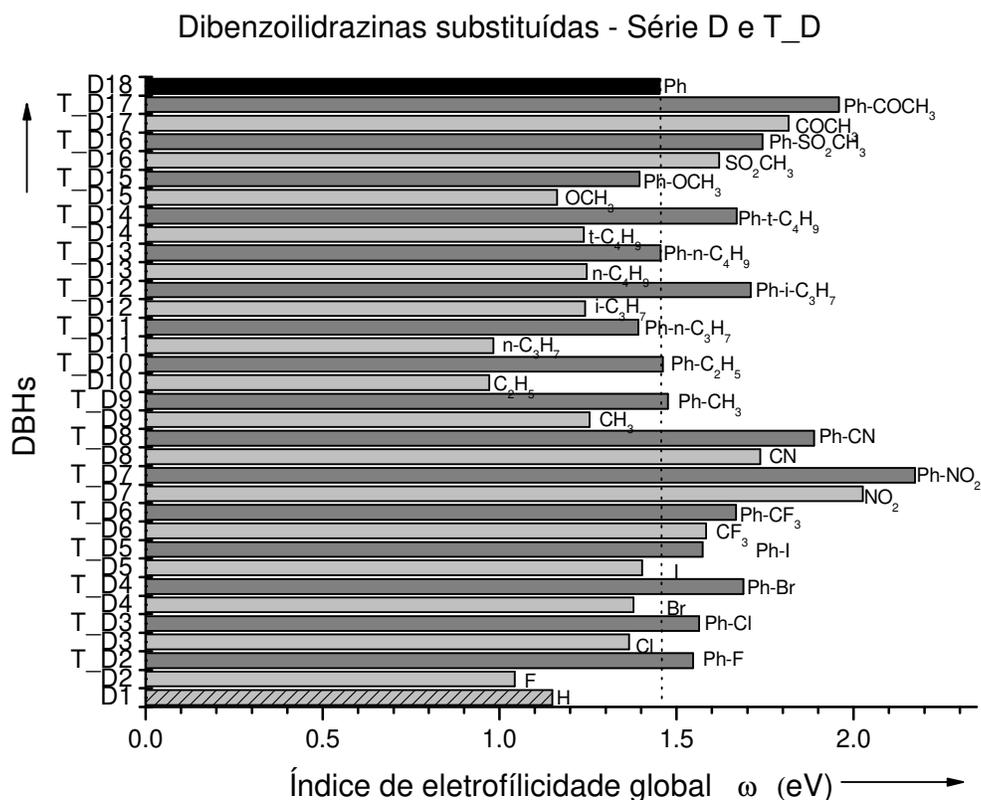
**Figura 83.** Bioafinidades comparativas das dibenzoilidrazinas série D (substituídas no anel B) e série T\_D (inclusão de Ph-Y substituído, anel C). Notar que D18 está contido na série T\_D definido pela baixa bioafinidade.

Faz-se necessário observar que, para todo substituinte no anel B, seu igual no anel C tem bioafinidade mais baixa, segundo a predição.

### 5.5.3.5 – Estudo de Reatividade

Em função dos resultados obtidos na análise dos compostos comparando a série D e T\_D, foi feita uma análise empregando o índice global eletrofílico  $\omega$  (v.

detalhes em Métodos, sub-seção 5.5.2), comparando a série D com a série T\_D (valores de  $\omega$  tirados da Tabela 50 e Tabela 51), ilustrado no gráfico da Figura 84 a seguir.



**Figura 84.** Eletrofilicidade ( $\omega$ ) comparativa das dibenzoilidrazinas da série D e série T\_D.

Em termos de eletrofilicidade global, os compostos da série T\_D têm valores maiores do que a série D de dibenzoilidrazinas, e dentro desta, verifica-se que o substituinte fenil (D18) não é o causador desta diferença, pois várias dibenzoilidrazinas (barras cinzas claras) têm valor maior ou menor, dependendo do substituinte. Assim empregaram-se os valores das Tabelas 50 e 51 para avaliar as contribuições nucleofílicas e eletrofílicas de grupo, gerando os resultados das Tabelas 52 e 53.

**Tabela 50.** Valores usados para análise de contribuições eletrofílica e nucleofílica de grupo nas dibenzoilidrazinas contra *S. frugiperda*.

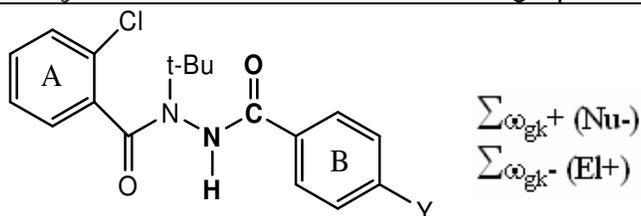
MolID	E(0) <sup>a</sup> (eV)	E(N+1) <sup>b</sup> (eV)	E(N-1) <sup>c</sup> (eV)	IP <sup>d</sup> (eV)	EA <sup>e</sup> (eV)	S <sup>f</sup> (eV)	ω <sup>g</sup> (eV)	η <sup>h</sup> (eV)
D1	-281,669	-273,244	-281,915	-0,246	-8,425	0,122	1,149	4,090
D2	-282,429	-274,198	-282,467	-0,039	-8,231	0,122	1,043	4,096
D3	-280,428	-272,270	-281,233	-0,805	-8,158	0,136	1,366	3,677
D4	-279,752	-271,642	-280,593	-0,841	-8,110	0,138	1,378	3,634
D5	-279,082	-271,092	-280,002	-0,920	-7,990	0,141	1,403	3,535
D6	-301,761	-293,353	-302,928	-1,167	-8,408	0,138	1,583	3,620
D7	-298,600	-289,910	-300,445	-1,845	-8,690	0,146	2,026	3,423
D8	-295,508	-287,092	-296,940	-1,432	-8,416	0,143	1,736	3,492
D9	-299,233	-291,205	-299,836	-0,603	-8,028	0,135	1,254	3,713
D10	-316,638	-308,599	-316,547	0,091	-8,039	0,123	0,971	4,065
D11	-334,064	-326,059	-334,013	0,050	-8,005	0,124	0,982	4,028
D12	-334,085	-326,073	-334,666	-0,581	-8,012	0,135	1,242	3,715
D13	-351,461	-343,485	-352,060	-0,599	-7,976	0,136	1,246	3,689
D14	-351,544	-343,552	-352,121	-0,577	-7,992	0,135	1,238	3,707
D15	-306,142	-298,346	-306,604	-0,463	-7,796	0,136	1,163	3,667
D16	-317,228	-308,988	-318,496	-1,268	-8,239	0,143	1,621	3,486
D17	-316,012	-306,602	-317,400	-1,389	-9,410	0,125	1,817	4,011
D18	-355,190	-347,446	-356,256	-1,067	-7,744	0,150	1,453	3,338

(a) Molécula neutra; (b) ânion; (c) cátion; (d) potencial de ionização:  $IP=E(N-1)-E(0)$ ; (e) afinidade eletrônica  $EA=E(0)-E(N+1)$ ; (f) maciez:  $S=1/(IP-EA)$ ; (g) filicidade global; (h) dureza:  $\eta=(IP-EA)/2$ .

**Tabela 51.** Valores usados para análise comparativa de contribuições eletrofílica e nucleofílica de grupo com as dibenzoilidrazinas contra *S. frugiperda* e análise do composto D18.

MolID	E(0) <sup>a</sup> (eV)	E(N+1) <sup>b</sup> (eV)	E(N-1) <sup>c</sup> (eV)	IP <sup>d</sup> (eV)	EA <sup>e</sup> (eV)	S <sup>f</sup> (eV)	ω <sup>g</sup> (eV)	η <sup>h</sup> (eV)
T_D3	-353,785	-345,792	-355,000	-1,215	-7,993	0,148	1,564	3,389
T_D4	-353,113	-344,199	-354,370	-1,257	-8,914	0,131	1,689	3,828
T_D5	-352,449	-344,722	-353,730	-1,281	-7,727	0,155	1,573	3,223
T_D6	-375,147	-367,148	-376,538	-1,391	-7,999	0,151	1,668	3,304
T_D7	-372,461	-363,694	-374,506	-2,045	-8,767	0,149	2,174	3,361
T_D8	-368,921	-360,419	-370,584	-1,663	-8,502	0,146	1,889	3,419
T_D9	-372,589	-364,536	-373,631	-1,043	-8,052	0,143	1,475	3,505
T_D10	-389,982	-382,009	-391,014	-1,033	-7,973	0,144	1,461	3,470
T_D11	-407,357	-399,801	-408,349	-0,992	-7,556	0,152	1,392	3,282
T_D12	-407,417	-397,490	-408,492	-1,075	-9,927	0,113	1,709	4,426
T_D13	-424,823	-416,945	-425,866	-1,043	-7,877	0,146	1,455	3,417
T_D14	-424,863	-414,996	-425,873	-1,010	-9,866	0,113	1,670	4,428
T_D15	-379,478	-371,783	-380,449	-0,971	-7,696	0,149	1,396	3,362
T_D16	-390,774	-382,782	-392,285	-1,511	-7,992	0,154	1,742	3,240
T_D17	-389,370	-379,722	-390,960	-1,589	-9,648	0,124	1,959	4,029

(a) Molécula neutra; (b) ânion; (c) cátion; (d) potencial de ionização:  $IP=E(N-1)-E(0)$ ; (e) afinidade eletrônica  $EA=E(0)-E(N+1)$ ; (f) maciez:  $S=1/(IP-EA)$ ; (g) filicidade global; (h) dureza:  $\eta=(IP-EA)/2$ .

**Tabela 52.** Contribuição nucleofílica e eletrofílica de grupo

MolID	Substituinte Y	pIC <sub>50</sub> (M) Observado	Anel B+Y <sup>a,b,c</sup> (eV)		Grupo CONH <sup>d</sup> (eV)	
			$\omega_{gk+}$	$\omega_{gk-}$	$\omega_{gk+}$	$\omega_{gk-}$
D1	H	7,08	0,2691	<b>0,4185</b>	0,1935	0,1232
D2	F	7,29	<b>0,2893</b>	0,1480	0,1048	0,0292
D3	Cl	6,89	0,4475	<b>0,6210</b>	0,1206	0,1815
D4	Br	7,69	0,4865	<b>0,6438</b>	0,1171	0,1840
D5	I	7,30	0,5636	<b>0,6844</b>	0,1097	0,1716
D6	CF <sub>3</sub>	6,96	0,3222	<b>0,7846</b>	0,1706	0,2115
D7	NO <sub>2</sub>	6,42	<b>0,9232</b>	0,9022	0,2795	0,0928
D8	CN	5,74	0,4128	<b>0,9690</b>	0,1700	0,2257
D9	CH <sub>3</sub>	7,28	0,3993	<b>0,5113</b>	0,1296	0,1548
D10	C <sub>2</sub> H <sub>5</sub>	7,89	<b>0,3208</b>	0,0851	0,0930	0,0200
D11	n-C <sub>3</sub> H <sub>7</sub>	7,64	<b>0,3363</b>	0,1253	0,0907	0,0218
D12	i-C <sub>3</sub> H <sub>7</sub>	7,82	0,4246	<b>0,5238</b>	0,1226	0,1565
D13	n-C <sub>4</sub> H <sub>9</sub>	8,27	0,4425	<b>0,5315</b>	0,1250	0,1527
D14	t-C <sub>4</sub> H <sub>9</sub>	7,66	0,4393	<b>0,5223</b>	0,1232	0,1530
D15	OCH <sub>3</sub>	7,21	<b>0,4803</b>	0,4534	0,1010	0,1460
D16	SO <sub>2</sub> CH <sub>3</sub>	5,54	0,4590	<b>0,7874</b>	0,2285	0,1919
D17	COCH <sub>3</sub>	6,37	<b>1,4427</b>	1,1214	0,0860	0,2046
D18	C <sub>6</sub> H <sub>5</sub>	5,17	0,6991	<b>0,9183</b>	0,1061	0,1678

(a) Valor experimental; (b) inclui hidrogênios; (c) em negrito o maior valor de  $\omega_{gk}$ ; (d) contribuição do grupo CONH.

Na Tabela 52, ressaltados em negrito os maiores valores de  $\omega_{gk+}$  (Nu-) e  $\omega_{gk-}$  (El+). O índice nucleofílico(Nu-)/eletrofílico(EI+) de grupo corresponde à uma fração da energia produzida pela possível transferência de elétrons entre doador e receptor especificamente na região do anel B incluindo o substituinte Y. Esta transferência eletrônica está também caracterizada pelo  $\sigma$  de Hammett no modelo de QSAR. Apesar deste conceito estar muito associado à reações químicas, no presente estudo deve ser visto também pelo conceito de densidade eletrônica presente em uma certa região do ligante, que favorece ou desfavorece uma provável interação entre moléculas, como no caso de ligante-proteína. A eletrofilicidade global ( $\omega$ ) de uma molécula se conserva, fornecendo um valor que abrange toda a molécula, como

mostrado na Figura 84, sendo que o interesse da presente análise é associar uma fração de  $\omega$ , dada pelos valores  $\omega_{gk+}$  (Nu-) e  $\omega_{gk-}$  (El+) aos valores experimentais da bioafinidade e verificar se existe correlação. Caso ela exista, pode se obter uma tendência de influência destes índices na bioafinidade de dibenzoilidrazinas contra *S. frugiperda*.

Verifica-se que há alguma tendência de correlação do índice eletrofílico mas, nenhuma correlação do índice nucleofílico, o que causa estranheza se olharmos os valores da Tabela 52:

$$\begin{aligned} \text{pIC}_{50} &= - 2,025 (\pm 1,126) \omega_{gk-} + 8,222 (\pm 0,745) \\ n &= 18 \quad R = 0,690 \quad s = 0,640 \quad F = 14 \end{aligned} \quad (5.5-19)$$

Graficamente a reta associada à Eq.(5.5-19) tem apenas cinco pontos contidos próximos à reta, com os demais pontos longe desta (não mostrado). Para poder analisar as contribuições  $\omega_{gk+}$  (Nu-) e  $\omega_{gk-}$  (El+) lançou-se mão de um procedimento utilizado por Verma e Hansch<sup>146</sup>, que empregaram camptotecinas da literatura, ensaiadas experimentalmente com células de carcinoma pulmonar humano H460, para construir um modelo de QSAR. Constataram que seria impossível construí-lo pois 15 dos 33 compostos eram *outliers*. Resolveram o problema, dividindo o conjunto em três subconjuntos, com 18, 10 e 5 compostos, respectivamente. Nenhum dos três subconjuntos continha *outliers* e todos os modelos de QSAR tinham consistência estatística. A interpretação ou justificativa para o procedimento dá-se em função de mecanismos diferentes associados à bioatividade.

Assim, seguindo o mesmo conceito e procedimento, admite-se no presente estudo a hipótese de que as contribuições tanto eletrofílicas como nucleofílicas na região do anel B substituído atuam de forma diferente na bioafinidade, não havendo nenhum motivo para se supor que o parâmetro eletrofílico (ou nucleofílico) deva contribuir linearmente para todos os 18 compostos simultaneamente, ou seja, que se possa traduzir as contribuições em uma única equação.

Iniciou-se a análise apenas com os compostos da coluna  $\omega_{gk}^+$  em negrito (D2, D7, D10, D11, D15, D17) e obteve-se a seguinte equação, mostrando correlação de  $\omega_{gk}^+$  (**Nu-**), no caso, satisfatória:

$$\begin{aligned} \text{pIC}_{50} &= - 1,347 (\pm 0,574) \omega_{gk}^+ + 7,773 (\pm 0,357) \\ n &= 6 \quad R = 0,956 \quad s = 0,205 \quad F = 42 \end{aligned} \quad (5.5-20)$$

Analogamente, apenas a seleção dos compostos em negrito da coluna  $\omega_{gk}^-$  não gera resultado tão bom, contudo, mostrando certa coerência na tendência eletrofílica:

$$\begin{aligned} \text{pIC}_{50} &= - 4,473 (\pm 2,390) \omega_{gk}^- + 9,900 (\pm 1,626) \\ n &= 12 \quad R = 0,797 \quad s = 0,617 \quad F = 17 \end{aligned} \quad (5.5-21)$$

Segue-se uma análise de correlação  $\text{pIC}_{50}/\omega_{gk}^+$  (**Nu-**) e  $\text{pIC}_{50}/\omega_{gk}^-$  (**EI+**) dividindo os 18 compostos em grupos cujo resultados seguem:

**(a)** Compostos D4, D5, D7, D10, D12, D14 mostraram boa correlação como (**Nu-**):

$$\begin{aligned} \text{pIC}_{50} &= - 2,585 (\pm 0,596) \omega_{gk}^+ + 8,824 (\pm 0,334) \\ n &= 6 \quad R = 0,986 \quad s = 0,101 \quad F = 145 \end{aligned} \quad (5.5-22)$$

**(b)** Compostos D1, D3, D6, D17 mostraram boa correlação como (**Nu-**):

$$\begin{aligned} \text{pIC}_{50} &= - 0,561 (\pm 0,235) \omega_{gk}^+ + 7,173 (\pm 0,184) \\ n &= 4 \quad R = 0,991 \quad s = 0,052 \quad F = 106 \end{aligned} \quad (5.5-23)$$

**(c)** Compostos D2, D11, D13 mostraram boa correlação como (**Nu-**):

$$\begin{aligned} \text{pIC}_{50} &= + 6,318 (\pm 4,502) \omega_{gk}^+ + 5,484 (\pm 1,628) \\ n &= 3 \quad R = 0,998 \quad s = 0,039 \quad F = 318 \end{aligned} \quad (5.5-24)$$

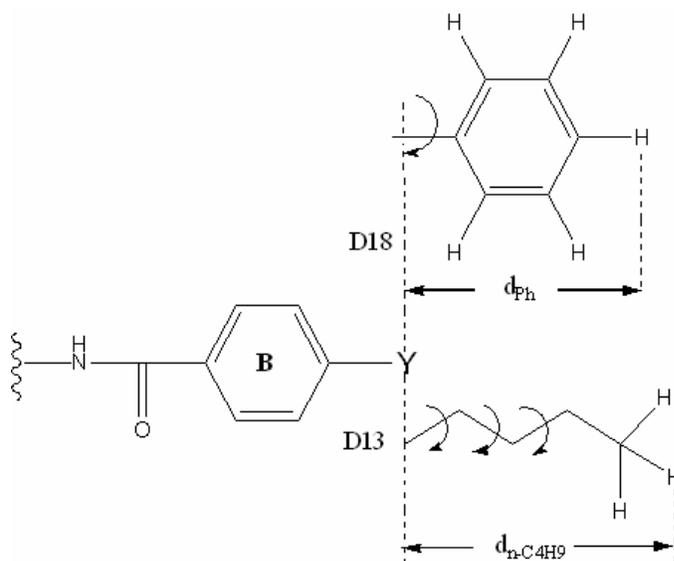
**(d)** Compostos D9, D15, D8, D16, D18 mostraram correlação como (**EI+**):

$$\begin{aligned} \text{pIC}_{50} &= - 3,894 (\pm 2,916) \omega_{gk}^- + 9,023 (\pm 2,209) \\ n &= 5 \quad R = 0,926 \quad s = 0,430 \quad F = 18 \end{aligned} \quad (5.5-25)$$

Neste último grupo estão os três compostos com energias maiores e valores de  $pIC_{50}$  menores, na ordem: D18 (0,918 eV), D16 (0,787 eV) e D8 (0,969 eV) todos com uma acentuada contribuição eletrofílica na região em estudo. D9 (0,511 eV) e D15 (0,453 eV) têm energias menores com respectivos valores  $pIC_{50}$  maiores.

O que se pôde observar, quando os grupos de compostos foram montados para gerar as equações (5.5-22 a 5.5-25), como tendência, a valores baixos de  $\omega_{gk^+}$  e  $\omega_{gk^-}$ , correspondem valores altos de  $pIC_{50}$  e vice-versa. Nada se consegue concluir em relação ao composto D18 utilizando apenas os seus valores eletrofílicos ou nucleofílicos.

O anel fenílico de D18 se estende por aproximadamente  $d_{Ph} = 3,90 \text{ \AA}$  que comparado ao composto de maior atividade (D13,  $Y=n-C_4H_9$ ) tem um alcance de  $4,70 \text{ \AA}$ , porém, três ângulos de torção, um número maior de graus de liberdade que favorecem uma acomodação da molécula na cavidade, observando que o sistema ligante-receptor não é estático, sofrendo variações entrópicas e entálpicas (Figura 85).



**Figura 85.** Comparativo de distância e posicionamento do composto D18 de pior bioafinidade ( $d_{Ph} = 3,90 \text{ \AA}$ ), substituinte “rígido”, com o composto D13 com a maior bioafinidade ( $d_{n-C_4H_9} = 4,70 \text{ \AA}$ ), substituinte flexível.

Como anteriormente comentado, Ogura *et al.*<sup>26</sup> argumentam que uma repulsão estérica poderia ser um possível agente da baixa bioafinidade de D18 conjugado com a aromaticidade e rigidez do grupo fenil. Como foi predita uma baixa bioafinidade em toda a série T\_D comparada às demais dibenzoilidrazinas, é possível que esta hipótese esteja correta. Outro detalhe já citado na ref.[26], é o efeito elétron-doador do substituinte favorecendo uma interação do grupo CONH com o receptor. Este efeito é representado no modelo de QSAR pela constante eletrônica  $\sigma$  cuja correlação também foi avaliada em base ao índice nucleofílico correspondente ao grupo CONH da Tabela 52:

**(a)** Compostos D2, D3, D4, D5, D6, D10, D11:

$$\sigma = + 8,419 (\pm 3,221) \omega_{gk}^+ - 0,833 (\pm 0,380)$$

$$n = 7 \quad R = 0,949 \quad s = 0,082 \quad F = 45 \quad (5.5-26)$$

**(b)** Compostos D1, D9, D12, D13, D14, D15:

$$\sigma = + 2,726 (\pm 1,006) \omega_{gk}^+ - 0,519 (\pm 0,136)$$

$$n = 6 \quad R = 0,966 \quad s = 0,026 \quad F = 56 \quad (5.5-27)$$

**(c)** Compostos D7, D8, D16, D17:

$$\sigma = + 1,433 (\pm 0,663) \omega_{gk}^+ + 0,391 (\pm 0,135)$$

$$n = 4 \quad R = 0,989 \quad s = 0,022 \quad F = 87 \quad (5.5-28)$$

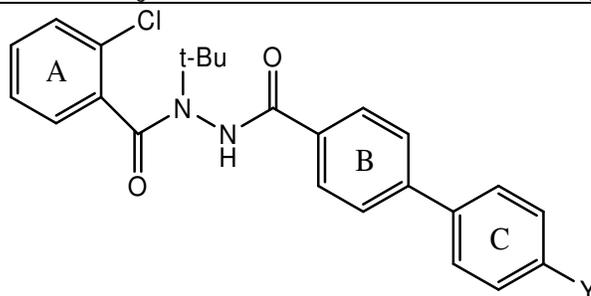
Exceto pelo composto D18 que tende a preservar sua característica de *outlier* também na análise do grupo CONH, as equações 5.5-26 a 5.5-28, mostram uma correlação coerente, entre  $\sigma$  de Hammett e o índice nucleofílico  $\omega_{gk}^+$ , ambos parâmetros caracterizando transferência eletrônica, e cada uma das equações caracteriza a peculiaridade relacionada ao grupo de substituintes. Estes dados indicam que o  $\sigma$  de Hammett é um descritor adequado no modelo de QSAR mas não explica o comportamento de D18.

A seguir foi feita uma análise dos índices  $\omega_{gk+}$  (Nu-) e  $\omega_{gk-}$  (EI+) para a série de compostos T\_D, construídos para análise em conjunto com o composto D18, uma

vez que ficou caracterizado que este composto supostamente pertence a uma classe própria como ilustrado na Figura 83. Por outro lado, os valores de  $pIC_{50}$  não são valores experimentais e foram preditos pelo modelo de QSAR da Eq. (5.5-16).

A expectativa desta análise é que os resultados a serem obtidos auxiliem na melhor compreensão da baixa bioafinidade destes compostos.

**Tabela 53** Contribuição nucleofílica e eletrofílica de grupo



MolID	Substituinte Y	$pIC_{50}(M)$ Predito	Anel B+C <sup>a,b</sup> (eV)		Grupo CONH <sup>c</sup> (eV)	
			$\omega_{gk+}$	$\omega_{gk-}$	$\omega_{gk+}$	$\omega_{gk-}$
T_D2	F	5,38	0,2664	<b>0,8422</b>	0,1586	0,1970
T_D3	Cl	5,05	<b>0,9338</b>	0,9036	0,0665	0,1908
T_D4	Br	6,14	0,2914	<b>0,9885</b>	0,3008	0,2105
T_D5	I	4,62	0,9224	<b>0,9392</b>	0,1416	0,1934
T_D6	CF <sub>3</sub>	5,05	0,7292	<b>1,0879</b>	0,1905	0,2005
T_D7	NO <sub>2</sub>	3,82	<b>1,6977</b>	1,6193	0,1876	0,2130
T_D8	CN	4,36	0,2791	<b>1,2830</b>	0,2877	0,2025
T_D9	CH <sub>3</sub>	5,66	0,4121	<b>0,8013</b>	0,2517	0,1851
T_D10	C <sub>2</sub> H <sub>5</sub>	6,00	<b>1,3366</b>	0,9541	0,0095	0,0780
T_D11	n-C <sub>3</sub> H <sub>7</sub>	6,22	0,7468	<b>0,7879</b>	0,1365	0,1744
T_D12	i-C <sub>3</sub> H <sub>7</sub>	5,20	<b>1,2725</b>	0,9537	0,1970	0,2218
T_D13	n-C <sub>4</sub> H <sub>9</sub>	4,87	<b>1,0507</b>	0,8438	0,0431	0,0847
T_D14	t-C <sub>4</sub> H <sub>9</sub>	5,36	<b>0,1207</b>	0,9000	-0,0927	0,2253
T_D15	OCH <sub>3</sub>	5,64	<b>1,1711</b>	0,7472	0,1085	0,1856
T_D16	SO <sub>2</sub> CH <sub>3</sub>	4,03	0,9255	<b>1,1869</b>	0,0857	0,0617
T_D17	COCH <sub>3</sub>	5,30	<b>1,3700</b>	1,3000	0,1090	0,1940
D18	C <sub>6</sub> H <sub>5</sub>	5,17 <sup>a</sup>	0,6991	<b>0,9183</b>	0,1061	0,1678

(a) Inclui Hidrogênios; (b) em negrito o maior valor de  $\omega_{gk}$ ; (c) contribuição do grupo CONH.

Em base à Tabela 53, obtém-se uma equação com baixa correlação considerando todos os 17 valores, e especificamente para o índice  $\omega_{gk-}$  (EI+) sem ocorrência de *outliers*:

$$pIC_{50} = - 2,053 (\pm 1,210) \omega_{gk^-} + 7,229 (\pm 1,243)$$

$$n = 17 \quad R = 0,683 \quad s = 0,516 \quad F = 13 \quad (5.5-29)$$

Selecionando-se os compostos T\_D2, T\_D3, T\_D6, T\_D7, T\_D8, T\_D9, T\_12, TD14, TD15, D18, obtém-se uma correlação significativamente melhor, sem *outliers*, contendo o composto D18, alvo desta análise e com boa predição de seu valor experimental, sugerindo um modelo de QSAR alternativo para esta série de compostos:

$$pIC_{50} = - 2,117 (\pm 0,379) \omega_{gk^-} + 7,198 (\pm 0,393)$$

$$n = 10 \quad R^2 = 0,954 \quad s = 0,130 \quad F = 166 \quad p < 0,0001 \quad Q^2 = 0,930 \quad SPress = 0,161$$

$$SDEP = 0,152 \quad (5.5-30)$$

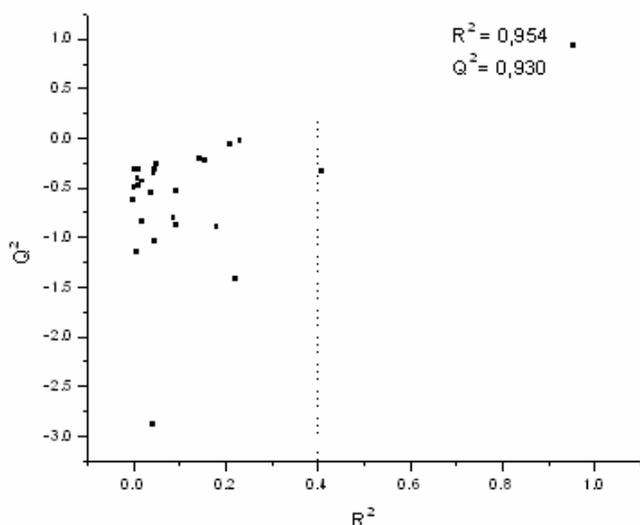
A Tabela 54 mostra os valores preditos  $pIC_{50}$  calculados pelo modelo de QSAR Eq. (5.5-30), empregando os valores preditos de  $pIC_{50}$  pelo modelo de QSAR Eq.(5.5-16), e utilizados como se fossem os experimentais.

**Tabela 54.** Valores que geram o modelo de QSAR Eq.(5.5-30)

MolID	Substituinte Y	$pIC_{50}^a$ (M)	$\omega_{gk^-}$ (eV)	$pIC_{50}^b$ (M)	res <sup>d</sup>
T_D2	F	5,38	0,842	5,42	-0,04
T_D3	Cl	5,05	0,904	5,29	-0,24
T_D6	CF <sub>3</sub>	5,05	1,088	4,89	0,16
T_D7	NO <sub>2</sub>	3,82	1,619	3,77	0,05
T_D8	CN	4,36	1,283	4,48	-0,12
T_D9	CH <sub>3</sub>	5,66	0,801	5,50	0,16
T_D12	i-C <sub>3</sub> H <sub>7</sub>	5,20	0,954	5,18	0,02
T_D14	t-C <sub>4</sub> H <sub>9</sub>	5,36	0,900	5,29	0,07
T_D15	OCH <sub>3</sub>	5,64	0,747	5,62	0,02
D18	C <sub>6</sub> H <sub>5</sub>	5,17 <sup>c</sup>	0,918	5,25	-0,08

(a) Valor predito pela Eq. (5.5-16); (b) valor predito pela Eq.(5.5-30);  
(c) valor experimental; (d) res =  $pIC_{50}^a - pIC_{50}^b$

Pelos resultados, foram feitas duas avaliações básicas: o teste de randomização (*Y-randomization*)<sup>61</sup> com 25 repetições (Figura 86), e a validação da equação (5.5-30) pelos parâmetros RQK<sup>53</sup> constantes da Tabela 55.



**Figura 86.** Y-randomization dos valores  $pIC_{50}$  em relação à variável  $\omega_{gk}$ .

**Tabela 55.** Validação do modelo de QSAR-MLR (n = 10) pelos parâmetros RQK

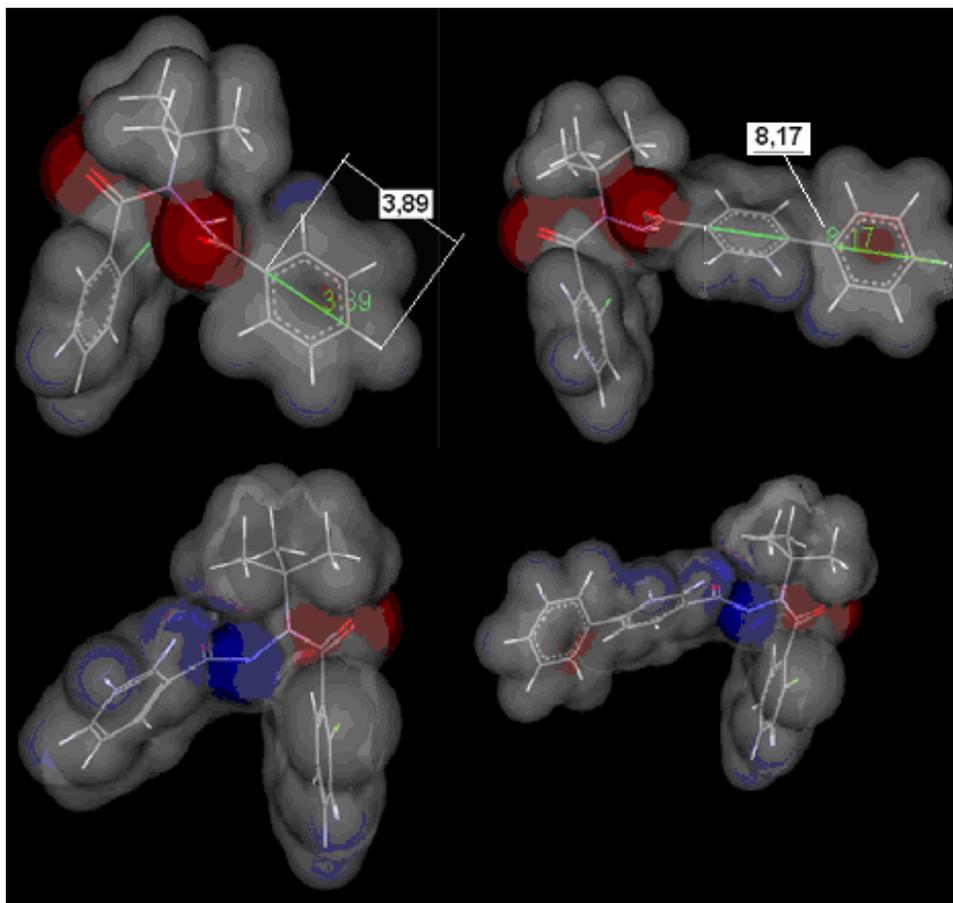
Critérios de aprovação (*)				
1. Colinearidade entre variáveis $K_{xy} - K_x > DK$	$K_{xy}$ 0,977	$K_x$ 0,00	$K_{xy} - K_x$ 0,977	DK 0,000
2. Capacidade de predição $Q^2_{LOO} > Q^0$ (Referência $Q^0 > 0,5$ )	0,930	> 0,5		
3. Capacidade de predição do modelo $Q^2_{LOO} - Q^2_{ASYM} > DQ$	$Q^2_{LOO}$ 0,930	$Q^2_{ASYM}$ 0,943	$Q^2_{LOO} - Q^2_{ASYM}$ -0,013	DQ -0,025
4. Redundância nas variáveis $R^P > t^P$	$R^P$	$t^P$	Não aplicável para uma única variável	
5. Sobre ajuste (overfitting) $R^N > t^N$	$R^N$	$t^N$	Não aplicável para uma única variável	

(\*) Caso contrário rejeitar modelo. Critérios devem ser atendidos simultaneamente

Em ambos os testes, ocorre a validação interna do modelo. Três aspectos importantes podem ser destacados destes resultados: **(a)** o composto D18 se confirma pertencer a uma classe própria, isto é, a um domínio de aplicabilidade específico, figurando como *outlier* na série D das dibenzoilidrazinas, não tendo este comportamento quando analisado na série congênere T\_D; **(b)** considerando o valor residual obtido do composto D18 (0,08) e valores residuais quase todos relativamente baixos, pode-se considerar isto como um indicador de que a Eq.(5.5-16) tem boa consistência na predição externa, já que os resultados foram reproduzidos com valores próximos, através de metodologias independentes; **(c)** o

estudo utilizando a filicidade localizada, ou de grupo, consegue identificar D18 com valor eletrofílico  $\omega_{gk^-}$  (El+) prevalecendo sobre o nucleofílico  $\omega_{gk^+}$  (Nu-) na região estudada, mas não se consegue identificar a causa precisa da baixa bioafinidade. Todos os compostos da série T\_D possuem um índice favorável à hidrofobicidade, e o baixo valor  $\sigma$  de Hammett favorece a bioafinidade nos compostos T\_D9 até T\_D15. Mor25v é calculado pelo volume de van der Waals da molécula, e tem valores altos para toda a série T\_D. Mostrou-se que valores baixos de Mor25v são mais favoráveis à bioafinidade (v. Figura 73). Associado ao volume de van der Waals temos o potencial eletrostático de superfície (ESP – *Eletrostatic Potential Surface*), e a ele, a distribuição da densidade eletrônica, que em caso da ‘rigidez’ molecular suposta por Ogura *et al.*<sup>26</sup>, pode estar localizada desfavoravelmente em relação ao receptor na região de interação. Como não foi possível obter uma estrutura homóloga de *S. frugiperda*, não se tem a oportunidade de fazer uma avaliação teórica da qualidade do alojamento destas estruturas na cavidade receptora através de *docking* e identificar os receptores (aminoácidos) e os átomos de provável interação.

A Figura 87 mostra o comparativo dos compostos D1 e D18 representados pelo volume de van der Waals associado ao potencial eletrostático de superfície, com algumas distâncias em Å como referência. Regiões em vermelho escuro se referem a regiões eletronegativas (em especial relacionadas à ligação =O) e regiões em azul escuro são regiões eletropositivas, relacionadas aos nitrogênios.



**Figura 87.** Volume de van der Waals e o potencial eletrostático de superfície dos compostos D1 e D18. Distâncias em Å para informação dimensional.

#### 5.5.4 – Conclusão

Por este estudo, pôde-se verificar que o método de similaridade molecular ShaEP produziu resultados favoráveis na seleção de confôrmeros gerados pelos campos de força MMFF94 e Dreiding modificado, tornando possível construir-se um modelo de QSAR equivalente ao da literatura.<sup>26</sup> Aproveitando dois dos descritores da literatura (LogP e  $\sigma$ ) e introduzindo um terceiro (Mor25v) sensível às variações conformacionais, construiu-se um modelo de QSAR utilizando-se confôrmeros gerados pelos citados campos de força, obtendo maior proximidade dos valores preditos aos valores experimentais. Pela avaliação de cada um dos descritores,

verificou-se que compostos com valores de LogP mais altos, e  $\sigma$  de Hammett e Mor25v baixos geram a melhor bioafinidade, característica útil para a proposição *in silico* de novos compostos.

Construindo compostos *in silico* com um fenil substituído, mostrou-se que o composto D18 classificado como *outlier* pertence a uma classe específica, ou, utilizando a terminologia de QSAR, domínio de aplicabilidade distinto da série em que foi testado.

Empregando o conceito de densidade eletrônica da Teoria Funcional de Densidade, funções de Fukui e o conceito de filicidade de grupo eletrofílico (EI+) e nucleofílico (Nu-), foram calculados valores das dibenzoilidrazinas tanto da série testada experimentalmente quanto da série construída *in silico*, obtendo-se através dos índices  $\omega_{gk+}$  (Nu-) e  $\omega_{gk-}$  (EI+), a prevalência de cada índice na região substituída em cada uma das séries. O procedimento poderá ser útil em novos estudos *in silico* envolvendo pesquisa de prováveis interações na cavidade da proteína, que atualmente é desconhecida, mas possivelmente poderá ser construída por homologia se os aminoácidos do receptor da ecdisona (EcR) forem conhecidos.

O estudo complementa e fortalece a conclusão obtida por Ogura *et al.*<sup>26</sup> na hipótese de que o composto D18 tem baixa bioafinidade causada pela sua ‘rigidez’ na região substituída em função de alguns aspectos mencionados. Para este composto e seus análogos ainda há espaço para pesquisas adicionais para a melhor compreensão da causa.

