

**UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

ESTUDOS DAS RELAÇÕES ESTRUTURA-TOXICIDADE DE AGROTÓXICOS

CESAR AUGUSTO KUROHANE

SÃO CARLOS

2012

**UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

ESTUDOS DAS RELAÇÕES ESTRUTURA-TOXICIDADE DE AGROTÓXICOS

CESAR AUGUSTO KUROHANE

**Dissertação apresentada ao Programa
de Pós-Graduação em Química para
obtenção do título de Mestre
Profissional em química.
Área de concentração: Química
Tecnológica.**

**Orientação: Prof. Dr. Carlos Alberto
Montanari**

SÃO CARLOS

2012

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

K96er

Kurohane, Cesar Augusto.

Estudos das relações estrutura-toxicidade de agrotóxicos / Cesar Augusto Kurohane. -- São Carlos : UFSCar, 2012.
138 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2012.

1. Química orgânica. 2. Agrotóxicos. 3. Reconhecimento de padrões. 4. Herbicidas. 5. Inseticidas. I. Título.

CDD: 547 (20^a)

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Departamento de Química
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA
Curso de Mestrado Profissional

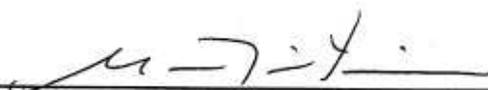
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de dissertação de Mestrado Profissional do candidato Cesar Augusto Kurohane, realizada em 23 de março de 2012:



Prof. Dr. Carlos Alberto Montanari



Dra. Háya Oliveira Souza Dória



Prof. Dr. Moacir Rossi Forim

"Tente fazer algo que vá além do que você já domina
caso contrário você nunca crescerá."
(Ronald E. Osborn)

AGRADECIMENTOS

- Aos meus pais Mauro e Marie, pelo apoio e orientações nas horas em que precisei;
- À minha noiva Patrícia, pelo incentivo e amor;
- Aos companheiros do Grupo Nequimed, pela colaboração e atenção dada no decorrer do mestrado;
- Ao meu orientador e amigo Prof. Dr. Carlos Alberto Montanari, pela paciência em me orientar, pelas discussões, conselhos e incentivos;
- Ao Raphael Montanari do ICMC/USP pelo desenvolvimento das redes neurais no ambiente MATLAB;
- À minha gerente Thais Balbão Clemente, pelo suporte, incentivo e apoio para essa pesquisa;
- Aos meus colegas de trabalho da empresa Ouro Fino, pelo suporte e ajuda nos momentos em que precisei;
- À Ouro Fino Química Ltda, empresa que me deu todo o suporte para a realização do mestrado;
- E a todos aqueles que de alguma forma auxiliaram na realização deste trabalho.

LISTA DE FIGURAS

Figura 1. Terras disponíveis para expansão em nível mundial.....	01
Figura 2. Projeção de vendas de defensivos agrícolas.....	02
Figura 3. Dados para desenvolvimento de uma nova molécula de agrotóxicos	03
Figura 4. Proposta de desenvolvimento molecular de (Q)SAR/(Q)STR com base na estrutura molecular	07
Figura 5. Decomposição da matriz X em componentes principais	14
Figura 6. Exemplo de gráfico de variância versus PC.....	14
Figura 7. Representação esquemática de um neurônio biológico	17
Figura 8. Esquema de unidade de processamento proposto por McCulloch-Pitts	19
Figura 9. Algoritmo backpropagation	21
Figura 10. Fluxograma ilustrativo do processo para o cálculo as similaridade 3D dos bancos de dados	28
Figura 11. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos herbicidas da classe das Uréias	29
Figura 12. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos herbicidas da classe das Uréias	30
Figura 13. Gráfico de PC1xPC2 da estabilidade metabólica dos herbicidas da classe das Uréias.....	31
Figura 14. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos herbicidas da classe das Uréias.....	32
Figura 15. Gráfico PC1xPC2 da permeabilidade em células CaCO2 dos herbicidas da classe das Uréias	33
Figura 16. Gráfico PC1xPC2 demonstra o volume de distribuição herbicidas da classe das Uréias.....	34
Figura 17. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos herbicidas da classe das triazinas.....	37
Figura 18. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos herbicidas da classe das triazinas.....	38
Figura 19. Gráfico de PC1xPC2 da estabilidade metabólica dos herbicidas da classe das triazinas.....	39
Figura 20. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos herbicidas da classe das triazinas.....	40
Figura 21. Gráfico PC1xPC2 da permeabilidade em células CaCO2 dos herbicidas da classe das triazinas	41
Figura 22. Gráfico PC1xPC2 demonstra o volume de distribuição dos herbicidas da classe das triazinas.....	42
Figura 23. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos inseticidas da classe dos neonicotinóides.....	44
Figura 24. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos inseticidas da classe dos neonicotinóides.....	45
Figura 25. Gráfico de PC1xPC2 da estabilidade metabólica dos inseticidas da classe dos neonicotinóides	46
Figura 26. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos inseticidas da classe dos neonicotinóides	47
Figura 27. Gráfico PC1xPC2 da permeabilidade em células CaCO2 dos inseticidas da classe dos neonicotinóides	48
Figura 28. Gráfico PC1xPC2 demonstra o volume de distribuição dos inseticidas da classe dos	

neonicotinóides	49
Figura 29. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos inseticidas da classe dos carbamatos	51
Figura 30. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos inseticidas da classe dos carbamatos	52
Figura 31. Gráfico de PC1xPC2 da estabilidade metabólica dos inseticidas da classe dos carbamatos	53
Figura 32. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos inseticidas da classe dos carbamatos	54
Figura 33. Gráfico PC1xPC2 da permeabilidade em células CaCO2 dos inseticidas da classe dos carbamatos	55
Figura 34. Gráfico PC1xPC2 demonstra o volume de distribuição dos inseticidas da classe dos carbamatos	56
Figura 35. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos inseticidas da classe das benzoiluréias	60
Figura 36. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos inseticidas da classe das benzoiluréias	61
Figura 37. Gráfico de PC1xPC2 da estabilidade metabólica dos inseticidas da classe das benzoiluréias	62
Figura 38. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos inseticidas da classe das benzoiluréias	63
Figura 39. Gráfico PC1xPC2 da permeabilidade em células CaCO2 dos inseticidas da classe das benzoiluréias	64
Figura 40. Gráfico PC1xPC2 demonstra o volume de distribuição dos inseticidas da classe das benzoiluréias	65
Figura 41. Gráfico de valores preditos x experimentais para DL50 oral aguda em ratos para inseticidas da classe das benzoiluréias	67
Figura 42. Erro quadrático médio da série de treinamento, teste e validação de 325 herbicidas	70
Figura 43. Ajuste linear para a série completa de herbicidas (treinamento + validação + teste)	70
Figura 44. Erro quadrático médio da série de treinamento, teste e validação de 292 inseticidas	71
Figura 45. Ajuste linear para a série completa de inseticidas (treinamento + validação + teste)	71
Figura 46. Dados de literatura do Glifosato Técnico	74
Figura 47. Projeção do Glifosato (azul), 7 Impurezas (verde) e 12 Herbicidas diversos (vermelho) sobre o modelo de PCA gerado a partir de 614 herbicidas descritos no banco de dados PAN (preto)	84
Figura 48. Gráfico de Cooman. Glifosato e Impurezas (vermelho); Herbicidas (preto): 95,4 % e 92,3 % de classificação correta, respectivamente	87
Figura 49. Gráfico de escores referente à análise de regressão por PLS Discriminante obtido a partir de um conjunto contendo 23 substâncias com valores de DL ₅₀ inferiores a 500 mg kg ⁻¹ (vermelho), 19 substâncias com DL ₅₀ superior a 500 mg kg ⁻¹	90
Figura 50. Estrutura química da Cipermetrina Técnica	95
Figura 51. Interação a proteína. Em amarelo, análise qualitativa que implica serem a cipermetrina e sua impureza 1 de elevada ligação proteica	97
Figura 52. Modelo hERG. Os círculos em vermelho definem as regiões em que os compostos não são bloqueadores do canal de potássio hERG. Em amarelo, observam-se a cipermetrina e a impureza 1 em estudo	98

Figura 53. Estabilidade metabólica. Pontos em vermelho representam compostos estáveis; os pontos azuis indicam os compostos instáveis metabolicamente	100
Figura 54. Interação a proteína (Albumina Sérica Humana, HSA). Substâncias do SciFinder similares à impureza 1 em níveis de agrupamento similar. Todas anotadas em amarelo.....	101
Figura 55. Modelo hERG de interação com a identificação em amarelo das substâncias químicas similares à impureza 1 da cipermetrina para as quais há relatos na literatura corrente sobre efeitos toxicológicos regulamentados por agências reguladoras	102
Figura 56. Perfil de estabilidade metabólica que demonstra similaridade química da cipermetrina e sua impureza 1 com agrupamento a substâncias que que apresentam perfil toxicológico conhecido. Todas estão anotadas em amarelo	103
Figura 57. Análise qualitativa das interações de moléculas com a proteína. Em amarelo, substâncias com moderada ligação proteica; em vermelhos, elevada ligação proteica e em azuis substâncias que apresentam baixa ligação proteica.....	112
Figura 58. Modelo hERG. Os círculos em vermelho definem as regiões em que os compostos não são bloqueadores do canal de potássio hERG. Em amarelo, observam-se os componentes da mistura em estudo	113
Figura 59. Estabilidade metabólica. Em amarelo estão codificadas as substâncias em estudo. As impurezas L, M e N apresentam estabilidade metabólica similar àquela demonstrada pelo Carbosulfano, enquanto que as impurezas E e K são estáveis metabolicamente	115

LISTA DE TABELAS

Tabela 1. Similaridade 3D Tanimoto Combo dos herbicidas das classes das Uréias.....	35
Tabela 2. Legenda para as Figuras 11-16 e Tabela 1 dos herbicidas das classes das Uréias....	36
Tabela 3. Similaridade 3D Tanimoto Combo dos herbicidas das classes das Triazinas	43
Tabela 4. Legenda para as Figuras 17 - 22 e Tabela 3 dos herbicidas das classes das Triazinas	43
Tabela 5. Similaridade 3D Tanimoto Combo dos inseticidas da classe dos neonicotinóides ..	50
Tabela 6. Legenda para as Figuras 23-28 e Tabela 5 dos inseticidas da classe dos neonicotinóides	50
Tabela 7. Similaridade 3D Tanimoto Combo dos inseticidas da classe dos carbamatos	57
Tabela 8. Legenda para as Figuras 29 - 34 e Tabela 7 dos inseticidas da classe dos carbamatos	58
Tabela 9. Similaridade 3D Tanimoto Combo dos inseticidas da classe das benzoiluréias.....	66
Tabela 10. Legenda para as Figuras 35-40 e Tabela 9 dos inseticidas da classe das benzoiluréias	66
Tabela 11. Dados para os inseticida da classe das benzoiluréias	68
Tabela 12. Classificação de toxicidade segundo a EPA	78
Tabela 13. Similaridade 3D combinada (Tanimoto/Combo) entre os pares de substâncias contidas no conjunto de dados: Glifosato (CAS: 1071-83-6, em azul); Impurezas em amarelo; Outros herbicidas (laranja). Similaridades (S) $1 \leq S < 1,4$ (verde claro); $S \geq 1,4$ (verde escuro)	81
Tabela 14. Estruturas moleculares e dados toxicológicos dos doze herbicidas utilizados nas análises de similaridade 3D	82
Tabela 15. Dados Toxicológicos e Ecotoxicológicos do Pesticide Manual	106
Tabela 16. Similaridade 3D combinada (Tanimoto/Combo) entre os pares de substâncias contidas no conjunto de dados: Carbosulfano (azul escuro); Impurezas (azul claro); carbamatos com DL_{50} (oral em ratos) conhecidos (laranja), Similaridades $> 1,2$ (verde); $\leq 1,2$ (amarelo)	107
Tabela 17. Similaridade 2D combinada (Tanimoto) entre os pares de substâncias contidas no conjunto de dados: Carbosulfano (azul escuro); Impurezas (azul claro); carbamatos com DL_{50} (oral em ratos) conhecidos (laranja), Similaridades $> 0,6$ (verde)	108
Tabela 18. Valores preditos a partir do melhor modelo de HQSAR. Carbosulfano (azul); serie treinamento (laranja escuro); serie teste (laranja claro); valores preditos de pDL_{50} (verde escuro)	117
Tabela 19. Dados estatísticos do modelo.....	118
Tabela 20. Resultados para os seis endpoints testados dos bancos: Benzoiluréias, Carbamatos, Neonicotinóides, Triazinas e Uréias	119

SUMÁRIO

RESUMO

1. INTRODUÇÃO	1
1.1. Perspectiva para o aumento da utilização de agrotóxicos no Brasil.....	1
1.2. Setor regulatório de defensivos agrícola e a busca de novas moléculas	2
1.3. A quimioinformática na identificação de ligantes	4
1.4. Ensaios virtuais.....	5
1.5. QSAR: Relações quantitativas entre estrutura e atividade	6
1.6. Volsurf	8
1.6.1. Descritores do Volsurf.....	9
1.6.1.1. Descritores de forma e tamanho.....	9
1.6.1.2. Descritores de regiões hidrofílicas	9
1.6.1.3. Descritores de regiões hidrofóbicas	10
1.6.1.4. Momentos de energia de interação	10
1.6.1.5. Descritores Mistos.....	10
1.6.2. Modelos do Volsurf (ENDPOINTS).....	11
1.6.2.1. ENDPOINT BBB	11
1.6.2.2. ENDPOINT hERG	12
1.6.2.3. ENDPOINT interação com albumina sérica de humanos	12
1.6.2.4. ENDPOINT CaCO ₂	12
1.6.2.5. ENDPOINT estabilidade metabólica (CYP3A4).....	12
1.6.2.6. ENDPOINT volume de distribuição	13
1.7. Análise de Componentes Principais (PCA).....	13
1.8. Regressão por análise dos mínimos múltiplos parciais (PLS).....	15
1.9. SIMCA (Soft Independent Modelling of Class Analogies).....	15
1.10. Estudos de similaridade química	16
1.11. Redes Neurais	16
2. OBJETIVO.....	22
3. MATERIAIS E MÉTODOS.....	23
3.1. Materiais	23
3.2. Métodos	23
3.2.1. Seleção de Banco de Dados	23
3.2.1.1. Busca no site da ANVISA por monografias autorizadas e não autorizadas.....	23
3.2.1.2. Busca no banco de dados da União Européia	23
3.2.1.3. Busca no banco de dados PPDB	24
3.2.1.4. Banco Selecionado	24
3.2.1.5. Outros bancos gerados	25
3.2.2. Estrutura 3D	25
3.2.3. Projeção nos modelos do Volsurf.....	25
3.2.4. SIMCA aplicado para Reconhecimento de moléculas.....	26
3.2.5. Modelo de PLS	26
3.2.6. Similaridade 3D	27
3.2.6.1. Construção do banco multiconformacional.....	27
3.2.6.2. Cálculo de similaridade através do ROCS	28
3.2.6.3. Estudo de similaridade dos compostos	28
4. RESULTADO E DISCUSSÃO	29
4.1. Uréias.....	29
4.2. Triazinas	37

4.3. Neonicotinóides	44
4.4. Carbamatos	51
4.5. Benzoiluréias	60
4.5.1. Modelo PLS para benzoiluréias	67
4.6. Método SIMCA no reconhecimento de herbicidas e inseticidas em um banco de dados compostos por 617 moléculas dos dois grupos	68
4.7. Aplicação de Redes Neurais para o banco de dados de inseticidas e herbicidas.....	69
5. ESTUDO DE CASOS	72
5.1. Caso 1: Determinação das Relações entre Estrutura e Toxicidade (STR) do Glifosato e suas Impurezas	72
5.2. Caso 2: Determinação das Relações entre Estrutura e Toxicidade (STR) da Cipermetrina	94
5.3. Caso 3: Determinação das Relações entre Estrutura e Toxicidade (STR) do Carbosulfano e suas Impurezas	105
6. CONCLUSÃO	119
7. REFERÊNCIAS BIBLIOGRÁFICAS	121

RESUMO

ESTUDOS DAS RELAÇÕES ESTRUTURA-TOXICIDADE DE AGROTÓXICOS

Através do uso de métodos computacionais para avaliação da similaridade estrutural e eletrostática, determinou-se um padrão de similaridade para agrotóxicos presentes no banco de dados do Grupo de Química Medicinal do IQSC/USP. Esse banco é constituído de uma coleção de herbicidas e inseticidas permitidos e banidos, segundo critérios da Organização Mundial da Saúde. Várias análises quimiométricas foram realizadas com o objetivo de determinar os processos de reconhecimento de padrões e classificar as substâncias dentro de suas respectivas classes químicas que foram associadas aos níveis de 61-86 % de predição correta. Três estudos de casos foram realizados com o objetivo de validar as hipóteses estabelecidas com resultados satisfatórios foram encontrados em todos os casos. As definições do espaço químico associado ao biológico através de métodos robustos de análise propiciaram o estabelecimento de regras de seleção e identificação de padrões que podem agora ser usados para o planejamento de novos herbicidas e inseticidas.

Palavras-chave: Agrotóxicos, Herbicidas, Inseticidas, Métodos de Reconhecimento de Padrões.