Outra hipótese formulada na literatura<sup>26</sup> se relaciona ao efeito doador de elétrons identificado pelo  $\sigma$  de Hammett na região CONH, aspecto confirmado pela análise dos grupos nucleofílicos  $\omega_{gk+}$  (Nu-) bem correlacionados a  $\sigma$  como foi demonstrado para subgrupos dos 17 compostos, excluindo-se o composto D18.

A análise de correlação com a bioafinidade dos 16 compostos construídos *in silico* (série T\_D), gerou um modelo de QSAR alternativo construído com 10 compostos e tendo  $\omega_{gk-}$  (EI+) como descritor único. Através dos resultados de predição interna obtidos, foi possível atestar a qualidade de predição externa do modelo de QSAR principal (Eq. 5.5-16).



## 6. CONCLUSÃO GERAL E RECOMENDAÇÕES

---

### QSAR-MLR – QSAR-SVM – Confôrmeros – Similaridade Molecular

Em termos gerais a literatura corrente trata as estruturas dos compostos submetidos à otimização da geometria, como sendo a bioativa. Demonstrou-se nos cinco estudos de caso, que se os compostos possibilitarem a geração de um espaço conformacional significativo com seus integrantes na energia mínima, existem boas possibilidades de pelo menos um confôrmero melhor reproduzir um valor próximo à atividade biológica experimental.

Apesar de se saber que o confôrmero escolhido não representa necessariamente a conformação bioativa assumida na sua interação com o receptor, propõe-se a utilização do método como uma alternativa para obter modelos de QSAR com maior qualidade de predição interna. Outros aspectos associados ao método: reintegração de supostos *outliers* assim classificados por apresentarem grande valor residual; maior chance de se obter uma validação interna e de se obter um modelo de QSAR robusto. O aspecto da escolha de descritores tem um grande peso na aplicação do método. Sugere-se que os descritores selecionados para análise conformacional devam ter, além de alguma correlação aceitável com a atividade, uma tendência de crescimento ou decaimento com a atividade, senão para todos os compostos, pelo menos quando analisados dentro do espaço conformacional do mesmo composto. Isto pode auxiliar na interpretabilidade físico-química do descritor no modelo e na predição do mecanismo de ação.

Conhecer a tendência do descritor pode servir de indicador na proposição de novos compostos, conforme foi investigado no estudo de dibenzoilidrazinas.

Os métodos de similaridade molecular (distância euclidiana, ShaEP e índice de Carbó) foram empregados em confôrmeros de compostos com bioatividade experimental conhecida (pseudo *blind set*). Foi verificada a sua empregabilidade na identificação de confôrmeros que melhor representam a bioatividade. Os resultados foram arbitrados como aceitáveis em 60%, ou seja, os confôrmeros

## 6. CONCLUSÃO GERAL E RECOMENDAÇÕES

---

selecionados tiveram sua atividade predita próxima ao valor observado. No presente estudo foi possível obter resultados favoráveis com acerto em um caso específico de até 80%, mas identificando também falsos positivos. Foram feitas também avaliações, em que nenhum dos três métodos agregou valor ao estudo para orientar a continuidade da pesquisa.

A teoria estatística do aprendizado (SLT – *Statistical Theory of Learning*) tem incorporadas metodologias de aprendizado de máquinas, como Máquinas de Vetores de Suporte (SVM), para resolver problemas com um reduzido número de amostras<sup>36a,b</sup>, mostrando emprego favorável nos estudos de QSAR em que SVM foi utilizado. No estudo de diterpenóides, para exemplificar, os resultados produzidos pelas duas metodologias podem orientar o pesquisador na escolha de quais compostos deva sintetizar e testar, em base à equivalência dos valores preditos, como primeiro critério. A escolha dos dois métodos, MLR e SVM, mostrou também, que podem ocorrer predições com valores muito diferentes para um mesmo composto pelos respectivos modelos de QSAR validados para predição externa.

### ***Docking – Dibenzoilidrazinas contra *Spodoptera exigua****

Este trabalho faz, no melhor de nosso conhecimento, duas contribuições à literatura: **(a)** partindo de um complexo cristalizado ligante-proteína da Lepidoptera *Heliothis virescens*, obtido por Billas *et al.*<sup>96</sup>, foi feito um estudo de *docking* de treze dibenzoilidrazinas halogenadas construídas dentro da cavidade a partir do cristal de benzoilidrazina de referência BYI06830. Obteve-se assim, informações de quais átomos, de quais aminoácidos, possivelmente interagem com os átomos halogênicos de forma não-covalente. **(b)** Construiu-se uma proteína por homologia à de *Heliothis virescens*, partindo do domínio ligante de *S. exigua* SeEcR-LBD, praticando-se *docking* de 24 dibenzoilidrazinas halogenadas testadas experimentalmente.<sup>27</sup> Verificou-se que a ancoragem das dibenzoilidrazinas se dá

## 6. CONCLUSÃO GERAL E RECOMENDAÇÕES

---

com os mesmos aminoácidos (Met380, Met381, Tyr403) comuns aos dois domínios ligantes, respectivamente, HvEcR-LBD e SeEcR-LBD. Analogamente ao caso anterior, foram identificados os aminoácidos e respectivos átomos que provavelmente interagem com os átomos halogênicos das dibenzoilidrazinas. As distâncias e os ângulos de interação foram comparados com biomoléculas halogenadas, da literatura, mostrando boa concordância. Como já mencionado, apenas uma validação experimental (fora do escopo deste trabalho) pode confirmar a estrutura homóloga de *S. exigua*, assim como as interações entre as dibenzoilidrazinas com os domínios ligantes de *Heliothis virescens* e *Spodoptera exigua*, respectivamente.

### **Reatividade – Dibenzoilidrazinas contra *Spodoptera frugiperda*.**

As contribuições maiores deste estudo são: **(a)** empregando compostos *para*-substituídos no anel-C das dibenzoilidrazinas, obteve-se um indicador que sugere a hipótese do composto D18 ter o comportamento de *outlier* por pertencer a uma classe própria de dibenzoilidrazinas. **(b)** A hipótese se verifica válida após um modelo de QSAR alternativo ter sido construído e validado com um descritor eletrofílico de grupo  $\omega_{gk}^-(EI+)$ , utilizando apenas compostos similares ao D18. **(c)** O presente estudo confirma a hipótese levantada por Ogura *et al.*<sup>18</sup> do efeito doador de elétrons associado ao  $\sigma$  de Hammett na região CONH da molécula, o que foi identificado através do índice nucleofílico de grupo  $\omega_{gk}^+(Nu-)$ . Além disto, foi mostrado que há boa correlação entre o  $\sigma$  de Hammett com o índice  $\omega_{gk}^+$  para os compostos em questão.

## 6. CONCLUSÃO GERAL E RECOMENDAÇÕES

---

### Estudo de Proposta de novos Compostos

Foram preditos valores de bioatividade de 113 diterpenóides clerodanos, conforme literatura, extraídos de órgãos de árvores de diversas partes do mundo. Foram feitas algumas simulações com diterpenóides propostos *in silico* com foco em espécies de árvores existentes no Brasil: 13 Casearinas derivadas da Casearina G (*Casearia sylvestris* Sw.); 12 compostos propostos a partir da Casearborina A (*Casearia arborea* (L.C.Rich.) Urban); 21 compostos derivados das Casearvestrinas A, B e C (*Casearia sylvestris* Sw.) e 7 compostos da Caseamembrina A, cujo nome deriva da *Casearia membranacea* Hance (encontrada em Taiwan), também encontrada na casca da raiz da *Casearia sylvestris* Sw. var. *lingua*.

Os modelos de QSAR com dibenzoilidrazinas geraram proposta de dois grupos de novos compostos: 13 compostos contra *Spodoptera exigua*; 30 compostos contra *Spodoptera frugiperda*.

### Oportunidades

A literatura<sup>16</sup> de produtos naturais contém mais de 800 diterpenóides clerodanos identificados nestes últimos 24 anos aos quais se poderia aplicar a metodologia proposta. Aqui um esforço conjunto envolvendo a parte computacional e biológica experimental seria recomendável.

Dibenzoilidrazinas foram pouco estudadas no Brasil, segundo levantamento feito pelo *Web of Science*.<sup>147</sup> Estudos experimentais acompanhados de estudos de QSAR poderiam possibilitar a criação de inseticidas locais, ajustados às condições ecológicas exigidas pela legislação brasileira, e com possibilidades de proteção da aplicação dos compostos, por patente.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzolidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

1. Drivers of Pharmaceutical Industry Investment. Drug Development and Costs, pg 13. <http://medicinesaustralia.com.au/files/2010/01/Drivers-of-Pharmaceutical-Industry-Investment-July-20061.pdf> Acesso: 20-Out-2010.
2. Purdue University – The Pesticide Market Place (2006)  
F. Whitford *et al.* *Purdue Pesticides Program* –<http://www.ces.purdue.edu/new> Acesso em 20-Out-2010.
3. Kubinyi, Hugo. QSAR: Hansch Analysis and Related Approaches. ED. VHC, Weinheim, 1993, vol. 1, 240 p.
4. (a) <http://infochim.u-strasbg.fr/CS3/program/material/Fujita.pdf>, Acesso: Novembro-2010; (b) Mathematical Challenges from Theoretical/Computational Chemistry. Ed. National Academy of Sciences, Washington, 1995, p. 19, 144p.
5. Foye, William O.; Lemke, Thomas L.; Williams, David A. Principles of Medicinal Chemistry. 4<sup>th</sup>. Ed., Williams & Wilkins, Baltimore, 1996, p.7, 995 p.
6. Benfenati, Emilio. Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes. Ed. Elsevier, Amsterdam, 2007, 510 p.
7. Pettit, Georg R.; Pierson, Fiona Hogan; Herald, Cherry L. Anticancer Drugs From Animals, Plants and Microorganisms. Ed. John Wiley & Sons, Inc., New York, 1994, p.68, 670 p.
8. Corrêa, Arlene G. Feromônios e Aplicação no Controle de Pragas. In: Corrêa, Arlene G.; Vieira, Paulo C. (organizadores). 2<sup>a</sup>. Ed. EdUFSCar, São Carlos, 2007, Capítulo 2, 150 p.
9. Robbers, James E.; Speedie, Marilyn K.; Tyler, Varro E. Pharmacognosy and Pharmacobiotechnology. Williams & Wilkins, Baltimore, 1996, (a) p.170; (b) p.88. 337 p
10. Evans, Charles William. Trease and Evans' Pharmacognosy. WB Saunders Co. Ltd., London, p. 417, 612 p.
11. Costa, Aloísio Fernandes. Farmacognosia. 5<sup>a</sup>.ed. Fundação Calouste Gulbenkian, Lisboa, 1994, vol. I, p.801, 1031 p.
12. Breitmeier, Ebehard. Terpene, Aromen, Düfte, Pharmaka, Pheromone. B.G. Teubner Suttgart, 1999. p.62, 227 p.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

13. Barroso, G. M. Sistemática de Angiospermas no Brasil. Livros Técnicos e Científicos S.A. São Paulo, 1978, vol. 1, p. 171.
14. Cramer, Bruno. Caracterização Farmacognóstica Comparativa das espécies *Casearia sylvestris* Swartz e *Casearia decandra* Jacquin Flacourtiaceae. Dissertação de Mestrado. Universidade São Francisco, **2005**.
15. Chen, Tong-Bin; Wiemer, David F. Corymbotins A-I: Highly Oxidized Kovalene Derivatives from *Casearia corymbosa*. *Journal of Natural Products*, 54, 1612-1618, **1991**.
16. Handson, J. R. *Natural Product Reports*. Review Article. Artigos de revisão bibliográfica com o título 'Diterpenoids' abrangendo os anos de 1983-2007.
17. Busch, T.; Kirsching, A. Recent Advances in the Total Synthesis of Pharmaceutical Relevant Diterpenoids. *Nat. Prod. Rep.*, 25, 318-341, **2008**.
18. Cramer, Bruno; Takahata, Yuji. QSAR of Clerodane Diterpenoids with Cytotoxic and Antitumoral Activity Against V-79 Cells. *QSAR & Comb. Sci.*, 28, 1176-1186, **2009**.
19. Dhadialla, Tarlochan S.; Carlson, Glenn R.; Le, Dat P. New Insecticides with Ecdysteroidal and Juvenile Hormone Activity. *Annu. Rev. Entomol.*, 43, 545-569, **1998**.
20. Wing, Keith D.; Slawecki, Richard A; Carlson, Glenn R. RH 5849, a Nonsteroidal Ecdysone Agonist: Effects on Larval Lepidoptera. *Science*, 241, 470-472, **1988**.
21. Mori, Kenji. Chemical Synthesis of Hormones, Pheromones and Other Bio regulators. John Wiley & Sons Ltd., 2010, p.81-82, 298 p.
22. Chan, T.H.; Ali, A.; Britten, J.F.; Thomas, A. W.; Strunz, G. M.; Saloni, A. The crystal structure of 1,2-dibenzoil-1-*tert*-butylhydrazine, a nonsteroidal ecdysone agonist, and its effects on spruce budworm (*Choristoneura fumiferana*). *Canadian Journal of Chemistry*, 68, 1178-1181, **1990**.
23. Fujita, Toshio; Nakagawa Yoshiaki. QSAR and mode of action studies of insecticidal ecdysone agonists. *SAR and QSAR in Environmental Research*, 18, 77-88, **2007**.
24. Oikawa, Nobuhiro; Nakagawa, Yoshiaki; Nishimura, Keiichiro; Ueno, Tamio; Fujita, Toshio. Quantitative Structure-Activity Studies of Insect Growth Regulators. *Pesticide Biochemistry and Physiology*, 48, 135-144, **1994**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

25. Nakagawa, Yoshiaki; Smagghe, Guy; Van Paemel, Marleen; Tirry, Luc; Fujita, Toshio. Quantitative structure-activity studies of insect growth regulators: XVIII. *Pest Management Science*, 57, 858-865, **2001**.
26. Ogura, Takehico; Nakagawa, Yoshiaki; Minakuchi, Chieka; Miyagawa, Hisashi. QSAR for Binding Affinity of Substituted Dybenzoylhydrazines to Intact Sf-9 Cells. *J. Pest. Sci.*, 30, 1-6, **2005**.
27. Nakagawa, Yoshiaki; Smagghe, Guy; Tirry, Luc; Fujita, Toshio. Quantitative structure-activity studies of insect growth regulators: XIX. *Pest Management Science*, 58, 131-138, **2001**.
28. Santos, Ana Cristina Souza dos. Sant'Anna, Carlos Mauricio. 20-hydroxyecdysone receptor ligand-binding domain: 1. A semiempirical study of Dibenzoylhydrazines selectivity. *Journal of Molecular Structure (Theochem)*, 585, 61-68, **2002**.
29. Böhm, H.J.; Klebe, G.; Kubinyi, Hugo. *Wikstoffdesign, der weg zum Arzneimittel*. Spektrum Akademischer Verlag, Heidelberg, 1996, p. 363-380, 599p.
30. Liao, Quan; Yao, Jianhua; Yuan, Shengang. SVM approach for predicting LogP. *Molecular Diversity*, 10, 301-309, **2006**.
31. Yao, X.J.; Panaye, A.; Doucet, J.P.; Zhang, R.S.; Chen, H.F.; Liu, M.C.; Hu, Z.D.; Fan, B.T. Comparative Study of QSAR/QSPR Correlations Using Support Vector Machines, Radial Basis Function Neural Networks, and Multiple Linear Regression. *J. Chem. Inf. Comput. Sci.*, 44, 1257-1266, **2004**.
32. Oloff, Scott, Mailman, Richard B., Tropsha, Alexander. Application of Validated QSAR Models of D<sub>1</sub> Dopaminergic Antagonists for Database Mining. *J. Med. Chem.*, 48, 7322-7332, **2005**.
33. Ferrão, F. Marco; Mello, César; Borin, Alessandra; Maretto, Danilo, A.; Poppi, Ronei J. LS-SVM: Uma nova Ferramenta Quimiométrica para Regressão Multivariada. Comparação de Modelos de Regressão LS-SVM e PLS na Quantificação de Adulterantes em Leite em Pó Empregando NIR. *Quim. Nova*, 30, 852-859, **2007**.
34. Chatterjee, Samprit; Price, Bertram. *Regression Analysis by Example*. John Willey & Sons, New York, 1977, p. 51-73, 228 p
35. Gaudio, Anderson Coser; Zandonate, Eliana. Proposição, Validação e Análise dos Modelos que Correlacionam Estrutura Química e Atividade Biológica. *Quim. Nova*, 24, 568-671, **2001**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

36. (a) Vapnik, V. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1995; (b) Chen, Nianyi; Lu, Wencong; Yang, Jie; Li, Guozheng. *Support Vector Machine in Chemistry*, World Scientific Co., New Jersey, 2004, 331p.
37. Molegro Data Modeller (MDM), V. 2.1.0 Copyright Molegro, 2009. [www.molegro.com](http://www.molegro.com).
38. Chang, Chih-Chung; Lin, Chih-Jen. LIBSVM: a Library for Support Vector Machines. [www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf](http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf). Acesso: Março 2011.
39. SciFinder Scholar, v. 2007, Copyright 2007, American Chemical Society.
40. DSViewerLite 5.0 Copyright 2002, Accelrys Inc.
41. ChemDraw 9.0/Chem 3D v. 9.0 Copyright 1986-2004, CambridgeSoft.
42. Allinger, N. L. Conformational Analysis. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms. *J. Am. Chem. Soc.*, 99, 8127-8134, **1977**.
43. Dewar, M.J.S; Zoebisch, E. G.; Healy, E. F.; Steward, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.*, 107, 3902-3909, **1985**.
44. Gaussian 03, Revision B.05, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 2003.
45. Neese, Frank. ORCA, An *ab initio*, DFT and semiempirical SCF-MO package. v. 2.6.0. Linux.
46. ADF – Amsterdam Density Functional. V. 2009.01-Linux. Scientific Computing & Modelling NV. Copyright 1993-2009 SCM.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

47. Talete srl., DRAGON for Windows (Software for Molecular Descriptor Calculation), v. 5.4, 2006.
48. N. Sukumar, N.; Breneman, C.M.; Program RECON, v. 5.3, Copyright, 2001
49. CaChe WorsystemPro v. 6.1.12.33, Copyright 2000-2004, Fujitsu Limited. 1989-2000, Oxford Molecular Ltd.
50. Molinspiration : <http://www.molinspiration.com/services>. Acesso: July, 2008
51. Gasteiger, J. PETRA Online Service-Parameter Estimation for the Treatment of Reactivity Applications - Copyright Johann Gasteiger, 1998.  
<http://www2.ccc.uni-erlangen.de/services/petra/smiles.html>. Acesso: Novembro, 2008. Site cancelado pelo autor em 2010.
52. Hansch, Corwin; Leo, Albert; Hoekman, David. Exploring QSAR. American Chemical Society, 1995, 348 p.
53. Talete srl, Moby Digs Professional, v. 1.1, 2009.
54. Todeschini, Roberto; Consonni, Viviana; Mauri, Andrea; Pavan, Manuela. Detecting 'bad' regression models: multi criteria fitness functions in regression analysis. *Anal. Chim. Acta*, 515, 199-208, **2004**.
55. De Oliveira, D.B.; Gaudio, Anderson Coser. Quantitative Structure Activity-Relationships, 19 (6), 599-601, **2001**. BuildQSAR, v. 2.0.0.0, Universidade Federal do Espírito Santo. Copyright 2007.
56. Eriksson, Lennart; Jaworska, Joanna; Worth, P. Andrew; Cronin, Mark T.D.; McDowell, Robert M.; Gramatica, Paola. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environmental Health Perspectives*, 111, 1361-1375, **2003**.
57. R. Wehrens; H. Putter; L.M.C. Buydens. The bootstrap: a tutorial. *Chemom. Intell. Lab. Syst.*, 54, 35-52, **2000**.
58. Wold, S.; Eriksson, L. Statistical Validation of QSAR results. Em: Chemometric Methods in Molecular Design; van de Waterbeemd, H. Ed. VHC: Weinheim, 1995, p.309.
59. Lindgren, L.; Hansen, B.; Karcher, W.; Sjostrom, M.; Eriksson, L. Model Validation by Permutation Tests: Applications to Variable Selection. *J. Chemom.*, 10, 521-532, **1996**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzolidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

60. Kiralj, Rudolf; Ferreira, Márcia M. C. Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *J. Braz. Chem. Soc.*, 20, 770-787, **2009**.
61. Teófilo, Reinaldo F.; Martins, João Paulo A.; Ferreira, Márcia M.C. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. *Journal of Chemometrics*, 23, 32-48, **2009**.
62. Myatt, Glenn. Making Sense of Data. A Practical Guide to Exploratory Analysis and Data Mining. John Wiley & Sons Inc., New Jersey, 2007, p. 167, 280 p.
63. Jaworska, Joanna; Nikolova-Jeliazkova, Nina; Aldenberg, Tom. QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *ATLA*, 33, 445-459, **2005**.
64. Golbraikh, Alexander; Tropsha, Alexander. Beware of Q2! *J. Mol. Graphics Modell.*, 20, 269-276, **2002**.
65. (a) Perola, Emanuele; Charifson, Paul S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.*, 47, 2499-2510, **2004**; (b) Ercanli, Tulay; Boyd, Donald B. Evaluation of Computational Chemistry Methods: Crystallographic and Cheminformatics Analysis of Aminothiazole Methoximes. *J. Chem. Inf. Model.*, 45, 591-601, **2005**.
66. Diller, D. J.; Merz, K. M. Jr.; Can we separate active from inactive conformations? *J. Comput.-Aided Mol. Des.*, 16, 105-112, **2002**.
67. Kirchmair, J.; Laggner, C; Wolber, G.; Langer, T. Comparative analysis of protein-bound ligand with respect to Catalyst's conformational space sub sampling algorithms. *J. Chem. Inf. Model.*, 45, 422, **2005**.
68. (a) MACROMODEL v.7.0 Copyright 2001, Shroedinger Inc. (b) MMFF implementado conf. Halgren, T.A, *J. Comp. Chem.*, 17, 490-512; 520-552; 553-586; 587-615; 616-641, **1996**.
69. (a) FROG v.2.1 Disponível em <http://bioserv.rpbs.jussieu.fr/cgi-bin/Frog>; (b) Leite, T.B.; Gomes, D.; Miteva, M.A.; Chomilier, J.; Villoutreix, B.O.; Tufféry, P., Frog: a FRee Online druG 3D conformation generator; *Nucleic Acids Res.*, 35, 569-572, **2007**.

Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS

---

70. (a) Vconf.v.2.0.Copyright @ Verachem. VeraChem's Optimized Dreiding Parameters PDF:<http://www.verachem.com/downloads.html> ; (b) Chang, C.E.; Gilson, M.K.Tork: Conformational analysis method for molecules and complexes. *J. Comput. Chem.*, 24,1987-1998, **2003**.
71. Nikolova, Nina; Jaworska, Joanna. Approaches to Measure Chemical Similarity – a Review. *QSAR & Comb. Sci.*, 22, 1006-1025, **2003**.
72. Vainio, M.; Puranen, J.S.; Johnson, M.S. ShaEP: molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.*, 49, 492-502. **2009**.
73. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S., Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **25**, 1605-1612, **2004**.
74. (a) Wang J.; Wang, W.; Kollman, P.A.; Case, D.A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247-260, **2006**. (b) Jakalian, Arazaz; Busch, Bruce L.; Jack, David B.; Byly, Christopher. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *Journal of Computation of Chemistry*, 21,136-146, **2000**.
75. Itokawa; H.; Totsuka, N.; Morita, H.; Takeya, K.; Itaka, Y., Schenkel, E. P.; Motidome, M. New Antitumor Principles, Casearins A-F, for *Casearia sylvestris* Sw. (Flacourtiaceae). *Chem. Pharm. Bull.*, 38 (12), 3384-3388, **1990**.
76. Morita, H.; Nakayama, M.; Kojima, H.; Takeya, K.; Itokawa, H.; Schenkel, E. P.; Motidome, M. Structures and Cytotoxic Activity Relationship of Casearins, New Clerodane Diterpenes from *Casearia sylvestris* Sw. *Chem. Pharm. Bull.* 39 (12), 693-697, **1991**.
77. Isanbo, C.; O'Hagan, D. Fluorine in Medicinal Chemistry: A Review of Anticancer Agents. *J. of Fluorine Chem.*, 127, 303-319, **2006**.
78. Bégué, Jean-Pierre; Bonnet-Delpon, Danièle. Chimie Bioorganique et Médicinale du Fluor. *EDP Science*, Paris, 2005, p. 2, 12-13, 366p.
79. Bégué, J.-P.; Bonnet-Delpon. Recent advances ( 1995-2005) in Fluorinated Pharmaceuticals based on Natural Products. *J. Fluorine Chem.*,127, 992-1012, **2006**.
80. Hemmer, M.C.; Steinhauer, V.; Gasteiger, J. Deriving the 3D Structure of Organic Molecules from their Infrared Spectra. *J.Vibr. Spect.*, 19, 151-164, **1999**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzolidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

81. Consonni, V.; Todeschini, R.; Pavan, M. Structure/Response Correlations and Similarity/ Diversity Analysis by GETAWAY Descriptors. 1. Theory of the novel 3D Molecular Descriptors. *Chem. Inf. Sci.*, 42, 682-692, **2002**.
82. Breneman, C. M.; Sundling, C. M.; Sukumar, N.; Shen, L.; Katt, W.P; Embrechts, M.J. New developments in PEST shape/property hybrid descriptors. *J. Comput.-Aided Mol. Des.*, 17, 231-240, **2003**.
83. Breneman, C. M.; Thomson, T. R.; Rhem, M., Dung, M. Electron Density Modeling of Large Systems using the Transferable Atom Equivalent Method. *Comp. and Chem.*, 19, 161-179, **1995**, e correspondência pessoal.
84. Shen, Ya-Ching; Wang, Chi-Hsi; Cheng, Yuan-Bin; Wang, Li-Tung; Guh Jih-Hwa; Chien, Chin-Te; Khalil, Ashraf Taha. New Cytotoxic Clerodane Diterpenoids from the Leaves and Twigs of *Casearia membranacea*. *J. Nat. Prod.*, 67, 316-321, **2004**.
85. Espindola, Laila S.; Vasconcelos Júnior, José Rossy e; Mesquita, Mariana L. de; Marquié, Patrick; Paula, José E. de; Mambu, Lengo; Santana, Jaime M. Trypanocidal Activity of a New Diterpene from *Casearia sylvestris* var. *lingua*. *Planta Medica*, 70, 1093-1095, **2004**.
86. Beckel, Helenara dos Santos; Lorini, Irineu; Lazzari, Sonia M.N. *Revista Brasileira de Entomologia*, 50, 110-114, **2006**.
87. Stanton, David T.; Jurs, Peter C. Development and Use of Charges Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies. *Anal. Chem.*, 62, 2323-2329, **1990**.
88. Consonni, Viviana; Todeschini, Roberto; Pavan, Manuela. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *J. Chem. Inf. Comput. Sci.*, 42, 682-692, **2002**.
89. Oikawa, Nobuhiro; Nakagawa, Yoshiaki; Nishimura, Keiichiro; Ueno, Tamio; Fujita, Toshio. Quantitative Structure-Activity Analysis of Larvicidal 1-(Substituted benzoyl)-2-benzoyl-1-*tert*-butylhydrazines against *Chilo suppressalis*. *Pestic. Sci.*, 41, 139-148, **1994**.
90. Carbó, R.; Leyda, L.; Arnau M. How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *International Journal of Quantum Chemistry*, 17, 1185-1189, **1980**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

91. APBS - Adaptive Poisson-Boltzmann Solver (APBS) -- Software for evaluating the electrostatic properties of nanoscale biomolecular systems. Acesso: 20-10-2010.  
*Open source:* <http://apbs.sf.net/>
92. Shuur, Jan, H.; Selzer, Paul; Gasteiger, Johan. The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity. *J. Chem. Inf. Comput. Sci.*, 36, 334-344, **1996**.
93. Soltzberg, Leonard, J.; Wilkins, Charles, L. Molecular Transforms: a Potential Tool for Structure-Activity Studies. *J. Am. Chem. Soc.*, 99, 439-443, **1976**.
94. Azidah, A.A.; Sofian-Azirun, M. Life history of *Spodoptera exigua* (Lepidoptera: Noctuidae) on various host plants. *Bulletin of Entomological Research*, 96, 613-618, **2006**.
95. Biotecnologia, Ciência e Desenvolvimento, ano IV, no. 24, Jan/Fev **2002**.  
Acesso em 14-04-2011: [www.biotecnologia.com.br](http://www.biotecnologia.com.br)
96. Billas, Isabelle M.L.; Iwema, Thomas; Garnier, Jean-Marie; Mitschler, André; Rochel, Natacha; Moras, Dino. Structural adaptability in the ligand-binding pocket of the ecdyson hormone receptor. *Nature*, 426, 91-96, **2003**.
97. Kasuya, Atsushi; Sawada, Yoshihiro; Tsukamoto, Yoshihisa; Tanaka, Keiji; Toya, Tetsuya; Yanagi, Mikio. Binding mode of ecdysone agonists to receptor: comparative modeling and docking studies. *J. Mol. Model.*, 9, 58-65, **2003**.
98. Wurz, Jean-Marie; Guillot, Benoit; Fagart, Jérôme; Moras, Dino; Tietjen, Klaus; Schindler, Michael. A new model for 20-hydroxyecdysone and dibenzoylhydrazine binding: A homology modeling and docking approach. *Protein Science*, 9, 1073-1084, **2000**.
99. Maple, Jon, R.; Ewig, Carl S. An ab initio procedure for deriving atomic polarizability tensors in molecules. *Journal of Chemical Physics*, 115, 4981-4988, **2001**;
100. Levine, Ira, N. Quantum Chemistry, Prentice Hall, Englewood Cliffs, 1991, p.401, 629p.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

- 101.** (a) Sebera, Donal K. Electronic Structure and Chemical Bonding. Blaisdell Publishing Co. London, Waltham, 1964, p. 260-267, 298p; (b) Companion, Audrey L. Ligação Química. Ed. Edgar Bluecher Ltda, 1964, p. 96,140p (c) Atkins, Peter; de Paula, Julio. ATKINS' Physical Chemistry, 8th Ed. Oxford University Press, Oxford, 2006, p. 634, 1064 p.
- 102.** (a) Fluege, Siegfried. Practical Quantum Mechanics. Springer-Verlag, New York, 1974, p.74,78, 287 p. (b) Schiff, Leonard I. Quantum Mechanics, 2<sup>nd</sup> Ed. McGraw-Hill Bok Co., New York, 1965, p. 176-178, 417 p.
- 103.** Tipler, Paul A. Physik. Spektrum Akademischer Verlag, Heidelberg, **1998**, p. 1316, 1522 p.
- 104.** Moris, Garrett M.; Lim-Wilby, Marguerita. Molecular Docking. In: Kukol Andreas, Editor. Molecular Modeling of Proteins, Human Press, New Jersey, 2008, 365-378, 390 p.
- 105.** PDB Data Bank: [www.pdb.org](http://www.pdb.org).
- 106.** Hormann, Robert E.; Smagghe, Guy; Nakagawa, Yoshiaki. Multidimensional Quantitative Structure-Activity Relationships of Diacylhydrazine Toxicity to Lepidopteran and Coleopteran Insects Pests. *Qsar & Combinatorial Science*, 27, 1098-1112, **2008**.
- 107.** Wheelock, Craig E.; Nakagawa, Yoshiaki; Harada, Toshiyuki; Oikawa, Nobuhiro; Akamatsu, Miki; Smagghe, Guy; Stefanou, Dimitra; Iatrou, Kostas; Swevers, Luc. High-throughput screening of ecdyson agonists using a reporter gene assay followed by 3-D QSAR analysis of the molting hormonal activity. *Bioorganic & Medicinal Chemistry*, 14, 1143-1159, **2006**.
- 108.** Carmichael, Jennifer A.; Lawrence, Michael C.; Graham, Lloyd D.; Pilling, Patricia A.; Epa,V. Chandana; Noyce, Leonie; Lovrecz, George; Winkler, David A.; Pawlak-Skrzecz, Anna; Eaton; Ruth E.; Hannan, Garry N.; Hill, Ronald J. The X-ray Structure of a Hemipteran Ecdysone Receptor Ligand-binding Domain. COMPARISON WITH A LEPIDOPTERAN ECDYSONE RECEPTOR LIGAND-BINDING DOMAIN AND IMPLICATIONS FOR INSECTICIDE DESIGN. *The Journal of Biological Chemistry*, 280, 22258-22269, **2005**.
- 109.** Swevers, Luc; Soin, Thomas; Mosallanejad, Hadi; Iatrou, Kostas; Smagghe, Guy. Ecdysteroid signaling in ecdysteroid-resistant cell lines from the polyphagous noctuid pest *Spodoptera exigua*. *Insect Biochemistry and Molecular Biology*, 38, 825–833, **2008**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

110. Cache & BioMedCache, Molecular Modeling in Drug Design. V. 6.1.12.33 Fujitsu Ltd, 2003.
111. Muegge, Ingo; Martin, Yvonne C. A general and fast Scoring function for protein-ligand interaction: A simplified potential approach. *J. Med. Chem.*, 42, 791-804, **1999**.
112. Getting Started with BioMedCache, Fujitsu, Ltd., 2003.
113. Madan, B.; Sharp, K.A. Heat Capacity Changes Accompanying Hydrophobic and Ionic Solvation: A Monte Carlo and Random Network Model Study. *J. Phys. Chem.*, 100, 7713-7721, **1996**.
114. Muegge, I; Rarey, M. Small Molecule Docking and Scoring. In: Reviews in Computational Chemistry, v. 17, Editor, B. Lipkowitz, Donald B. Boyd, John Wiley & Sons, 2001, p.1-60.
115. Wang, Renxiao; Lai, Luhua; Wang, Shaomeng. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, 16, 11-26, **2002**.
116. (a) UniProt. <http://www.uniprot.org/> acesso: 15-5-2011. (b) Roy, Ambrish; Kucukural, Alper; Zhang; Yang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5, 725-738, **2010**;  
(c) Zhang, Yang. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins, Prediction Report*, 108-117, **2007**;  
[Servidor: http://zhanglab.ccmb.med.umich.edu/I-TASSER/](http://zhanglab.ccmb.med.umich.edu/I-TASSER/), Acesso: 21-05-2011.
117. Martinez, L.; Andreani, R.; Martinez, J. M. Convergent algorithms for protein structural alignment. *BMC Bioinformatics*, 8, 306, **2007**.  
Acesso: 22/05/2011 <http://www.ime.unicamp.br/~martinez/lovoalign/software.html>
118. Lu, Yunxiang; Wang, Yong; Zhu, Weiliang. Nonbonding interactions of organic halogens in biological systems: implications for drug discovery and biomolecular design. *Physical Chemistry Chemical Physics*, 12, 4543-4551, **2010**.
119. Kortagere, Sandhya; Ekins, Sean; Welsh, William J. Halogenated ligands and their interactions with amino acids: Implications for structure-activity and structure-toxicity relationships. *Journal of Molecular Graphics and Modelling*, 27, 170-177, **2008**.
120. Politzer, P; Murray, J.S.; Concha, M.C. Halogen bonding and the design of new materials: organic bromides, chlorides and perhaps even fluorides as donors. *J Mol. Model.*, 13, 643-650, **2007**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoilidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

- 121.** Auffinger, Pascal; Hays, Franklin A.; Westhof, Eric; Ho, P. Shin. Halogen bond in biological molecules. The National Academy of Science. PNAS, v.101, no. 48, 16789-16794, **2004**. Acesso: Fevereiro-2011.  
[www.pnas.org/cgi/doi/10.1073/pnas.0407607101](http://www.pnas.org/cgi/doi/10.1073/pnas.0407607101).
- 122.** (a) Gonnade, Rajesh G.; Shashidhar, Mysore S.; Bhadbhade, Mohan M. 'Halogen Bonding' Interactions in Molecular Crystals: From Early Recognition to Recent Developments. *Journal of the Indian Institute of Science*, 82, 149 -165, **2007**. (b) CSD – Cambridge Crystall Structure Database.  
[www.ccdc.cam.ac.uk/products/csd](http://www.ccdc.cam.ac.uk/products/csd)
- 123.** Fricke, B. On the correlation between electric polarizabilities and the Ionization potential of atoms. *J. Chem. Phys.*, 84, 862-866, **1986**.
- 124.** (a) Fujiwara, Hideaki; Da, Yong-Zhong; Ito, Katsuhiko. Energy aspects of Oil/Water Partition Leading to the Novel Hydrophobic Parameters for the Analysis of Quantitative Structure-Activity Relationships. *J. Med. Chem.*, 35, 3382-3387, **1992**. (b) Freire, Ernesto. Do Enthalpy and Entropy Distinguish First in Class from Best in Class? *Drug Discovery Today*, 13, 869-874, **2008**. (c) Luque, Irene; Freire, Ernesto. Structural Parameterization of the Binding Enthalpy of Small Ligands. *PROTEINS: Structure, Function and Genetics*, 49, 181-190, **2002**.
- 125.** Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot. Eng.*, 8, 127-134, **1995**.
- 126.** Tauer, Anthony P. Theoretical Investigations Of Pi-Pi And Sulfur-Pi Interactions and their Roles in Biomolecular Systems. Tese de Mestrado, Georgia Institute of Technology, **2005**, Capitulo V.
- 127.** Saraogi, Ishu; Vijay, V.G.; Das, Soma; Sekar, K.; Row, T.N. Guru. C-halogen...Pi interactions in proteins: a database study. *Crystal Engineering* 6, 69-77, **2003**.
- 128.** Bordas, B.; Belai, I.; Lopata, A.; Szanto, Z. Interpretation of Scoring Functions Using 3D Molecular Fields. Mapping the Diacyl-Hydrazine-Binding Pocket of an Insect Ecdysone Receptor. *J. Chem. Inf. Model.*, 47, 176-185, **2007**.
- 129.** Molegro Molecular Viewer, v. 2.2.0 Copyright 2005 –**2011**, Molegro ApS.
- 130.** Smagghe, Guy; Nakagawa, Yoshiaki; Mourad, Ahmed Kamal; Fujita, Toshio; Tirry, Luc. Comparative Ecdysteroid Action of Ring-Substituted Dibenzoilhydrazines in *Spodoptera exigua*. *Archives of Insect Biochemistry and Physiology*, 41, 42-53, **1999**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzoidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

- 131.** Cruz & Monteiro, 2004 Cruz, I.; Monteiro, M.A.R. Controle biológico da lagarta-do-cartucho do milho *Spodoptera frugiperda* utilizando o parasitóide de ovos *Trichogramma pretiosum*. Sete Lagoas: Embrapa Milho e Sorgo, **2004**. 4p. (Comunicado Técnico).
- 132.** Williams, W.P.; Davis, F.M. Response of corn to artificial infestation with fall armyworm and southwestern corn borer larvae. *Southwestern Entomologist, Dallas*, v.15, p.163-166, **1990**.
- 133.** Cruz, I. et al. Efeito do nível de saturação de alumínio em solo ácido sobre os danos de *Spodoptera frugiperda* (J. E. Smith) em milho. *Anais da Sociedade Entomológica do Brasil*, Londrina, v.25, p.293-297, **1996**
- 134.** Lima, José Farid Maia; Grützmacher, Anderson Dionei; da Cunha, Uemerson Silva; Porto, Marilda Pereira; Martins, José Francisco da Silva; Dalmazo, Gabriel Olle. Ação de inseticidas naturais no controle de *Spodoptera frugiperda* (J.E. Smith, 1797)(Lepidoptera: Noctuidae) em milho cultivado em agroecossistema de várzea. *Ciência Rural*, Santa Maria, v.38, 607-613, **2008**.
- 135.** Parr R.G.; Szentpaly, L.V.; S. Liu. Electrophilicity Index. *J. Am. Chem. Soc.*, 121, 1922, **1999**.
- 136.** Roy, R.K.; Pal, S.; Hirao, K.; On non-negativity of Fukui Function indices. *J. Chem. Phys.* 110, 8236-8245, **1999**
- 137.** Parthasarathi, R.; Padmanabhan, J.; Elango, M.; Subramaniam, V.; Chattaraj, P.K.; Intermolecular reactivity through the generalized philicity concept. *Chem. Phys. Letters*, 394, 225-230. **2004**.
- 138.** Duarte, Hélio Anderson. Índice de Reatividade Química a Partir da Teoria do Funcional de Densidade: Formalismo e Perspectivas. *Quim. Nova*, 24, 501-508, **2001**.
- 139.** Parr, R.G.; Pearson, R.G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.*, 105, 7512, **1983**.
- 140.** Parr, R.G.; Yang, Weitao. Density Functional *Approach to the Frontier-Electron Theory of Chemical Reactivity*. *J. Am. Chem. Soc.*, 106, 4049-4050, **1984**.
- 141.** Yang, Weitao; Mortier, Wilfried J. The Use of Global and Local Molecular Parameters for the Analysis of Gas-Phase Basicity of Amines. *J. Am. Chem. Soc.*, 108, 5708-5711, **1986**.