ABSTRACT**STUDIES OF RELATIONSHIP STRUCTURE-TOXICITY OF AGROCHEMICALS**

Computational methods such as electrostatic and structural similarity of compounds were used to determine a similarity pattern for pesticides present in the database of the Medicinal Chemistry Group IQSC/USP. A collection consisting of herbicides and insecticides allowed and banned, according to the World Health Organization was used in this study. Several chemometric analyzes were performed to determine the recognition pattern processes in the classification of substances into their chemical classes that were associated with levels of 61-86% of correct prediction. Three case studies were carried out to validate the established hypotheses which showed good results in all cases. The disclosure of chemical space associated with the biological space using robust methods of analysis led to the establishment of rules for the selection and identification of patterns that can now be used as the foundation for the design of new herbicides and insecticides.

Keywords: Pesticides, Herbicides, Insecticides, Methods of Pattern Recognition.

1. INTRODUÇÃO

1.1. Perspectiva para o aumento da utilização de agrotóxicos no Brasil

O Brasil é privilegiado por ter enorme potencialidade produtiva na área da agricultura, por apresentar uma vasta extensão de terras de boa qualidade para cultivo, abundância de água, clima favorável para se obter várias colheitas no ano, um enorme mercado consumidor e mão de obra abundante. Para se comprovar a potencialidade produtiva da América Latina, a Figura 1 apresenta a distribuição das terras disponíveis para expansão da agricultura em âmbito mundial. De um total disponível de 41% para a América Latina, a maior área agricultável pertence ao Brasil, que possui para a expansão da sua agricultura uma área de 60 a 200 milhões de hectares.[1]

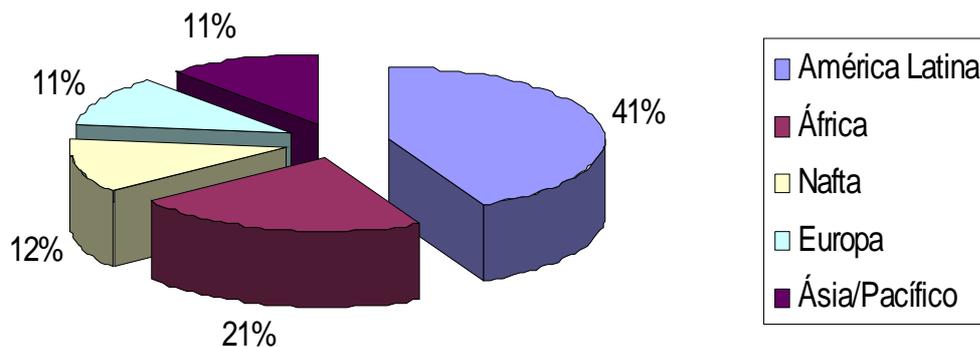


Figura 1. Terras disponíveis para expansão em nível mundial [2].

Apesar de toda a potencialidade agrária, o Brasil ainda não foi capaz de demonstrar uma agricultura orgânica sustentável, com produtividade agrícola e benefícios consistentes e significativos para a saúde humana. Além disso, este tipo de agricultura não oferece soluções para os problemas urgentes, decorrentes da atual tendência de produção e consumo mundial de alimentos, tornando necessário o uso de agrotóxicos. Conforme atesta a FAO, Organização das Nações Unidas para Agricultura e Alimentação, sem a tecnologia eficiente no controle de pragas a produção de alimentos reduziria em cerca de 40%. Ou seja, não proteger as plantações acrescentaria uma situação de calamidade mundial ao quadro, hoje, já dramático, conforme o relatório recente da ONU: apesar da crescente oferta de alimentos, a fome atinge mais de 925 milhões de pessoas [3]. A agricultura no estágio atual ainda não pode prescindir do uso de agrotóxicos, que se constitui em insumo necessário para a competitividade da produção agrícola.

Os fatores apresentados evidenciam o potencial brasileiro de se tornar o país mais importante no mercado de agroquímicos nos próximos dois anos, ultrapassando os Estados Unidos, cujo consumo está em torno de US\$ 7 bilhões em média por ano [4] e a perspectiva é de aumento, conforme mostra a Figura 2. No entanto, na utilização desses produtos precisa-se respeitar o que dispõe a legislação pertinente, procurando minimizar os impactos ecológicos, buscando a preservação da saúde do trabalhador rural e a segurança alimentar, que se tornaram uma exigência dos mercados consumidores.

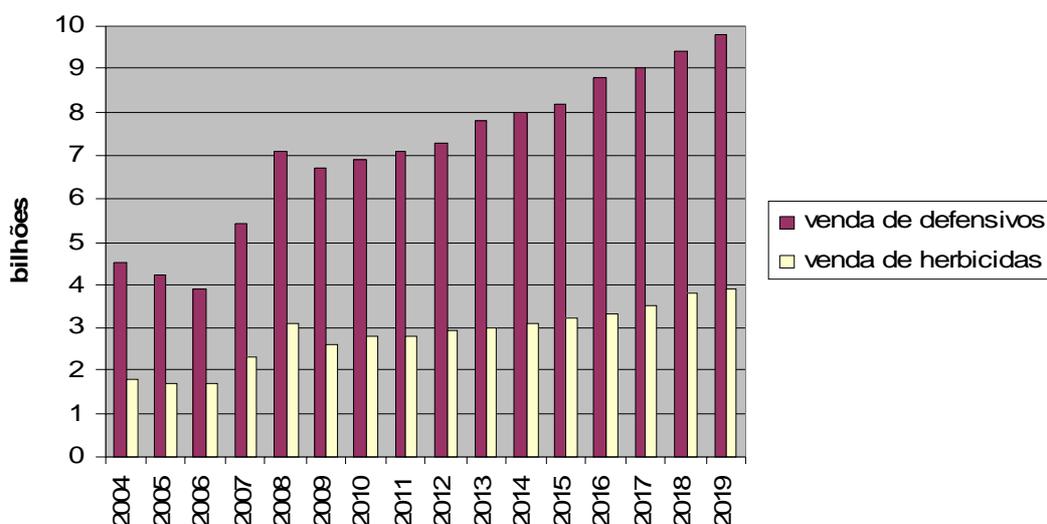


Figura 2. Projeção de vendas de defensivos agrícolas. Fonte: MAPA 2010. Projeções do Agronegócio Brasil 2009/2010 a 2019/2019. Baseado na média do crescimento anual nas produtividades das culturas: algodão, arroz, milho, trigo, soja (grão e óleo), cana-de-açúcar (açúcar e etanol), laranja (suco de laranja e laranja) e celulose.

1.2. Setor regulatório de defensivos agrícola e a busca de novas moléculas

O setor regulatório de defensivos agrícolas brasileiro é um dos mais rígidos do mundo [5], sendo que a aprovação de um agrotóxico deve passar por avaliação de três ministérios: MAPA (Ministério da Agricultura, Pecuária e Abastecimento), ANVISA (Agência Nacional de Vigilância Sanitária) e IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis), nos quais os pleitos de registro de agrotóxicos devem conter diversos tipos de estudos toxicológicos, físico-químicos e ecotoxicológicos e também teste de resíduos. Com esses dados em mãos, os técnicos especializados de cada órgão se dedicam por até três anos [5] na análise de um requerimento de registro de um produto genérico e, em caso de registro de um novo produto, este tempo de análise se prolonga. O produto poderá então ser

colocado no mercado após prévia avaliação dos estudos, que buscam comprovar que quando os agrotóxicos são utilizados da forma que preconizam os órgãos reguladores, não causarão danos ao trabalhador e nem problemas de intoxicação ao consumidor.

Testes toxicológicos que incluem irritação cutânea aguda, irritação ocular aguda, irritação dermal aguda entre outros são realizados em animais. Trabalhos estão sendo desenvolvidos visando minimizar testes desnecessários dos animais, respeitando leis internacionais como a diretiva 86/609EC [6] que trata da experimentação animal na União Européia, a qual afirma que: "Um experimento com animais não devem ser realizado se outro método cientificamente satisfatório de obtenção do resultado pretendido, que não implique a utilização de um animal, é razoável e praticamente possível." e também a declaração universal dos direitos dos animais que no Art. 8º diz que: "As técnicas de substituição devem ser utilizadas e desenvolvidas" [7].

Métodos alternativos como QSAR (Relação Quantidade Estrutura Atividade) e QSTR (Relação Quantidade Estrutura Toxicidade) são alternativas para minimização de testes em animais, uma vez que utilizam métodos computacionais robustos para se predizer toxicidades e desenvolver novos produtos químicos com base na similaridade entre compostos já exaustivamente estudados. Estes métodos visam também otimizar tempo e custo no desenvolvimento de novas moléculas, uma vez que o processo de desenvolvimento de novas molécula é um processo caro e demorado, conforme mostra a Figura 3.

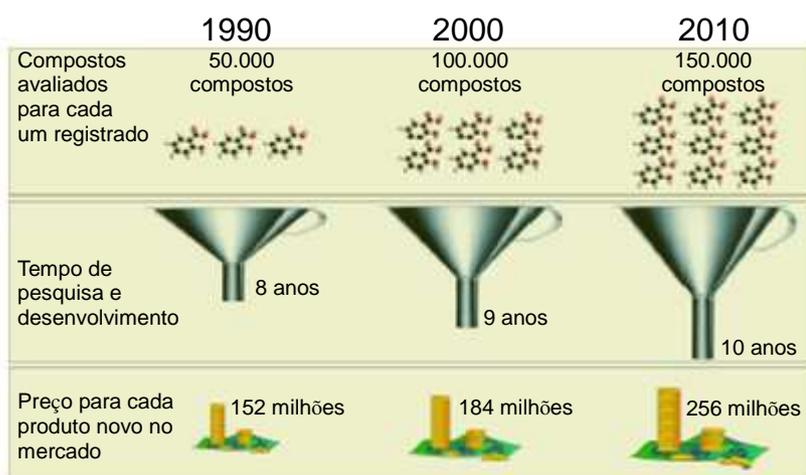


Figura 3. Dados para desenvolvimento de uma nova molécula de agrotóxicos [8].

1.3. A quimioinformática na identificação de ligantes

A exploração sistemática do espaço químico e o delineamento das regiões estabelecidas pelo espaço químico-biológico só são possíveis através da criação de bancos de dados *in silico* para o armazenamento adequado das informações sobre as pequenas moléculas. Como consequência disso, os métodos e ferramentas computacionais tem-se tornado parte integral durante as etapas identificação e otimização de ligantes para o armazenamento, gerenciamento, análise de dados e modelagem molecular de compostos, tanto dentro das indústrias químicas quanto na academia. O desenvolvimento de estratégias e a aplicação de métodos computacionais para a solução dos problemas associados à seleção e otimização de agrotóxicos fazem parte do escopo de uma disciplina da química medicinal que se chama quiminformática.

O termo “quiminformática” foi definido pela primeira vez por Brown (1998) [9] como “uma combinação de recursos para transformar dados em informação e informação em conhecimento, com o propósito de se tomar melhores decisões mais rapidamente dentro das tecnologias de informação”. Uma definição mais ampla foi dada por Gasteiger [10]: “A quiminformática consiste do uso dos métodos de informática para resolver problemas químicos”.

Nos últimos anos, a quiminformática tem emergido como um avanço tecnológico guiado pelas inovações na área de informática para alavancar as pesquisas na indústria agroquímica, e tem como finalidade a manipulação de um grande banco de dados comuns nas empresas que desenvolvem novas moléculas [11], no intuito de desenvolver produtos mais eficazes e seguros. A quiminformática une todas as frentes de pesquisa, dos ensaios virtuais à toxicologia animal, por via de um único elemento central: a estrutura química. Os protocolos de química computacional descritos a seguir fazem parte das atividades da quiminformática no contexto da identificação de toxicóforos: descrição, manipulação e análise de bancos de dados de estruturas químicas (virtuais ou reais); SAR (qualitativos ou quantitativos) para a identificação de séries de compostos-líderes; pós-processamento de dados obtidos em ensaios reais ou virtuais; ensaios *in silico* para a predição de propriedades de absorção, distribuição, metabolismo e excreção (ADME), predição do metabolismo e toxicologia; mineração e visualização de dados; ensaios virtuais para a determinação da similaridade química.

1.4. Ensaios virtuais

Os ensaios virtuais (VS, do inglês, *Virtual Screening*) consistem da aplicação de modelos computacionais para selecionar ou priorizar compostos para ensaios experimentais ou para a descrição do espaço químico-toxicológico [12]. Através de ensaios virtuais é possível pontuar, classificar ou “filtrar” moléculas contidas em grandes bancos de dados de acordo com as suas probabilidades de possuir alguma propriedade caracteristicamente desejada. Neste sentido, o ensaio virtual pode ser considerado como uma tentativa de estender o conceito por detrás do QSAR, que está focado em apenas pequenos conjuntos de compostos que consistem de moléculas existentes dentro de conceitos retrospectivos.

Em contraste com as técnicas de ensaios automatizados em larga escala (HTS, do inglês, *High-Throughput Screening*) [13] que testam uma grande quantidade de moléculas e são orientadas pela tecnologia utilizada, os ensaios virtuais são métodos orientados pelo conhecimento, ou seja, requerem que alguma informação estrutural seja conhecida. Assim, existem diferentes critérios, pelos quais as estruturas dos compostos podem ser pontuados, ordenados ou filtrados através de um experimento de ensaio virtual. Por exemplo, um modelo matemático tal como de regressão linear múltipla pode ser construído para prever a toxicidade de cada estrutura contida num banco de dados, ou então uma série de subestruturas moleculares de referência pode ser utilizada para eliminar moléculas que contenham funcionalidades indesejadas de uma coleção de compostos.

O número de estruturas que podem ser consideradas no ensaio virtual é bastante grande. Por isso, muitas vezes é mais efetivo utilizar os métodos de ensaios virtuais sucessivamente e em ordem de complexidade crescente. Assim, cada método atua como um filtro para remover estruturas que não seriam interessantes. Obtendo-se ao final do processo uma série de estruturas candidatas, as quais estarão em um conjunto final para a seleção.

Todas essas estratégias dependem do princípio da similaridade, a qual estabelece que moléculas similares devam exibir propriedades biológicas similares, e são chamadas de ensaios virtuais baseados no ligante (LBVS, do inglês, *Ligand-Based Virtual Screening*).

O método de similaridade química tem-se mostrado uma robusta ferramenta na gênese planejada de fármacos que inclui fenômenos de toxicidade [14]. Tipicamente, pequenas coleções de moléculas são selecionadas a partir de coleções maiores baseadas nas suas similaridades estruturais em relação a compostos de toxicidade conhecida [15].

1.5. QSAR: Relações quantitativas entre estrutura e atividade

A ação de agrotóxicos resulta das interações destes com sistemas biológicos e é dependente de fatores relacionados com sua estrutura química [16] e, conseqüentemente, de suas propriedades físico-químicas. Estes fatores sejam eles de caráter eletrônico, hidrofóbico ou estereoquímico, influenciam a interação do agrotóxico com a biofase e a sua distribuição nos compartimentos que compõem o sistema biológico. Assim, dois agrotóxicos estruturalmente semelhantes, diferenciando-se apenas por um átomo ou posição que este ocupa na molécula, podem apresentar diferenças quanto às suas propriedades físico-químicas e, conseqüentemente, quanto à atividade toxicológica, tanto do ponto de vista quantitativo como qualitativo.

As propriedades físico-químicas de compostos, reflexo de sua estrutura química, podem ser descritas quantitativamente se forem expressas por meio de parâmetros físico-químicos ou descritores estruturais adequados. Assim, com base neste conceito, foi possível desenvolver área do conhecimento, que se preocupa com o estudo das relações entre a estrutura química e a atividade biológica/toxicológica. Estas relações são expressas por modelos matemáticos, que correlacionam descritores estruturais de séries de compostos análogos com a atividade toxicológica por eles desempenhada.

Entre as propriedades físico-químicas mais importantes para o estabelecimento da toxicidade de agrotóxicos citam-se a massa molecular, solubilidade, logP octanol entre outras [17]. Cada uma destas propriedades contribui com maior ou menor intensidade para o estabelecimento da resposta toxicológica e podem ser representadas por descritores estruturais, que expressam de forma qualitativa e quantitativa a sua influência na toxicidade de compostos químicos. Um estudo de QSAR parte do princípio de que as subestruturas de um dado composto químico contribuem de modo individual e aditivo, para suas propriedades toxicológicas. Assim, é possível estabelecer uma metodologia de otimização de tais propriedades a partir de estudos sistemáticos, como pode ser visualizado na Figura 4 a seguir:

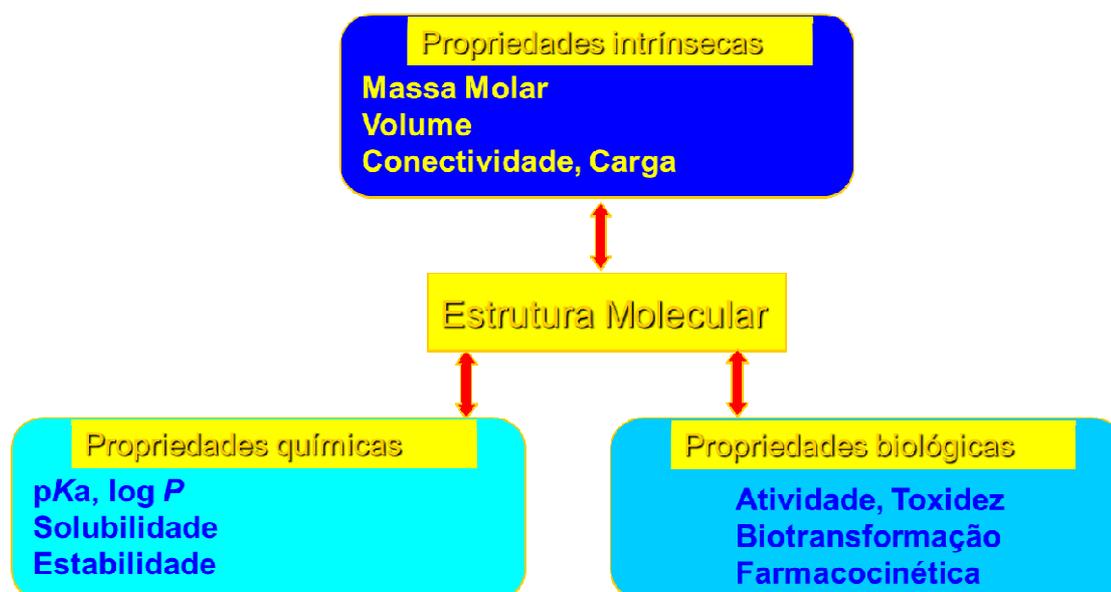


Figura 4. Proposta de desenvolvimento molecular de (Q)SAR/(Q)STR com base na estrutura molecular.