**Estudos de QSAR-2 aplicados a  
Diterpenóides Clerodanos e Dibenzolidrazinas  
REFERÊNCIAS BIBLIOGRÁFICAS**

---

- 142.** Saha, Soumen; Roy, Ram Kinkar; Ayers, Paul W. Are the Hirshfeld and Mulliken Population Analysis Schemes Consistent with Chemical Intuition? *International Journal of Quantum Chemistry*, 109, 1790-1806, **2006**
- 143.** Guerra, Celia Fonseca; Handgraaf, Jan-Willem; Baerends, Evert Jan; Bickelhaupt, F. Matthias. Voronoi Deformation Density (VDD) Charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD Methods for Charge Analysis. *J. Comp. Chem*, 25, 188-209, **2003**.
- 144.** Chattaraj, Pratim Kumar; Maiti, Buddhadev; Sarkar, Utpal. Philicity: A Unified Treatment of Chemical Reactivity and Selectivity. *J. Phys. Chem. (A)*, 107, 4973-974, **2003**.
- 145.** Verma, Rajeshwar P.; Hansch, Corwin. An Approach toward the problem of outliers in QSAR. *Bioorganic & Medicinal Chemistry*, 13, 4597-4621, **2005**.
- 146.** Web of Science – diversos acessos via Unicamp em outubro de 2010.

**Referências utilizadas como meios auxiliares durante a execução da tese não citadas no texto.**

- 147.** OriginPro v. 7.5 Copyright 1991-2008. OriginLab Corporation.  
Análises gráficas em geral e gráficos da tese.
- 148.** VegaZZ, v. 2.3.2.28 – Copyright Alessandro Pedretti & Giulio Vistoli.  
Análises auxiliares.
- 149.** OpenBabelGUI v. 2.2.3. Copyright 2006 por Chris Morley. Disponibilidade:  
[http://openbabel.sourceforge.net/wiki/Windows\\_GUI](http://openbabel.sourceforge.net/wiki/Windows_GUI).

## **A N E X O S**



**Tabela 3.** Predição interna e externa do QSAR – Métodos MLR e SVM

No.	MolID	CAS No. Reg.	$Y_{Obs}^a$	Descritores			$Y_{MLR}^b$	Res <sup>c</sup>	$Y_{SVM}^b$	Res <sup>d</sup>
				RDF060m	R8p	Fuk5				
1	A	134174-73-5	6,00	29,874	0,617	119,230	5,84	0,16	5,86	0,14
2	B	133577-59-0	5,07	28,128	0,702	119,330	5,04	0,03	5,10	-0,03
3	C	133577-60-3	6,11	35,602	0,576	132,500	6,05	0,07	5,94	0,17
4	D	133557-53-6	5,74	28,914	0,579	126,070	5,74	0,01	5,77	-0,03
5	E	133535-78-1	5,33	19,374	0,506	129,190	5,33	0,00	5,50	-0,17
6	G	134955-62-7	6,77	31,047	0,610	105,670	6,61	0,16	6,60	0,17
7	H	134955-63-8	6,41	27,228	0,580	108,220	6,40	0,01	6,48	-0,07
8	I	134955-64-9	6,29	31,794	0,583	114,280	6,49	-0,20	6,45	-0,16
9	J	134955-65-0	5,96	31,524	0,568	125,300	6,08	-0,12	6,03	-0,07
10	O	134955-70-7	5,22	31,519	0,670	125,390	5,30	-0,08	5,32	-0,10
11	P	134955-71-8	5,11	29,160	0,717	113,260	5,30	-0,19	5,28	-0,17
12	Q	134955-72-9	5,37	30,480	0,693	121,870	5,20	0,17	5,21	0,16
13	R	134955-73-0	5,27	24,239	0,569	129,270	5,27	0,00	5,44	-0,17
1	Aa <sup>T</sup>	134955-74-1	6,26	29,562	0,607	111,380	6,25	0,01	6,29	-0,03
2	K <sup>T</sup>	134955-66-1	6,28	37,110	0,634	121,870	6,22	0,07	6,03	0,25
3	L <sup>T</sup>	134955-67-2	5,80	27,584	0,596	119,230	5,81	-0,01	5,87	-0,08
4	M <sup>T</sup>	134955-68-3	5,75	24,939	0,573	127,840	5,36	0,38	5,50	0,24
5	N <sup>T</sup>	134955-69-4	5,23	33,333	0,703	125,390	5,21	0,02	5,20	0,03

(T) Compostos do conjunto de teste; **(a)**  $pIC_{50}$  experimental ( $IC_{50} = \mu\text{mol/L}$ ); **(b)**  $pIC_{50}$  calculados e método; **(c)**  $Y_{Obs} - Y_{MLR}$ ; **(d)**  $Y_{Obs} - Y_{SVM}$

**Tabela 7a.** Valores preditos  $IC_{50}$  de diterpenóides clerodanos contra células tumorais V79 com os modelos de QSAR-MLR Eq. (5.1-6) e QSAR-SVM ( $n = 13$ )

Item	Composto	CAS Reg. Nr.	Descritores			Métodos e valores preditos			
						$Y^a$		$IC_{50}^b$	
			RDF060m	R8p	Fuk5	MLR	MLR	SVM	SVM
1	Casearborin A	272771-14-9	25,294	0,533	91,404	7,35	0,045	7,07	0,085
2	Casearborin B	272771-15-0	26,967	0,510	100,840	7,24	0,058	7,07	0,086
3	Caseargrewiin H	945607-74-9	24,403	0,510	98,199	7,14	0,073	6,97	0,106
4	Casearborin D	272771-17-2	29,539	0,547	104,960	6,99	0,102	6,94	0,114
5	Caseargrewiin E	945607-68-1	24,030	0,544	94,658	7,01	0,098	6,92	0,121
6	Casearin S	220178-32-5	25,802	0,570	97,506	6,84	0,146	6,85	0,143
7	Caseargrewiin G	945607-72-7	27,365	0,560	103,030	6,80	0,160	6,82	0,152
8	Caseargrewiin D	847841-95-6	19,575	0,500	88,496	7,24	0,058	6,81	0,156
9	Cgrayi 1	947672-97-1	21,581	0,498	106,010	6,63	0,233	6,58	0,264
10	Cgrewiifolia 5	1006868-63-8	24,546	0,559	105,420	6,45	0,353	6,55	0,281
11	Casearlucina F	405507-88-2	20,268	0,557	98,141	6,43	0,372	6,53	0,293
12	Casearinol A	196090-44-5	28,084	0,607	103,840	6,46	0,343	6,53	0,294
13	Cgrewiifolia 4	1006868-61-6	23,851	0,572	103,650	6,37	0,423	6,50	0,313
14	Casearborin C	272771-16-1	30,066	0,621	104,960	6,48	0,333	6,49	0,322
15	Ctremula 3	173324-52-2	21,284	0,528	105,300	6,41	0,388	6,48	0,331
16	Zguidonia 8	128486-53-3	13,351	0,499	83,108	6,96	0,111	6,45	0,359
17	Casearborin E	272771-18-3	34,830	0,654	105,060	6,63	0,233	6,44	0,359
18	Corymbolusina B	326473-52-3	30,738	0,539	118,840	6,53	0,297	6,44	0,362
19	Caseargrewiin J	945607-78-3	21,065	0,490	109,290	6,50	0,317	6,44	0,363
20	Ccorymbosa 5	131766-54-6	15,539	0,492	97,670	6,54	0,291	6,43	0,372
21	Casearlucina A	405507-78-0	27,067	0,561	111,700	6,37	0,428	6,43	0,373
22	Cgrayi 6	947672-99-3	15,053	0,515	95,445	6,42	0,379	6,40	0,398
23	Caseargrewiin I	945607-76-1	23,075	0,511	111,930	6,39	0,404	6,40	0,401
24	Casearinone B	196090-47-8	17,969	0,585	93,282	6,24	0,576	6,39	0,404
25	Casearinol B	196090-45-6	30,153	0,596	111,870	6,36	0,435	6,38	0,418
26	Caseargrewiin F	945607-70-5	23,040	0,535	110,160	6,29	0,514	6,37	0,428
27	Cgrayi 7	947673-00-9	12,336	0,517	88,670	6,48	0,332	6,34	0,458
28	Zguidonia 7	128486-52-2	11,797	0,505	76,005	7,10	0,080	6,34	0,460
29	Intrapetacina A	343944-35-4	22,120	0,544	109,430	6,17	0,669	6,29	0,507
30	Zuelanina 13	130827-60-0	15,269	0,591	90,262	6,10	0,797	6,28	0,525
31	Corymbolusina A	326473-51-2	24,464	0,470	120,190	6,45	0,356	6,26	0,550
32	Zuelanina 9	130827-56-4	22,456	0,597	103,820	6,06	0,878	6,26	0,556
33	Zuelanina 3	128572-38-3	23,619	0,571	109,640	6,09	0,814	6,24	0,581
34	Corymbotina F	138950-31-9	24,757	0,589	109,000	6,08	0,832	6,22	0,596
35	Casearlucina G	405507-90-6	14,265	0,540	98,141	6,04	0,910	6,22	0,602
36	Zuelanina 1	128486-31-7	23,010	0,567	109,640	6,07	0,857	6,22	0,603
37	Casearvestrina C	402935-06-2	20,089	0,511	111,760	6,14	0,718	6,20	0,626
38	Caseargrewiin L	945607-81-8	19,130	0,526	108,890	6,08	0,836	6,20	0,629

(a)  $Y = \text{Log}(1/IC_{50})$ ; (b)  $\mu\text{mol/L}$

**Tabela 7b.** Valores preditos IC<sub>50</sub> de diterpenóides clerodanos contra células tumorais V79 com os modelos de QSAR-MLR Eq. (5.1-6) e QSAR-SVM (n = 13)

Item	Composto	CAS Reg. Nr.	Descritores			Métodos e valores preditos			
						Y <sup>a</sup>		IC <sub>50</sub> <sup>b</sup>	
			RDF060m	R8p	Fuk5	MLR	MLR	SVM	SVM
39	Laetiaprocerin C	871217-64-0	36,162	0,651	113,470	6,39	0,408	6,19	0,644
40	Ctremula 2	173324-51-1	18,667	0,561	105,240	5,94	1,152	6,16	0,689
41	Ccorymbosa 8	131740-59-5	18,156	0,570	103,650	5,90	1,263	6,15	0,710
42	Cgrewiifolia 2	947681-66-5	16,566	0,608	95,867	5,83	1,490	6,14	0,730
43	Cgrayi 4	947672-96-0	13,951	0,520	102,550	5,97	1,083	6,14	0,732
44	Casearinone A	196090-46-7	14,160	0,584	95,539	5,82	1,526	6,13	0,746
45	Caseargrewiin A	847841-92-3	28,140	0,591	115,910	6,04	0,907	6,10	0,797
46	Intrapetacin B	343944-36-5	22,594	0,564	112,070	5,94	1,138	6,10	0,798
47	Zguidonia 13	128572-39-4	23,501	0,591	109,730	5,92	1,193	6,10	0,799
48	Caseamembrin A	664351-45-5	28,559	0,526	124,870	6,16	0,685	6,10	0,801
49	Cgrewiifolia 1	947672-98-2	19,329	0,600	102,640	5,82	1,521	6,10	0,803
50	Corymbolusin C	326473-53-4	25,499	0,542	118,840	6,05	0,885	6,09	0,805
51	Ccorymbosa 11	131740-55-1	10,353	0,535	95,773	5,85	1,414	6,09	0,821
52	Casearluцин C	405507-82-6	27,452	0,615	111,790	5,99	1,028	6,08	0,840
53	Ccorymbosa 4	131740-57-3	7,484	0,312	87,138	7,69	0,021	6,07	0,851
54	Zguidonia 12	128486-37-3	20,046	0,561	109,730	5,85	1,402	6,06	0,873
55	Ccorymbosa 13	131740-56-2	18,140	0,576	105,420	5,77	1,694	6,05	0,897
56	Cgrewiifolia 3	1006868-59-2	21,095	0,556	112,190	5,87	1,349	6,04	0,909
57	Zuelanin 10	130827-57-5	22,216	0,580	110,920	5,84	1,439	6,03	0,931
58	Caseamembrin B	664351-46-6	23,370	0,552	116,160	5,92	1,214	6,03	0,942
59	Casearnigrescens D	925693-22-7	20,241	0,540	113,280	5,87	1,355	6,02	0,953
60	Zuelanin 4	130858-02-5	29,974	0,615	116,840	5,98	1,057	5,99	1,019
61	Cgrayi 8	947673-01-0	7,541	0,305	64,534	8,77	0,002	5,97	1,071
62	Laetiaprocerin A	871217-62-8	20,205	0,585	109,160	5,71	1,946	5,96	1,101
63	Casearvestrin A	402937-50-2	19,118	0,568	109,930	5,71	1,944	5,96	1,106
64	Zuelanin 5	40185-32-8	4,900	0,358	61,245	8,29	0,005	5,94	1,140
65	Zuelanin 8	130858-01-4	19,344	0,526	115,300	5,81	1,566	5,94	1,141
66	Ccorymbosa 12	131830-93-8	13,135	0,615	95,867	5,48	3,324	5,94	1,144
67	Casearvestrin B	402935-05-1	19,317	0,560	111,700	5,71	1,956	5,93	1,167
68	Caseamembrin E	664351-50-2	24,923	0,518	124,870	5,91	1,225	5,91	1,238
69	Caseamembrol A	723302-01-0	24,923	0,518	124,870	5,91	1,225	5,91	1,238
70	Ctremula 5	173324-53-3	21,571	0,496	122,400	5,90	1,253	5,90	1,261
71	Zuelanin 2	128486-32-8	16,505	0,524	112,910	5,68	2,068	5,89	1,293
72	Corymbotin D	139015-30-8	15,862	0,608	102,150	5,48	3,304	5,89	1,299
73	Zguidonia 6	128486-33-9	24,789	0,588	116,840	5,73	1,844	5,86	1,375
74	Caseargrewiin K	945607-80-7	19,826	0,475	122,850	5,89	1,287	5,86	1,396
75	Zguidonia 9	128486-34-0	19,989	0,538	117,130	5,69	2,057	5,85	1,427
76	Laetiaprocerin D	871217-65-1	15,117	0,441	119,200	5,91	1,234	5,84	1,459

(a) Y = Log(1/IC<sub>50</sub>); (b) μmol/L.

**Tabela 7c.** Valores preditos IC<sub>50</sub> de diterpenóides clerodanos contra células tumorais V79 com os modelos de QSAR-MLR Eq. (5.1-6) e QSAR-SVM (n = 13)

Item	Composto	CAS Reg. Nr.	Descritores			Métodos e valores preditos			
			RDF060m	R8p	Fuk5	Y <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	Y <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>
77	Caseargrewiin B	847841-93-4	24,532	0,511	126,810	5,84	1,435	5,84	1,460
78	Casearlucin D	405507-84-8	28,375	0,614	118,900	5,75	1,769	5,81	1,565
79	Casearlucin E	405507-86-0	23,045	0,585	116,260	5,63	2,328	5,80	1,582
80	Casearlucin I	405507-94-0	29,948	0,547	132,040	5,80	1,589	5,79	1,608
81	Casearlucin H	405507-92-8	30,892	0,560	132,040	5,78	1,653	5,78	1,651
82	Zguidonia 10	128486-35-1	3,889	0,328	112,100	6,12	0,757	5,77	1,714
83	Ctremula 4	173324-54-4	21,545	0,484	127,720	5,75	1,782	5,74	1,829
84	Zuelanin 11	130827-58-6	29,182	0,712	103,820	5,76	1,723	5,73	1,847
85	Zuelanin 6	130827-54-2	16,411	0,642	102,150	5,27	5,365	5,72	1,889
86	Laetiaprocerin B	871217-63-9	29,954	0,678	111,690	5,73	1,860	5,71	1,948
87	Corymbotin A	138967-76-7	18,050	0,680	99,605	5,24	5,767	5,66	2,177
88	Casearlucin B	405507-80-4	17,986	0,634	106,710	5,26	5,500	5,65	2,247
89	Corymbotin G	138950-32-0	15,416	0,600	109,890	5,15	7,052	5,60	2,506
90	Ctremula 6	173324-55-5	27,247	0,538	135,470	5,48	3,320	5,58	2,642
91	Zuelanin 7	130827-55-3	17,567	0,567	116,940	5,27	5,409	5,58	2,643
92	Casearlucin K	405507-98-4	25,316	0,543	132,040	5,43	3,712	5,55	2,843
93	Caseamembrin C	664351-47-7	27,330	0,532	138,430	5,40	4,008	5,54	2,908
94	Corymbotin E	138950-30-8	19,674	0,632	111,570	5,20	6,318	5,52	3,006
95	Ccorymbosa 9	131740-60-8	21,180	0,611	116,430	5,27	5,400	5,52	3,038
96	Ccorymbosa 10	131740-61-9	20,006	0,616	114,660	5,21	6,180	5,50	3,129
97	Caseamembrin M	870538-04-8	26,275	0,604	124,960	5,37	4,246	5,49	3,218
98	Caseamembrin D	664351-49-9	28,147	0,554	138,520	5,30	5,052	5,48	3,299
99	Caseamembrin B	723302-02-1	25,752	0,529	138,100	5,30	5,025	5,48	3,347
100	Caseamembrin F	664351-51-3	13,285	0,455	128,840	5,21	6,210	5,46	3,466
101	Caseargrewiin C	847841-94-5	20,266	0,554	125,010	5,23	5,869	5,46	3,468
102	Casearnigrescens B	925693-17-0	19,790	0,542	126,840	5,20	6,334	5,43	3,683
103	Caseamembrin O	870538-07-1	26,524	0,587	132,110	5,20	6,343	5,38	4,198
104	Zuelanin 12	130827-59-7	23,424	0,580	128,360	5,15	7,013	5,36	4,347
105	Corymbotin B	138950-28-4	17,570	0,667	109,030	4,87	13,543	5,33	4,698
106	Caseamembrin N	870538-05-9	24,144	0,591	128,970	5,11	7,852	5,32	4,827
107	Casearlucin J	405507-96-2	21,654	0,552	132,040	5,05	8,978	5,32	4,833
108	Casearnigrescens C	925693-19-2	25,321	0,614	126,930	5,12	7,505	5,30	4,984
109	Casearnigrescens A	925693-15-8	20,926	0,587	126,930	4,95	11,198	5,25	5,662
110	Casearin T	220178-33-6	32,143	0,715	119,330	5,29	5,111	5,22	5,957
111	Corymbotin I	138950-33-1	13,696	0,617	116,740	4,56	27,315	5,20	6,292
112	Corymbotin H	138967-77-8	18,419	0,641	119,320	4,67	21,321	5,10	8,009
113	Corymbotin C	138950-29-5	17,340	0,666	117,850	4,46	35,035	4,98	10,529

(a)Y=Log(1/IC<sub>50</sub>); (b) μmol/L.

**Tabela 8.** Valores preditos IC<sub>50</sub> da Casearina G substituída em R<sup>1</sup>, R<sup>4</sup> e R<sup>5</sup> contra células tumorais V79 com QSAR-MLR Eq. (5.1-7) e QSAR-SVM (n=18)

Item	Substituintes					Composto	Descritores			Métodos empregados			
	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	R <sup>5</sup>		RDF060m	R8p	Fuk5	Y <sup>b</sup>	IC50 <sup>c</sup>	Y <sup>b</sup>	IC50 <sup>c</sup>
1	OMe	Ac	Ac	H	OBu	Casearina G <sup>a</sup>	31,047	0,610	105,67	6,69	0,20	6,68	0,21
2	OH	Ac	Ac	Cl	H	Gv8	31,047	0,610	105,67	6,69	0,20	6,68	0,21
3	OH	Ac	Ac	H	Cl	Gv10	31,794	0,583	114,28	6,55	0,28	6,51	0,31
4	OH	Ac	Ac	F	H	Gv9	27,228	0,580	108,22	6,45	0,35	6,41	0,39
5	OAc	Ac	Ac	H	OMe	Gv3	29,562	0,607	111,38	6,31	0,49	6,30	0,50
6	F	Ac	Ac	H	OBu	Gv12	37,110	0,634	121,87	6,29	0,52	6,28	0,53
7	OMe	Ac	Ac	H	F	Gv5	35,602	0,576	132,50	6,07	0,85	6,11	0,78
8	OH	Ac	Ac	H	F	Gv11	31,524	0,568	125,30	6,11	0,78	6,07	0,85
9	OMe	Ac	Ac	H	ODc	Gv1	29,874	0,617	119,23	5,90	1,26	5,87	1,34
10	Cl	Ac	Ac	H	OBu	Gv13	27,584	0,596	119,23	5,85	1,42	5,80	1,58
11	OH	Ac	Ac	H	ODc	Gv2	28,914	0,579	126,07	5,77	1,72	5,74	1,82
12	OMe	Ac	Ac	Cl	H	Gv6	24,939	0,573	127,84	5,77	1,72	5,74	1,82
13	OMe	Ac	Ac	F	H	Gv7	19,374	0,506	129,19	5,31	4,95	5,33	4,68
14	OMe	Ac	Ac	H	Cl	Gv4	28,128	0,702	119,33	5,13	7,44	5,12	7,52

(a) Referência: Casearina G maior IC<sub>50</sub> observado: 0,17 µmol/L, [pIC<sub>50</sub>=Log (1/IC<sub>50</sub>) = 6,77]; (b) Y=Log(1/IC<sub>50</sub>); (c) µmol/L.

**Tabela 11a.** Valores preditos  $IC_{50}$  de diterpenóides clerodanos contra células tumorais V79 com os modelos de QSAR-MLR Eq. (5.1-7) e QSAR-SVM ( $n = 18$ )

Item	Composto	CAS	Descritores			Métodos e valor predito			
			RDF060m	R8p	Fuk5	Y <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	Y <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>
						MLR	MLR	SVM	SVM
1	Casearborin A	272771-14-9	25,294	0,53	91,40	7,41	0,039	7,03	0,094
2	Casearborin B	272771-15-0	26,967	0,51	100,84	7,28	0,052	7,00	0,099
3	Casearborin D	272771-17-2	29,539	0,55	104,96	7,05	0,090	6,93	0,117
4	Caseargrewiin H	945607-74-9	24,403	0,51	98,20	7,18	0,066	6,87	0,135
5	Casearinol A	196090-44-5	29,888	0,58	103,84	6,89	0,128	6,84	0,144
6	Caseargrewiin E	945607-68-1	24,030	0,54	94,66	7,07	0,085	6,83	0,148
7	Casearia S	220178-32-5	25,802	0,57	97,51	6,90	0,125	6,79	0,164
8	Caseargrewiin G	945607-72-7	27,365	0,56	103,03	6,85	0,141	6,76	0,173
9	Zguidonia 6	128486-33-9	32,410	0,61	107,48	6,75	0,176	6,74	0,182
10	Caseargrewiin D	847841-95-6	19,575	0,50	88,50	7,28	0,052	6,72	0,190
11	Casearborin E	272771-18-3	34,830	0,65	105,06	6,74	0,184	6,71	0,195
12	Caseargrewiin I	945607-76-1	27,052	0,50	111,76	6,85	0,143	6,63	0,235
13	Cgrewiifolia 1	947672-98-2	25,597	0,56	102,64	6,70	0,199	6,61	0,243
14	Casearinol B	196090-45-6	29,070	0,55	111,70	6,68	0,209	6,59	0,256
15	Casearborin C	272771-16-1	30,066	0,62	104,96	6,56	0,276	6,56	0,275
16	Corymbotin F	138950-31-9	22,858	0,58	95,44	6,65	0,225	6,56	0,278
17	Corymbotin D	139015-30-8	20,387	0,59	88,59	6,70	0,201	6,53	0,295
18	Zuelanin 4	130858-02-5	20,941	0,59	90,26	6,67	0,216	6,53	0,296
19	Laeciaprocerin C	871217-64-0	34,469	0,63	113,47	6,52	0,304	6,51	0,313
20	Zguidonia 8	128486-52-3	15,131	0,51	83,11	7,10	0,080	6,48	0,328
21	Intrapetacin B	343944-36-5	18,195	0,50	96,28	6,78	0,165	6,46	0,345
22	Corymbotin E	138950-30-8	21,300	0,57	96,37	6,57	0,270	6,46	0,345
23	Corymbolusin_B	326473-52-3	30,738	0,54	118,84	6,55	0,279	6,45	0,353
24	Cgrewiifolia 4	1006868-61-6	24,892	0,57	103,65	6,50	0,313	6,45	0,357
25	Zuelanin 2	128486-32-8	27,456	0,57	109,64	6,51	0,312	6,44	0,360
26	Caseargrewiin J	945607-78-3	23,116	0,49	109,12	6,70	0,198	6,44	0,361
27	Ctremula 3	173324-52-2	22,574	0,52	105,42	6,62	0,240	6,44	0,367
28	Cgrayi 1	947672-97-1	21,581	0,50	106,01	6,65	0,224	6,41	0,392
29	Zguidonia 13	128572-39-4	29,332	0,60	109,73	6,39	0,405	6,38	0,419
30	Caseargrewiin L	945607-81-8	26,523	0,56	109,93	6,44	0,364	6,37	0,429
31	Corymbotin A	138967-76-7	25,107	0,63	99,61	6,33	0,473	6,34	0,459
32	Corymbotin C	138950-29-5	27,647	0,62	106,46	6,32	0,480	6,32	0,483
33	Zguidonia 7	128486-52-2	11,231	0,52	76,01	7,00	0,101	6,31	0,487
34	Corymbotin B	138950-28-4	26,972	0,60	107,38	6,33	0,466	6,31	0,487
35	Ctremula 2	173324-51-1	23,528	0,56	105,48	6,37	0,426	6,30	0,503
36	Intrapetacin A	343944-35-4	17,400	0,50	100,07	6,54	0,288	6,29	0,510
37	Zuelanin 3	128572-38-3	32,480	0,61	116,84	6,31	0,495	6,29	0,516
38	Zuelanin 5	40185-32-8	26,629	0,57	110,99	6,32	0,480	6,26	0,548

(a)  $Y = pIC_{50} = \text{Log}(1/IC_{50})$ ; (b)  $\mu\text{mol/L}$

**Tabela 11b.** Valores preditos IC<sub>50</sub> de diterpenóides clerodanos contra células tumorais V79 com os modelos de QSAR-MLR Eq. (5.1-7) e QSAR-SVM (n = 18)

Item	Composto	CAS	Descritores			Métodos e valor predito			
			RDF060m	R8p	Fuk5	Y <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	Y <sub>a</sub>	IC <sub>50</sub> <sup>b</sup>
39	Casearinone B	196090-47-8	17,969	0,59	93,28	6,31	0,495	6,25	0,557
40	Laeciaprocerin B	871217-63-9	34,965	0,68	111,69	6,27	0,538	6,25	0,559
41	Caseargrewiin C	847841-94-5	32,114	0,55	125,01	6,31	0,487	6,25	0,568
42	Zuelanin 12	130827-59-7	27,313	0,59	110,92	6,28	0,523	6,24	0,572
43	Cgrayi 7	947673-00-9	12,336	0,52	88,67	6,52	0,304	6,24	0,580
44	Cgrayi 6	947672-99-3	15,053	0,52	95,45	6,45	0,354	6,24	0,581
45	Zguidonia 12	128486-37-3	27,008	0,64	103,92	6,22	0,607	6,23	0,584
46	Zuelanin 1	128486-31-7	21,070	0,50	109,64	6,45	0,356	6,23	0,595
47	Casearlucin F	405507-88-2	18,990	0,56	100,56	6,26	0,548	6,18	0,658
48	Caseargrewiin F	945607-70-5	22,251	0,54	108,22	6,29	0,507	6,18	0,659
49	Caseargrewiin K	945607-80-7	26,944	0,49	122,68	6,38	0,418	6,18	0,665
50	Corymbolusin_A	326473-51-2	24,464	0,47	120,19	6,44	0,366	6,17	0,681
51	Zuelanin 11	130827-58-6	23,077	0,60	103,82	6,18	0,663	6,16	0,690
52	Laeciaprocerin A	871217-62-8	23,846	0,57	109,16	6,21	0,616	6,14	0,723
53	Cgrayi 8	947673-01-0	7,541	0,31	64,53	8,76	0,002	6,13	0,739
54	Caseargrewiin B	847841-93-4	35,399	0,59	130,50	6,07	0,853	6,10	0,794
55	Ctremula 6	173324-55-5	24,926	0,47	122,58	6,34	0,454	6,10	0,800
56	Ccorymbosa 4	131740-57-3	3,873	0,31	87,14	7,36	0,043	6,10	0,802
57	Corymbotin I	138950-33-1	24,744	0,59	109,00	6,13	0,739	6,09	0,808
58	Cgrewiifolia 2	947681-66-5	15,902	0,57	95,87	6,12	0,764	6,09	0,809
59	Ctremula 4	173324-53-3	21,409	0,47	116,26	6,37	0,429	6,09	0,809
60	Cgrewiifolia 3	1006868-59-2	12,561	0,52	95,77	6,18	0,660	6,07	0,857
61	Zuelanin 6	130827-54-2	2,792	0,36	61,25	8,13	0,007	6,07	0,859
62	Caseamembrin A	664351-45-5	28,559	0,53	124,87	6,17	0,672	6,06	0,869
63	Corymbotin H	138967-77-8	23,935	0,58	109,93	6,12	0,763	6,06	0,873
64	Ccorymbosa 11	131749-55-1	12,351	0,52	95,77	6,16	0,688	6,06	0,879
65	Corymbotin G	138950-32-0	20,864	0,60	102,15	6,03	0,925	6,04	0,917
66	Ccorymbosa 12	131830-93-8	16,457	0,59	95,87	6,00	1,000	6,03	0,925
67	Zguidonia 9	128486-34-0	26,797	0,56	117,13	6,11	0,778	6,03	0,932
68	Casearvestrin_C	402935-06-2	20,089	0,51	111,76	6,15	0,702	5,99	1,018
69	Cgrewiifolia 5	1006868-63-8	16,770	0,52	105,42	6,07	0,843	5,97	1,078
70	Corymbolusin_C	326473-53-4	25,499	0,54	118,84	6,07	0,846	5,96	1,091
71	Zguidonia 10	128486-35-1	10,111	0,34	112,10	6,48	0,333	5,96	1,105
72	Casearinone A	196090-46-7	14,160	0,58	95,54	5,87	1,345	5,94	1,140
73	Casearlucin G	405507-90-6	26,157	0,60	113,14	5,97	1,061	5,93	1,164
74	Ccorymbosa 8	131740-59-5	18,156	0,57	103,65	5,94	1,145	5,92	1,194
75	Ccorymbosa 5	131766-54-6	10,736	0,52	97,67	5,92	1,192	5,92	1,209
76	Cgrayi 4	947672-96-0	13,951	0,52	102,55	5,98	1,039	5,92	1,209

(a)  $Y = \text{pIC}_{50} = \text{Log}(1/\text{IC}_{50})$ ; (b)  $\mu\text{mol/L}$ .

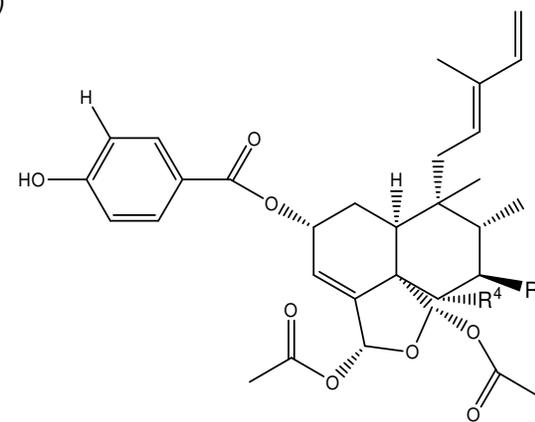
**Tabela 11c.** Valores preditos IC<sub>50</sub> de diterpenóides clerodanos contra células tumorais V79 com os modelos de QSAR-MLR Eq. (5.1-7) e QSAR-SVM (n = 18)

Item	Composto	CAS	Descritores			Métodos e valor predito			
			RDF060m	R8p	Fuk5	Y <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	Y <sub>a</sub>	IC <sub>50</sub> <sup>b</sup>
77	Caseamembrin B	664351-46-6	23,370	0,55	116,16	5,94	1,149	5,84	1,436
78	Ccorymbosa 13	131740-56-2	18,140	0,58	105,42	5,81	1,538	5,81	1,546
79	Zuelanin 8	130858-01-4	24,772	0,68	103,82	5,76	1,743	5,81	1,559
80	Caseamembrin E	664351-50-2	24,923	0,52	124,87	5,91	1,229	5,79	1,613
81	Caseamembrin A	723302-01-0	24,923	0,52	124,87	5,91	1,229	5,79	1,613
82	Zuelanin 13	130827-60-0	24,634	0,68	103,82	5,74	1,823	5,79	1,623
83	Zuelanin 10	130827-57-5	17,990	0,50	115,30	5,91	1,229	5,76	1,718
84	Caseargrewiin A	847841-92-3	23,890	0,56	117,68	5,83	1,464	5,75	1,773
85	Ctremula 5	173324-54-4	21,429	0,46	127,89	5,91	1,217	5,74	1,838
86	Zuelanin 9	130827-56-4	16,264	0,48	115,30	5,87	1,359	5,73	1,877
87	CasearLucin B	405507-80-4	17,944	0,62	102,72	5,64	2,279	5,72	1,902
88	Casearvestrin_A	402937-50-2	19,118	0,57	109,93	5,74	1,802	5,71	1,949
89	Casearvestrin_B	402935-05-1	19,317	0,56	111,70	5,74	1,837	5,69	2,064
90	Laeciaprocerin D	871217-65-1	11,670	0,41	119,20	5,80	1,568	5,68	2,073
91	Zuelanin 7	130827-55-3	15,523	0,61	102,15	5,52	3,012	5,65	2,248
92	CasearLucin A	405507-78-0	22,483	0,59	114,12	5,66	2,211	5,61	2,449
93	Caseamembrin C	664351-47-7	27,330	0,53	138,43	5,38	4,147	5,53	2,949
94	Casearnigrescens A	925693-15-8	28,744	0,60	128,97	5,48	3,279	5,51	3,083
95	Caseamembrin D	664351-49-9	28,147	0,55	138,52	5,29	5,112	5,48	3,308
96	Casearnigrescens D	925693-22-7	16,155	0,54	115,32	5,47	3,409	5,46	3,466
97	Caseamembrin B	723302-02-1	25,752	0,53	138,10	5,28	5,240	5,44	3,642
98	CasearLucin C	405507-82-6	25,209	0,59	125,43	5,41	3,869	5,39	4,078
99	Caseamembrin M	870538-04-8	26,275	0,60	124,96	5,41	3,928	5,39	4,088
100	Casearia T	220178-33-6	32,143	0,72	119,33	5,39	4,096	5,39	4,107
101	Casearnigrescens C	925693-19-2	26,010	0,58	128,97	5,36	4,374	5,38	4,206
102	Casearnigrescens B	925693-17-0	23,128	0,55	128,88	5,35	4,491	5,35	4,466
103	Caseamembrin F	664351-51-3	13,285	0,46	128,84	5,15	7,015	5,33	4,638
104	CasearLucin D	405507-84-8	22,073	0,59	120,67	5,35	4,428	5,32	4,806
105	Ccorymbosa 9	131740-60-8	21,180	0,61	116,43	5,31	4,904	5,29	5,087
106	CasearLucin J	405507-96-2	26,206	0,56	143,71	4,85	13,978	5,29	5,172
107	Ccorymbosa 10	131740-61-9	20,006	0,62	114,66	5,25	5,576	5,26	5,452
108	Caseamembrin N	870538-05-9	24,144	0,59	128,97	5,12	7,532	5,17	6,779
109	CasearLucin K	405507-98-4	24,384	0,56	143,71	4,67	21,599	5,16	6,921
110	Caseamembrin O	870538-07-1	26,442	0,61	132,11	5,01	9,739	5,13	7,367
111	CasearLucin H	405507-92-8	23,122	0,57	133,81	4,98	10,520	5,13	7,433
112	CasearLucin E	405507-86-0	23,239	0,60	127,47	5,07	8,492	5,11	7,801
113	CasearLucin I	405507-94-0	20,581	0,56	133,81	4,78	16,742	4,99	10,146

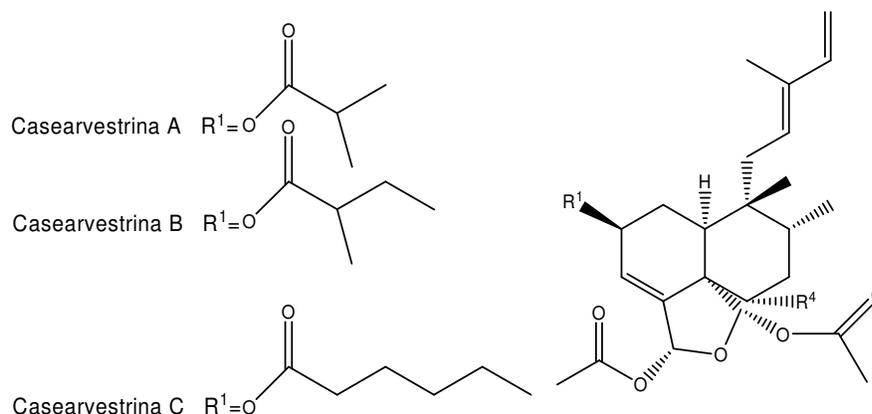
(a)  $Y = pIC_{50} = \text{Log}(1/IC_{50})$ ; (b)  $\mu\text{mol/L}$ .

**Tabela 14.** Predição de compostos derivados da Casearborina A substituída em R<sup>4</sup> e R<sup>5</sup>

Item	MolID*	Descritores			QSAR-MLR		QSAR-SVM		Substituintes em R <sup>4</sup> e R <sup>5</sup> e Estrutura
		RDF060m	R8p	Fuk5	pIC <sub>50</sub> <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	pIC <sub>50</sub> <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	
1	Casearborina-A	25.294	0,533	91,4	7,41	0,039	7,03	0,094	R <sup>4</sup> , R <sup>5</sup> (=H)
2	AR <sup>4</sup> -5	23.993	0,539	93,5	7,16	0,070	6,87	0,135	F, H
3	AR <sup>5</sup> -2	31.339	0,621	105,1	6,67	0,216	6,66	0,217	H, OAc
4	AR <sup>5</sup> -5	21.753	0,551	93,5	6,87	0,134	6,66	0,221	H, F
5	AR <sup>4</sup> -1	22.599	0,572	95,6	6,71	0,196	6,59	0,260	Me, H
6	AR <sup>4</sup> -2	30.284	0,62	105,1	6,58	0,262	6,58	0,262	OAc, H
7	AR <sup>4</sup> -4	25.982	0,582	102,4	6,60	0,251	6,55	0,279	OMe, H
8	AR <sup>5</sup> -4	26.061	0,586	102,4	6,58	0,264	6,54	0,289	H, OMe
9	AR <sup>4</sup> -3	20.551	0,543	98,18	6,60	0,249	6,45	0,358	=O, H
10	AR <sup>5</sup> -3	20.661	0,549	98,27	6,57	0,271	6,43	0,373	H, =O
11	AR <sup>5</sup> -1	18.644	0,585	95,36	6,27	0,541	6,23	0,590	H, Me
12	AR <sup>4</sup> -6	26.477	0,563	113,7	6,25	0,562	6,17	0,670	Cl, H
13	AR <sup>5</sup> -6	25.368	0,601	113,7	5,88	1,322	5,83	1,462	H, Cl

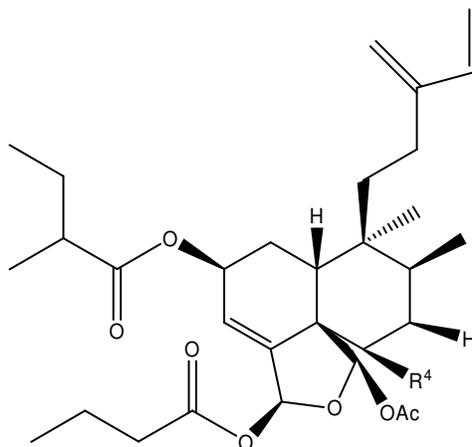


(\*) Composto derivado AR<sup>4</sup> ou AR<sup>5</sup> e número seqüencial; **(a)** Log(1/IC<sub>50</sub>); **(b)** µmol/L.

**Tabela 15.** Predição de compostos substituídos derivados das Casearvestrinas A, B e C

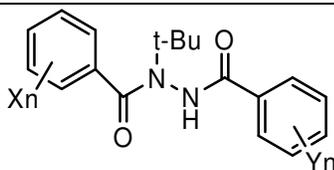
Item	MolID <sup>*</sup>	Descritores			QSAR-MLR		QSAR-SVM		Subs. R <sup>4</sup>
		RDF060m	R8p	Fuk5	pIC <sub>50</sub> <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	pIC <sub>50</sub> <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	
1	<b>Casearvestrina_A</b>	<b>19,118</b>	<b>0,568</b>	<b>109,9</b>	<b>5,74</b>	<b>1,802</b>	<b>5,71</b>	<b>1,949</b>	<b>OH</b>
2	CVA-6	29,005	0,620	98,46	6,78	0,166	6,75	0,177	F
3	CVA-1	25,452	0,601	96,37	6,70	0,198	6,65	0,226	H
4	CVA-2	24,971	0,611	100,3	6,40	0,397	6,39	0,403	Me
5	CVA-4	25,589	0,638	103,1	6,13	0,743	6,15	0,707	=O
6	CVA-3	35,272	0,709	110,0	6,14	0,717	6,11	0,769	OAc
7	CVA-5	26,540	0,667	107,4	5,80	1,571	5,82	1,519	OMe
8	CVA-7	27,041	0,634	118,7	5,55	2,792	5,52	3,027	Cl
9	<b>Casearvestrina-B</b>	<b>19,317</b>	<b>0,560</b>	<b>111,7</b>	<b>5,74</b>	<b>1,837</b>	<b>5,69</b>	<b>2,063</b>	<b>OH</b>
10	CVB-6	32,018	0,598	100,2	7,12	0,076	7,05	0,090	F
11	CVB-1	27,919	0,584	98,4	6,96	0,110	6,87	0,134	H
12	CVB-5	33,768	0,626	109,2	6,65	0,223	6,64	0,228	OMe
13	CVB-2	29,505	0,626	102,1	6,61	0,246	6,61	0,246	Me
14	CVB-4	23,309	0,606	104,9	6,08	0,841	6,07	0,854	=O
15	CVB-3	27,374	0,660	111,8	5,72	1,906	5,71	1,960	OAc
16	CVB-7	29,960	0,657	120,4	5,56	2,744	5,55	2,849	Cl
17	<b>Casearvestrina-C</b>	<b>20,089</b>	<b>0,511</b>	<b>111,8</b>	<b>6,15</b>	<b>0,702</b>	<b>5,99</b>	<b>1,018</b>	<b>OH</b>
18	CVC-3	42,027	0,674	111,9	6,90	0,125	6,74	0,183	OAc
19	CVC-2	28,651	0,599	102,2	6,73	0,188	6,70	0,200	Me
20	CVC-6	26,516	0,599	100,3	6,63	0,237	6,60	0,254	F
21	CVC-1	22,672	0,567	98,2	6,62	0,242	6,52	0,305	H
22	CVC-5	27,949	0,610	109,2	6,25	0,560	6,24	0,577	OMe
23	CVC-4	22,639	0,579	105,0	6,21	0,619	6,17	0,684	=O
24	CVC-7	30,365	0,624	120,5	5,83	1,472	5,81	1,545	Cl

(\*) Composto derivado e número seqüencial; (a) Log(1/IC<sub>50</sub>); (b) μmol/L

**Tabela 16.** Predição de compostos substituídos derivados da Caseamembrina A

Item	MoIID*	Descritores			QSAR-MLR		QSAR-SVM		Substituintes R <sup>4</sup>
		RDF060m	R8p	Fuk5	pIC <sub>50</sub> <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	pIC <sub>50</sub> <sup>a</sup>	IC <sub>50</sub> <sup>b</sup>	
1	<b>Caseamembrina-A</b>	<b>28,559</b>	<b>0,526</b>	<b>124,9</b>	<b>6,17</b>	<b>0,672</b>	<b>6,06</b>	<b>0,869</b>	OH
2	A-5	29,005	0,620	113,4	6,08	0,841	6,06	0,881	F
4	A-1	24,971	0,611	115,3	5,70	2,007	5,65	2,215	Me
5	A-2	35,272	0,709	125,0	5,44	3,626	5,50	3,183	OAc
6	A-3	25,589	0,638	118,1	5,42	3,763	5,39	4,069	=O
7	A-4	26,540	0,667	122,3	5,10	7,955	5,10	7,979	OMe
8	A-6	27,041	0,634	133,6	4,85	14,128	5,04	9,193	Cl

(\*) Composto derivado e número seqüencial: (a) Log(1/IC<sub>50</sub>); (b) µmol/L.

**Tabela 17.** QSAR de DBHs contra *Chilo suppressalis* antes e após análise conformacional

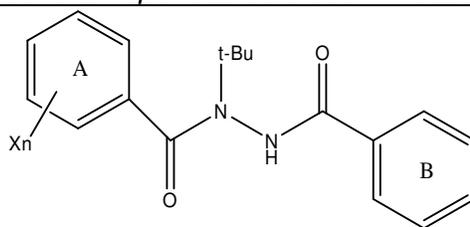
No.	MolID <sup>T,+</sup>	X <sub>n</sub>	Y <sub>n</sub>	Y <sub>obs</sub> <sup>a</sup>	Descritores			Y <sub>pred</sub> <sup>a,b</sup>	Res.c <sup>b</sup>	Y <sub>pred</sub> <sup>b,e</sup>	Res.c,d
					PCD	RNCG	H8u				
1	D1	2-Cl	H	6,83±0,36	31,437	0,160	0,534	6,19	0,64	6,82	0,01
2	D2	2-Cl	2-F	6,12	34,697	0,150	0,467	5,87	0,25	6,13	-0,01
3	D3	2-Cl	2-Cl	5,55	34,697	0,159	0,402	5,84	-0,29	5,64	-0,09
4	D4 <sup>T</sup>	2-Cl	2_Br	5,48±0,07	34,697	0,161	0,363	5,79	-0,31	5,38	0,10
5	D5	2-Cl	2-I	5,33±0,31	34,697	0,162	0,353	5,71	-0,38	5,32	0,01
6	D6	2-Cl	2-CF <sub>3</sub>	5,94	37,511	0,127	0,466	5,88	0,06	5,98	-0,04
7	D7	2-Cl	2-NO <sub>2</sub>	5,11±0,08	41,169	0,145	0,402	5,20	-0,09	5,07	0,04
8	D8	2-Cl	2-Me	6,41	34,697	0,158	0,511	5,92	0,49	6,35	0,06
9	D10 <sup>T</sup>	2-Cl	2-OMe	4,04±0,04	37,841	0,238	0,286	4,12	-0,08	4,04	0,00
10	D12	2-Cl	3-F	6,93	31,092	0,153	0,539	6,25	0,68	6,95	-0,02
11	D13	2-Cl	3-Cl	6,44±0,07	31,092	0,161	0,468	6,12	0,32	6,42	0,02
12	D14	2-Cl	3-Br	6,34±0,25	31,092	0,164	0,461	6,06	0,28	6,38	-0,04
13	D15	2-Cl	3-I	5,71±0,13	31,092	0,164	0,364	6,04	-0,33	5,74	-0,03
14	D16	2-Cl	3-CF <sub>3</sub>	6,09	32,964	0,128	0,414	6,07	0,02	6,09	0,00
15	D17 <sup>T</sup>	2-Cl	3-NO <sub>2</sub>	5,28	32,642	0,150	0,311	6,01	-0,73	5,31	-0,03
16	D18 <sup>T</sup>	2-Cl	3-CN	5,44	37,028	0,151	0,405	5,40	0,04	5,50	-0,06
17	D19	2-Cl	3-Me	5,78	31,092	0,158	0,369	5,97	-0,19	5,78	0,00
18	D20	2-Cl	3-OMe	5,80	33,440	0,244	0,495	5,25	0,55	5,84	-0,04
19	D22 <sup>T</sup>	2-Cl	4-F	6,68	31,032	0,154	0,494	6,45	0,23	6,66	0,02
20	D23 <sup>T</sup>	2-Cl	4-Cl	6,61±0,09	31,032	0,162	0,510	6,26	0,35	6,70	-0,09
21	D33	2-Cl	4-Bu-t	6,81±0,19	29,313	0,147	0,483	7,11	-0,30	6,76	0,05
22	D38	2-Cl	2,3-Me <sub>2</sub>	6,61±0,16	34,412	0,156	0,530	6,17	0,44	6,50	0,11
23	D41	2-Cl	2,5-Cl <sub>2</sub>	5,40±0,10	34,403	0,160	0,362	5,88	-0,48	5,41	-0,01
24	D43	2-Cl	2-Br-5-OMe <sub>2</sub>	4,24	33,972	0,246	0,265	4,69	-0,45	4,22	0,02
25	D48	2-Cl	3,4-Me <sub>2</sub>	6,09	30,794	0,156	0,418	6,23	-0,14	6,13	-0,04
26	D55	H	H	6,27	28,045	0,161	0,384	6,67	-0,40	6,17	0,10
27	D58	3,5-Cl <sub>2</sub>	4-Me	6,96	27,120	0,162	0,493	7,13	-0,17	6,97	-0,01
28	D61 <sup>T</sup>	3,5-Br <sub>2</sub>	4-Et	7,52	27,120	0,167	0,566	7,78	-0,26	7,52	0,00
29	D63 <sup>T</sup>	3,5-Me <sub>2</sub>	4-Cl	6,96	27,120	0,159	0,498	6,64	0,32	7,01	-0,05
30	D65	3,5-Me <sub>2</sub>	4-Et	7,32	26,405	0,151	0,523	7,38	-0,10	7,30	0,02

(T) Compostos do conjunto de teste; (+) compostos não incluídos no estudo com substituintes em Y<sub>n</sub>, e especificados na literatura<sup>24</sup> com pLD<sub>50</sub> < 4,00 ou < 5,00: 2-Ph; 2-SMe; 3-OPh; 4-Ph; 4-OPh; 2,6-Cl<sub>2</sub>; 3,5-(OMe)<sub>2</sub>; 3-OMe-5-OBu; 3,5-(OBu<sup>h</sup>)<sub>2</sub>; (a) Y=pLD<sub>50</sub> (LD<sub>50</sub> = mmol/inseto); (b) antes da análise conformacional; (c) resíduo = Y<sub>Obs</sub> - Y<sub>Pred</sub>; (d) após análise conformacional.

**Tabela 20.** Valores observados pLD<sub>50</sub>, preditos pelo QSAR e selecionados pelo índice de similaridade euclidiano

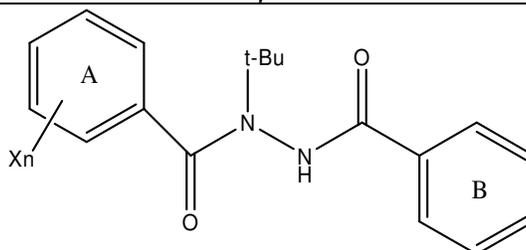
No.	MolID	Substituintes <sup>a</sup>		Y <sub>obs</sub> <sup>b</sup>	Descritores			Y <sub>pred</sub> <sup>b,c</sup>	res <sup>d</sup>	Y <sub>pred</sub> <sup>b,e</sup>	res <sup>d</sup>	Y <sub>sim</sub> <sup>b,f</sup>	res <sup>g</sup>
		X <sub>n</sub>	Y <sub>n</sub>		PCD	RNCG	H8u <sup>f</sup>						
1	D24	2-Cl	4-Br	6,88±0,29	31,032	0,164	0,542	6,42	0,46	6,88	0,00	6,61	0,27
2	D25	2-Cl	4-I	6,88±0,11	31,032	0,164	0,539	6,43	0,45	6,86	0,02	6,68	0,20
3	D26	2-Cl	4-CF <sub>3</sub>	6,79±0,04	29,313	0,128	0,466	7,17	-0,38	6,79	0,00	6,81	-0,02
4	D27	2-Cl	4-NO <sub>2</sub>	5,62	32,018	0,154	0,431	6,59	-0,97	6,13	-0,51	6,83	-1,21
5	D28	2-Cl	4-CN	5,32±0,08	32,784	0,152	0,497	6,50	-1,18	5,80	-0,48	5,80	-0,48
6	D29	2-Cl	4-Me	6,70	31,032	0,158	0,509	6,81	-0,11	6,71	-0,01	6,71	-0,01
7	D30	2-Cl	4-Et	6,64±0,34	30,266	0,153	0,478	7,28	-0,64	6,61	0,03	5,54	1,10
8	D31	2-Cl	4-Pr <sup>n</sup>	5,85	32,322	0,149	0,419	7,37	-1,52	6,05	-0,20	6,05	-0,20
9	D32	2-Cl	4-Pr <sup>i</sup>	6,25	29,730	0,150	0,426	7,85	-1,60	6,35	-0,10	6,73	-0,48
10	D35	2-Cl	4-OMe	6,15±0,02	30,266	0,245	0,503	6,01	0,14	6,17	-0,02	6,03	0,12
11	D37	2-Cl	2,3-Cl <sub>2</sub>	6,71±0,02	34,412	0,157	0,569	6,05	0,66	6,77	-0,06	6,49	0,22
12	D39	2-Cl	2,4-Cl <sub>2</sub>	6,59	34,388	0,160	0,542	6,12	0,47	6,58	0,01	6,50	0,09
13	D40	2-Cl	2,4-Me <sub>2</sub>	6,54	34,404	0,156	0,538	9,23	-2,69	6,58	-0,04	6,49	0,05
14	D42	2-Cl	2,5-Me <sub>2</sub>	5,20	34,403	0,156	0,409	5,73	-0,53	5,73	-0,53	6,52	-1,32
15	D44	2-Cl	2,6-F <sub>2</sub>	5,78±0,02	34,825	0,140	0,403	5,99	-0,21	5,76	0,02	5,54	0,24
16	D45	2-Cl	2-F-6-Cl	5,69	34,825	0,148	0,386	6,05	-0,36	5,59	0,10	5,50	0,19
17	D47	2-Cl	3,4-Cl <sub>2</sub>	6,47	30,805	0,160	0,469	6,91	-0,44	6,45	0,02	6,27	0,20
18	D49	2-Cl	3-Me-4-Pr <sup>n</sup>	6,21	32,298	0,147	0,443	6,95	-0,74	6,22	-0,01	6,08	0,13
19	D50	2-Cl	3,5-Cl <sub>2</sub>	5,95±0,10	30,843	0,163	0,399	6,18	-0,23	5,97	-0,02	6,08	-0,13
20	D51	2-Cl	3,5-Me <sub>2</sub>	5,98	30,843	0,156	0,307	5,42	0,56	5,42	0,56	6,16	-0,18
21	D56	3,5-Cl <sub>2</sub>	H	7,07	27,470	0,164	0,516	6,95	0,12	7,07	0,00	6,94	0,13
22	D57	3,5-Cl <sub>2</sub>	4-Cl	6,80	27,120	0,166	0,479	7,34	-0,54	6,84	-0,04	6,92	-0,12
23	D59	3,5-Cl <sub>2</sub>	4-Et	7,67	26,405	0,157	0,586	7,84	-0,17	7,67	0,00	7,24	0,43
24	D60	3,5-Br <sub>2</sub>	4-Me	6,89	27,120	0,167	0,491	8,09	-1,20	6,92	-0,03	7,01	-0,12
25	D62	3,5-Me <sub>2</sub>	H	6,43	27,470	0,157	0,412	6,69	-0,26	6,43	0,00	6,25	0,18
26	D64	3,5-Me <sub>2</sub>	4-Me	7,54	27,120	0,155	0,533	7,27	0,27	7,27	0,27	7,01	0,53
27	D66	3,5-Me <sub>2</sub>	4-Pr <sup>n</sup>	6,94	28,753	0,146	0,498	7,78	-0,84	6,94	0,00	6,35	0,59
28	D67	3,5-Me <sub>2</sub>	4-Pr <sup>i</sup>	7,32	25,931	0,153	0,546	7,70	-0,38	7,49	-0,17	7,52	-0,20

(a) Vide fig. Tabela 17 em Anexos; (b) pLD<sub>50</sub> (LD<sub>50</sub> = mmol/inseto); (c) antes da busca conformacional; (d) Y<sub>obs</sub> – Y<sub>pred</sub>;(e) após busca conformacional; (f) conformero selecionado por similaridade molecular; (g) resíduo = Y<sub>obs</sub> – Y<sub>sim</sub>.

**Tabela 21.** Compostos DBHs usados para desenvolvimento do modelo de QSAR de Dibenzoilidrazinas contra *Leptinotarsa decemlineata*

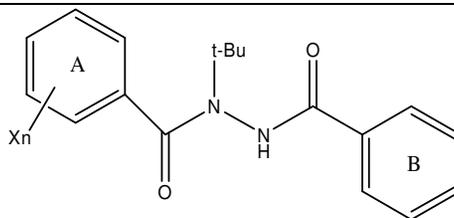
Item	MolID	X <sub>n</sub>	Y <sub>Obs</sub> <sup>a</sup>	V <sub>orto</sub> <sup>b</sup>	π <sup>meta</sup> <sup>b</sup>	ΔB <sub>5</sub> <sup>b</sup>	I <sub>2356</sub> <sup>b</sup>	Mor23p	Y <sub>Pred</sub> <sup>c</sup>	Res <sup>d</sup>	Res <sup>e</sup>
1	D1	H	5,34±0,07	0,00	0	0	0	-0,524	5,43	-0,09	-0,14
2	D2	2-F	5,52±0,04	0,33	0	0	0	-0,527	5,48	0,04	-0,01
3	D3	2-Cl	5,71±0,41	0,95	0	0	0	-0,582	5,75	-0,04	-0,06
4	D4 <sup>T</sup>	2-Br	5,94±0,11	1,26	0	0	0	-0,616	5,92	0,02	0,07
5	D5	2-I	6,06±0,08	1,71	0	0	0	-0,642	6,07	-0,01	0,14
6	D6	2-CF <sub>3</sub>	6,08±0,11	1,94	0	0	0	-0,632	6,06	0,02	0,09
7	D7 <sup>T</sup>	2-NO <sub>2</sub>	5,68±0,04	1,43	0	0	0	-0,565	5,75	-0,07	-0,13
8	D8	2-CH <sub>3</sub>	5,91±0,13	1,12	0	0	0	-0,612	5,89	0,02	0,26
9	D9 <sup>T</sup>	2-C <sub>2</sub> H <sub>5</sub>	5,77±0,14	2,14	0	0	0	-0,546	5,76	0,01	-0,03
10	D10	2-OCH <sub>3</sub>	5,70±0,11	1,44	0	0	0	-0,545	5,67	0,03	-0,12
11	D11	2-OBu <sup>sec</sup>	6,31±0,09	4,50	0	0	0	-0,628	6,34	-0,03	0,05
12	D12	2-SCH <sub>3</sub>	5,75±0,11	2,20	0	0	0	-0,528	5,70	0,05	-0,16
13	D13	3-F	6,19±0,16	0,00	0,33	0	0	-0,673	6,19	0,00	0,65
14	D14	3-Br	6,14±0,08	0,00	1,04	0	0	-0,530	6,10	0,04	0,21
15	D15	3-I	5,71±0,07	0,00	1,27	0	0	-0,394	5,73	-0,02	-0,26
16	D16	3-CF <sub>3</sub>	6,02±0,08	0,00	1,16	0	0	-0,491	6,03	-0,01	-0,07
17	D17	3-CN	5,18±0,05	0,00	0	0	0	-0,459	5,19	-0,01	-0,17
18	D18	3-CH <sub>3</sub>	6,08±0,21	0,00	0,34	0	0	-0,619	5,99	0,09	0,50
19	D19 <sup>T</sup>	4-Cl	4,49±0,09	0,00	0	0,80	0	-0,522	4,48	0,01	-0,08
20	D20 <sup>T</sup>	4-Br	4,04±0,04	0,00	0	0,95	0	-0,427	3,95	0,09	-0,39
21	D21	4-I	4,32±0,08	0,00	0	1,15	0	-0,585	4,30	0,02	0,19
22	D22	4-CF <sub>3</sub>	3,44±0,09	0,00	0	1,61	0	-0,508	3,47	-0,03	-0,25
23	D23	4-CN	4,22±0,00	0,00	0	0,60	0	-0,480	4,56	-0,34	-0,54
24	D24	4-NO <sub>2</sub>	4,46±0,04	0,00	0	0,70	0	-0,455	4,35	0,11	-0,12
25	D25	4-CH <sub>3</sub>	3,88±0,18	0,00	0	1,04	0	-0,411	3,78	0,10	-0,45
26	D27 <sup>T</sup>	4-Ph	3,59±0,04	0,00	0	2,11	0	-0,698	3,59	0,00	0,35
27	D28	4-OCH <sub>3</sub>	3,53±0,03	0,00	0	2,07	0	-0,681	3,57	-0,04	0,32
28	D30	2,3-Cl <sub>2</sub>	5,12±0,04	0,95	0,83	0	1	-0,556	5,14	-0,02	-0,04
29	D44	2,5-Cl <sub>2</sub> ,3-CF <sub>3</sub>	5,13±0,04	0,95	1,16	0	1	-0,508	5,16	-0,03	-0,20
30	D45	2,3,4,5-F <sub>4</sub>	4,77±0,03	0,33	0,33	0,35	1	-0,659	4,72	0,05	0,39

(T) Compostos do conjunto de teste; (a) Y<sub>Obs</sub> = pLD<sub>50-Observado</sub> (LD<sub>50</sub> = mmol/inseto); (b) valores da ref. [25]; (c) Y<sub>pred</sub> = pLD<sub>50-Predito</sub>; (d) Res = Y<sub>Obs</sub> - Y<sub>Pred</sub> após análise conformacional; (e) Res = Y<sub>Obs</sub> - Y<sub>Pred</sub> antes da análise conformacional.

**Tabela 22.** Comparativo de resultados obtidos entre modelos de QSAR-MLR e QSAR-SVM Dibenzoilidrazinas contra *Leptinotarsa decemlineata*

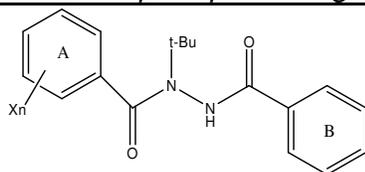
Item	MolID	X <sub>n</sub>	pLD <sub>50</sub> <sup>a</sup>	V <sub>ortho</sub> <sup>b</sup>	π <sup>meta</sup> <sup>b</sup>	ΔB <sub>5</sub> <sup>b</sup>	I <sub>2356</sub> <sup>b</sup>	Mor23p <sup>c</sup>	Y <sub>-MLR</sub> <sup>d</sup>	Y <sub>-SVM</sub> <sup>e</sup>	Δ <sup>f</sup>
1	D1	H	5,34±0,07	0	0	0	0	-0,524	5,43	5,41	0,02
2	D2	2-F	5,52±0,04	0,33	0	0	0	-0,527	5,48	5,47	0,01
3	D3	2-Cl	5,71±0,41	0,95	0	0	0	-0,582	5,75	5,76	-0,01
4	D4	2-Br	5,94±0,11	1,26	0	0	0	-0,616	5,92	5,93	-0,01
5	D5	2-I	6,06±0,08	1,71	0	0	0	-0,642	6,07	6,08	-0,01
6	D6	2-CF <sub>3</sub>	6,08±0,11	1,94	0	0	0	-0,632	6,06	6,08	-0,02
7	D7	2-NO <sub>2</sub>	5,68±0,04	1,43	0	0	0	-0,565	5,75	5,77	-0,03
8	D8	2-CH <sub>3</sub>	5,91±0,13	1,12	0	0	0	-0,612	5,89	5,89	-0,01
9	D9	2-C <sub>2</sub> H <sub>5</sub>	5,77±0,14	2,14	0	0	0	-0,546	5,76	5,81	-0,05
10	D10	2-OCH <sub>3</sub>	5,70±0,11	1,44	0	0	0	-0,545	5,67	5,70	-0,03
11	D11	2-OBu <sup>sec</sup>	6,31±0,09	4,50	0	0	0	-0,628	6,34	6,31	0,03
12	D12	2-SCH <sub>3</sub>	5,75±0,11	2,20	0	0	0	-0,528	5,70	5,74	-0,05
13	D13	3-F	6,19±0,16	0	0,33	0	0	-0,673	6,19	6,19	0,00
14	D14	3-Br	6,14±0,08	0	1,04	0	0	-0,530	6,10	6,14	-0,04
15	D15	3-I	5,71±0,07	0	1,27	0	0	-0,394	5,73	5,71	0,02
16	D16	3-CF <sub>3</sub>	6,02±0,08	0	1,16	0	0	-0,491	6,03	6,02	0,00
17	D17	3-CN	5,18±0,05	0	0	0	0	-0,459	5,19	5,18	0,01
18	D18	3-CH <sub>3</sub>	6,08±0,21	0	0,34	0	0	-0,619	5,99	6,07	-0,07
19	D19	4-Cl	4,49±0,09	0	0	0,80	0	-0,522	4,48	4,49	-0,01
20	D20	4-Br	4,04±0,04	0	0	0,95	0	-0,427	3,95	4,02	-0,07
21	D21	4-I	4,32±0,08	0	0	1,15	0	-0,585	4,30	4,27	0,03
22	D22	4-CF <sub>3</sub>	3,44±0,09	0	0	1,61	0	-0,508	3,47	3,44	0,03
23	D23	4-CN	4,22±0,00	0	0	0,60	0	-0,480	4,56	4,59	-0,04
24	D24	4-NO <sub>2</sub>	4,46±0,04	0	0	0,70	0	-0,455	4,35	4,39	-0,05
25	D25	4-CH <sub>3</sub>	3,88±0,18	0	0	1,04	0	-0,411	3,78	3,88	-0,10
26	D27	4-PH	3,59±0,04	0	0	2,11	0	-0,698	3,59	3,59	0,00
27	D28	4-OCH <sub>3</sub>	3,53±0,03	0	0	2,07	0	-0,681	3,57	3,56	0,01
28	D30	2,3-Cl <sub>2</sub>	5,12±0,04	0,95	0,83	0	1	-0,556	5,14	5,12	0,01
29	D44	2,5-Cl <sub>2</sub> ,3-CF <sub>3</sub>	5,13±0,04	0,95	1,16	0	1	-0,508	5,16	5,13	0,03
30	D45	2,3,4,5-F <sub>4</sub>	4,77±0,03	0,33	0,33	0,35	1	-0,659	4,72	4,77	-0,05

(a) Y<sub>Obs</sub> = pLD<sub>50</sub>-Observado (LD<sub>50</sub> = mmol/inseto); (b) descritores constantes; (c) descritor variável com a conformação; (d) Y<sub>MLR</sub> = pLD<sub>50</sub>-Predito; (e) Y<sub>SVM</sub> = pLD<sub>50</sub>-Predito; (f) Δ = Y<sub>MLR</sub> - Y<sub>SVM</sub>

**Tabela 25.** Conjunto de compostos (pseudo *blind set*) simulando a proposta de novos compostos contra *Leptinotarsa decemlineata*

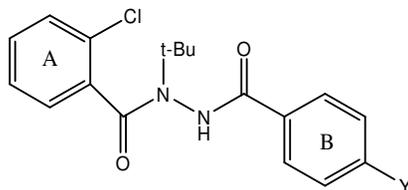
Item	MolID	Xn	Y <sub>Obs</sub> <sup>a</sup>	Y <sub>Pred-MLR</sub> <sup>b</sup>	res <sup>c,*</sup>	Y <sub>Pred-SVM</sub> <sup>d</sup>	res <sup>e,*</sup>	QSAR-MLR <sup>b</sup> Sim <sup>f</sup>	res <sup>g,*</sup>	QSAR-SVM <sup>d</sup> Sim <sup>f</sup>	res <sup>h</sup>
1	D26	4-Bu <sup>tert</sup>	3,71±0,04	3,73	-0,02	3,77	-0,06	3,61	0,10	3,64	0,07
2	D29	4-O(CH <sub>2</sub> ) <sub>3</sub> C <sub>6</sub> H <sub>5</sub>	3,35±0,05	2,18	1,17	3,33	0,02	2,18	1,17	3,32	0,03
3	D31	2,4-Cl <sub>2</sub>	5,22±0,06	5,22	0,00	5,26	-0,04	4,81	0,41	4,86	0,36
4	D32	2,5-Cl <sub>2</sub>	4,55±0,05	4,70	-0,15	4,93	-0,38	4,84	-0,29	4,99	-0,44
5	D33	3,5-Cl <sub>2</sub>	4,87±0,08	4,95	-0,08	5,02	-0,15	4,95	-0,08	5,02	-0,15
6	D34	2,3-(CH <sub>3</sub> ) <sub>2</sub>	4,46±0,06	4,47	-0,01	4,73	-0,27	4,47	-0,01	4,73	-0,27
7	D35	2,4-(CH <sub>3</sub> ) <sub>2</sub>	4,45±0,06	4,43	0,02	4,46	-0,01	4,49	-0,04	4,53	-0,08
8	D36	2,5-(CH <sub>3</sub> ) <sub>2</sub>	4,54±0,05	4,52	0,02	4,75	-0,21	3,29	1,25	4,54	0,00
9	D37	3,4-(CH <sub>3</sub> ) <sub>2</sub>	4,65±0,06	4,61	0,04	4,72	-0,07	4,46	0,19	4,58	0,07
10	D3	3,5-(CH <sub>3</sub> ) <sub>2</sub>	4,67±0,02	4,66	0,01	4,77	-0,10	4,66	0,01	4,77	-0,10
11	D39	2-CH <sub>2</sub> ,3-Cl	4,96±0,12	4,95	0,01	5,04	-0,08	4,83	0,13	4,99	-0,03
12	D40	2,6-F <sub>2</sub>	4,61±0,02	4,51	0,10	4,62	-0,01	4,91	-0,30	4,82	-0,21
13	D41	2-F,6-Cl	4,92±0,01	4,93	-0,01	4,87	0,05	5,18	-0,26	4,99	-0,07
14	D42	2-OCH <sub>3</sub> ,5Pr <sup>n</sup>	4,72±0,03	4,70	0,02	4,93	-0,21	4,83	-0,11	4,97	-0,25
15	D43	2,3,4-Cl <sub>3</sub>	5,79±0,12	5,57	0,22	5,74	0,05	4,86	0,93	5,13	0,66
16	D46	2,3,4,5,6-F <sub>5</sub>	4,34±0,07	4,37	-0,03	4,60	-0,26	3,43	0,91	4,32	0,02

[a] Valor experimental de pLD<sub>50</sub> (LD<sub>50</sub>=mmol/inseto); [b] predição pLD<sub>50</sub> com QSAR-MLR após busca conformacional; [c] res = Y<sub>Obs</sub>-Y<sub>Pred,MLR</sub>; [d] Predição pLD<sub>50</sub> com QSAR-SVM após busca conformacional; [e] res = Y<sub>Obs</sub>-Y<sub>Pred-SVM</sub>; [f] predição pLD<sub>50</sub> empregando conformêros indicados pelo índice de Carbó; [g] res = Y<sub>Obs</sub>-Y<sub>Pred-MLR</sub>; [h] res = Y<sub>Obs</sub>-Y<sub>Pred-SVM</sub>;

**Tabela 30.** Compostos usados para desenvolvimento dos modelos de QSAR pelos métodos MLR e SVM de dibenzoilidrazinas contra *Spodoptera exigua*

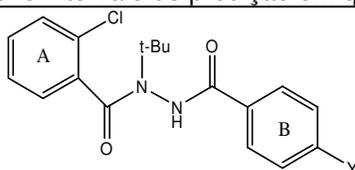
Item	MolID <sup>T</sup>	Xn	Y <sub>Obs</sub> <sup>a</sup>	Descritores <sup>b</sup>			Y <sub>MLR</sub> <sup>c</sup>	Métodos e valores preditos			
				α <sub>YY</sub>	μ <sub>x</sub>	RDF095p		res <sup>d</sup>	Y <sub>SVM</sub> <sup>e</sup>	res <sup>f</sup>	res <sup>g</sup>
1	D1	H	4,92	34,442	-6,795	0,189	4,93	-0,01	4,92	0,00	0,01
2	D2	2-F	6,51	30,160	-1,103	0,000	6,59	-0,08	6,51	0,00	0,08
3	D3	2-Cl	6,37	40,876	0,404	0,148	6,34	0,03	6,34	0,03	-0,01
4	D4	2-Br	6,21	47,994	1,481	0,190	6,21	0,00	6,21	0,00	0,00
5	D5	2-I	6,24	48,840	1,726	0,347	6,18	0,06	6,18	0,06	0,00
6	D6	2-CF <sub>3</sub>	6,19	40,120	0,278	0,636	6,21	-0,02	6,23	-0,04	-0,02
7	D7	2-NO <sub>2</sub>	5,95	33,966	-2,240	0,549	5,95	0,00	5,95	0,00	0,00
8	D8	2-CH <sub>3</sub>	5,85	41,766	-1,104	0,147	5,93	-0,08	5,96	-0,11	-0,03
9	D10 <sup>T</sup>	2-OCH <sub>3</sub>	6,10	40,050	-0,855	0,150	6,08	0,02	6,10	0,00	-0,02
10	D13	3-F	6,15	46,863	1,620	0,372	6,25	-0,10	6,25	-0,10	0,00
11	D14	3-Br	6,57	38,944	0,868	0,191	6,54	0,03	6,52	0,05	0,02
12	D15	3-I	6,11	38,082	0,151	1,233	6,11	0,00	6,14	-0,03	-0,03
13	D16 <sup>T</sup>	3-CF <sub>3</sub>	5,97	43,650	0,617	1,150	5,95	0,02	5,98	-0,01	-0,02
14	D17 <sup>T</sup>	3-CN	4,87	49,152	-1,134	2,335	4,90	-0,03	4,87	0,00	0,03
15	D18	3-CH <sub>3</sub>	6,06	32,408	0,196	2,831	5,97	0,09	5,99	0,07	-0,02
16	D19	4-Cl	6,12	41,343	-0,290	0,238	6,12	0,00	6,14	-0,02	-0,02
17	D20	4-Br	5,62	42,440	-0,135	2,022	5,59	0,03	5,60	0,02	-0,01
18	D21	4-I	5,43	35,265	0,123	4,082	5,44	-0,01	5,44	-0,01	0,00
19	D22	4-CF <sub>3</sub>	5,47	43,077	-1,915	0,948	5,43	0,04	5,45	0,02	-0,02
20	D24	4-NO <sub>2</sub>	3,93	34,847	-6,537	3,588	3,99	-0,06	3,94	-0,01	0,05
21	D25	4-CH <sub>3</sub>	5,55	34,235	-1,953	1,853	5,63	-0,08	5,64	-0,09	-0,01
22	D26	4-Bu <sup>tert</sup>	3,56	52,912	-1,636	5,564	3,65	-0,09	3,63	-0,07	0,02
23	D27	4-Ph	3,38	43,787	-4,375	6,147	3,30	0,08	3,38	0,00	-0,08
24	D28	4-OCH <sub>3</sub>	5,31	33,218	-1,513	3,529	5,31	0,00	5,31	0,00	0,00
25	D29 <sup>T</sup>	4-O(CH <sub>2</sub> ) <sub>3</sub> C <sub>6</sub> H <sub>5</sub>	3,78	54,259	-0,857	5,435	3,80	-0,02	3,78	0,00	0,02
26	D30	2,3-Cl <sub>2</sub>	6,01	44,054	1,679	1,778	6,01	0,00	6,01	0,00	0,00
27	D31 <sup>T</sup>	2,4-Cl <sub>2</sub>	6,54	42,009	1,689	0,205	6,57	-0,03	6,54	0,00	0,03
28	D32	2,5-Cl <sub>2</sub>	5,99	50,049	0,916	0,235	5,95	0,04	5,98	0,01	-0,03
29	D33	3,5-Cl <sub>2</sub>	6,64	28,965	-0,532	0,612	6,61	0,03	6,55	0,09	0,06
30	D35	2,4-(CH <sub>3</sub> ) <sub>2</sub>	4,57	53,145	1,899	5,189	4,60	-0,03	4,57	0,00	0,03
31	D36	2,5-(CH <sub>3</sub> ) <sub>2</sub>	5,01	51,354	-1,691	0,943	5,05	-0,04	5,09	-0,08	-0,04
32	D37	3,4-(CH <sub>3</sub> ) <sub>2</sub>	5,05	44,916	-2,924	1,293	4,99	0,06	4,99	0,06	0,00
33	D38 <sup>T</sup>	3,5-(CH <sub>3</sub> ) <sub>2</sub>	6,08	42,521	2,021	1,933	6,13	-0,05	6,12	-0,04	0,01
34	D39 <sup>T</sup>	2-CH <sub>2</sub> ,3-Cl	5,31	47,556	-1,336	1,078	5,29	0,02	5,32	-0,01	-0,02
35	D42	2-OCH <sub>3</sub> ,5 <sup>Pr-n</sup>	4,64	52,655	-1,260	2,541	4,62	0,02	4,60	0,04	0,02
36	D43	2,3,4-Cl <sub>3</sub>	5,67	45,418	-0,641	0,988	5,60	0,07	5,63	0,04	-0,03

(T) Compostos do conjunto de teste; (a)  $Y = pLD_{50}^{Observado}$  ( $LD_{50} = \text{mmol/inseto}$ ); Desvios não citados na ref. [27]; (b) valores pós análise conformacional; (c)  $Y_{MLR} = pLD_{50}^{Predito}$ ; (d)  $res = Y_{Obs} - Y_{MLR}$ ; (e)  $Y_{SVM} = pLD_{50}^{Predito}$ ; (f)  $res = Y_{Obs} - Y_{SVM}$ ; (g)  $res = Y_{MLR} - Y_{SVM}$ .