De modo simplificado, a estrutura líder é determinada a partir de alguma informação previamente disponível que relaciona um dado alvo e uma substância testada contra ele. Com o conhecimento do mecanismo de ação e dos fatores termodinâmicos envolvidos, como por exemplo: as contribuições entálpicas e entrópicas na ligação alvo-ligante viabilizam elaborar um estudo sistemático no qual se inclui uma etapa de planejamento molecular, onde é possível realizar modificações estruturais *in silico*.

A adequação das propriedades toxicológicas está fortemente relacionada com os mecanismos de absorção, distribuição e atividade intrínseca. Um dos pontos importantes – absorção – depende diretamente da lipofilia da molécula, uma vez que é necessário que estas se dissolvam e atravessem por membranas que formam os tecidos e sistemas multi-compartimentados de seres vivos, até alcançar os respectivos sítios de ação. Embora seja apenas uma das várias propriedades físico-químicas que determinam a toxicidade, a lipofilia desempenha papel fundamental na toxicidade de substâncias químicas. É bem conhecido que a lipofilia de um composto interferem na extensão e velocidade de absorção, distribuição tissular e capacidade de ligação ao sítio receptor.

Pode-se perceber com clareza a importância da etapa de planejamento estrutural, na qual as propriedades supracitadas podem sofrer alterações significativas, e as diretrizes úteis nesta etapa podem ser fornecidas por técnicas de QSAR.

Em linhas gerais, o significado estatístico das correlações obtidas em estudos de QSAR será maiores quanto mais precisas forem as medidas da toxicidade, pois descritores estruturais podem ser obtidos com alto grau de precisão.

A medida da toxicidade pode ser expressa em termos de potência, e esta, por sua vez, pode ser definida como o inverso da dose ou concentração de substância bioativa necessária para obter uma dada resposta biológica, o que significa que quanto menor for a concentração necessária para obter o efeito, mais potente será o efeito toxicológico.

O método de QSAR é utilizado amplamente na busca por novas moléculas no setor de agroquímicos, verificação de novos alvos biológicos e determinação da toxicidade de impurezas [18-22]. Hoje em dia poucas moléculas novas são lançadas no mercado e moléculas novas com modo de ação diferente dos já existentes são ainda mais raros. Por exemplo, na área de herbicidas, há 15 anos um novo mecanismo de ação para esta classe de uso não é inserido no mercado [23], o que ocasiona problemas de resistência por parte de ervas daninhas. Com isso, verifica-se que há necessidade de pesquisa e desenvolvimento de novas moléculas e novos mecanismos de ação.

1.6. Volsurf

O método Volsurf (do inglês *volume and surface*) é um método simples que produz descritores moleculares 2D a partir de campos de interação 3D, esses campos 3D são gerados pela energia de interação molecular [24]. A interação de uma molécula de agrotóxico com uma membrana biológica ocorre por um reconhecimento tridimensional mediada por propriedades de superfícies como forma, forças de Van-der-Waals, eletrostática, ligação de hidrogênio e hidrofobicidade. Portanto as propriedades moleculares calculadas pela energia de interação do campo molecular 3D (MIF do inglês *molecular interaction fields*) correlaciona propriedade físico-química e farmacocinética com a estrutura molecular 3D dos agrotóxicos a serem estudados.

Os MIFs são determinados através do uso de sondas químicas virtuais descritivas da superfície molecular, gerado pelo GRID. O GRID é um campo de força amplamente utilizado para mapear superfícies moleculares de agrotóxicos, fármacos e macromoléculas baseado na energia total de interação entre uma molécula alvo e uma sonda, por meio do movimento da sonda ao longo da superfície dessa molécula [25-27].

As sondas são seletivas e permitem coletar informações das propriedades do campo molecular e com isso pode-se obter modelos de QSAR-3D [28], os MIFs obtidos na interação da molécula com a sonda serão convertidos em descritores físico-químicos utilizados para estudos relacionados com propriedades farmacocinéticas ou com atividade biológica. As sondas mais utilizadas para simular as interações entre as moléculas e a membrana biológica,

são as sondas de água (OH₂), hidrofóbica (DRY) e de oxigênio carbonílico (O). As sondas de água são utilizadas para a simulação de processos de solvatação e dessolvatação e as sondas DRY e de O são utilizadas para simular as interações entre grupos apolares da membrana biológica [24].

A obtenção dos descritores físico-químicos pode ser resumida em duas etapas, na primeira etapa o campo molecular 3D é gerado pela interação entre a molécula alvo e a sonda virtual e na segunda etapa é realizado o cálculo dos descritores moleculares pelo método VolSurf a partir dos métodos 3D da primeira etapa.

1.6.1. Descritores do Volsurf

1.6.1.1. Descritores de forma e tamanho

A estrutura molecular é importante para se determinar similaridade, verificar regularidade e correlações no planejamento e na otimização de agrotóxicos e fármacos. Os descritores de forma e tamanho são volume molecular (V, volume acessível pela sonda de água), superfície molecular (S, superfície acessível a sonda de água), rugosidade (R, é a medida da superfície rugosa da molécula) e globularidade (G, classifica as molécula em níveis de globularidade comparando-as com moléculas de formato esférico perfeito, a propriedade globularidade é relacionada com a flexibilidade molecular [29]).

1.6.1.2. Descritores de regiões hidrofílicas

As regiões hidrofílicas são definidas como um envelope molecular que é acessível e atrai moléculas de água. O volume deste envelope varia com nível de energia de interação entre a água e a molécula. Os campos moleculares de -0,2 a -1 kcal mol⁻¹ considera forças de dispersão e polarizabilidade e campos de -1,0 a -6,0 kcal mol⁻¹ considera interações polares e doadores e receptores de hidrogênio. O fator de capacidade é outro descritor referente às regiões hidrofílicas, este parâmetro descreve a razão entre a região hidrofílica e superfície molecular total, e também é obtido em oito níveis de energia [29].

1.6.1.3. Descritores de regiões hidrofóbicas

Analogamente aos descritores de regiões hidrofílicas, os descritores hidrofóbicos podem ser definidos como o envelope molecular que gera interações hidrofóbicas, o Volsurf calcula esse descritor em oito níveis de energia adaptado para a faixa de energia usual para interações hidrofóbicas (0,0 a -2 kcal mol^{-1}) [29].

1.6.1.4. Momentos de energia de interação

Semelhantes ao momento dipolo, o momento de interação expressa o desequilíbrio entre o centro de massa e o baricentro das regiões hidrofílicas de uma molécula. Para regiões hidrofílicas os vetores apontam do centro de massa para o centro das regiões hidrofílicas. E quando o momento de energia de interação é alto, há uma clara concentração de regiões hidratadas em apenas uma parte da molécula, já se o momento de energia de interação é pequeno as porções polares ou estão próximos do centro de massas ou em posições opostas da molécula

Quando se referem a regiões hidrofóbicas, estes momentos medem o desequilíbrio entre o centro de massa e o baricentro de regiões hidrofóbicas de uma molécula em diferentes níveis de energia [29].

1.6.1.5. Descritores Mistos

Os descritores mistos levam em conta as interações hidrofílicas e lipofílicas. Os descritores de energia de distância mínima representam as distâncias entre os três locais de energia mínima, quando uma sonda de água interage com a molécula alvo. Já o descritor de balanço hidrofílico-lipofílico, representa a razão entre as regiões hidrofílicas e lipofílicas, determinando qual o efeito predominante na molécula ou se ambos estão igualmente balanceados. O momento anfifílico (A) é definido como um vetor que aponta do centro do domínio hidrofóbico para o centro do domínio hidrofílico, o comprimento do vetor é proporcional à força do momento anfifílico, e este pode determinar a habilidade do composto em permear a membrana. Os parâmetros de empacotamento crítico (CP) definem a razão entre a parte hidrofílica e lipofílica da molécula, em contraste com o balanço hidrofílico-lipofílico, refere-se somente à forma molecular. As ligações de hidrogênio, esse parâmetro descreve a capacidade da molécula alvo em realizar ligações de hidrogênio com uma sonda

polar. A sonda de água apresenta uma ótima habilidade para ligações de hidrogênio, atuando tanto como doadora como receptora. A polarizabilidade é o único descritor molecular que não é calculado através de campos moleculares 3D e é uma estimativa da polarizabilidade média molecular baseada na estrutura dos compostos, e é independente do número e tipos de sondas utilizadas [29].