**Tabela 43.** Valores utilizados no modelo de QSAR de Ogura *et al.*<sup>26</sup>, e no modelo de QSAR desenvolvido por este estudo de dibenzoilidrazinas contra *Spodoptera frugiperda*

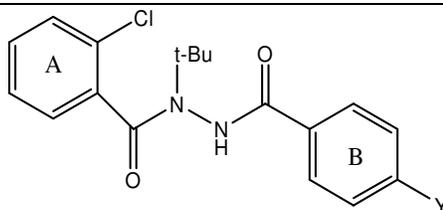
No.	MolID	Y	Y <sub>Obs</sub> <sup>b</sup> (M)	Y <sub>pred</sub> <sup>c</sup> (M)	Y <sub>Pred</sub> <sup>d</sup> (M)	Y <sub>pred</sub> <sup>e</sup> (M)	Res <sup>f,c</sup>	Res <sup>f,d</sup>	Res <sup>f,e</sup>	Descritores <sup>g</sup>			Mor25v <sup>d</sup>	Mor25v <sup>e</sup>
										LogP	σ	ΔB <sub>1</sub>		
1	D1	H	7,08±0,13	7,08	7,13	7,11	0,00	-0,05	-0,03	2,59	0	0,00	0,552	0,552
2	D2	F	7,29±0,08	7,07	7,36	7,34	0,22	-0,07	-0,05	2,87	0,06	0,35	0,519	0,519
3	D3	Cl	6,89±0,01	7,15	7,02	6,82	-0,26	-0,13	0,07	3,51	0,23	0,80	0,586	0,586
4	D4	Br	7,69±0,16	7,23	7,53	7,65	0,46	0,16	0,04	3,73	0,23	0,95	0,510	0,476
5	D5	I	7,30±0,09	7,34	7,43	7,26	-0,04	-0,13	0,04	3,96	0,18	1,15	0,546	0,546
6	D6	CF <sub>3</sub>	6,96±0,06	6,93	6,74	6,94	0,03	0,22	0,02	3,68	0,54	0,99	0,600	0,528
7	D7	NO <sub>2</sub>	6,42±0,01	6,29	5,88	6,41	0,13	0,54	0,01	2,78	0,78	0,70	0,668	0,540
8	D8 <sup>T</sup>	CN	5,74±0,09	6,22	6,06	5,77	-0,48	-0,32	-0,03	2,44	0,66	0,60	0,637	0,637
9	D9	CH <sub>3</sub>	7,28±0,06	7,36	7,26	7,21	-0,08	0,02	0,07	3,15	-0,17	0,52	0,581	0,581
10	D10	C <sub>2</sub> H <sub>5</sub>	7,89±0,04	7,62	7,50	7,93	0,27	0,39	-0,04	3,59	-0,15	0,52	0,561	0,490
11	D11	n-C <sub>3</sub> H <sub>7</sub>	7,64±0,03	7,88	7,34	7,64	-0,24	0,30	0,00	4,06	-0,13	0,52	0,608	0,541
12	D12	i-C <sub>3</sub> H <sub>7</sub>	7,82±0,02	7,79	7,93	7,87	0,03	-0,11	-0,05	4,11	-0,15	0,90	0,513	0,513
13	D13 <sup>T</sup>	n-C <sub>4</sub> H <sub>9</sub>	8,27±0,07	8,24	8,12	8,22	0,03	0,15	0,05	4,60	-0,16	0,52	0,507	0,480
14	D14 <sup>T</sup>	t-C <sub>4</sub> H <sub>9</sub>	7,66±0,13	7,8	7,57	7,76	-0,14	0,09	-0,10	4,48	-0,20	1,60	0,600	0,548
15	D15	OCH <sub>3</sub>	7,21±0,04	7,31	7,17	7,15	-0,10	0,04	0,06	2,82	-0,27	0,35	0,594	0,594
16	D16 <sup>T</sup>	SO <sub>2</sub> CH <sub>3</sub>	5,54±0,07	5,42	5,19	5,53	0,12	0,35	0,01	1,46	0,72	1,03	0,727	0,633
17	D17 <sup>T</sup>	COCH <sub>3</sub>	6,37±0,01	6,34	7,21	6,43	0,03	-0,84	-0,06	2,42	0,50	0,60	0,462	0,568
18	D18 <sup>a</sup>	C <sub>6</sub> H <sub>5</sub>	5,17±0,03	7,98	5,77	5,18	-2,81	-0,60	-0,01	4,49	0	0,71	0,881	0,881

(a) *Outlier* ref.[26]; (T) compostos do conjunto de teste; (b) pIC<sub>50</sub> experimental da ref. [26]; (c) pIC<sub>50</sub> conforme QSAR da ref.[26], (d) calculado usando estruturas com geometria otimizada pelo método AM1; (e) calculado com estruturas selecionadas pela análise conformacional; (f) Y<sub>Obs</sub>-Y<sub>Pred</sub>; (g) conforme ref.[26]. LogP e σ foram utilizados com Mor25v no cálculo dos valores das colunas Y<sub>Pred</sub><sup>d,e</sup>.

**Tabela 47.** Proposta *in silico* de DBHs contra *Spodoptera frugiperda*. Compostos selecionados pelo índice de similaridade molecular ShaEP e respectivo intervalo de predição em que o composto está contido.

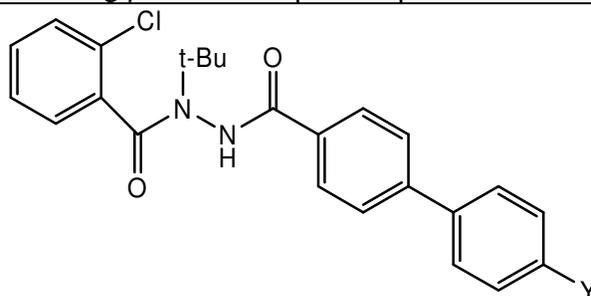
No.	MolID	Y	Descritores			Y <sub>pred</sub> <sup>b,c</sup>	Mor25v <sup>d</sup>	Y <sub>pred</sub> <sup>c,e</sup>	Y <sub>pred</sub> <sup>c,f</sup>
			LogP <sup>a</sup>	σ	Mor25v <sup>b</sup>				
1	T1	NHCOCH <sub>3</sub>	2,243	-0,05	0,363	5,76	0,587	6,86	7,34-6,02
2	T2	OCOCH <sub>3</sub>	2,399	-0,27	0,287	5,70	0,510	7,66	8,18-6,36
3	T3	OC <sub>2</sub> H <sub>5</sub>	3,498	-0,28	0,731	7,38	0,299	6,34	7,92-6,22
4	T4	CH <sub>2</sub> CN	1,950	-0,07	-0,049	4,82	0,515	7,33	7,67-6,74
5	T5	OC <sub>4</sub> H <sub>9</sub>	4,556	-0,30	0,380	8,14	0,625	7,33	8,79-5,83
6	T6	NH <sub>2</sub>	2,037	-0,55	0,296	6,94	0,503	7,93	8,20-6,96
7	T7	CH <sub>2</sub> Cl	3,326	-0,04	0,130	7,30	0,563	7,24	8,37-6,46
8	T8	CH <sub>2</sub> F	2,886	0,07	0,376	7,51	0,495	7,52	8,17-6,37
9	T9	CHOCH <sub>2</sub> NH(CH <sub>3</sub> ) <sub>2</sub>	2,039	0,25	-0,016	4,86	0,483	7,24	8,96-5,87
10	T10	(CH <sub>2</sub> ) <sub>3</sub> COOH	2,904	0,27	0,266	5,82	0,398	7,99	9,33-7,18
11	T11	SO <sub>3</sub> H	2,253	0,74	0,042	5,65	0,488	6,74	7,23-5,76
12	T12	CF <sub>3</sub>	3,832	0,43	0,209	7,09	0,583	6,72	7,78-6,46
13	T13	SCH <sub>3</sub>	3,423	0,13	0,196	7,47	0,279	9,09	9,09-6,83
14	T14	n-C <sub>5</sub> H <sub>11</sub>	5,379	-0,18	0,375	8,84	0,672	7,04	9,74-7,04
15	T15	i-C <sub>5</sub> H <sub>11</sub>	5,249	-0,25	0,199	6,94	0,496	8,33	8,92-7,10
16	T16	n-C <sub>6</sub> H <sub>13</sub>	5,908	-0,18	0,359	8,30	0,512	8,28	9,22-7,30
17	T17	i-C <sub>6</sub> H <sub>13</sub>	5,778	-0,16	0,334	7,42	0,768	6,43	9,08-6,43
18	T18	n-C <sub>7</sub> H <sub>15</sub>	6,437	-0,18	0,433	8,37	0,501	8,46	9,56-6,81
19	T19	CHOCHCH <sub>3</sub> NH <sub>2</sub>	0,897	-0,18	0,216	6,04	0,435	7,79	9,50-6,95
20	T20	CHONHC(CH <sub>3</sub> ) <sub>3</sub>	2,735	-0,18	0,021	5,28	0,276	9,29	10,40-8,01
21	T21	n-C <sub>9</sub> H <sub>19</sub>	7,495	-0,18	0,537	8,74	0,440	9,11	10,54-7,63
22	T22	COHCH <sub>2</sub> NH <sub>2</sub>	0,897	-0,28	0,336	6,26	0,479	7,58	8,15-6,69
23	T23	COCH <sub>2</sub> Cl	2,648	0,61	0,071	6,30	0,514	6,77	7,88-6,04
24	T24	COCH <sub>2</sub> F	2,208	0,52	0,204	6,31	0,536	6,62	7,99-5,65
25	T25	COCH <sub>2</sub> Br	2,788	0,61	-0,045	6,16	0,489	6,98	8,22-6,26
26	T26	COCH <sub>2</sub> I	3,178	0,45	-0,448	5,65	0,561	6,71	7,98-5,70
27	T27	COCH <sub>2</sub> CH <sub>3</sub>	3,064	0,39	0,396	7,09	0,547	6,85	7,96-5,87
28	T28	COCH <sub>2</sub> OCH <sub>3</sub>	2,195	0,38	0,265	6,30	0,626	6,13	7,60-5,65
29	T29	COCH <sub>2</sub> NHCH(CH <sub>3</sub> ) <sub>2</sub>	2,669	0,76	0,341	5,81	0,617	5,89	7,39-5,10
30	T30	CHOCH <sub>2</sub> NHCH(CH <sub>3</sub> ) <sub>2</sub>	2,039	0,27	0,271	5,37	0,437	7,54	8,32-5,76

(a) Calculado pelo ClogP de Chem3D ; (b) geometria otimizada com AM1; (c) pIC<sub>50</sub> (M) uso da Eq.(5.5-16); (d) valor do confôrmero selecionado por ShaEP; (e) melhor confôrmero selecionado pelo índice de similaridade ShaEP; (f) intervalo de predição entre confôrmeros (maior pIC<sub>50</sub>-menor pIC<sub>50</sub>).

**Tabela 48.** Compostos propostos contra *Spodoptera frugiperda* e predição dos valores pIC<sub>50</sub> pelos modelos de QSAR MLR e SVM

No.	MolID	Substituintes Y	LogP	$\sigma$	Mor25v	QSAR e Método		
						Y <sub>Pred</sub> <sup>a</sup> MLR	Y <sub>Pred</sub> <sup>a</sup> SVM	Y <sub>MLR</sub> -Y <sub>SVM</sub> <sup>b</sup>
1	T1	NHCOCH <sub>3</sub>	2,243	-0,05	0,587	6,84	6,86	-0.02
2	T2	OCOCH <sub>3</sub>	2,399	-0,27	0,510	7,65	7,40	0.25
3	T3	OC <sub>2</sub> H <sub>5</sub>	3,498	-0,28	0,731	6,33	6,53	-0.20
4	T4	CH <sub>2</sub> CN	1,950	-0,07	0,515	7,31	7,01	0.30
5	T5	OC <sub>4</sub> H <sub>9</sub>	4,556	-0,30	0,625	7,33	7,09	0.25
6	T6	NH <sub>2</sub>	2,037	-0,55	0,503	7,92	7,38	0.54
7	T7	CH <sub>2</sub> Cl	3,326	-0,04	0,563	7,23	7,29	-0.05
8	T8	CH <sub>2</sub> F	2,886	0,07	0,495	7,50	7,41	0.09
9	T9	CHOCH <sub>2</sub> NH(CH <sub>3</sub> ) <sub>2</sub>	2,039	0,25	0,483	7,22	6,91	0.31
10	T10	(CH <sub>2</sub> ) <sub>3</sub> COOH	2,904	0,27	0,398	7,99	7,73	0.25
11	T11	SO <sub>3</sub> H	2,253	0,74	0,488	6,71	6,66	0.05
12	T12	CF <sub>3</sub>	3,832	0,43	0,583	6,70	6,69	0.01
13	T13	SCH <sub>3</sub>	3,423	0,13	0,279	9,09	8,61	0.48
14	T14	n-C <sub>5</sub> H <sub>11</sub>	5,379	-0,18	0,672	7,05	6,48	0.57
15	T15	i-C <sub>5</sub> H <sub>11</sub>	5,249	-0,25	0,496	8,35	8,12	0.23
16	T16	n-C <sub>6</sub> H <sub>13</sub>	5,908	-0,18	0,512	8,30	7,90	0.40
17	T17	i-C <sub>6</sub> H <sub>13</sub>	5,778	-0,16	0,768	6,43	5,60	0.83
18	T18	n-C <sub>7</sub> H <sub>15</sub>	6,437	-0,18	0,501	8,49	7,93	0.56
19	T19	CHOCHCH <sub>3</sub> NH <sub>2</sub>	0,897	-0,18	0,435	7,77	6,68	1.09
20	T20	CHONHC(CH <sub>3</sub> ) <sub>3</sub>	2,735	-0,18	0,276	9,29	8,26	1.04
21	T21	n-C <sub>9</sub> H <sub>19</sub>	7,495	-0,18	0,440	9,15	8,28	0.87
22	T22	COHCH <sub>2</sub> NH <sub>2</sub>	0,897	-0,28	0,479	7,56	6,68	0.88
23	T23	COCH <sub>2</sub> Cl	2,648	0,61	0,514	6,75	6,69	0.05
24	T24	COCH <sub>2</sub> F	2,208	0,52	0,536	6,59	6,48	0.11
25	T25	COCH <sub>2</sub> Br	2,788	0,61	0,489	6,95	6,91	0.05
26	T26	COCH <sub>2</sub> I	3,178	0,45	0,561	6,70	6,69	0.00
27	T27	COCH <sub>2</sub> CH <sub>3</sub>	3,064	0,39	0,547	6,83	6,83	0.01
28	T28	COCH <sub>2</sub> OCH <sub>3</sub>	2,195	0,38	0,626	6,10	6,10	0.00
29	T29	COCH <sub>2</sub> NHCH(CH <sub>3</sub> ) <sub>2</sub>	2,669	0,76	0,617	5,86	5,83	0.03
30	T30	CHOCH <sub>2</sub> NHCH(CH <sub>3</sub> ) <sub>2</sub>	2,039	0,27	0,437	7,52	7,09	0.43

(a) Y<sub>Pred</sub> = pIC<sub>50</sub>; (b) diferença de predição entre modelos de QSAR MLR e SVM.

**Tabela 49.** Dibenzoilidrazinas com introdução de um fenil substituído contra *Spodoptera frugiperda*. Compostos produzidos *in silico*.

MolID	Substituinte Y	Descritores			Y <sub>pred</sub> <sup>d</sup>	Y <sub>pred</sub> <sup>e</sup>	res <sup>f</sup>
		LogP <sup>a</sup>	$\sigma^b$	Mor25v <sup>c</sup>			
T_D2	F	4,81	0,06	0,856	5,36	5,38	-0,02
T_D3	Cl	5,38	0,23	0,894	5,03	5,05	-0,02
T_D4	Br	5,53	0,23	0,744	6,13	6,14	-0,01
T_D5	I	5,79	0,18	0,974	4,61	4,62	-0,01
T_D6	CF <sub>3</sub>	5,56	0,54	0,853	5,04	5,05	-0,01
T_D7	NO <sub>2</sub>	2,89	0,78	0,915	3,77	3,81	-0,04
T_D8	CN	4,11	0,66	0,892	4,32	4,36	-0,04
T_D9	CH <sub>3</sub>	5,15	-0,17	0,860	5,65	5,66	-0,01
T_D10	C <sub>2</sub> H <sub>5</sub>	5,68	-0,15	0,824	6,00	6,00	0
T_D11	n-C <sub>3</sub> H <sub>7</sub>	6,21	-0,13	0,806	6,22	6,22	0
T_D12	i-C <sub>3</sub> H <sub>7</sub>	6,08	-0,15	0,950	5,19	5,20	-0,01
T_D13	n-C <sub>4</sub> H <sub>9</sub>	6,74	-0,16	1,017	4,87	4,87	0
T_D14	t-C <sub>4</sub> H <sub>9</sub>	6,48	-0,20	0,946	5,36	5,36	0
T_D15	OCH <sub>3</sub>	4,63	-0,27	0,862	5,63	5,64	-0,01
T_D16	SO <sub>2</sub> CH <sub>3</sub>	4,94	0,72	0,954	3,99	4,03	-0,04
T_D17	COCH <sub>3</sub>	4,14	0,50	0,783	5,27	5,30	-0,03

(a) Calculado pelo ClogP do programa Chem3D<sup>41</sup>; (b) valores da ref.[26,52];  
 (c) estruturas com geometria otimizada AM1; (d) QSAR Eq. (5.5-16); (e) QSAR Eq. (5.5-17);  
 Y<sub>Pred</sub> = pIC<sub>50</sub>; (f) res = Y<sub>Pred</sub><sup>d</sup> - Y<sub>Pred</sub><sup>e</sup>.



## **APÊNDICES**



# QSAR of Clerodane Diterpenoids with Cytotoxic and Antitumoral Activity Against V-79 Cells

Bruno Cramer<sup>a\*</sup>, Yuji Takahata<sup>b</sup>

<sup>a</sup> Department of Chemistry, University of Campinas-UNICAMP, Cidade Universitária Zeferino Vaz, Distrito de Barão Geraldo, Campinas 13084-862, São Paulo, Brazil

e-mail: cramer@iqm.unicamp.br; bruno.cramer@gmail.com

<sup>b</sup> School of Engineering, Amazonas State University, Av. Darcy Vargas, 1200, Parque 10 – CEP 69065-020, Manaus, Amazonas, Brazil

**Keywords:** Clerodane diterpenoids, Cytotoxicity, Antitumoral, QSAR-2D, Linear free energy relationships, Medicinal chemistry

Received: April 14, 2009; Accepted: July 2, 2009

DOI: 10.1002/qsar.200960038

## Abstract

This work deals with diterpenoids of the clerodane type, compounds isolated from the extracts of different plant organs from trees mainly of the *Flacourtiaceae* family. The aim of the present work is to study Quantitative Structure-Activity Relationship (QSAR) on Casearins isolated from the crude plant extract of the leaves of the shrub tree *Casearia sylvestris* Swartz, *Flacourtiaceae*, that show cytotoxic activities against V-79 Chinese Hamster lung cells. In total, ca. 2080 descriptors and indices were calculated for the QSAR study using various softwares. The obtained QSARs with three descriptors were developed using two independent methods, Multiple Linear Regression (MLR) and Support Vector Machine (SVM), using a small set of molecules. For external validation of the MLR QSAR we employed a training/test set that was evaluated through criteria ruled by *RQK fitness functions*. Both MLR and SVM QSARs were validated using the same test set, where the SVM QSAR achieved the best results. Both QSARs were applied to a data set with 113 clerodane diterpenoids never tested before against the Chinese Hamster lung fibroblast V-79 cells, and also applied to thirteen new compounds which were build in silico, adding fluorine and chlorine atoms as substituents at specific positions of the clerodane diterpenoids. Encouraging results were obtained in both cases which serve as an indicator for further studies in silico and/or bioassay of these compounds as anticancer agents.

## 1 Introduction

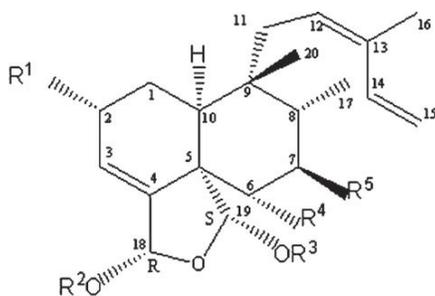
The botanical family *Flacourtiaceae* is a good source of diterpenoids, compounds known by their cytotoxic and antitumoral properties. The *Flacourtiaceae* family has ca. 1300 tropical and subtropical species subordinated to 86 genus [1, 2]. More than eight hundred isolated diterpenoids and nor-diterpenoids with the clerodane skeleton have been reported [3, 4], also involving other botanical families than *Flacourtiaceae*. Structures of several compounds were analyzed and elucidated by NMR and X-ray spectroscopy. Many of these compounds had their bioassay done with different protocols and many have never been biologically tested.

Whether by extraction process or through synthesis, both ways are troublous to obtain pure substances from natural products. The challenges of the synthetic chemist in the synthesis of pharmaceutical relevant diterpenes has

been emphasized in a recent review article [5] and corroborates with the concept that a validated QSAR predictive model can be of valuable aid in the choice of which compounds of a data source may receive priority attention for biological tests. To our knowledge, this is the first QSAR study that makes available a compound set with predicted  $IC_{50}$  values of clerodane diterpenoids. The study uses clerodane diterpenoids named Casearins, tested in vitro by Itokawa et al. [6] and Morita et al. [7], exhibiting cytotoxic activity against Chinese Hamster V-79 lung cells and antitumoral activity against Sarcoma 180 ascites in mice. DNA damage and repair, mutagenicity and toxicity are, to a great extend, studied with the V-79 cell line [8]. Engineered V-79 cell lines are also used as an alternative to animal experimentation in drug development. Another component in this QSAR study is based on the assumption that natural products can be synthetically fluorinated and chlorinated. These compounds are present in marketed

pharmaceuticals that are used as antimitotic and antitumor agents, used in cancer treatment [9]. A diversified use of fluorinated compounds and entailed biological properties can be found in the literature [10]. Synthetic fluorinated analogues of natural products are used in chemotherapy, e.g. Camptothecin, an alkaloid extracted from a tree in the South of China *Camptotheca acuminata* Decne., Nyssaceae; Podophyllo lignane toxin, isolated from *Podophyllum peltatum* L. as other plants of the Podophyllum species, is a cytotoxic agent against various cell lines, but with numerous side effects [11]. We used the clerodane diterpenoid Casearin G (Fig. 1) modifying it in silico by introducing fluorine or chlorine atoms at specific positions. Special attention was given to the carbon C6 mentioned in literature [7] as a sensible center to bioactivity changes.

The aim of the present work is establish Quantitative Structure-Activity Relationship (QSAR) on Casearins iso-



Casearin	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	R <sup>5</sup>	IC <sub>50</sub> (μmol/l)
A	OMe	Ac <sup>a</sup>	Ac	OH	OBu <sup>c</sup>	1
B	OMe	Ac	Ac	OAc	OBu	8.5
C	OH	Ac	Ac	OAc	ODc	0.77
D	OH	Bu	Ac	OH	OBu	1.8
E	OH	Et <sup>b</sup>	Ac	OH	ODc <sup>d</sup>	4.7
F	OH	Et	Ac	OH	OBu	29
G	OMe	Ac	Ac	H	OBu	0.17
H	OH	Ac	Ac	H	OBu	0.37
I	OH	Ac	Bu	H	OBu	0.51
J	OMe	Bu	Ac	OH	OBu	1.1
K	OAc	Ac	Ac	OH	OBu	0.52
L	OMe	Bu	Ac	OAc	OH	1.6
M	OH	Bu	Bu	OAc	OH	1.8
N	OMe	Ac	Bu	OAc	OBu	5.9
O	OMe	Bu	Ac	OAc	OBu	6
P	OMe	Ac	Ac	OAc	OAc	7.8
Q	OH	Ac	Ac	OAc	OBu	4.3
R	=O	Ac	Ac	OH	OBu	5.4
Aa	OMe	Ac	Ac	=O	OBu	0.55
Ab	OMe	Ac	Ac	OPr	OBu	17
Ac	OMe	Ac	Ac	OBu	OBu	38
Da	=O	Bu	Ac	=O	OBu	19

(a) OAc=acetate; (b) Et=ethyl; (c) OBu=butanoate; (d) ODc=decanoate

**Figure 1.** Structure of Casearins and derivatives, and cytotoxic activities against V-79 Chinese Hamster lung cells according to [7].

lated from the crude plant extract of the leaves of the shrub tree *Casearia sylvestris* Swartz, *Flacourtiaceae*, that show cytotoxic activities against V-79 Chinese Hamster lung cells [6, 7]. The obtained QSAR model will be used to predict inhibitory activity of one hundred and thirteen diterpenoids of the clerodane type isolated from crude plant extracts and from different plant organs (leaves, flowers, fruits, seeds, stems, barks and roots) of the genus *Casearia*, *Laetia* and *Zuelania*, *Flacourtiaceae* family, and diterpenoids from one species belongs to the *Chrysobalanaceae* family. The obtained QSARs were also applied to generate the predicted IC<sub>50</sub> values of Casearin G compound fluorine and chlorine substituted.

## 2 Materials and Methods

QSAR studies have to follow some rules and recommendations mentioned in literature [12, 13] and two topics were chosen and are mentioned here considering their impact on final QSAR results: (a) the biological IC<sub>50</sub> data must be generated by the same bioassay protocol. Our IC<sub>50</sub> data meet this rule as can be verified in the references [6, 7]. (b) The structures must be correct. Chirality centers (R/S) were considered based on the source literature [7] combined with informations based on SciFinder Scholar [14]. All structures were then build in 2D [15] and converted to 3D [16]. Chirality requires attention as the calculated descriptor's value, depending on descriptor category, may significantly change. Also structure correctness is mandatory in case of QSAR results reproduction or its use to predict IC<sub>50</sub> values for new compounds imported from data banks. Translation problems may occur during this process as mentioned in [13].

SciFinder Scholar [14] was also used to get the CAS registry numbers of the compounds used in the QSAR. Furthermore, the study used chemical structures of clerodane diterpenoids from literature that were supported by structural assignments established by NMR and/or X-Ray elucidation techniques.

Figure 1 lists the structures of twenty two oxygenated tricyclic clerodane diterpenoids, Casearins, with *cis* configuration and their cytotoxic activities against V-79 Chinese Hamster lung cell taken from literature [7]. The Casearins are identified by the capital letters A–R. The compounds Aa, Ab, Ac and Da correspond to synthesized clerodane diterpenoids from Casearin-A and D. Casearins F, Ab, Ac and Da were jackknifed from the set due to their low bioactivity (high IC<sub>50</sub> value) reducing the set to eighteen compounds. We limit our QSAR study to the eighteen compounds. The IC<sub>50</sub> values were converted to Log (1/IC<sub>50</sub>). All structures had the geometry optimized, using the Molecular Mechanics method MM2 as a first optimization step with Chem3D [15]. This was followed by the semi-empirical AM1 quantum chemical method with the Gaussian03W package [17], providing also the electronic de-

scriptors  $\epsilon_{\text{HOMO}}$  and  $\epsilon_{\text{LUMO}}$ , respectively energy of the highest occupied and lowest unoccupied molecular frontier orbital, and the energy gaps  $\epsilon_{\text{HOMO-HOMO}}$ ,  $\epsilon_{\text{LUMO-LUMO}}$  and  $\epsilon_{\text{HOMO-LUMO}}$ . MOPAC 7 [18] was used to calculate Wang-Ford charges that were obtained from molecular electrostatic surface of non-hydrogen and common C and O atoms of the clerodane skeleton [19]; CaChe WorkSystem-Pro [20] was used to calculate, e.g. LogP, molar refractivity, dielectric energy; DRAGON [21] was used to calculate all 20 classes, containing 1664 descriptors, for example, radial distribution function (RDF), topological, 3D WHIM and GETAWAY (Geometry, Topology, and Atom-Weights Assembly) descriptors. MarvinView [22] was employed to produce 21 descriptors (e.g., isoelectric point, polarizability); MolInspiration [23] for miLogP, polar surface area (PSA), Lipinski's Rule of 5 parameters. The RECON software [24] was used to compute 248 Transferable Atom Equivalent (TAE) descriptors [25, 26]. A few selected descriptors associated to molecular properties were calculated using the PETRA program [27]. In total, ca. 2080 descriptors and indices were calculated for the QSAR study. DRAGON calculated 1664 and through its constant and near constant selection/elimination criteria, ca. 450 were jackknifed. Descriptors calculated by other means presenting constant, near constant and null values were eliminated through visual screening. 1570 descriptors were available for the QSAR development.

Two independent methods were used: MLR (Multiple Linear Regression) and SVM (Support Vector Machine, specifically SVR-Support Vector Regression). The QSAR models using MLR were constructed using software MobyDigs [28] and BuildQSAR [29] as main software, both using genetic algorithms (GA) in the variable selection. Then MobyDigs 'Tabu List' reduces the amount of variables according to specific statistical properties provided by specific parameter settings which may be selected by the user. We used default values of the Tabu List. Details can be found in literature [30].

The regression models were chosen according to the following rules of the *RQK fitness functions*: (a) The *QUIK rule* that evaluates the collinearity between descriptors:  $(K_{xy} - K_x) > \delta K$ ; (b) The asymptotic  $Q^2$  rule that evaluates the difference between fitting and predictive ability of the model:  $Q^2_{\text{ASYM}} - Q^2_{\text{LOO}} < \delta Q$ ; (c) evaluation of redundancy in explanatory variables:  $R^p > t^p$  and (d) overfitting caused by noisy variables:  $R^N > t^N$  ( $\epsilon$ ). If the results of any of these tests are not met, the model is rejected. Details can be found elsewhere [31]. The sample/variable ratio was chosen as 5:1 [12, 32] which constraints the GA to select three variables to build the regression models. Genetic Algorithms can not necessarily find all the best combinations when large variable sets are used. Based on this assumption 1570 descriptors were split into specific classes forming smaller sets (300 or less). Different splitting criteria were used. One criterion employed classes of variables based on specific characteristics mentioned in the litera-

ture of the compounds with relation to their bioactivity, e.g. Casearins membrane affinity to V79 cells based on hydrophobicity evaluation and influence of bulkiness of substituents in the  $IC_{50}$  value. Another criterion adopted was the exclusion of variables with low correlation to the bioactivity, removing variables with values below to 0.100 using the correlation matrix of the Molegro Data Modeller (MDM) software [33], reducing the set to ca. 580 descriptors. In the first step, an internally validated QSAR model was obtained using the whole set of the 18 compounds, including outlier and leverage evaluation. QSAR internal validation was done using the MobyDigs's cross-validation leave-one-out (CV-LOO) and bootstrap. To verify presence or absence of chance-correlation, the Y-scrambling method was used along the whole QSAR study, in which the  $Y$  ( $= \text{Log}(1/IC_{50})$ ) values are shuffled leaving the variables set unchanged. We used the MobyDigs intercept values criteria  $a(R^2) < 0.3$  and  $a(Q^2) < 0.05$  running 10 loops of the scrambling test using the default value of  $\times 300$ , totaling 3000 times. In the second step, we aim at external validation of a QSAR model. The compounds were split into training and test set. The training set consists of 13 compounds, whereas the test set consists of 5 compounds. The judicious choice of the five compounds was based on the use of a dendrogram produced by Hierarchical Cluster Analysis (HCA) using the software Pirouette [34]. A QSAR model with 13 compounds as training set was constructed using exactly the same set of descriptors as those selected in the first step. Activity of the 5 compounds of the test set was predicted using the QSAR model. The value of  $Q_{\text{ext}}$ , externally validated CV-LOO, was obtained to estimate predictive power. The sample/variable ratio in this case reduces to 4:1, which is acceptable if no abnormality is detected through parameters evaluation, e.g. overfitting. A complementary study was done using Support Vector Machine (SVM), employing the software MDM [33]. The choice of these two independent methods, respectively MLR and SVM were to provide a robust assessment level of the proposed QSAR. In both methods, the variables were previously auto scaled.

### 3 Results and Discussion

#### 3.1 Internal and External Validation

Eq. (1) is the best QSAR model obtained for the set of the whole Casearins compound ( $n=18$ ) using two DRAGON [21] calculated descriptors (RDF060m, R8p) and one RECON [24] descriptor (Fuk5). It is intended to assess internal validation of the model.

$$\begin{aligned} \text{Log}(1/IC_{50}) = & 0.083 (\pm 0.021) \text{RDF060m} \\ & -7.636 (\pm 1.486) \text{R8p} - 0.043 (\pm 0.010) \text{Fuk5} \\ & + 13.219 (\pm 1.664) \end{aligned} \quad (1)$$

( $n = 18$ ;  $R^2 = 0.931$ ;  $s = 0.148$ ;  $F = 64.10$ ;  $p < 0.0001$ ;  $Q^2 = 0.889$ ;  $R^2_{\text{adj}} = 0.918$ ;  $SPress = 0.189$ ;  $SDEP = 0.172$ )

RDF060m is a Radial Distribution Function-6.0/weighted by atomic mass. Radial distribution function (RDF) of DRAGON [21] encodes geometrical and atomic features of chemical structures in 3D space according to atomic pair properties  $A_i$  and  $A_j$ , which in this work corresponds to the atomic mass. Equation 2 uses a Gaussian term as distance function

$$RDF(r) = f \sum_{i=1}^{N-1} \sum_{j=i+1}^N A_i A_j e^{-B(r-r_{ij})^2} \quad (2)$$

where  $f$  is a scaling factor,  $N$  the number of atoms  $i$  and  $j$  and  $r_{ij}$  the interatomic distances.  $A$  smoothing parameter,  $B$ , defines the probability distribution of individual distances of the atoms which may be regarded as their vibrations in the molecule. According to [35, 36] RDF can be interpreted as the probability distribution to find an atom in a spherical volume of radius  $r$ . Based on this definition it is plausible to presume that RDF060m encodes a partial contribution of the Casarin's bioactivity at regions where atoms are distanced  $r = 6.0 \text{ \AA}$  from the geometrical center of each molecule. The program calculated thirty atomic mass-weighted descriptors with a step size of  $r = 0.5 \text{ \AA}$  and only RDF060m presented relevance in the QSAR. The GETAWAY (Geometry, Topology, and Atom-Weights Assembly) are proposed as three-dimensional chemical structure descriptors, encoding structural and stereochemical information [36, 37]. The molecular structural representation is defined as Molecular Influence Matrix (MIM) denoted by  $H$  in Equation 3:

$$H = M \cdot (M^T \cdot M)^{-1} \cdot M^T \quad (3)$$

where  $M$  is the molecular matrix and the superscript T refers to the transposed matrix.

R8p belongs to R-GETAWAY descriptors class calculated according to Equation 4:

$$Rkw = \sum_{i=1}^{N-1} \sum_{j>i}^N \frac{\sqrt{h_{ii}h_{jj}}}{r_{ij}} w_i w_j \delta(k; d_{ij}) \quad (4)$$

$$k = 1, 2, \dots, 8$$

where  $Rkw$  in the present work corresponds to R8p: R autocorrelation of lag 8/weighted by atomic polarizabilities  $p (= w_i w_j)$ .  $N$  is the number of molecule atoms,  $d_{ij}$  is the topological distance between atoms  $i$  and  $j$ ,  $d$  is the topological diameter,  $r_{ij}$  is the atoms geometrical distance and  $\delta(k; d_{ij})$  is the delta Dirac function.  $h_{ii}$  and  $h_{ij}$  are the diagonal elements of the molecular influence matrix (MIM), encoding atomic information and represent the influence of

each molecule atom in determining the whole shape of the molecule.

The TAE descriptor Fuk5 relates to the Fukui  $F^+$  function scalar value [25, 26]

$$F^+(r) = \rho_{\text{HOMO}}(r) \quad (5)$$

where  $\rho_{\text{HOMO}}$  is the electron density of the highest occupied molecular orbital (HOMO) at a point  $r$  in the space around the molecule. It characterizes nucleophilic attack and describes radical reactivity entailed to a surface area and the 5<sup>th</sup> bin is an average value of this function. Resuming, the QSAR contains physicochemical information of the molecule atoms at positions favorably to reaction (RDF060m). It captures information about the molecular shape and atoms polarizabilities (R8p), and molecular reactivity at the molecular surface (Fuk5).

The descriptor values are listed in Table 1. The correlation coefficient between two descriptors out of the three is as follows: RDF060m/R8p (0.431), RDF060m/Fuk5 (0.096) and R8p/Fuk5 (0.230).

However these descriptors do not permit any clear interpretation of physico-chemical behavior of the molecule in relation to its biological activity. Also literature is missing experimental studies concerned to the mechanism of interaction amongst diterpenoids and the V79 cell line. A regression analysis of descriptor vs. bioactivity resulted that the Fukui function Fuk5 has the highest influence in the  $IC_{50}$  result followed by R-GETAWAY descriptor R8p, and RDF060m presented the lowest  $R^2$  and Fisher statistical  $F$  value. This result indicates that a partitioned contribution of the descriptors is responsible in the formation of the good result when they are combined (as in Eq. 1). Another aspect to be considered is that a two variable QSAR does not capture molecular information to meet the requirement of  $R^2 > 0.60$ .

A Williams plot (not shown) as provided by MobyDigs of the QSAR model of Eq. (1) indicates that all compounds are contained in the Applicability Domain (AD). An outlier evaluation was done employing MDM which uses an interquartile range (IQR), defined as the distance (in IQR units) between the first and third quartile [33]. Values above the score 3.0 are ranked as outlier compounds. No outlier was observed. The QSAR model tested on randomized response by Y-scrambling presented low intercept  $a(R^2)$ , respectively  $a(Q^2)$  values when compared to the threshold value  $a(R^2) < 0.3$  and  $a(Q^2) < 0.05$ . This demonstrates that the QSAR model (Eq. 1) is not obtained by chance correlation. After 5000 iterations,  $Q^2$  in bootstrap resulted in a high average value  $Q^2_{\text{boot}} = 0.876$ , which demonstrates the model robustness and internal predictability. As the main goal of the present QSAR study is to obtain a model with external predictive ability, an attempt has been done to search for such a model using the same variables selected by the GA which resulted in Equation 1.

**Table 1.** Log ( $1/IC_{50}$ ) and descriptor values used in the QSAR.

Casearin	CAS Reg. Nr.	Log ( $1/IC_{50}$ )	RDF060m [a]	R8p [b]	Fuk5 [c]
A	134174-73-5	6.000	29.874	0.617	119.23
Aa	134955-74-1	6.259	29.562	0.607	111.38
B	133577-59-0	5.070	28.128	0.702	119.33
C	133577-60-3	6.113	35.602	0.576	132.5
D	133557-53-6	5.744	28.914	0.579	126.07
E	133535-78-1	5.328	19.374	0.506	129.19
G	134955-62-7	6.769	31.047	0.61	105.67
H	134955-63-8	6.409	27.228	0.58	108.22
I	134955-64-9	6.292	31.794	0.583	114.28
J	134955-65-0	5.959	31.524	0.568	125.3
K	134955-66-1	6.284	37.11	0.634	121.87
L	134955-67-2	5.796	27.584	0.596	119.23
M	134955-68-3	5.745	24.939	0.573	127.84
N	134955-69-4	5.229	33.333	0.703	125.39
O	134955-70-7	5.222	31.519	0.67	125.39
P	134955-71-8	5.108	29.16	0.717	113.26
Q	134955-72-9	5.366	30.48	0.693	121.87
R	134955-73-0	5.268	24.239	0.569	129.27

QSAR Descriptors: [a] Radial Distribution Function; [b] R-GETAWAY; [c] Fukui Function

The data set was divided into training set ( $n=13$ ) and test set ( $n=5$ ) based on the Hierarchical Cluster Analysis (HCA) dendrogram (not shown). The compounds of the test set were chosen from alternative classes of the dendrogram avoiding the choice of the extreme max/min of the bioactivity values (Casearin G and B, respectively). The best model was selected after ten trials using  $Q_{\text{ext}}^2$  as reference but also making use of bootstrap, *QUIK rule* and Y-scrambling. Equation 2 is the best QSAR model obtained with the training set ( $n=13$ ).

$$\begin{aligned} \text{Log}(1/IC_{50}) = & 0.086 (\pm 0.025) \text{RDF060m} \\ & -7.581 (\pm 1.616) \text{R8p} - 0.045 (\pm 0.012) \text{Fuk5} \\ & + 13.363 (\pm 1.954) \end{aligned} \quad (6)$$

$$\begin{aligned} (n=13; R^2=0.952; s=0.141; F=59.52; p<0.0001; \\ Q^2=0.904; R^2_{\text{adj}}=0.936; Q^2_{\text{ext}}=0.816; SPress=0.199; \\ SDEP=0.172) \end{aligned}$$

It is worth to point out that the coefficients of Equations 1 and 6 are very similar, making the equations almost equivalent. The training set was also submitted to the leave- $N$ -out test (using  $N=3, 20$  to 25% is recommended for small sets) running three iterations, with the QsarModeling software which is a module of a recently in-house developed QSAR package. The difference between  $Q^2$  and the calculated  $Q^2_{\text{leave-}N\text{-out}}$  values of Equation 6 was always less than 0.1, which assures the model robustness until this stage.

An additional external validation according to [38] is applied solely to the test set. According to the recommended criteria, a predictive QSAR model, must attend the following conditions: (a) High  $Q^2$  cross-validated value  $>0.5$ ; (b) Correlation coefficient  $R^2 > 0.6$ ; (c) Correlation coefficients  $R^2_0$  or  $R^2_0$  for regressions through the origin (pre-

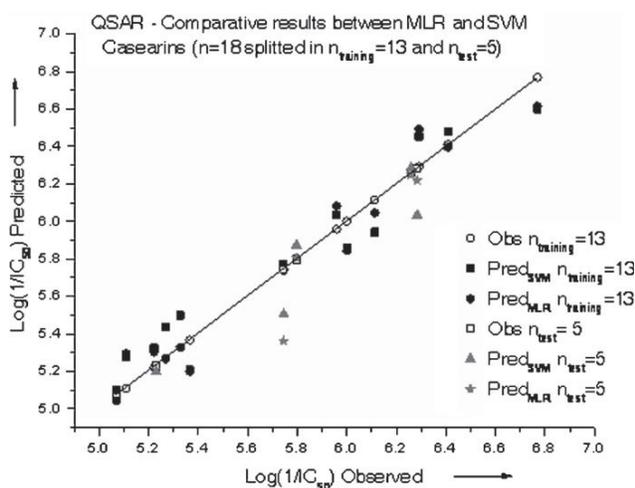
dicted vs. observed and observed vs. predicted activities) should be close to  $R^2$ , i.e.  $[(R^2 - R^2_0)/R^2] < 0.1$  or  $[(R^2 - R_0^2)/R^2] < 0.1$  and (d)  $0.85 \leq k \leq 1.15$  or  $0.85 \leq k' \leq 1.15$ , where  $k$  and  $k'$  are the slopes of the regression lines.

Using the test set  $n=5$  (Casearins Aa, K, L, M, N) the QSAR meets all the mentioned conditions.

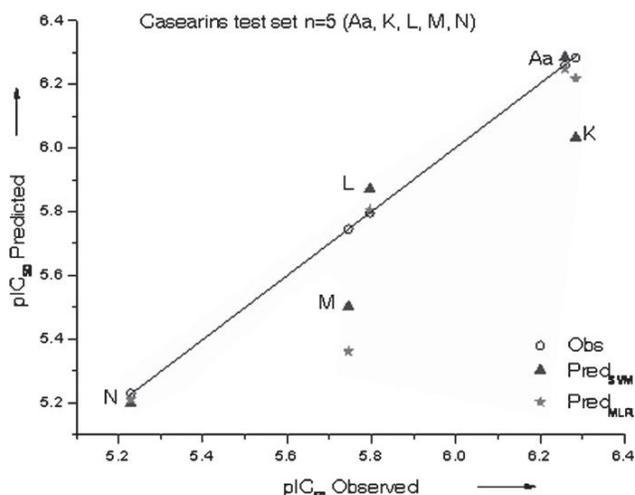
### 3.2 Complementary Study to the MLR Method: Support Vector Machine (SVM)

The above mentioned results were obtained by following the rules and principles for the development of a robust QSAR predictive model, using MLR method. But as we are working with a small set of compounds, a complementary analysis was performed to verify further an alternative of the obtained QSAR model with the independent method Support Vector Machine (SVM). It is a modeling tool used for classification and regression employing support vector methods. The theory was originally developed by Vapnik [39]. Details of the theory can also be found in [40–42]. SVM has been adopted in structure-activity relationship (SAR) [43] and in comparative quantitative structure property/activity relationship (QSPR/QSAR) studies. It has been presenting better performance than the MLR and Partial Least Squares (PLS) method [44, 45].

In our work the SVM model used the kernel Radial Basis Function (RBF). Through a fine-tuning procedure, we tried to obtain the lowest Root Mean Square Error (RMSE) related to the best regression parameter using the leave-one-out (LOO) and cross-validation (CV). The optimal value obtained for each of the SVM parameters is as follows: cost (or capacity) parameter,  $C=10$ ; width of the radial basis function,  $\gamma=0.100$  and  $\epsilon$ -insensitive loss function,  $\epsilon=0.200$ . The external predictive model



**Figure 2.** QSAR result using MLR and SVM - Predicted and observed values of  $\text{Log}(1/IC_{50})$  using the training and test set of the 18 compounds.



**Figure 3.** MLR and SVM predicted and observed results of  $\text{Log}(1/IC_{50})$  using only the test set (Casearins Aa, K, L, M, N).

( $n_{\text{training}}=13$ ) consists of five support vectors. When applying the recommended validation of [38] using the test set, SVM ( $n_{\text{test}}=5$ , Casearins Aa, K, L, M, N) predicted slightly better results as it can be seen in Table 2. The  $R^2$  linear

regression value is the same for both methods, with the difference that  $R_0^2$  and  $R_0'^2$  are slightly closer in SVM, and the Mean Square Error (MSE) of the training/test set are closer in SVM than in MLR.  $Q^2_{\text{ext}}$  is also slightly better in SVM. Based on the fulfillment of all the validation criteria, we consider the SVM QSAR model validated to be used for external prediction. The observed and predicted values of the training and test set of all the 18 compounds obtained through MLR and SVM are plotted in Figure 2. In Figure 3 only the test set comparative results are plotted.

### 3.3 Prediction of Activity

Since the QSAR obtained with MLR and SVM were both validated we used them to predict the cytotoxic and antitumoral activity of 113 clerodane diterpenoids on a comparative purpose. The best QSAR was provided by SVM and the  $IC_{50}$  predicted values are labeled in gray in descending order in Table 3. Some of the compounds were named in the literature only with the IUPAC nomenclature. We replaced it by (arbitrary) correlated names to the plant species name (e.g. Ccorymbosa 4; Zguidonia 6). All compounds are identified by their CAS registry number.

Applying the SVM QSAR, one hundred and four compounds (ca. 90%) are contained in the  $IC_{50}$  experimental range (0.17–8.5  $\mu\text{mol/L}$ ). Only 8 compounds (ca. 10%) have extrapolated values below the best experimental value (0.17  $\mu\text{mol/L}$ ), and 1 compound is above 8.5  $\mu\text{mol/L}$ .

By applying the QSAR obtained with MLR, ninety five compounds had their bioactivity predicted with the  $IC_{50}$  value laying within the observed range of the 18 compounds used for the QSAR development. Eighteen are extrapolated, where 13 are below 0.17  $\mu\text{mol/L}$  and 5 are above 8.5  $\mu\text{mol/L}$ , respectively, to the lowest (best) and highest (worst) experimental  $IC_{50}$  value.

Four diterpenoids have lower descriptor values mainly RDF060m and Rp8 compared to the average: Ccorymbosa 4, Cgrayi 8, Zuelanin 5 and Zguidonia 10. This is probably caused by the missing of the diacetal ring affecting the molecular geometry (less bulkiness). QSAR Applicability Domain (AD) is characterized by physicochemical, structural and/or biological informations of the corresponding intrinsic properties contained in the training set used to develop the QSAR model. Based on this concept, only interpolations are to be considered. In the present QSAR study, it is advisable not to discard the extrapolated values

**Table 2.** Comparative results of MLR and SVM.

Method	Training set $n=13$				Validation [a] set $n=5$ (Casearins Aa, K, L, M, N)						
	$R^2_{\text{train}}$	$Q^2$	$Q^2_{\text{ext}}$	MSE	$R^2_{\text{test}}$	$R_0^2$	$R_0'^2$	$ R_0^2 - R_0'^2  < 0.3$	$0.85 \leq k \leq 0.15$	$0.85 \leq k' \leq 0.15$	MSE
MLR	0.952	0.904	0.82	0.0137	0.879	0.957	0.942	0.014	1.015	0.948	0.0304
SVM	0.941	0.830	0.84	0.0182	0.879	0.954	0.951	0.003	1.014	0.985	0.0260

[a] Method used as in [38]

**Table 3.** Predicted  $IC_{50}$  values of clerodane diterpenoids against V-79 cell line with MLR and SVM QSAR.

Item	Compound	CAS Reg. Nr.	Descriptors [a]			Method and predicted values			
			RDF060m	R8p	Fuk5	MLR		SVM	
						Y [b]	$IC_{50}$ [c]	Y [b]	$IC_{50}$ [c]
1	Casearborin A	272771-14-9	25.294	0.533	91.404	7.35	0.045	7.07	0.085
2	Casearborin B	272771-15-0	26.967	0.510	100.840	7.24	0.058	7.07	0.086
3	Caseargrewiin H	945607-74-9	24.403	0.510	98.199	7.14	0.073	6.97	0.106
4	Casearborin D	272771-17-2	29.539	0.547	104.960	6.99	0.102	6.94	0.114
5	Caseargrewiin E	945607-68-1	24.030	0.544	94.658	7.01	0.098	6.92	0.121
6	Casearin S	220178-32-5	25.802	0.570	97.506	6.84	0.146	6.85	0.143
7	Caseargrewiin G	945607-72-7	27.365	0.560	103.030	6.80	0.160	6.82	0.152
8	Caseargrewiin D	847841-95-6	19.575	0.500	88.496	7.24	0.058	6.81	0.156
9	Cgrayi 1	947672-97-1	21.581	0.498	106.010	6.63	0.233	6.58	0.264
10	Cgrewiifolia 5	1006868-63-8	24.546	0.559	105.420	6.45	0.353	6.55	0.281
11	Casearluicin F	405507-88-2	20.268	0.557	98.141	6.43	0.372	6.53	0.293
12	Casearinol A	196090-44-5	28.084	0.607	103.840	6.46	0.343	6.53	0.294
13	Cgrewiifolia 4	1006868-61-6	23.851	0.572	103.650	6.37	0.423	6.50	0.313
14	Casearborin C	272771-16-1	30.066	0.621	104.960	6.48	0.333	6.49	0.322
15	Ctremula 3	173324-52-2	21.284	0.528	105.300	6.41	0.388	6.48	0.331
16	Zguidonia 8	128486-53-3	13.351	0.499	83.108	6.96	0.111	6.45	0.359
17	Casearborin E	272771-18-3	34.830	0.654	105.060	6.63	0.233	6.44	0.359
18	Corymbolusin B	326473-52-3	30.738	0.539	118.840	6.53	0.297	6.44	0.362
19	Caseargrewiin J	945607-78-3	21.065	0.490	109.290	6.50	0.317	6.44	0.363
20	Ccorymbosa 5	131766-54-6	15.539	0.492	97.670	6.54	0.291	6.43	0.372
21	Casearluicin A	405507-78-0	27.067	0.561	111.700	6.37	0.428	6.43	0.373
22	Cgrayi 6	947672-99-3	15.053	0.515	95.445	6.42	0.379	6.40	0.398
23	Caseargrewiin I	945607-76-1	23.075	0.511	111.930	6.39	0.404	6.40	0.401
24	Casearinone B	196090-47-8	17.969	0.585	93.282	6.24	0.576	6.39	0.404
25	Casearinol B	196090-45-6	30.153	0.596	111.870	6.36	0.435	6.38	0.418
26	Caseargrewiin F	945607-70-5	23.040	0.535	110.160	6.29	0.514	6.37	0.428
27	Cgrayi 7	947673-00-9	12.336	0.517	88.670	6.48	0.332	6.34	0.458
28	Zguidonia 7	128486-52-2	11.797	0.505	76.005	7.10	0.080	6.34	0.460
29	Intrapetacin A	343944-35-4	22.120	0.544	109.430	6.17	0.669	6.29	0.507
30	Zuelanin 13	130827-60-0	15.269	0.591	90.262	6.10	0.797	6.28	0.525
31	Corymbolusin A	326473-51-2	24.464	0.470	120.190	6.45	0.356	6.26	0.550
32	Zuelanin 9	130827-56-4	22.456	0.597	103.820	6.06	0.878	6.26	0.556
33	Zuelanin 3	128572-38-3	23.619	0.571	109.640	6.09	0.814	6.24	0.581
34	Corymbotin F	138950-31-9	24.757	0.589	109.000	6.08	0.8326	6.22	0.596
35	Casearluicin G	405507-90-6	14.265	0.540	98.141	6.04	0.910	6.22	0.602
36	Zuelanin 1	128486-31-7	23.010	0.567	109.640	6.07	0.857	6.22	0.603
37	Casearvestrin C	402935-06-2	20.089	0.511	111.760	6.14	0.718	6.20	0.626
38	Caseargrewiin L	945607-81-8	19.130	0.526	108.890	6.08	0.836	6.20	0.629
39	Laetiaprocerin C	871217-64-0	36.162	0.651	113.470	6.39	0.408	6.19	0.644
40	Ctremula 2	173324-51-1	18.667	0.561	105.240	5.94	1.152	6.16	0.689
41	Ccorymbosa 8	131740-59-5	18.156	0.570	103.650	5.90	1.263	6.15	0.710
42	Cgrewiifolia 2	947681-66-5	16.566	0.608	95.867	5.83	1.490	6.14	0.730
43	Cgrayi 4	947672-96-0	13.951	0.520	102.550	5.97	1.083	6.14	0.732
44	Casearinone A	196090-46-7	14.160	0.584	95.539	5.82	1.526	6.13	0.746
45	Caseargrewiin A	847841-92-3	28.140	0.591	115.910	6.04	0.907	6.10	0.797
46	Intrapetacin B	343944-36-5	22.594	0.564	112.070	5.94	1.138	6.10	0.798
47	Zguidonia 13	128572-39-4	23.501	0.591	109.730	5.92	1.193	6.10	0.799
48	Caseamembrin A	664351-45-5	28.559	0.526	124.870	6.16	0.685	6.10	0.801
49	Cgrewiifolia 1	947672-98-2	19.329	0.600	102.640	5.82	1.521	6.10	0.803
50	Corymbolusin C	326473-53-4	25.499	0.542	118.840	6.05	0.885	6.09	0.805
51	Ccorymbosa 11	131740-55-1	10.353	0.535	95.773	5.85	1.414	6.09	0.821
52	Casearluicin C	405507-82-6	27.452	0.615	111.790	5.99	1.028	6.08	0.840
53	Ccorymbosa 4	131740-57-3	7.484	0.312	87.138	7.69	0.021	6.07	0.851
54	Zguidonia 12	128486-37-3	20.046	0.561	109.730	5.85	1.402	6.06	0.873
55	Ccorymbosa 13	131740-56-2	18.140	0.576	105.420	5.77	1.694	6.05	0.897
56	Cgrewiifolia 3	1006868-59-2	21.095	0.556	112.190	5.87	1.349	6.04	0.909
57	Zuelanin 10	130827-57-5	22.216	0.580	110.920	5.84	1.439	6.03	0.931

**Table 3.** (cont.)

Item	Compound	CAS Reg. Nr.	Descriptors [a]			Method and predicted values			
			RDF060m	R8p	Fuk5	MLR		SVM	
						Y [b]	IC <sub>50</sub> [c]	Y [b]	IC <sub>50</sub> [c]
58	Caseamembrin B	664351-46-6	23.370	0.552	116.160	5.92	1.214	6.03	0.942
59	Casearnigrescens D	925693-22-7	20.241	0.540	113.280	5.87	1.355	6.02	0.953
60	Zuelanin 4	130858-02-5	29.974	0.615	116.840	5.98	1.057	5.99	1.019
61	Cgrayi 8	947673-01-0	7.541	0.305	64.534	8.77	0.002	5.97	1.071
62	Laetiaprocerin A	871217-62-8	20.205	0.585	109.160	5.71	1.946	5.96	1.101
63	Casearvestrin A	402937-50-2	19.118	0.568	109.930	5.71	1.944	5.96	1.106
64	Zuelanin 5	40185-32-8	4.900	0.358	61.245	8.29	0.005	5.94	1.140
65	Zuelanin 8	130858-01-4	19.344	0.526	115.300	5.81	1.566	5.94	1.141
66	Ccorymbosa 12	131830-93-8	13.135	0.615	95.867	5.48	3.324	5.94	1.144
67	Casearvestrin B	402935-05-1	19.317	0.560	111.700	5.71	1.956	5.93	1.167
68	Caseamembrin E	664351-50-2	24.923	0.518	124.870	5.91	1.225	5.91	1.238
69	Caseamembrin A	723302-01-0	24.923	0.518	124.870	5.91	1.225	5.91	1.238
70	Ctremula 5	173324-53-3	21.571	0.496	122.400	5.90	1.253	5.90	1.261
71	Zuelanin 2	128486-32-8	16.505	0.524	112.910	5.68	2.068	5.89	1.293
72	Corymbotin D	139015-30-8	15.862	0.608	102.150	5.48	3.304	5.89	1.299
73	Zguidonia 6	128486-33-9	24.789	0.588	116.840	5.73	1.844	5.86	1.375
74	Caseargrewiin K	945607-80-7	19.826	0.475	122.850	5.89	1.287	5.86	1.396
75	Zguidonia 9	128486-34-0	19.989	0.538	117.130	5.69	2.057	5.85	1.427
76	Laetiaprocerin D	871217-65-1	15.117	0.441	119.200	5.91	1.234	5.84	1.459
77	Caseargrewiin B	847841-93-4	24.532	0.511	126.810	5.84	1.435	5.84	1.460
78	Casearlucin D	405507-84-8	28.375	0.614	118.900	5.75	1.769	5.81	1.565
79	Casearlucin E	405507-86-0	23.045	0.585	116.260	5.63	2.328	5.80	1.582
80	Casearlucin I	405507-94-0	29.948	0.547	132.040	5.80	1.589	5.79	1.608
81	Casearlucin H	405507-92-8	30.892	0.560	132.040	5.78	1.653	5.78	1.651
82	Zguidonia 10	128486-35-1	3.889	0.328	112.100	6.12	0.757	5.77	1.714
83	Ctremula 4	173324-54-4	21.545	0.484	127.720	5.75	1.782	5.74	1.829
84	Zuelanin 11	130827-58-6	29.182	0.712	103.820	5.76	1.723	5.73	1.847
85	Zuelanin 6	130827-54-2	16.411	0.642	102.150	5.27	5.365	5.72	1.889
86	Laetiaprocerin B	871217-63-9	29.954	0.678	111.690	5.73	1.860	5.71	1.948
87	Corymbotin A	138967-76-7	18.050	0.680	99.605	5.24	5.767	5.66	2.177
88	Casearlucin B	405507-80-4	17.986	0.634	106.710	5.26	5.500	5.65	2.247
89	Corymbotin G	138950-32-0	15.416	0.600	109.890	5.15	7.052	5.60	2.506
90	Ctremula 6	173324-55-5	27.247	0.538	135.470	5.48	3.320	5.58	2.642
91	Zuelanin 7	130827-55-3	17.567	0.567	116.940	5.27	5.409	5.58	2.643
92	Casearlucin K	405507-98-4	25.316	0.543	132.040	5.43	3.712	5.55	2.843
93	Caseamembrin C	664351-47-7	27.330	0.532	138.430	5.40	4.008	5.54	2.908
94	Corymbotin E	138950-30-8	19.674	0.632	111.570	5.20	6.318	5.52	3.006
95	Ccorymbosa 9	131740-60-8	21.180	0.611	116.430	5.27	5.400	5.52	3.038
96	Ccorymbosa 10	131740-61-9	20.006	0.616	114.660	5.21	6.180	5.50	3.129
97	Caseamembrin M	870538-04-8	26.275	0.604	124.960	5.37	4.246	5.49	3.218
98	Caseamembrin D	664351-49-9	28.147	0.554	138.520	5.30	5.052	5.48	3.299
99	Caseamembrin B	723302-02-1	25.752	0.529	138.100	5.30	5.025	5.48	3.347
100	Caseamembrin F	664351-51-3	13.285	0.455	128.840	5.21	6.210	5.46	3.466
101	Caseargrewiin C	847841-94-5	20.266	0.554	125.010	5.23	5.869	5.46	3.468
102	Casearnigrescens B	925693-17-0	19.790	0.542	126.840	5.20	6.334	5.43	3.683
103	Caseamembrin O	870538-07-1	26.524	0.587	132.110	5.20	6.343	5.38	4.198
104	Zuelanin 12	130827-59-7	23.424	0.580	128.360	5.15	7.013	5.36	4.347
105	Corymbotin B	138950-28-4	17.570	0.667	109.030	4.87	13.543	5.33	4.698
106	Caseamembrin N	870538-05-9	24.144	0.591	128.970	5.11	7.852	5.32	4.827
107	Casearlucin J	405507-96-2	21.654	0.552	132.040	5.05	8.978	5.32	4.833
108	Casearnigrescens C	925693-19-2	25.321	0.614	126.930	5.12	7.505	5.30	4.984
109	Casearnigrescens A	925693-15-8	20.926	0.587	126.930	4.95	11.198	5.25	5.662
110	Casearin T	220178-33-6	32.143	0.715	119.330	5.29	5.111	5.22	5.957
111	Corymbotin I	138950-33-1	13.696	0.617	116.740	4.56	27.315	5.20	6.292
112	Corymbotin H	138967-77-8	18.419	0.641	119.320	4.67	21.321	5.10	8.009
113	Corymbotin C	138950-29-5	17.340	0.666	117.850	4.46	35.035	4.98	10.529

[a] Radial Distribution Function, R-GETAWAY and Fukui Function; [b]  $Y = \text{Log}(1/IC_{50})$ ; [c]  $\mu\text{mol/L}$ .

predicted by SVM QSAR. They should be considered as candidate compounds for further studies with the V-79 cell line. It must be taken into account that all of them have evidenced cytotoxicity in bioassay with other protocols. The challenge resides on the production of experimental cytotoxicity values and seeing how they would relate to the QSAR predicted  $IC_{50}$  results. For example, Casearborins A-E were tested in a routine cytotoxicity in vitro assay with SF539 human glioblastoma tumor cell line and human melanoma LOX IMVI cell line showing no differential cytotoxicity [46]; Caseargrewiin A-D and Caseargrewiin E-L exhibited cytotoxic activity against three cancer cell lines, namely KB (human oval epidermoid carcinoma), BC1 (human breast cancer) and NCI-H187 (human lung cancer cells) [47, 48]. Least but not last, Casearin S belongs to the *C. sylvestris* Sw. species which produced the Casearins used in the cytotoxic bioassay [6, 7], never tested against V-79 cells, but assayed in DNA-repair [49]. For compounds whose prediction is contained in the experimental range we have also some examples as follows. Casearinol A-B showed immunomodulatory activity inhibiting the binding of T-cell leukocyte function antigen 1 (LFA-1) to intercellular adhesion molecule 1 (ICAM-1) [50]. Caseamembrin A-F exhibited cytotoxicity against human prostate (PC-3) cancer cells and Hepatoma (Hep3B) cell line [51, 52]; Caseamembrin M-O exhibited bioactivity against KB, medulla (Med) and colon (DLD-1) cancer cell line [53]; Casearlucin A-K [54] as well as clerodane diterpenoids from *Casearia nigrescens* Tul. [55] exhibited cytotoxicity against the A2780 ovarian cancer cell line. Intrapetacin A-B are clerodane diterpenoids from *Licania intrapetiolaris* Spruce ex Hook (Chrysobalanaceae), displaying moderate cytotoxicity against the KB assay [56]. As compounds that have never been tested, to our knowledge, with any bioassay, we may mention clerodane diterpenoids from *Casea-*

*ria grayi* L. W. Jessup [57], *Casearia corymbosa* H. B. K. [58, 59], *Zuelania guidonia* (Sw.) Britton et Millsp. [60] and *Casearia tremula* Griesebach. [61].

### 3.4 Evaluation of New Compounds

The new compounds were named G1v to G13v and are in silico modifications of Casearin G substituents at position R<sup>1</sup> (carbon C2), R<sup>4</sup> (carbon C6) and R<sup>5</sup> (carbon C7), (Fig. 1). Table 4 contains the modifications and the respective results of predicted  $IC_{50}$  values with the MLR and SVM QSARs.

The MLR QSAR extrapolates the value of six compounds with higher  $IC_{50}$  value than the experimental value of Casearin G ( $IC_{50}$  = 0.170  $\mu$ mol/L). The SVM QSAR extrapolates the value of seven compounds. Except for compound Gv3, all others were not in the AD, namely, Gv5, Gv7, Gv9, Gv11, Gv12 and Gv13. The compounds substituted at R<sup>4</sup> and R<sup>5</sup> with fluorine, and at R1 with fluorine or chlorine, are all out of the AD. All other chlorine substituted compounds in R<sup>4</sup> and R<sup>5</sup> are contained in the AD (Gv4, Gv6, Gv8, and Gv10), with predicted  $IC_{50}$  values less bioactive than Casearin G. The compounds Gv1 and Gv2 were substituted with a bulkier substituent at R<sup>5</sup> (Dc = Decanoate) resulting less bioactive than Casearin G. Gv3 substituent at R<sup>5</sup> with a MeO resulted to be slightly close to the experimental  $IC_{50}$  value of Casearin G. Contrary to the preceding case, we have no known cytotoxic bioassay examples from literature employing fluorinated or chlorinated diterpenoids, it is reasonable to adopt the interpolated compounds for further bioassays. This in silico exercise was intended to evaluate a few possibilities in line with the objective of this work, exemplifying the use of the proposed QSAR model with new clerodane diterpenoids as anticancer agents.

**Table 4.** Predicted  $IC_{50}$  values of modified Casearin G at R<sup>1</sup>, R<sup>4</sup> and R<sup>5</sup> (see Fig. 1).

Item	Substituents					Compound	Descriptors [d]			Method and predicted values			
	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	R <sup>5</sup>		RDF060m	R8p	Fuk5	MLR		SVM	
										Y [b]	$IC_{50}$ [c]	Y [b]	$IC_{50}$ [c]
1	OMe	Ac	Ac	H	F	Gv5	24.474	0.550	88.046	7.30	0.050	7.00	0.101
2	OH	Ac	Ac	H	F	Gv11	22.054	0.493	90.590	7.41	0.039	6.96	0.109
3	Cl	Ac	Ac	H	OBu	Gv13	27.603	0.613	87.188	7.13	0.074	6.88	0.132
4	F	Ac	Ac	H	OBu	Gv12	28.341	0.573	70.954	8.24	0.006	6.85	0.141
5	OH	Ac	Ac	F	H	Gv9	19.609	0.494	90.590	7.19	0.064	6.80	0.157
6	OAc	Ac	Ac	H	OMe	Gv3	28.427	0.597	99.605	6.76	0.173	6.78	0.168
7	OMe	Ac	Ac	F	H	Gv7	20.685	0.551	88.046	6.97	0.107	6.77	0.170
8	OMe	Ac	Ac	H	ODc	Gv1	31.074	0.529	116.300	6.75	0.179	6.63	0.235
9	OMe	Ac	Ac	H	OBu	Casearin G [a]	31.047	0.610	105.670	6.60	0.251	6.61	0.245
10	OH	Ac	Ac	Cl	H	Gv8	25.053	0.519	110.780	6.56	0.278	6.55	0.282
11	OH	Ac	Ac	H	Cl	Gv10	26.320	0.541	110.780	6.50	0.318	6.53	0.293
12	OH	Ac	Ac	H	ODc	Gv2	28.791	0.500	118.840	6.66	0.221	6.50	0.318
13	OMe	Ac	Ac	H	Cl	Gv4	28.360	0.601	108.240	6.33	0.464	6.40	0.396
14	OMe	Ac	Ac	Cl	H	Gv6	26.169	0.588	108.240	6.24	0.571	6.36	0.439

[a] Casearin G highest observed  $IC_{50}$  value: 0.17  $\mu$ mol/L, [Log (1/ $IC_{50}$ ) = 6.77]; [b]  $Y = \text{Log}(1/IC_{50})$ ; [c]  $\mu$ mol/L; [d] Radial Distribution Function, R-GET-AWAY and Fukui Function.

## 4 Conclusions

Two independent methods, MLR and SVM were used for the development of externally validated QSARs. Three variables using a small data set of eighteen compounds have been shown to be feasible, meeting all test requirements recommended by literature. In our evaluation, the best QSAR was obtained with SVM method. It was applied to predict the bioactivity of 113 clerodane diterpenoids and 13 new compounds not tested in cytotoxic activity against Chinese Hamster V-79 lung cells and/or antitumoral activity against Sarcoma 180 ascites in mice. The predicted results of compounds with extrapolated  $IC_{50}$  values should be considered with care. It should not be discarded for further studies, in silico and/or experimental, as clerodane diterpenoids are known for their cytotoxic and antitumoral activity. The majority of the compounds that have been predicted by their  $IC_{50}$  values are contained in the range of the experimental  $IC_{50}$  values of the data set used to develop the QSAR, providing opportunities for experimental research of anticancer chemical agents with the V-79 cell line. The QSAR was applied to clerodane diterpenoids chosen from selected literature covering the last 20 years, concentrated mainly on plant species of the *Flacourtiaceae*. Considering that more than 800 clerodane diterpenoids are reported in literature, opportunities are open for additional in silico research, proposing new compounds in the aid of the development of new pharmaceuticals for cancer treatment.

## 5 Acknowledgements

We gratefully acknowledge to Prof. Dr. Anderson Coser Gaudio for providing the software BuildQSAR, to Mr. João Paulo A. Martins and Prof. Dr. Marcia M. C. Ferreira for providing the software QsarModeling, to FAPESP and CNPq.

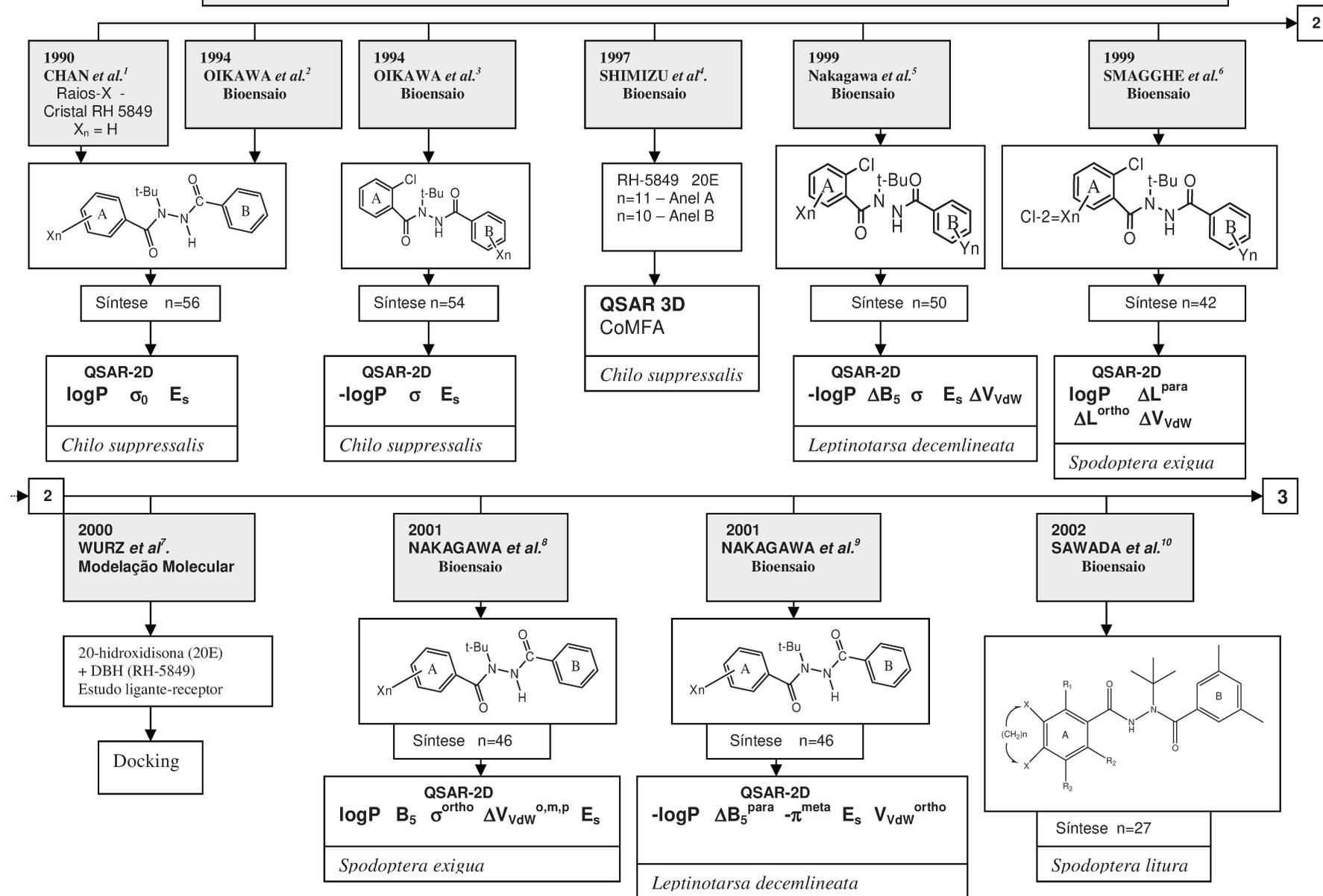
## 6 References

- [1] G. M. Barroso, *Sistemática de Angiospermas do Brasil*, Vol. 1, Livros Técnicos e Científicos Editora S. A., São Paulo **1978**, p. 171.
- [2] H. A. Melchior, *Engler's Syllabus der Pflanzenfamilien*, 12th. ed., Gebrüder Borntraeger, Berlin **1964**, pp. 27–28, 322–323.
- [3] A. T. Merrit, S. V. Ley, *Nat. Prod. Rep.* **1992**, *9*, 243–287.
- [4] J. R. Hanson, *Nat. Prod. Rep.* Review articles of Diterpenoids from **1983–2007**.
- [5] T. Busch, A. Kirsching, *Nat. Prod. Rep.* **2008**, *25*, 318–341.
- [6] H. Itokawa, N. Totsuka, H. Morita, K. Takeya, Y. Iitaka, E. P. Schenkel, M. Motidome, *Chem. Pharm. Bull.* **1990**, *38*, 3384–3388.
- [7] H. Morita, M. Nakayama, H. Kojima, K. Takeya, H. Itokawa, E. P. Schenkel, M. Motidome, *Chem. Pharm. Bull.* **1991**, *39*, 693–697.
- [8] W. Chaung, L. Mi, R. J. Boorstein, *Nucleic Acid Res.* **1997**, *25*, 992–994.
- [9] C. Isanbor, D. O'Hagan, *J. Fluorine Chem.* **2006**, *127*, 303–319.
- [10] J. -P. Begué, D. Bonnet-Delpon, *Chimie Bioorganique et Médicinal du Fluor*, CNRS, Paris **2005**.
- [11] J.-P. Begué, D. Bonnet-Delpon, *J. Fluorine Chem.* **2006**, *127*, 992–1012.
- [12] M. T. D. Cronin, T. W. Schultz, *J. Mol. Struct. (Theochem.)* **2003**, *622*, 39–51.
- [13] D. Young, T. Martin, R. Venkatapathy, P. Harten, *QSAR Comb. Sci.* **2008**, *27*, 1337–1345.
- [14] *SciFinder Scholar*, v. 2007, Copyright **2007**, American Chemical Society.
- [15] *ChemDraw v.5.0*, Copyright **1986–2004** CambridgeSoft.
- [16] *Chem3D v. 5.0*, Copyright **1986–2004** CambridgeSoft.
- [17] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, *Gaussian 03, Revision B.05*, Gaussian, Inc., Pittsburgh, PA **2003**.
- [18] *MOPAC 7.0*, Copyright Steward Computational Chemistry **2007**.
- [19] B. Wang, G. P. Ford, *J. Comp. Chem.* **1994**, *15*, 200–207.
- [20] *CaChe WorsystemPro*, v. 6.1.12.33, Copyright 2000–2004, Fujitsu Limited. 1989–2000 Oxford Molecular Ltd.
- [21] *Talete srl, DRAGON for Windows* (Software for Molecular Descriptor Calculations), V. 5.4 – 2006.
- [22] *MarvinView v. 5.0.5*, ChemAxon LTD, Copyright 1998–2008.
- [23] *Molinspiration*, <http://www.molinspiration.com/services>, Accessed July, **2008**.
- [24] N. Sukumar, C. M. Breneman, *Program RECON*, v. 5.3, Copyright, **2001**, Rensselaer Polytechnic Institute.
- [25] C. M. Breneman, C. M. Sundling, N. Sukumar, L. Shen, W. P. Katt, M. J. Embrechts, *J. Comput.-Aided Mol. Des.* **2003**, *17*, 231–240.
- [26] C. M. Breneman, T. R. Thomson, M. Rhem, M. Dung, *Computers Chem.* **1995**, *19*, 161–179.
- [27] J. Gasteiger, *PETRA Online Service-Parameter Estimation for the Treatment of Reactivity Applications*, Copyright Johann Gasteiger, 1998. <http://www2.ccc.uni-erlangen.de/services/petra/smiles.html>, access: November, **2008**.
- [28] *Talete srl, MobyDigs Professional*, v. 1.0, 2004.
- [29] *BuildQSAR*, v. 2.0.0.0, Federal University of Espírito Santo, A. C. Gaudio, Copyright **2007**.
- [30] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, *Mobydigs: Software for regression and classification models by genetic algorithms*, in: *Chemometrics: Genetic Algorithms and Artificial*