1.6.2. Modelos do Volsurf (ENDPOINTS)

Utilizando-se os descritores 2D obtidos pelo programa Volsurf e correlacionando-se com as propriedades farmacocinéticas, e por análises estatísticas é possível identificar padrões físico-químicos relevantes que estão associados à estas propriedades e gerar modelos qualitativo e/ou quantitativos entre as estruturas e propriedades de um conjunto de compostos.

O Volsurf possui modelos qualitativos que podem ser utilizados para se prever propriedades farmacocinéticas de novos compostos químicos e com isso determinar como essas novas moléculas se comportam em humanos, são propriedade fundamentais na determinação da toxicidade de agrotóxicos e suas impurezas.

1.6.2.1. ENDPOINT BBB

O modelo utilizado de BBB pode ser utilizado para prever a capacidade das substâncias químicas de penetrarem na barreira hemato-encefálica, penetração na BBB é um fenômeno complexo que dependem de múltiplos fatores.

O modelo apresenta 313 compostos quimicamente diversos e o pressuposto básico é de permeação passiva, no qual foram utilizadas as sondas DRY, OH2 e O. E utilizando análise estatística de regressão dos mínimos múltiplos parciais (PLS do inglês *Partial Least Square*). O modelo gerado é capaz de separar compostos que apresentam elevada permeação na BBB, permeação moderada e as que não atravessam a BBB, sendo o modelo gerado qualitativo [30].

Em casos que os agrotóxicos conseguem penetrar a BBB, podem provocar depressão do sistema nervoso central. As pessoas contaminadas com estes produtos químicos podem apresentar sintomas brandos, como sonolência, confusão mental e letargia à casos mais graves, como hipotonia, hipotensão, arreflexia, entre outros.

1.6.2.2. ENDPOINT hERG

O modelo de hERG utilizado é constituído por 167 compostos, no qual utiliza sondas químicas DRY, OH2 e O. Com isso o modelo gerado é capaz de separar moléculas que bloqueiam ou não o canal de hERG.

O gene hERG codifica uma proteína conhecida como canal iônico de potássio (canal de hERG). Este canal iônico é conhecido por sua contribuição para a atividade elétrica do coração, coordenando o batimento do mesmo ao mediar a corrente repolarizante de ação cardíaca potencial. Quando este canal perde capacidade de conduzir corrente elétrica através da membrana celular, uma doença potencialmente fatal chamada de síndrome de QT longo pode resultar em morte súbita e por isso o canal de hERG é conhecido como anti-alvo [31].

1.6.2.3. ENDPOINT interação com albumina sérica de humanos

O modelo de interação com albumina sérica de humanos é constituído de 408 compostos, no qual foram utilizadas as sondas químicas DRY, OH2 e O. A importância de se avaliar a interação dos agrotóxicos com proteínas está relacionada à facilidade que estas possuem em distribuir as substâncias bioativas aos respectivos tecidos alvos, o que ocorre através do sistema circulatório [32].

1.6.2.4. ENDPOINT CaCO2

O modelo de CaCO2 é constituído de 751 compostos químicos diversos, o qual utiliza as sondas químicas DRY, OH2 e O. A permeabilidade através do epitélio intestinal é um importante fator a ser determinado nas avaliações para o desenvolvimento de novas substâncias químicas[33]. A técnica mais empregada para essa determinação *in vitro* utiliza a cultura de células CaCO2 que é atualmente o modelo celular mais empregado para investigar os processos de absorção intestinal *in vitro* [34]. No caso de agrotóxicos é importante a avaliação desse parâmetro para se verificar o risco a saúde humana.

1.6.2.5. ENDPOINT estabilidade metabólica (CYP3A4)

O modelo de CYP3A4 é constituído por 1507 compostos, o qual utiliza as sondas DRY, OH2 e O. A estabilidade metabólica é uma propriedade importante dos xenobióticos,

uma vez que afeta parâmetros como a depuração, meia-vida e biodisponibilidade oral. Sendo assim, a estabilidade metabólica é determinada com o objetivo de avaliar o potencial tóxico do xenobiótico, verificar se o metabólito é farmacologicamente inativo devido à fase I do metabolismo ou acumula no corpo devido à falta ou degradação metabólica insignificante.

A determinação da estabilidade metabólica é, portanto, uma medida para descrever o destino metabólico do xenobiótico. A estabilidade metabólica foi acessada através de banco de dados indicativos do metabolismo pela isozima CYP3A4- o grupo mais proeminente de enzimas metabolizadoras de xenobióticos da superfamília de citocromo P450s [35].

1.6.2.6. ENDPOINT volume de distribuição

O modelo de volume de distribuição é constituído por 118 compostos químicos diversos, no qual foram utilizadas as sondas DRY, OH2 e O. O volume de distribuição é o parâmetro farmacocinético que avalia a extensão da distribuição da substância ativa além do plasma. Relacionando a concentração plasmática com a quantidade de agrotóxico existente no organismo [36].

1.7. Análise de Componentes Principais (PCA)

A Análise por Componentes Principais (PCA) emprega métodos matemáticos para agrupar moléculas com características semelhantes, além de ser capaz de identificar moléculas que não seguem o padrão da maioria, os chamados “*outliers*”.

Quando se tem muitas variáveis para uma mesma molécula (exemplo: 100 variáveis), não se consegue avaliar os dados simultaneamente sem perda de informações. O método PCA reduz a dimensionalidade dos dados, extraindo as principais contribuições, que explicam a maior parte da variância da matriz de dados [37-39]. Assim, a primeira componente principal (PC1) descreve a maior variância em relação às componentes subsequentes (PC2, PC3, etc). Esta maior variância está relacionada à maior quantidade de informação dos dados originais e é apresentada em porcentagem (%).

Com a aplicação do PCA, a matriz original é dividida em dois novos conjuntos: *scores* (T) e *loadings* (P), sendo que cada conjunto score/loading representa uma PC (Componente Principal), conforme mostra a Figura 5.

$$\boxed{X} = \begin{matrix} \overline{} \\ | \\ T_1 \end{matrix} + \begin{matrix} \overline{} \\ | \\ T_2 \end{matrix} + \dots + \begin{matrix} \overline{} \\ | \\ T_n \end{matrix}$$

Figura 5. Decomposição da matriz X em componentes principais.

Os *scores* podem ser interpretados como sendo a localização de cada amostra na nova dimensão. Eles também fornecem a intensidade de contribuição de cada *loading*.

Os *loadings* são as novas dimensões rotacionadas em relação às dimensões originais no sentido de maior variância dos dados. Eles mostram a importância de cada variável no conjunto de dados.

Para se determinar a quantidade de PC's que devem ser considerados para garantir que o modelo representa os dados originais, faz-se inicialmente um gráfico de variância *versus* PC. A Figura 6 apresenta o perfil deste gráfico.

Na prática, utiliza-se como último PC que representa uma porcentagem ainda significativa. No exemplo apresentado (Figura 6), a soma de PC1, PC2 e PC3 totaliza aproximadamente 99% e a soma dos demais apenas cerca de 1%. Assim, 3 PC's devem ser considerados.

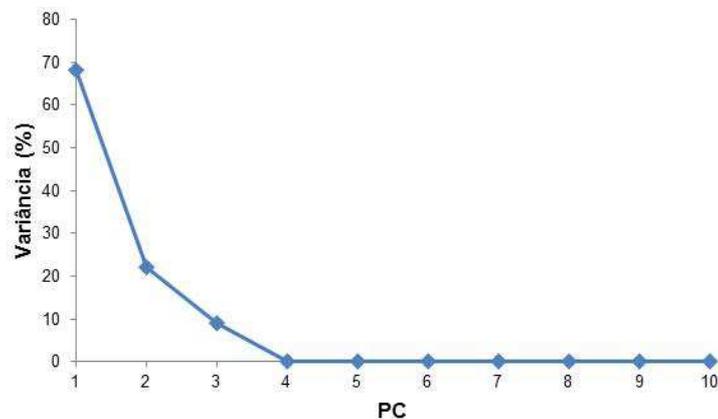


Figura 6: Exemplo de gráfico de variância *versus* PC.

1.8. Regressão por análise dos mínimos múltiplos parciais (PLS)

A técnica de PCA descrita acima é uma análise multivariada que utiliza somente variáveis independentes, sendo assim para se construir modelos de QSPR e QSAR é necessária a utilização de uma ferramenta estatística capaz de reduzir o número de variáveis, como a PCA, além disso, que consiga correlacionar às variáveis independentes com uma ou mais variáveis dependentes, essa técnica é o PLS. Por diminuir o número de variáveis e conseguir relacionar as variáveis independentes com as dependentes esse método é cada vez mais utilizada em diversas áreas [40-42].