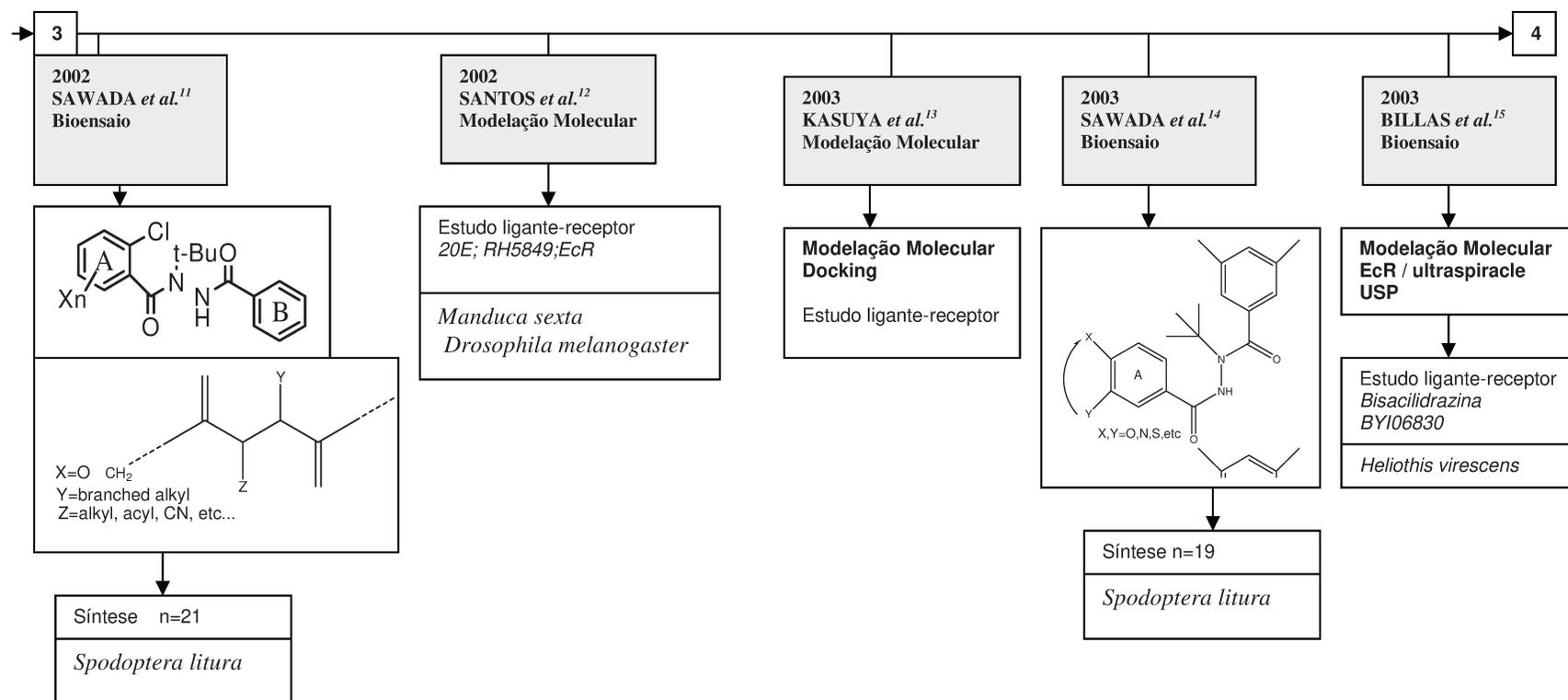
- cial Neural Networks* (Ed: R. Leardi), Elsevier, Amsterdam **2003**, ch. 5.
- [31] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, *Anal. Chim. Acta* **2004**, *515*, 199–208.
- [32] A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **2003**, *22*, 69–77.
- [33] *Molegro Data Modeller (MDM)*, v. 1.5.0, Copyright Molegro **2008**.
- [34] *Pirouette*, v. 3.11, Infometrix Inc., Copyright **1990–2003**.
- [35] M. C. Hemmer, V. Steinhauer, J. Gasteiger, *Vibrat. Spect.* **1999**, *19*, 151–164.
- [36] *Dragon Molecular Descriptor Theory*, in: *Talete srl, DRAGON for Windows*, v. 5.4, **2006**.
- [37] V. Consonni, R. Todeschini, M. Pavan, *J. Chem. Inf. Sci.* **2002**, *42*, 682–692.
- [38] A. Golbraikh, A. Tropsha, *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.
- [39] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, Berlin **1995**.
- [40] S. R. Gunn, *Support Vector Machines for Classification and Regression. Technical Report*, University of Southampton **1998**; available at <http://users.ecs.soton.ac.uk/srg/publications/pdf/SVM.pdf>; access: November, 2008.
- [41] C. W. Hsu, C. C. Chang, C. J. Lin, *A Practical Guide to Support Vector Classification* **2008**; available at: [www.csie.ntu.edu.tw/~cjlin](http://www.csie.ntu.edu.tw/~cjlin). Access: Oct., 2008
- [42] *Molegro Data Modeller User Manual*, Copyright Molegro **2008**.
- [43] B. Niu, W. Lu, S. Yang, Y. Cai, G. Li, *Acta Pharm. Sinica* **2007**, *28*, 1075–1086.
- [44] X. J. Yao, A. Panaye, J. P. Doucet, R. S. Zhang, H. F. Chen, M. C. Liu, Z. D. Hu, B. T. Fan, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1257–1266.
- [45] S. Oloff, R. B. Mailman, A. Tropsha, *J. Med. Chem.* **2005**, *48*, 7322–7332.
- [46] J. A. Beutler, K. L. McCall, K. Herbert, D. L. Herald, G. R. Pettit, T. Johnson, R. H. Shoemaker, M. R. Boyd, *J. Nat. Prod.* **2000**, *63*, 657–661.
- [47] S. Kanokmedhakul, K. Kanokmedhakul, T. Kanarsa, M. Buayairaksa, *J. Nat. Prod.* **2005**, *68*, 183–188.
- [48] S. Kanokmedhakul, K. Kanokmedhakul, M. Buayairaksa, *J. Nat. Prod.* **2007**, *70*, 1122–1126.
- [49] P. R. F. De Carvalho, M. Furlan, M. C. M. Young, D. G. I. Kingston, V. Da S. Bolzani, *Phytochemistry* **1998**, *49*, 1659–1662.
- [50] M. S. Hunter, D. G. Corley, C. P. Carron, E. Rowold, B. F. Kilpatrick, R. C. Durley, *J. Nat. Prod.* **1997**, *60*, 894–899.
- [51] Y. C. Shen, C. H. Wang, Y. B. Cheng, L. T. Wang, J. H. Guh, C. T. Chien, A. T. Khalil, *J. Nat. Prod.* **2004**, *67*, 316–321.
- [52] D. M. Huang, Y. C. Shen, C. Wu, Y. T. Huang, F. L. Kung, C. M. Teng, J. H. Guh, *Eur. J. Pharmacol.* **2004**, *503*, 17–24.
- [53] Y. C. Shen, Y. B. Chen, A. F. Ahmed, C. L. Lee, S. Y. Chen, C. T. Chien, Y. H. Kuo, G. L. Tzeng, *J. Nat. Prod.* **2005**, *68*, 1665–1668.
- [54] C. V. S. Prakash, J. M. Hoch, D. G. I. Kingston, *J. Nat. Prod.* **2002**, *65*, 100–107.
- [55] R. B. Williams, A. Norris, J. S. Mille, C. Birkinshaw, F. Ratooson, R. Andriansiferana, V. E. Rasamison, D. G. I. Kingston, *J. Nat. Prod.* **2007**, *760*, 206–209.
- [56] N. H. Oberlies, J. P. Burgess, H. A. Navarro, R. E. Pinos, D. D. Soejarto, N. R. Farnsworth, A. D. Kinghorn, M. C. Wani, M. E. Wall, *J. Nat. Prod.* **2001**, *64*, 497–501.
- [57] A. Mosaddik, R. Booth, P. G. Waterman, *Nat. Prod. Commun.* **2006**, *1*, 441–448.
- [58] T.-B. Chen, D. F. Wiemer, *J. Nat. Prod.* **1991**, *54*, 1612–1618.
- [59] M. R. Khan, A. I. Gray, I. H. Sadler, P. G. Waterman, *Phytochemistry* **1990**, *29*, 3591–3595.
- [60] M. R. Khan, A. I. Gray, P. G. Waterman, *Phytochemistry* **1990**, *29*, 2939–2942.
- [61] S. Gibbons, A. I. Gray, P. G. Waterman, *Phytochemistry* **1996**, *41*, 565–570.



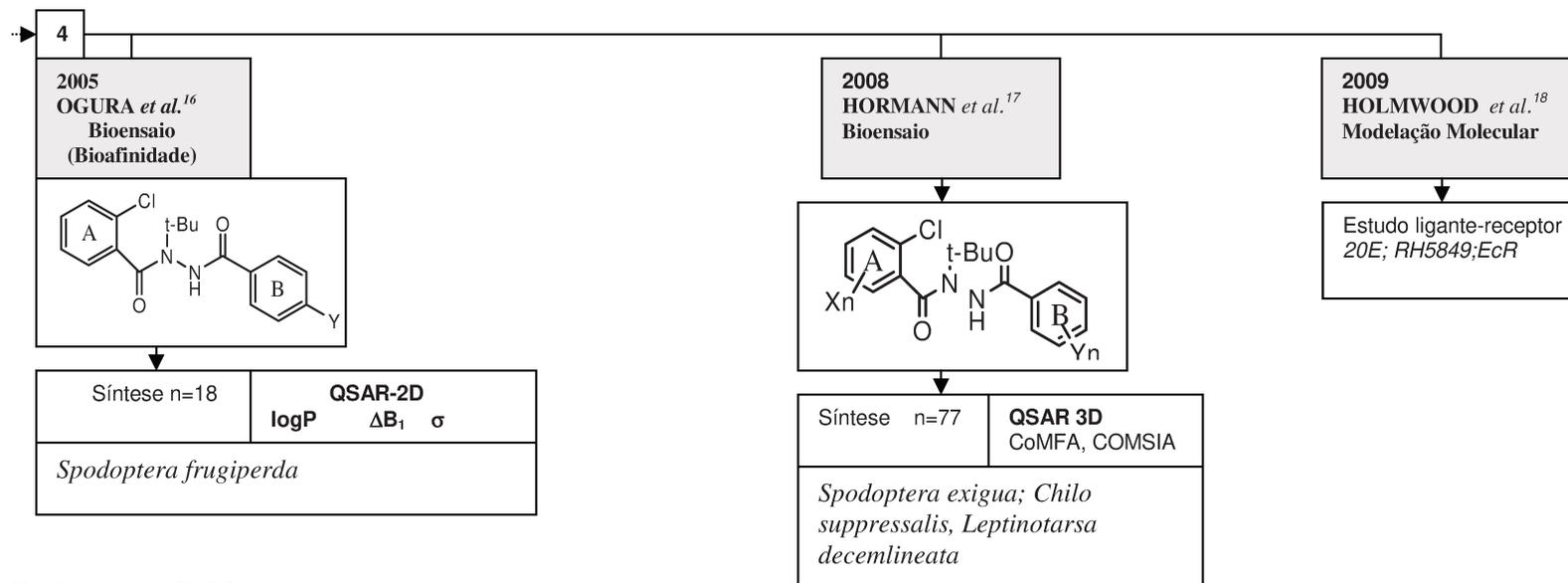
**Pesquisas realizadas com N,N'-dibenzoil-N-t-butilidrazinas (DBHs) Período: 1990 – 2009**  
**Estudos de DBHs contra insetos da ordem Lepidoptera e Coleoptera**



**Pesquisas realizadas com N,N'-dibenzoil-N-t-butildrazinas (DBHs) Período: 1990 – 2009**  
**Estudos de DBHs contra insetos da ordem Lepidoptera e Coleoptera**



**Pesquisas realizadas com N,N'-dibenzoil-N-t-butildrazinas (DBHs) Período: 1990 – 2009 - Estudos de DBHs contra insetos da ordem Lepidoptera e Coleoptera**



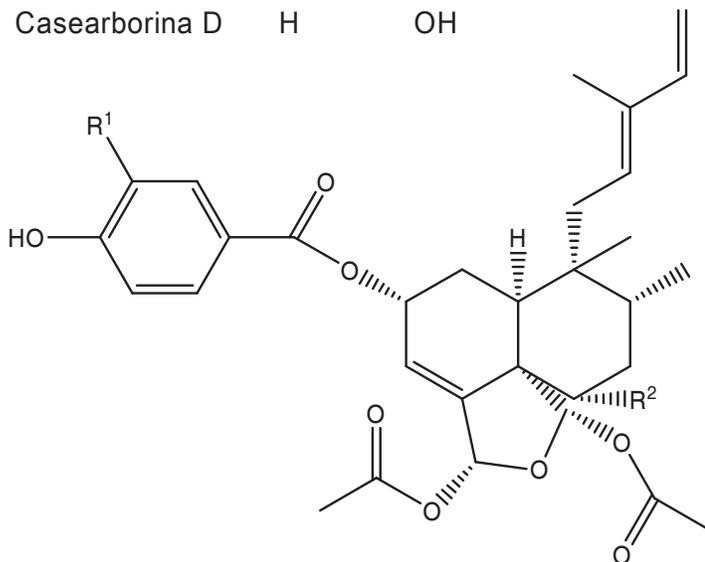
**Referências Bibliográficas**

- Chan, T.H.; Ali, A.; Britten, J.F. *Canadian Journal of Chemistry*, 68, 1178-1181, **1990**
- Oikawa,N.; Nakagawa,B. *Pesticide Science*. 41, 139-148, **1994**
- Oikawa, Nobuhiro; Nakagawa, Yoshiaki; Nishimura, Keiichiro; Ueno, Tamio; Fujita, Toshio. *Pesticide Biochemistry and Physiology*, 48, 135-144, **1994**.
- Shimizu,Bun-ichi; Nakagawa,Yoshiaki; Hattori,Kazunari; Nishimura,Keijchiro; Kurahara,Norio; Ueno,Tamio. *Steroids*, 62, 638-642, **1997**
- Nakagawa,Yoshiaki; Smagghe,Guy; Kugimiya,Soichi; Hattori,Kazunari; Ueno,Tamio; Tirry,Luc; Fujita,Toshio. *Pesticide Science*, 55, 909-918, **1999**
- Smagghe,Guy; Nakagawa,Yoshiaki; Carton,Bert; Mourad,Ahmed Kamal; Fujita,Toshio;Tirry,Luc. *Archives of Insect Biochemistry and Physiology*, 41, 42-53, **1999**
- Wurtz,Jean-Marie; Guillot,Benoit; Fagart,Jerome; Moras, Dino; Tietjen,Klaus; Schindler,Michael. *Protein Science*, 9, 1073-1084, **2000**
- Nakagawa,Yoshiaki; Smagghe,Guy; Tirry,Luc; Fujita,Toshio. *Pest Management Science*, 58, 131-138, **2001**
- Nakagawa, Yoshiaki; Smagghe, Guy; Van Paemel, Marleen; Tirry, Luc; Fujita, Toshio. *Pest Management Science*, 57, 858-865, **2001**.
- Sawada, Y.; Yanai, T.; Nakagawa, H.; Tsukamoto, Y.; Yokoi, S.; Yanagi, M.; Toya, T.; Sugizaki, H.; Kato, Y.; Shirakura, H.; Watanabe, T.; Yajjima, Y.; Kodama, S.; Masui, A. *Pest Management Science*, 59, 36- 48, **2002**
- Sawada Y.; Yanai, T.; Nakagawa, H.; Tsukamoto, Y.; Tamagawa, Y.; Yokoi, S.; Yanagi, M.; Toya, T.; Sugizaki, H.; Kato, Y.; Shirakura, H.; Watanabe, T.; Yajjima, Y.; Kodama, S.; Masui, A. *Pest Management Science*, 59, 49-57, **2002**
- Santos, Ana Cristina S.; Sant' Anna, Carlos Maurício R. *Journal of Molecular Structure (Theochem)*, 585, 61-68, **2002**
- Kasuya, Atsushi; Sawada, Yoshihiro; Tsukamoto, Yoshihisa; Tanaka, Keiji; Toya, Tetsuya; Yanagi, Mikio. *J. Mol. Model*, 9, 58-65, **2003**.
- Sawada, Y.; Yanai, T.; Nakagawa, H.; Tsukamoto, Y.; Yokoi, S.; Yanagi, M.; Toya, T.; Sugizaki, H.; Kato, Y.; Shirakura, H.; Watanabe, T.; Yajjima, Y.; Kodama, S.; Masui, A. *Pest Management Science*, 59, 25-35, **2003**.
- Billas, Isabelle M.L.; Iwema, Thomas; Garnier, Jean-Marie; Mitschler, A.; Rochel, N.; Moras, D. *Nature*, 426, 91-96, **2003**
- Ogura, Takehico; Nakagawa, Yoshiaki; Minakuchi, Chieka; Miyagawa, Hisashi. *J. Pest. Sci.*, 30, 1-6, **2005**.
- Hormann, R. E.; Snagghe, G.; Nakagawa, Y. *QSAR & Combinatorial Science*. 27, 1098-1112, **2008**
- Holmwood, Graham; Schindler, Michael. *Bioorganic & Medicinal Chemistry*, 17, 4064-4070, **2009**

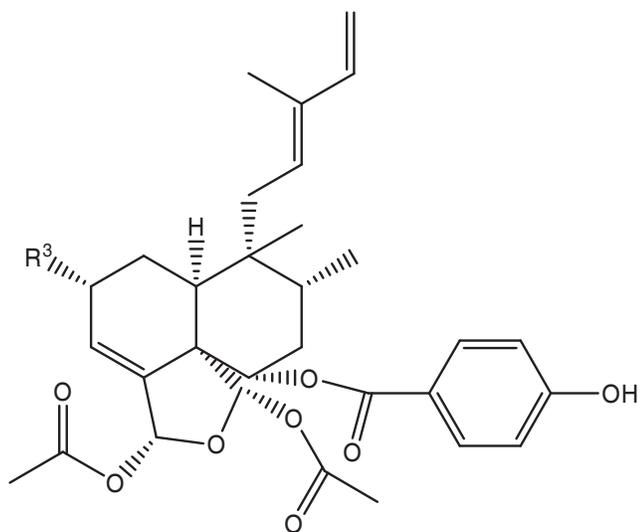


*Casearia arborea* (L.C.Rich.) Urban Flacourtiaceae/Salicaceae

Nome	R <sup>1</sup>	R <sup>2</sup>
Casearborina A	H	H
Casearborina B	MeO	H
Casearborina D	H	OH

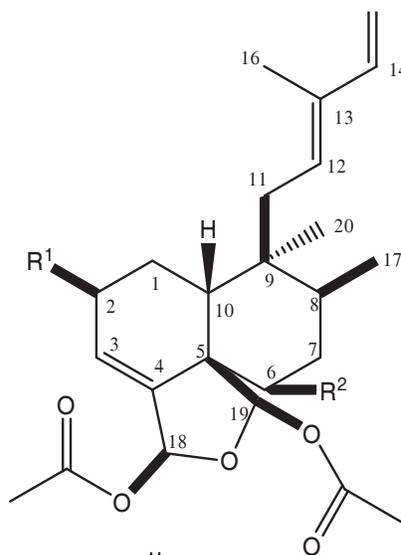


Nome	R <sup>3</sup>
Casearborina C	OH
Casearborina E	OAc

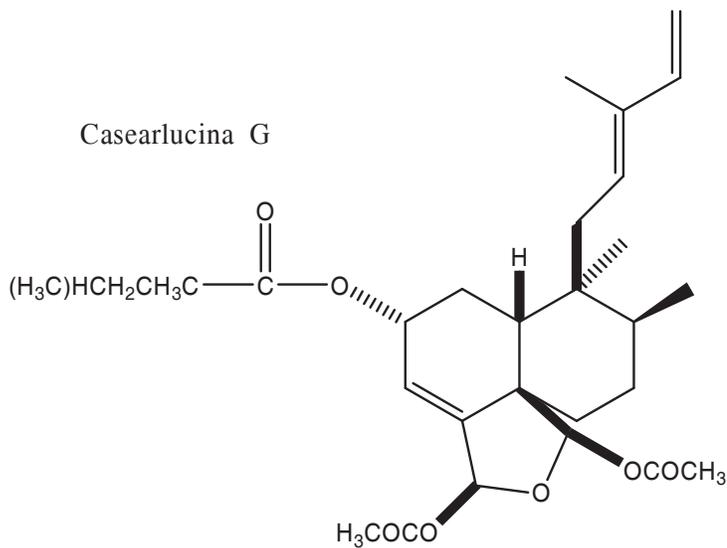


*Casearia lucida* Hils. & Bojer ex Tul. Flacourtiaceae/Salicaceae

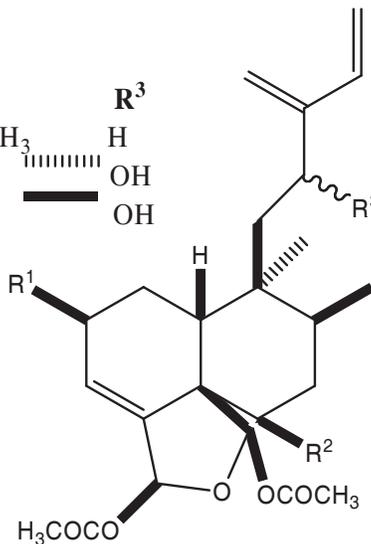
Nome	R <sup>1</sup>	R <sup>2</sup>
Casearlucina A	H	H
Casearlucina C	MeO	H
Casearlucina F	H	OH



Casearlucina G

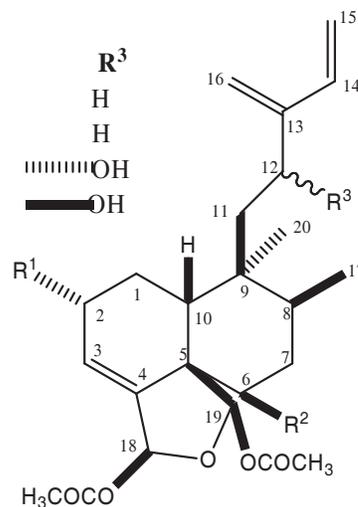


Nome	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>
Casearlucina D	CH <sub>3</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )COO	OCOCH <sub>3</sub>	H
Casearlucina H	CH <sub>3</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )COO	OH	OH
Casearlucina I	CH <sub>3</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )COO	OH	OH



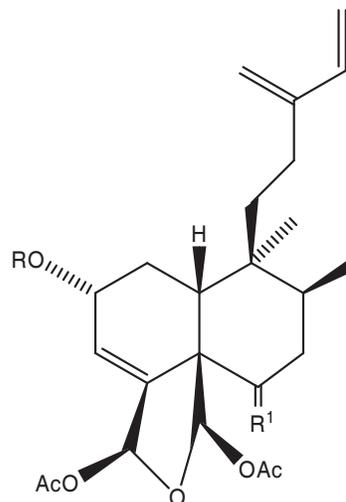
*Casearia lucida* Hils. & Bojer ex Tul. Flacourtiaceae/Salicaceae

Nome	R <sup>1</sup>	R <sup>2</sup>
Casearlucina B	CH <sub>3</sub> COO	OCH <sub>3</sub>
Casearlucina E	CH <sub>3</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )COO	OCH <sub>3</sub>
Casearlucina J	CH <sub>3</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )COO	OH
Casearlucina K	CH <sub>3</sub> CH <sub>2</sub> CH(CH <sub>3</sub> )COO	OH



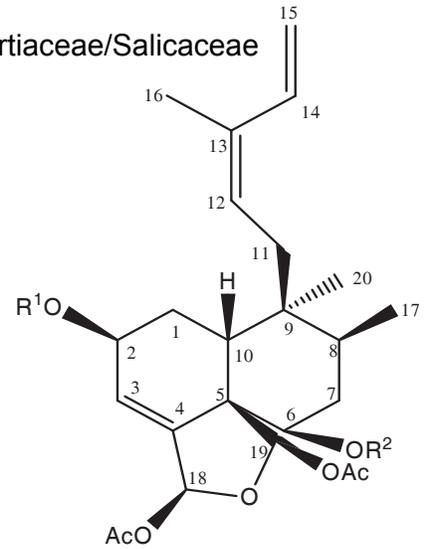
*Casearia grewiifolia* var. *gelonioides* (Blume) Sleumer Flacourtiaceae/Salicaceae *sensu lato*

Nome	R	R <sup>1</sup>
Cgrewiifolia 1	CH <sub>3</sub> CO	O
Cgrewiifolia 2	CH <sub>3</sub> CO	H <sub>2</sub>
Cgrewiifolia 3	EtCH(Me)CO	O
Cgrewiifolia 4	Me <sub>2</sub> CHCO	H <sub>2</sub>
Cgrewiifolia 5	EtCH(Me)CO	H <sub>2</sub>

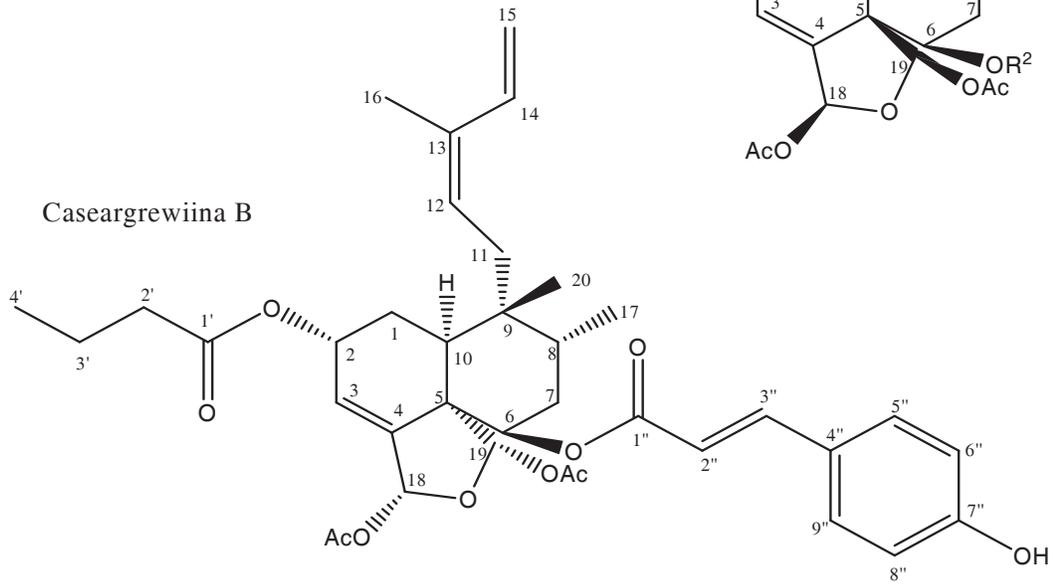


*Casearia grewiifolia* Vent. Flacourtiaceae/Salicaceae

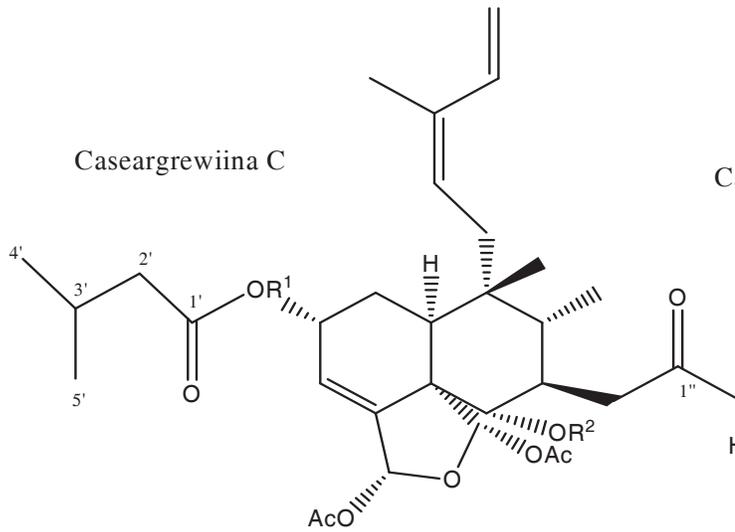
Nome R<sup>1</sup> R<sup>2</sup>  
Caseargrewiina A (CH<sub>3</sub>)<sub>2</sub>CHCH<sub>2</sub>CO CH<sub>3</sub>



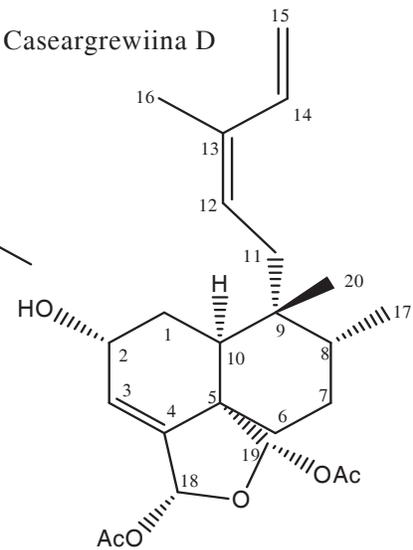
Caseargrewiina B



Caseargrewiina C

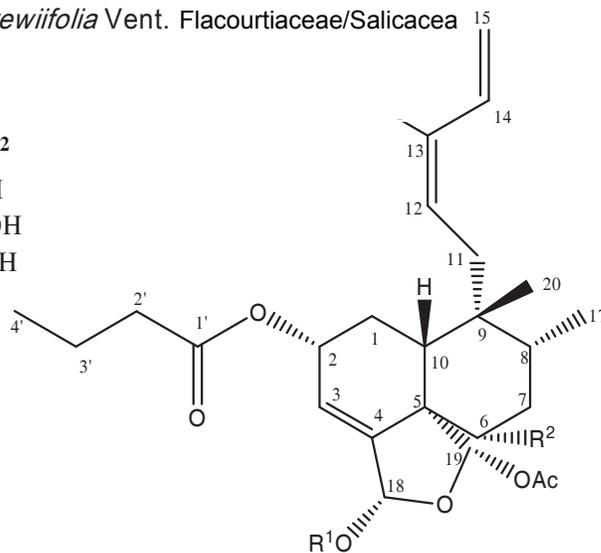


Caseargrewiina D

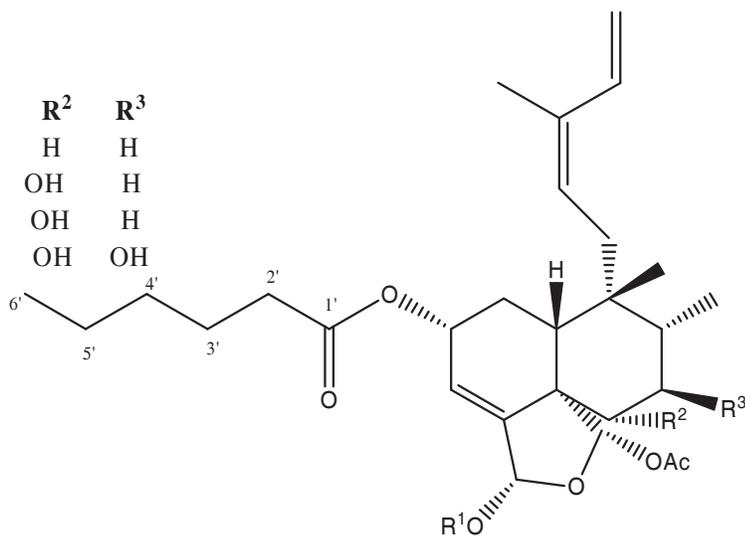


*Casearia grewiifolia* Vent. Flacourtiaceae/Salicacea

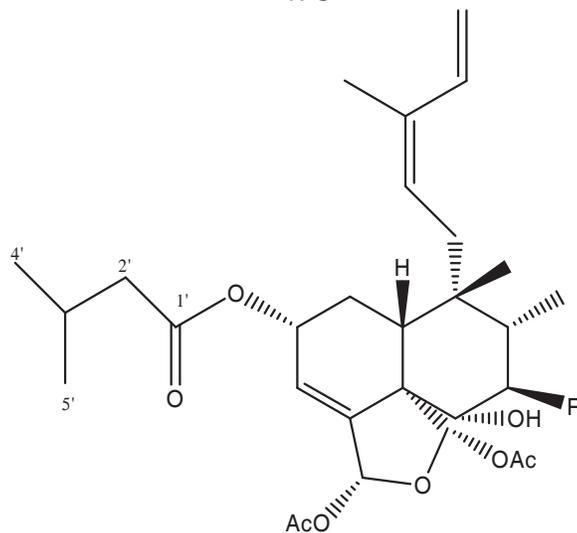
Nome	R <sup>1</sup>	R <sup>2</sup>
Caseargrewiina E	Ac	H
Caseargrewiina F	Ac	OH
Caseargrewiina G	Me	OH



Nome	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>
Caseargrewiina H	Ac	H	H
Caseargrewiina I	Ac	OH	H
Caseargrewiina J	Me	OH	H
Caseargrewiina K	Me	OH	OH

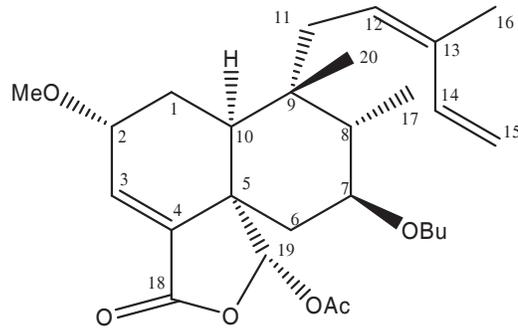


Nome	R
Caseargrewiina L	H

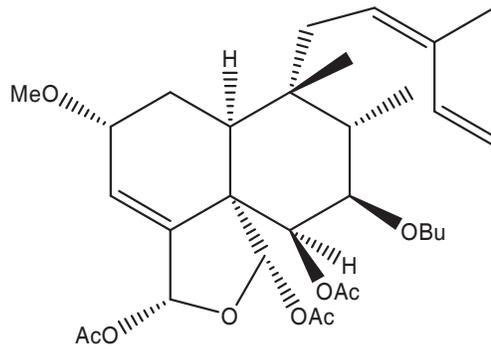


*Casearia sylvestris* Swartz | Flacourtiaceae/Salicaceae

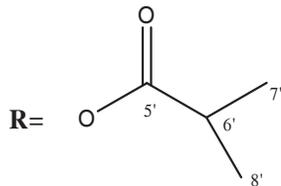
Casearina S



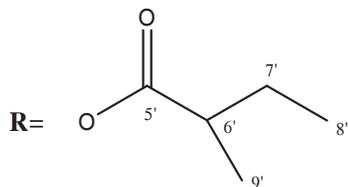
Casearina T



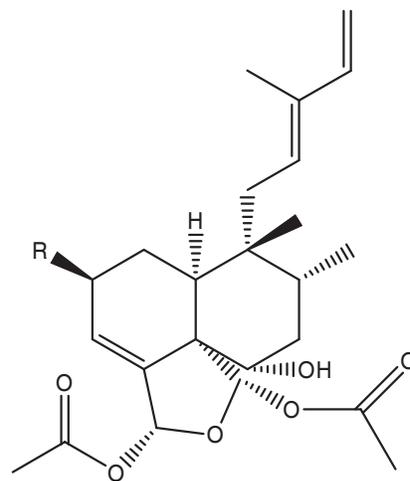
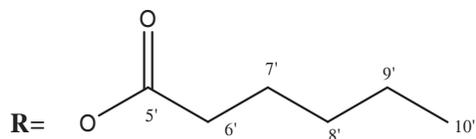
Casearvestrina A



Casearvestrina B

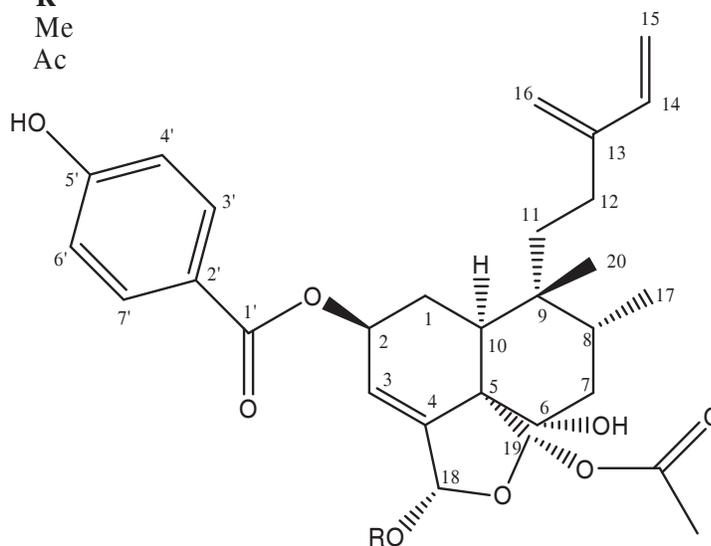


Casearvestrina C



*Licania intrapeciolaris* Spruce ex Hook Chrysobalanaceae

Nome	R
Intrapetacina A	Me
Intrapetacina B	Ac



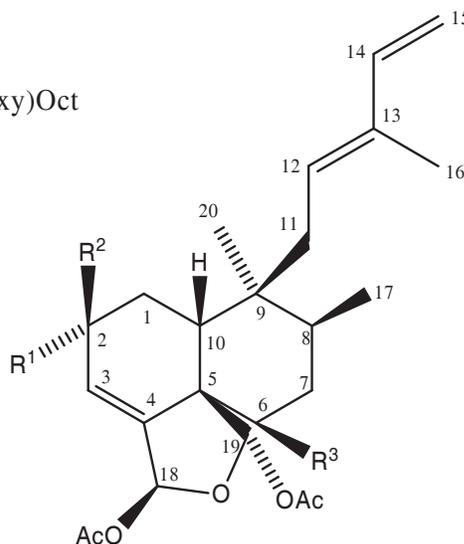
*Zuelania guidonia* (Sw.) Britton et Millsp. Flacourtiaceae/Salicaceae

Nome	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>
Zuelanina1	OH	H	O-Cin
Zuelanina2	O-Cin	H	OH
Zuelanina3	H	OH	O-Cin
Zuelanina6	OAc	H	OH
Zuelanina7	O-Oct	H	OH
Zuelanina8	H	O-Oct	OH
Zuelanina9	H	O-Bz	OH
Zuelanina11	OH	H	O-Bz
Zuelanina12	OH	H	O-(3-hydroxy)Oct
Zuelanina13	H	O-Bz	H