A principal diferença entre PCA e PLS é que os modelos gerados por PCA é útil para se ter uma visão geral dos dados de um modelo (qualitativo), enquanto o modelo gerado por PLS é utilizado para a construção de modelos quantitativos. Podendo-se dizer que os modelos gerados por PLS são preditivos e podem ser utilizados para prever variáveis dependentes como o pDL_{50} , para novas moléculas e/ou impurezas de agrotóxicos, antes de serem realizados algum tipo de experimento.

A habilidade preditiva de um modelo de PLS é atribuída à existência de uma relação verdadeira entre as propriedades medidas entre as variáveis independentes e dependentes.

1.9. SIMCA (*Soft Independent Modelling of Class Analogies*)

O método SIMCA (do inglês *Soft independent Modelling of Class analogies*) foi a primeira técnica de classificação utilizada na quimiometria [43]. Atualmente o método é bem conhecido e amplamente utilizado como uma técnica de classificação supervisionada [44]. Nesse método existe um conjunto de moléculas de treinamento que é modelado por análise de componentes principais (PCA do inglês *Principal Component Analysis*), onde hiper-caixas limitam espaço de cada categoria de moléculas. Esse conjunto chamado de treinamento são moléculas de classe conhecida como, por exemplo, herbicidas e inseticidas. Nesse conjunto já existente podem ser colocadas moléculas denominadas testes que serão classificadas como similares ou não similares quando projetadas dentro do espaço da componente principal de cada classe. A classificação de uma molécula teste se dá por meio da distância, sendo que esta molécula será alocada no grupo que apresentar a menor distância para as hiper-caixas das categorias previamente definidas [44].

Usualmente, quando se tem um conjunto treinamento com duas classes como herbicidas e inseticidas, normalmente o modelo gerado possuirá uma região de sobreposição

entre as moléculas, e as moléculas nessa região não poderão ser classificadas. O método SIMCA só poderá ser utilizado quando já se tem um conhecimento prévio das moléculas, e essas deverão ser utilizadas como padrões [45]. O método é sensível para a detecção de “outliers” que são facilmente identificadas das outras moléculas, pois estarão em uma área diferente na projeção [41].

1.10. Estudos de similaridade química

A similaridade química é um dos conceitos mais básico da quimioinformática. É amplamente utilizado para a realização de ensaios virtuais e busca de novas moléculas. Estes estudos têm como princípio que compostos similares possuem propriedades similares [46].

O método de similaridade química tem-se mostrado uma robusta ferramenta na gênese planejada de moléculas bioativas [47]. Tipicamente, pequenas coleções de moléculas são selecionadas a partir de coleções maiores baseadas nas suas similaridades estruturais em relação a compostos de atividade conhecida. Os métodos topológicos (2D) e de sobreposição (3D) são usados para a quantificação da semelhança estrutural entre os compostos analisados.

Métodos de similaridade 3D, como o ROCS [48], conseguem desenvolver melhores combinações estruturais possibilitando comparações entre as estruturas de um dado conjunto de compostos com sua estrutura de referência [15]. Como a sobreposição das estruturas é extremamente complicada, por estas apresentarem volume, faz-se necessário o uso de funções matemáticas para descrever a forma estrutural que é prontamente comparada com outra função gerada para outra molécula. A função usada é a gaussiana [49]. O uso destas funções faz com que o ROCS tenha um alto desempenho em termos de tempo de processo, e uma grande acurácia na sobreposição estrutural.

1.11. Redes Neurais

As redes neurais artificiais (ANNs) consistem em um método de solucionar problemas de inteligência artificial, construindo um sistema que simula o cérebro humano, inclusive seu comportamento, ou seja, aprendendo, errando e fazendo descobertas. São técnicas computacionais que apresentam um modelo inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência.

As ANNs são capazes de lidar com informações incompletas ou com ruídos e por esta razão são especialmente úteis para mapear os problemas que devem ser tolerantes a alguns

erros e têm muitos dados de exemplo disponíveis, mas para as quais regras fixas e rápidas não podem ser facilmente aplicadas.

Elas são feitas de muitas unidades de processamento simples, que são conectadas juntas em uma complexa rede de comunicação, no qual recebem e enviam sinais. Estas unidades são a simulação dos neurônios, recebendo e retransmitindo informações. Não existe qualquer unidade central de processamento seguindo uma sequência lógica de normas.

Em princípio, as ANNs podem calcular qualquer função computável, ou seja, elas podem fazer tudo que um computador digital normal pode fazer. Tudo o que pode ser representado como um mapeamento entre os espaços de um vetor pode ser aproximado, com uma precisão arbitrária, por Redes Neurais.

Redes Neurais podem ser classificadas com base nos seguintes atributos:

- 1) Aplicação (Classificação, Agrupamento de dados, Aproximação de funções, Predição);
- 2) Tipo de conexão [Estáticas (*Feedforward*), Dinâmicas (*Feedback*)];
- 3) Topologia (Camada única, Multicamadas, Recorrentes, Auto-organizáveis);
- 4) Método de aprendizado (Supervisionado, Não supervisionado).

Neste trabalho o modelo de redes neurais que foi usada foi a MLP (multicamadas).

Resumidamente, uma rede neural artificial é um sistema paralelo e distribuído composto por unidades de processamento simples que calculam funções matemáticas. Essas unidades são dispostas em uma ou mais camadas e são interligadas por um grande número de conexões. Para cada conexão são associados pesos para armazenar o conhecimento representado.

Para melhor compreender uma rede neural artificial, é preciso revisar o modelo biológico em que foi inspirado: o cérebro humano (Figura 7).

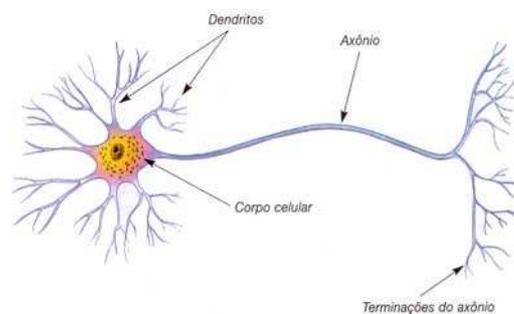


Figura 7. Representação esquemática de um neurônio biológico.

O neurônio é o principal componente do sistema nervoso, sendo o cérebro humano

composto por aproximadamente 86 bilhões destas células [50]. Eles têm um papel essencial na determinação do funcionamento, comportamento e do raciocínio do ser humano. Os neurônios estão conectados uns aos outros através de sinapses e, juntos, formam uma rede neural. As sinapses transmitem estímulos através de diferentes concentrações de sódio e potássio. Esta grande rede proporciona uma incrível capacidade de processamento e armazenamento de informação, sendo que todas as funções e movimentos do organismo estão relacionados ao funcionamento destas células.

O sistema nervoso é formado por um conjunto extremamente complexo de neurônios. Nos neurônios, a comunicação é realizada através de impulsos. Quando um impulso é recebido, o neurônio o processa e, passado um limite de ação, dispara um segundo impulso que produz uma substância neurotransmissora, a qual flui do corpo celular para o axônio (que pode ou não estar conectado a um dendrito de outra célula). O neurônio que transmite o pulso pode controlar a frequência de pulsos aumentando ou diminuindo a polaridade na membrana pós-sináptica.

Existem diversos tipos de neurônios, com diferentes funcionalidades dependendo da sua localização e estrutura morfológica, mas em geral constituem-se dos mesmos componentes básicos: o corpo da célula, o axônio e os dendritos. O corpo da célula (também conhecido como soma) é responsável por coletar e combinar informações vindas de outros neurônios. O axônio (fibra nervosa) é um prolongamento único e grande que aparece no corpo da célula. É responsável pela condução dos estímulos nervosos para as próximas células. Os dendritos (prolongamentos menores em forma de ramificações que emergem do final do axônio) são os responsáveis pela comunicação entre os neurônios através das sinapses. Basicamente, cada neurônio, possui uma região receptiva e outra efetora em relação à condução da sinalização.

A sinapse é a estrutura através da qual ocorrem os processos de comunicação entre os neurônios; ou seja, onde ocorre a passagem do sinal neural (transmissão sináptica) através de processos eletroquímicos. Em uma sinapse os neurônios não se tocam, permanecendo um espaço entre eles, denominado fenda sináptica. O sinal nervoso (impulso), que vem através do axônio da célula pré-sináptica chega a sua extremidade e provoca na fenda a liberação de neurotransmissores depositados em bolsas chamadas de vesículas sinápticas. Essa substância liga-se quimicamente aos receptores no neurônio pós-sináptico, dando continuidade à propagação do sinal.

Um neurônio pode receber ou enviar entre 1.000 e 100.000 conexões sinápticas em relação a outros neurônios, dependendo de seu tipo e localização no sistema nervoso. O

número e a qualidade das sinapses em um neurônio podem variar, entre outros fatores, pela experiência e aprendizagem.