Cin=Cinnamoyl

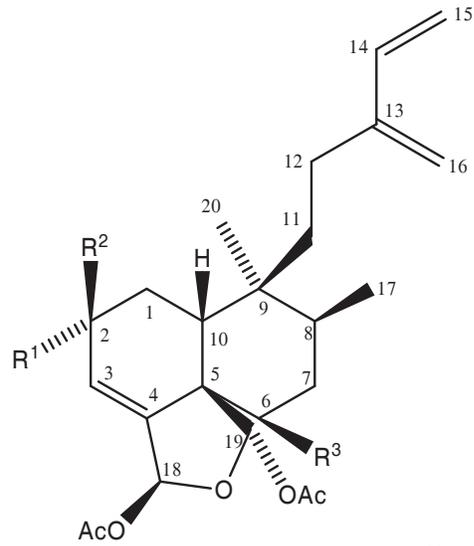
Bz=Benzoyl

Oct=Octahydroic acid

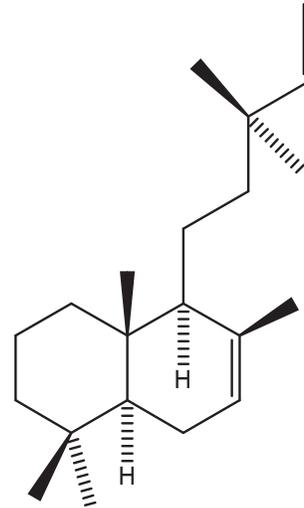


*Zuelania guidonia* (Sw.) Britton et Millsp. Flacourtiaceae/Salicaceae

Nome	R1	R2	R3
Zuelanina4	H	Oac	O-Cin
Zuelanina10	H	O-Bz	OH

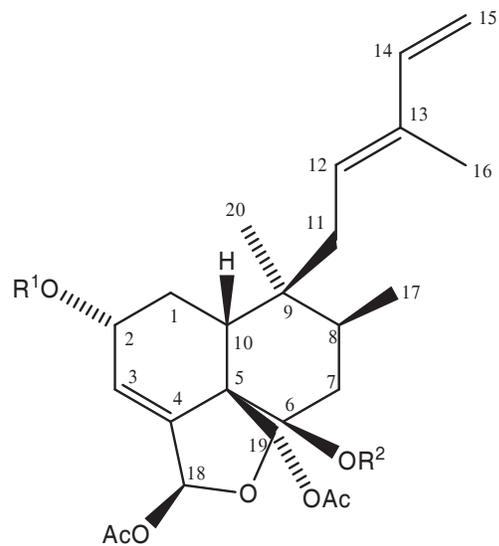


Zuelanina5



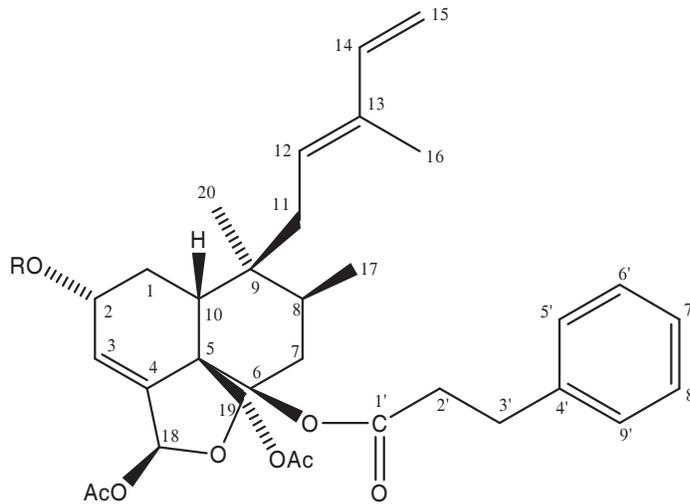
Nome	R1	R2
Zguidonia3	H	Cin
Zguidonia4	Cin	H
Zguidonia11	Ac	Cin
Zguidonia12	Cin	Ac

Cin=Cinnamoyl

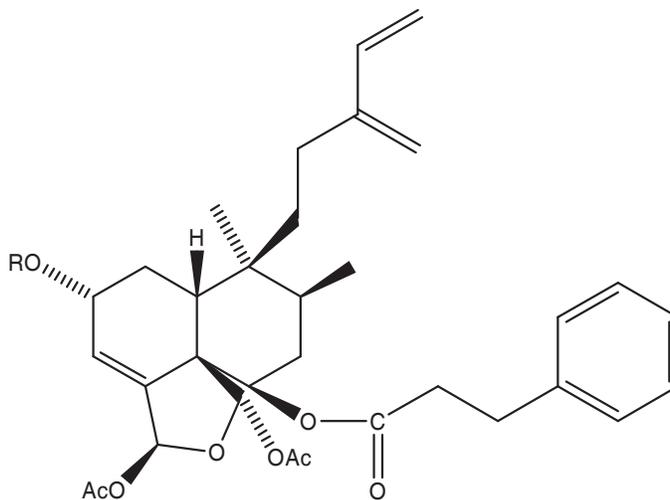


*Zuelania guidonia* (Sw.) Briton et Millsp. Flacourtiaceae/Salicaceae

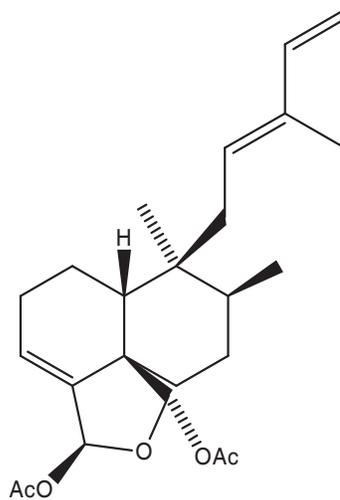
<b>Nome</b>	<b>R</b>
Zguidonia5	H
Zguidonia13	Ac



Zguidonia6



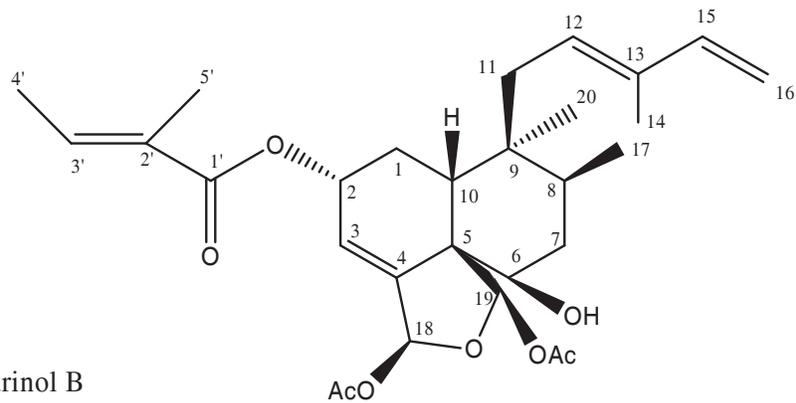
Zguidonia7



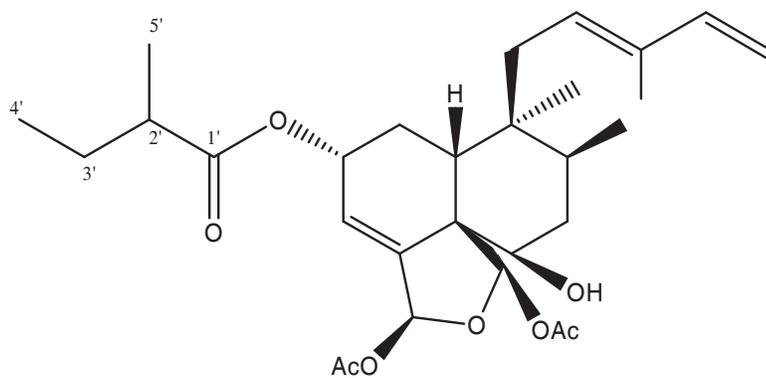


*Casearia guianensis* J. R. Johnston Flacourtiaceae/Salicaceae

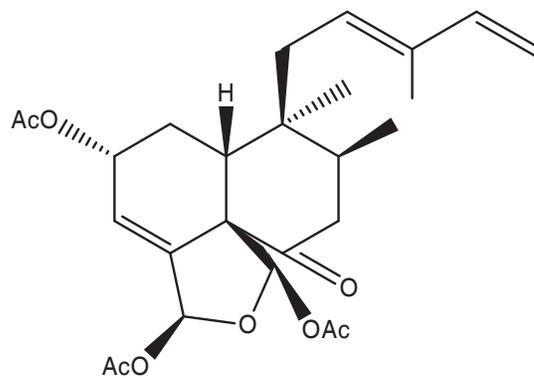
Casearinol A



Casearinol B

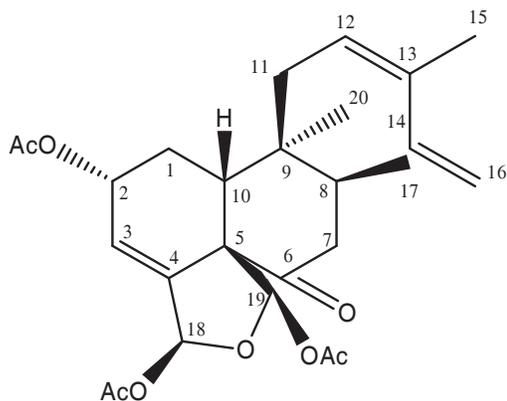


Casearinona A



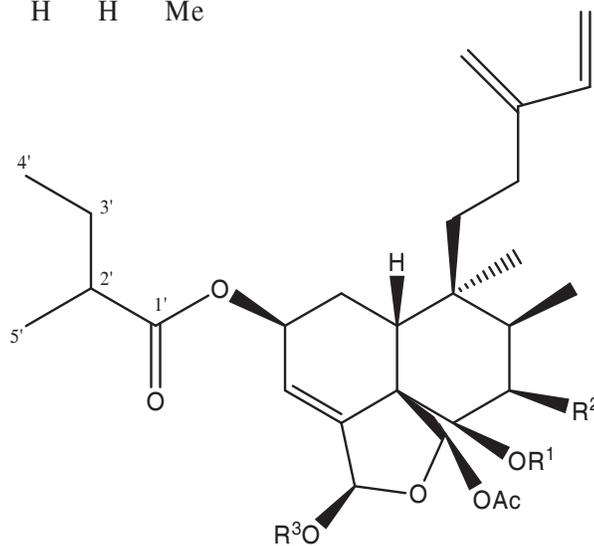
*Casearia guianensis* J. R. Johnston Flacourtiaceae/Salicaceae

Casearinona B



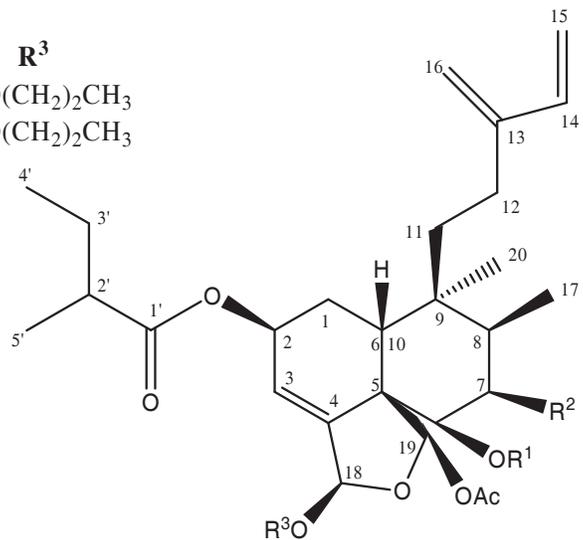
*Casearia membranacea* Hance Flacourtiaceae/Salicaceae

Nome	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>
Caseamembrina A	H	H	CO(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>
Caseamembrina B	H	H	Me

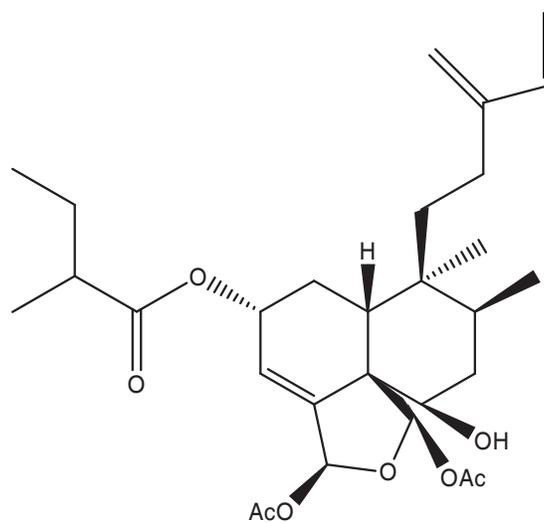


*Casearia membranacea* Hance Flacourtiaceae/Salicaceae

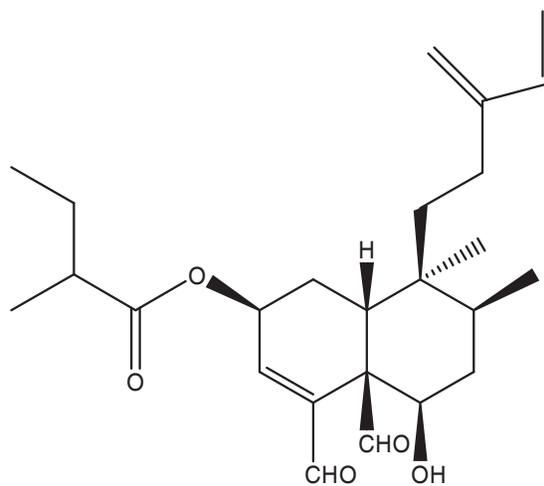
Nome	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>
Caseamembrina C	H	OH	CO(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>
Caseamembrina D	H	OAc	CO(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>



Caseamembrin E



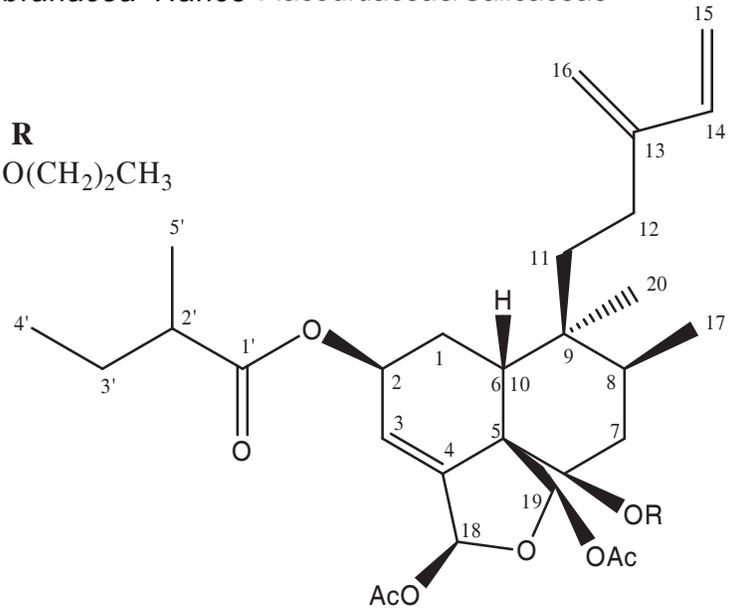
Caseamembrin F



*Casearia membranacea* Hance Flacourtiaceae/Salicaceae

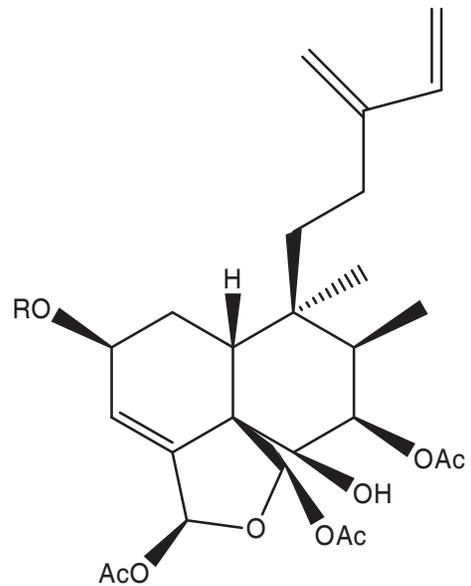
**Nome**  
Caseamembrina M

**R**  
 $\text{CO}(\text{CH}_2)_2\text{CH}_3$



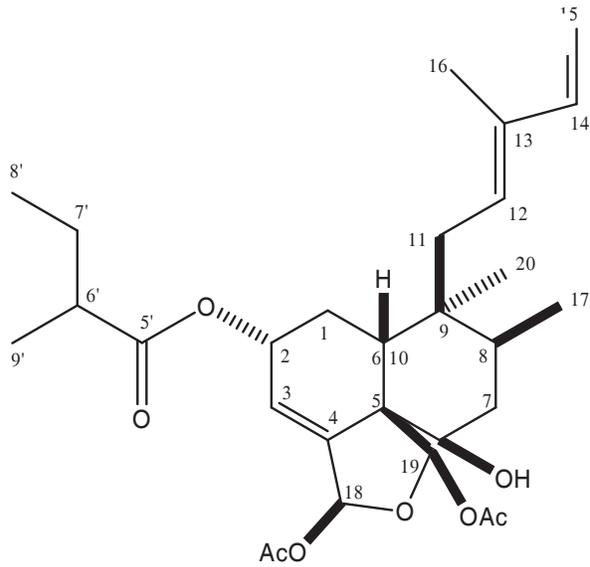
**Nome**  
Caseamembrina N  
Caseamembrina O

**R**  
 $\text{CO}(\text{CH}_2)_2\text{CH}_3$   
 $\text{COCH}_2\text{CH}(\text{CH}_3)_2$

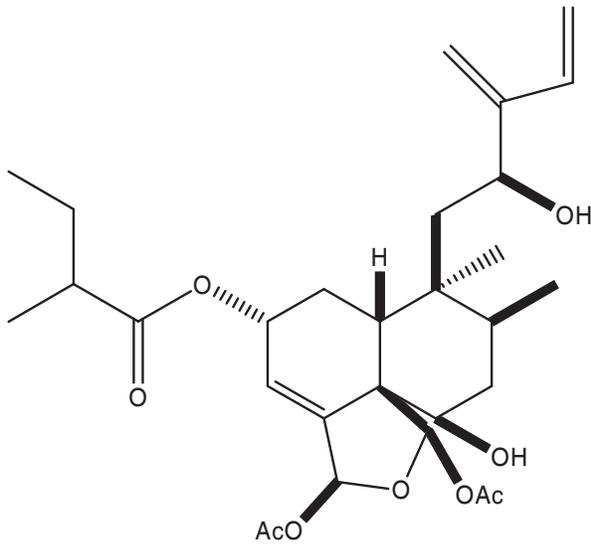


*Casearia membranacea* Hance Flacourtiaceae/Salicaceae

Caseamembrol A

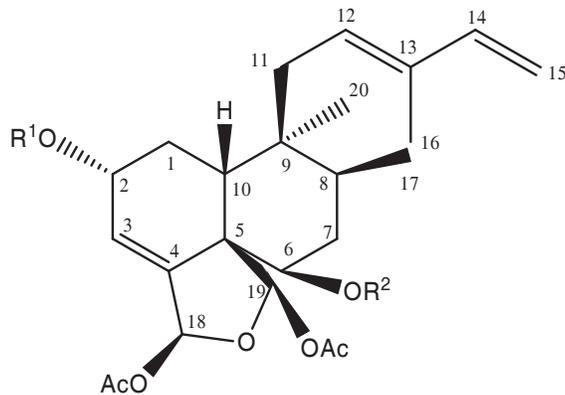


Caseamembrol B

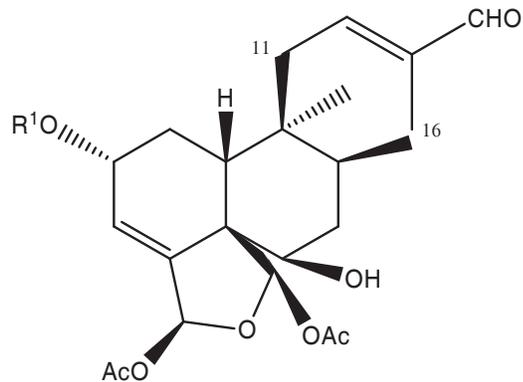


*Casearia corymbosa* H.B.K. Flacourtiaceae/Salicaceae

Nome	<b>R<sup>1</sup></b>	<b>R<sup>2</sup></b>
Corimbotina A	Ac	Me
Corimbotina B	(Me) <sub>2</sub> CHCO	Me
Corimbotina C	H <sub>2</sub> C=C(Me)CO	Me
Corimbotina D	Ac	H
Corimbotina E	(Me) <sub>2</sub> CHCO	H
Corimbotina F	H <sub>2</sub> C=C(Me)CO	H

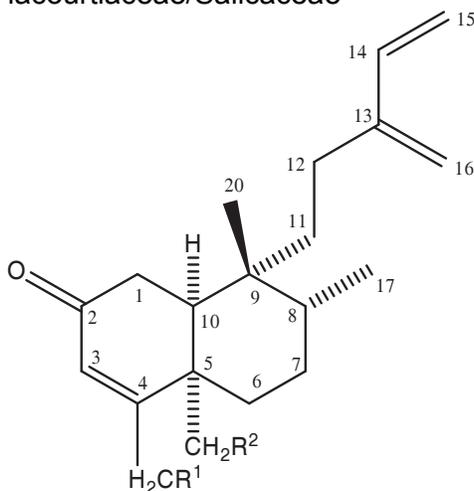


Nome	<b>R<sup>1</sup></b>
Corimbotina G	Ac
Corimbotina H	(Me) <sub>2</sub> CHCO
Corimbotina I	H <sub>2</sub> C=C(Me)CO

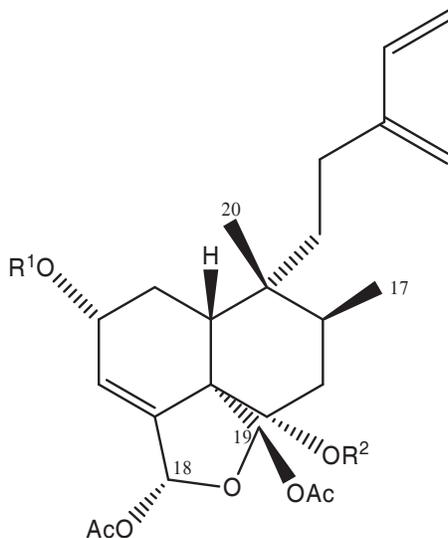


*Casearia corymbosa* H.B.K Flacourtiaceae/Salicaceae

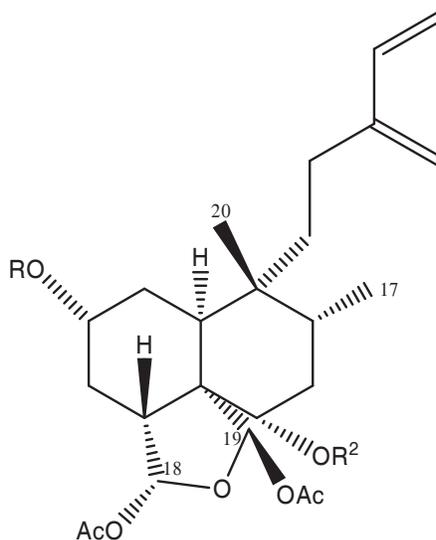
Nome	R <sup>1</sup>	R <sup>2</sup>
Ccorymbosa 4	OH	H
Ccorymbosa 5	OAc	OAc



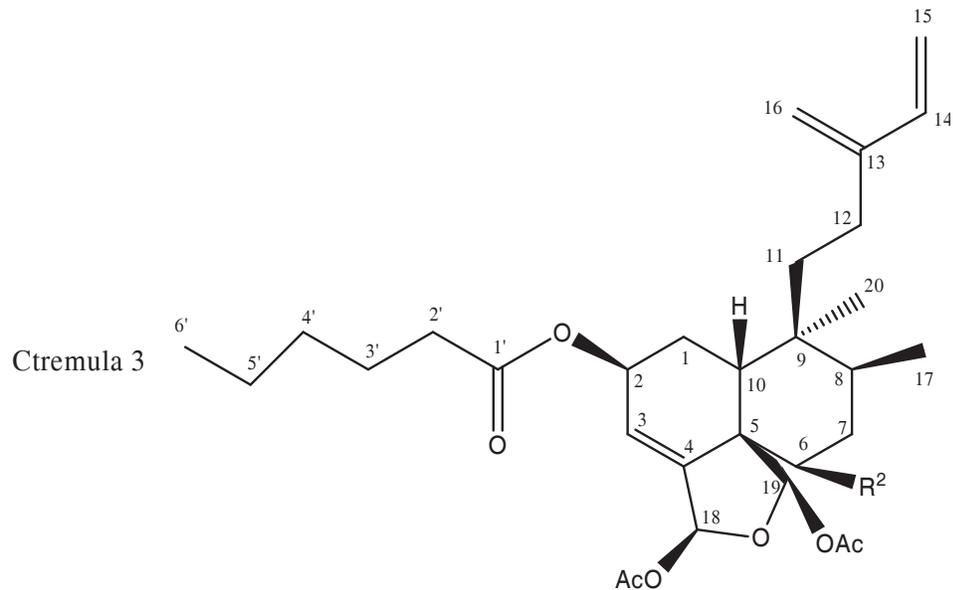
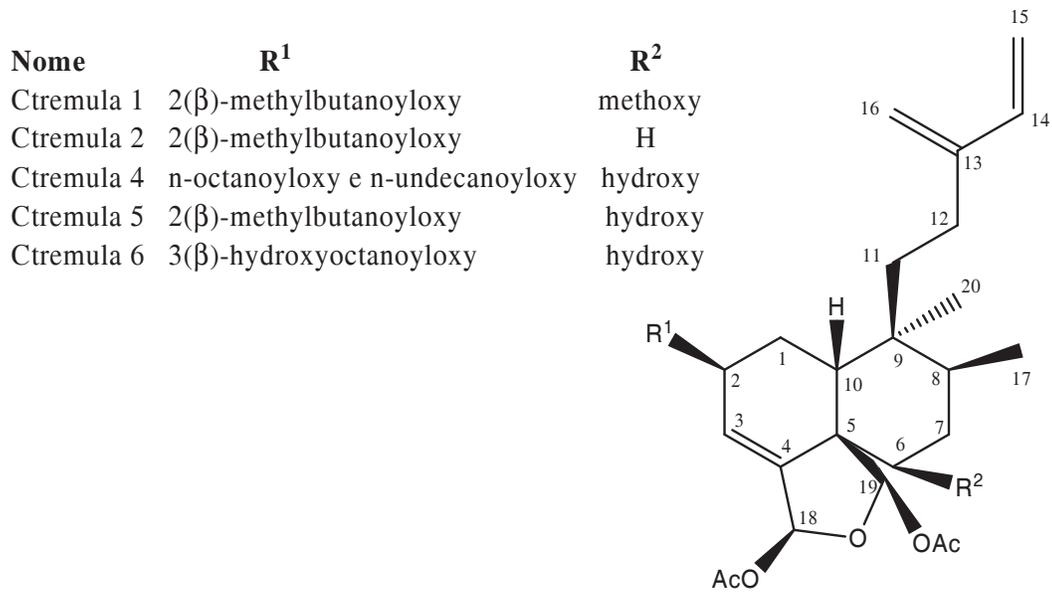
Nome	R <sup>1</sup>	R <sup>2</sup>
Ccorymbosa 8	COCH(Me) <sub>2</sub>	H
Ccorymbosa 9	COCH(Me)Et	Me
Ccorymbosa 10	COCH(Me) <sub>2</sub>	Me



Nome	R
Ccorymbosa 11	H
Ccorymbosa 12	Ac
Ccorymbosa 13	COCH(Me)Et

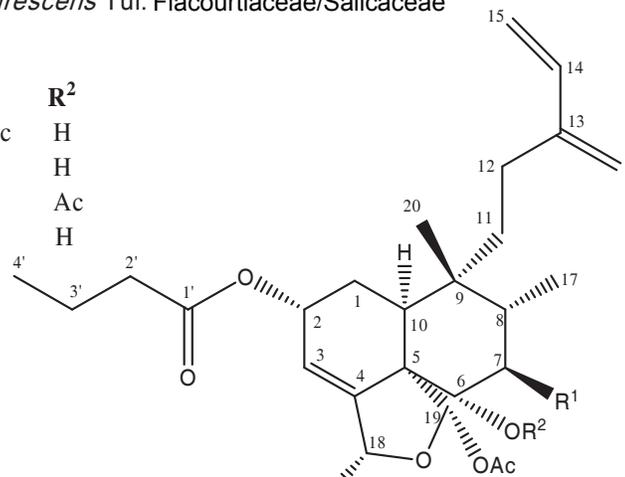


*Casearia tremula* Grisebach Flacourtiaceae/Salicaceae

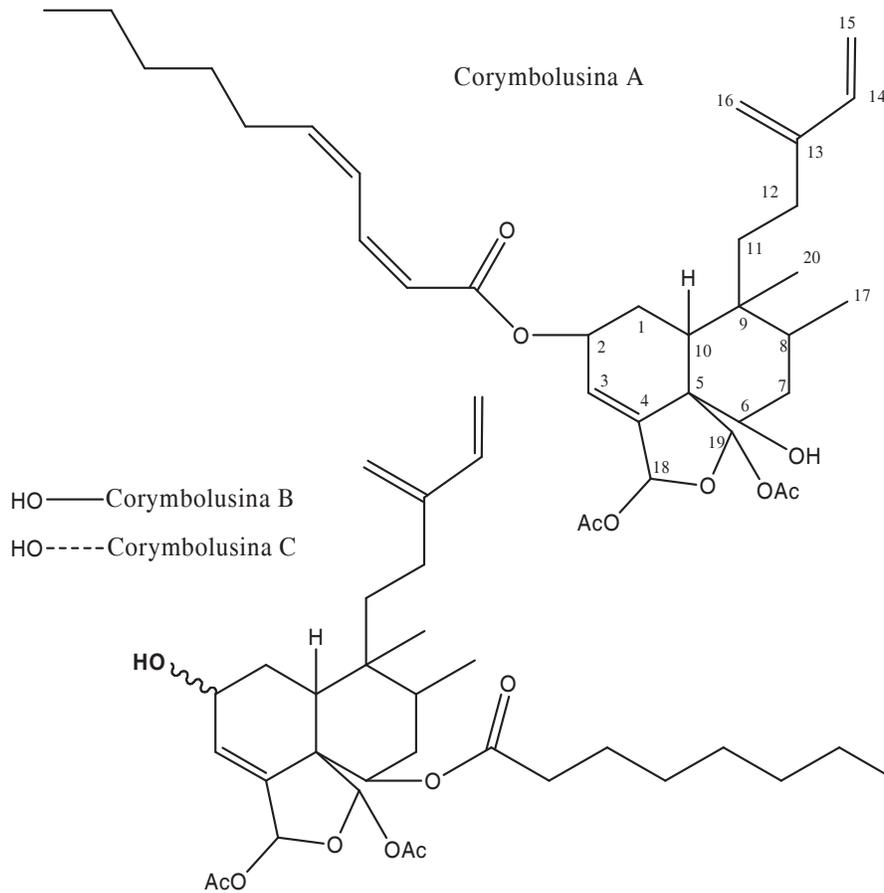


*Casearia nigrescens* Tul. Flacourtiaceae/Salicaceae

Nome	R <sup>1</sup>	R <sup>2</sup>
Casearnigrescens A	OAc	H
Casearnigrescens B	OH	H
Casearnigrescens C	OH	Ac
Casearnigrescens D	H	H

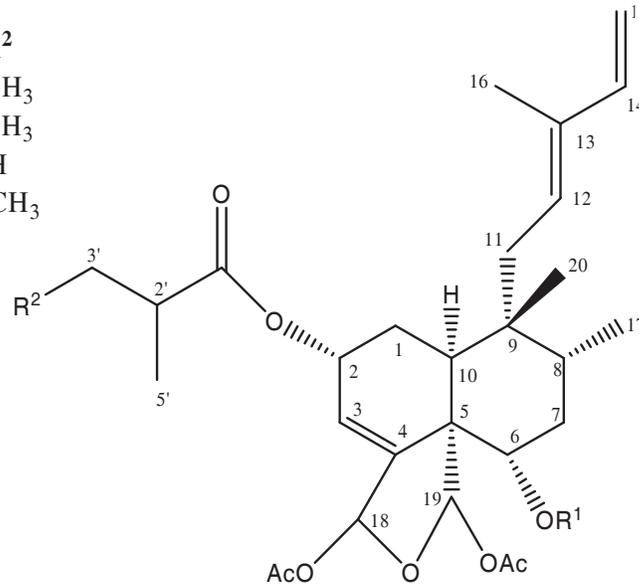


*Laetia conrymbulosa* Spruce ex Benth. Flacourtiaceae



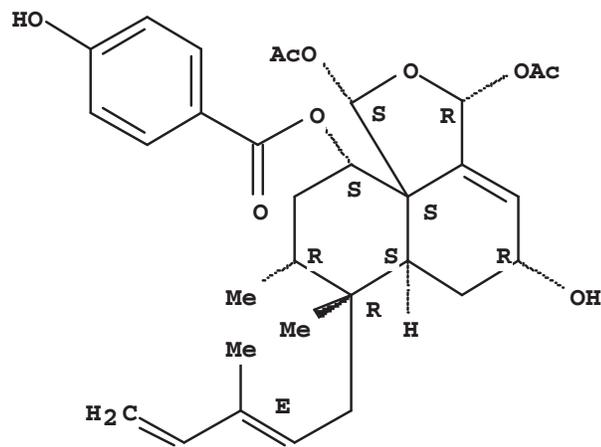
*Laetia procera* (Poepp.) Eichler Flacourtiaceae/Salicaceae

Nome	R <sup>1</sup>	R <sup>2</sup>
Laeciaprocerina A	H	CH <sub>3</sub>
Laeciaprocerina B	CH <sub>3</sub>	CH <sub>3</sub>
Laeciaprocerina C	COPh	H
Laeciaprocerina D	COPh	CH <sub>3</sub>



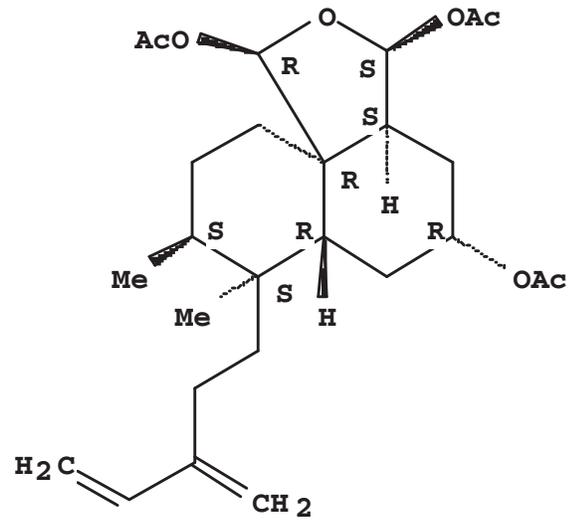
*Casearia grayi* Jessup Flacourtiaceae/Salicaceae

Cgrayi 1

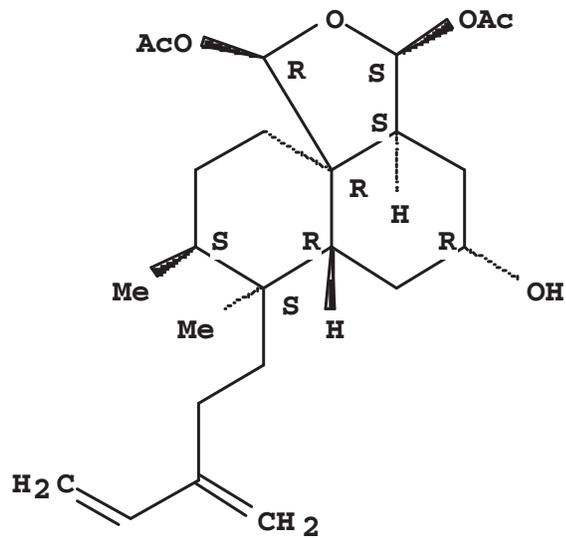


*Casearia grayi* Jessup Flacourtiaceae/Salicaceae

Cgrayi-2

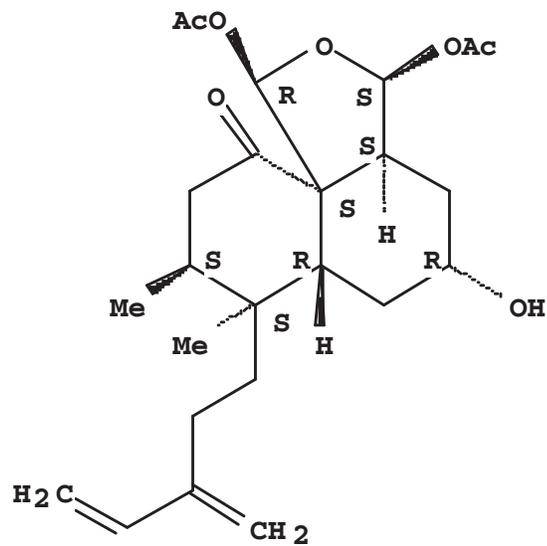


Cgrayi 3

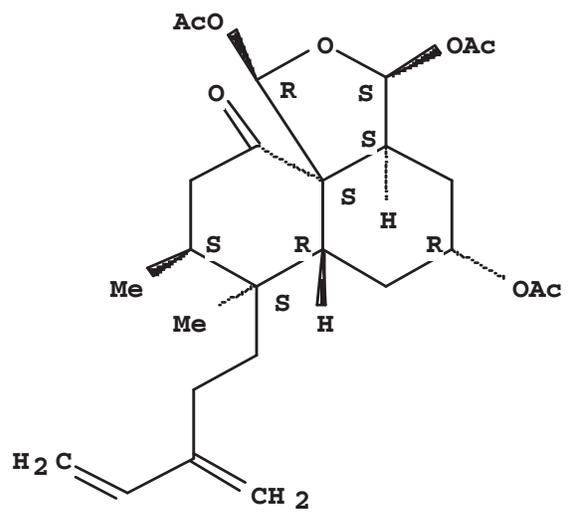


*Casearia grayi* Jessup Flacourtiaceae/Salicaceae

Cgrayi 4

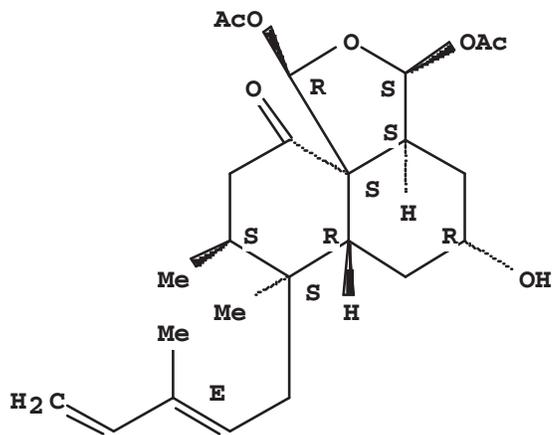


Cgrayi 5

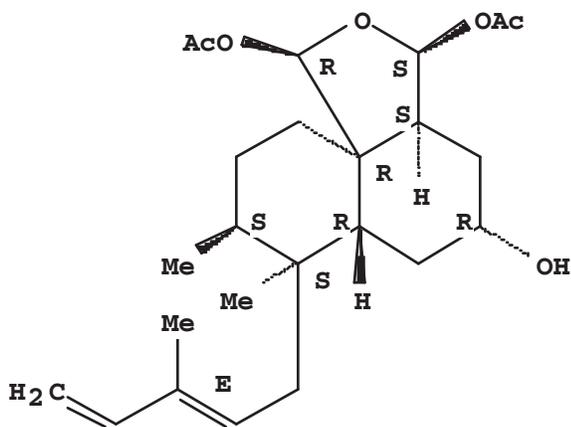


*Casearia grayi* Jessup Flacourtiaceae/Salicaceae

Cgrayi 6



Cgrayi7



Cgrayi 8