O primeiro modelo de neurônio artificial foi proposto por McCulloch e Pitts [51], um fisiologista e um matemático que interpretaram o funcionamento do neurônio biológico como sendo um circuito de entradas binárias combinadas por uma soma ponderada produzindo uma entrada efetiva. Nesse modelo, as entradas são combinadas para produzir um estado de ativação do neurônio (correspondente à frequência de descarga do neurônio biológico). As entradas chegam através dos dendritos e tem um peso atribuído pela sinapse. A função básica de um neurônio é somar as entradas e retornar uma saída, caso essa soma exceda um limite.

O neurônio artificial (Figura 8) é uma estrutura lógica e matemática que procura simular a forma, o comportamento e as funções de um neurônio biológico. Assim sendo, os dendritos foram substituídos por entradas, cujas ligações com o corpo celular artificial são realizadas através de elementos chamados de peso (simulando as sinapses). Essas ligações, em comparação com o sistema biológico, representam o contato dos dendritos com outros neurônios, formando assim as sinapses. Os estímulos captados pelas entradas são processados pela função de soma, e o limiar de disparo do neurônio biológico foi substituído pela função de ativação.

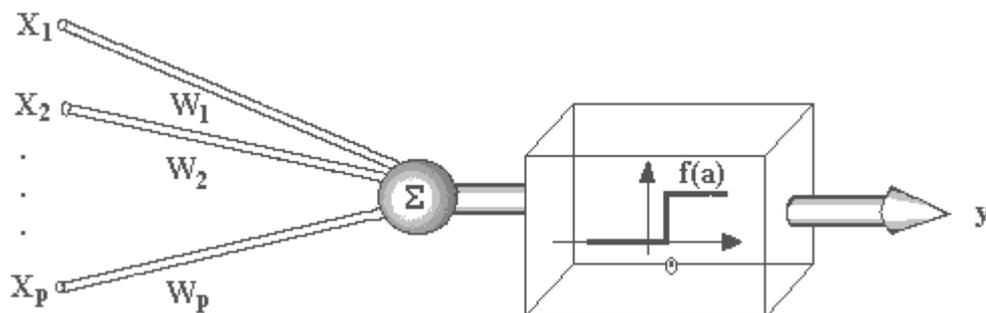


Figura 8. Esquema de unidade de processamento proposto por McCulloch-Pitts [3].

Combinando diversos neurônios artificiais podemos formar o que é chamado de rede neural artificial. Adotando este modelo, foi possível partir para a construção das mais básicas das redes neurais, o Perceptron. As redes Perceptrons são redes neurais de duas camadas usadas no reconhecimento de caracteres. Nesta rede, os nós de saída são conectados às entradas por conjuntos de pesos. A soma dos produtos dos pesos pelas entradas é calculada por cada nó de saída e, se o valor calculado ultrapassar certo limite, o neurônio dispara e ajusta a saída para o valor 1; se o valor calculado for menor que o limite, a saída é ajustada para o valor -1. Neurônios com esse comportamento são chamados de neurônios de McCulloch-Pitts [51]. Ao mesmo tempo, um algoritmo de aprendizado calcula a diferença

entre a saída calculada e os dados de entrada e usa o valor da diferença para ajustar os pesos da rede. Porém esse modelo é limitado, sendo que ele só é capaz de resolver problemas que sejam linearmente separáveis. Para resolver isso, faz-se necessário a utilização de mais camadas na rede. Por causa dessa limitação as pesquisas em ANNs foram deixadas de lado por 12 anos, até que Rumelhart lançou as bases para novas pesquisas com seu livro “*Parallel Distributed Processing*” [52].

Uma rede neural artificial MLP (*Multilayer Perceptron*) é um sistema de neurônios artificiais interligados e divididos em neurônios de entrada, que recebem estímulos do meio externo, neurônios intermediários e neurônios de saída, que se comunicam com o exterior. Os neurônios intermediários são importantes para as ANNs, pois se provou que sem estes, elas se tornam incapazes de resolver problemas linearmente não separáveis. O Multilayer Perceptron foi concebido para resolver esses problemas mais complexos, os quais não poderiam ser resolvidos pelo modelo de neurônio básico.

Este tipo de rede consiste de múltiplas camadas de unidades computacionais, geralmente interconectadas em uma forma *Feedforward*. Isso quer dizer que cada neurônio em uma camada tem conexões diretas a neurônios da próxima camada.

O teorema de aproximação universal dita que toda função contínua que mapeia intervalos de números reais de entrada a algum intervalo de números reais de saída, pode ser aproximada com precisão arbitraria por um Perceptron multicamadas com somente uma camada oculta. Esse resultado só é válido para classes restritas de funções de ativação, por exemplo, funções sigmóides.

Depois de escolhida a rede neural e definida a sua arquitetura, segue-se uma fase chamada de treinamento, ou seja, uma fase cuja tarefa é "treinar" a rede neural com uma coleção de estímulos que se deseja que a rede reconheça quando em operação. O aprendizado pode ser feito de modo supervisionado ou não supervisionado.

Em um treinamento supervisionado, ambas as entradas e saídas são dadas. A rede então processa as entradas e compara sua saída resultante com a saída desejada. Os erros são calculados e usados para ajustar os pesos que controlam a rede. Este processo repete-se várias vezes até que os pesos estejam adequadamente ajustados.

No treinamento não supervisionado, a rede não é alimentada com as saídas desejadas, somente com as entradas. O sistema deve decidir então quais características irá usar para agrupar os dados de entrada. Isto também é conhecido como auto-organização ou adaptação.

As redes multicamadas podem usar um grande número de técnicas de aprendizado, sendo que a mais popular é o de retro propagação do erro (*backpropagation*) proposto por

Rumelhart [52]. Nesse algoritmo, os valores de saída da rede são comparados com a resposta correta dada pela saída padrão esperada para computar o valor de uma função de erro. Usando essa informação, o erro é então alimentado de volta na rede, isto é, o algoritmo ajusta os pesos de cada conexão para reduzir o valor da função de erro. As etapas de propagação dos sinais para a saída e de retro propagação dos erros para ajustar os pesos das conexões vão se repetindo até que a rede apresente um erro aceitável. A Figura 9 mostra um pseudocódigo do algoritmo *backpropagation*.

```
1: Iniciar os pesos da rede com pequenos valores aleatórios
2: repita
3:   erro_total = 0
4:   para cada padrão de treinamento faça
5:     Apresentar padrão à rede
6:     Calcular saída_produzida pela rede
7:     erro = saída_produzida - saída_desejada
8:     erro_total = erro_total + erro
9:     para toda camada i da rede, i =n, n - 1,..., 1 faça
10:      para cada neurônio nij da i-ésima camada faça
11:        Ajustar pesos associados ao neurônio
12:      fim para
13:    fim para
14:  fim para
15: até erro_total < limiar
```

Figura 9. Algoritmo backpropagation [52].

2. OBJETIVO

A proposta do projeto é desenvolver um banco de dados corporativo para que este, no futuro, seja utilizado para o desenvolvimento de novas moléculas, otimizando tempo na pesquisa por utilização de métodos computacionais robustos, diminuindo o uso de animais utilizados em testes. Com o auxílio dos métodos computacionais poder-se-á selecionar moléculas com grande potencial a virar um novo agrotóxico incluindo-se a possibilidade de evitar testes de moléculas que teriam baixas chances de se tornar um novo agrotóxico e, conseqüentemente, com a diminuição do número de testes ocorrerá à diminuição do valor de descoberta e desenvolvimento de novas moléculas. A criação de banco de dados para análise de toxicidade de novos ingredientes ativos e impurezas também é escopo do projeto.

3. MATERIAIS E MÉTODOS

3.1. Materiais

Um microcomputador Core i7 870 2,91GHz, 4Gb DDR3 com OS Linux Fedora 13. Para a montagem do banco de dados foi utilizado o programa Instant JChem da ChemAxon [53].

Durante a etapa de cálculo da similaridade 3D dos compostos das classes das uréias, triazinas, neonicotinóides, carbamatos e benzoiluréias, os programas utilizados para se determinar a similaridade foram os do pacote da Openeye, os programas foram o OMEGA2 e o ROCS [54].

O programa para a geração de estrutura 3D, análise de SIMCA, PLS e utilização dos modelos existentes para verificação de toxicidade das moléculas a saúde humana foram utilizados o programa SYBYL [55] e suas ferramentas, esse programa pertence ao pacote da Tripos.

3.2. Métodos

3.2.1. Seleção de Banco de Dados

3.2.1.1. Busca no site da ANVISA por monografias autorizadas e não autorizadas

Num primeiro momento foram realizadas buscas nas monografias disponibilizadas pela Agência Nacional de Vigilância Sanitária (ANVISA) [56] para substâncias destinadas ao uso agrícola, domissanitário, não agrícola, ambientes aquáticos e preservante de madeira. Verificou-se que havia 472 (quatrocentos e setenta e duas) substâncias autorizadas e 93 (noventa e três) substâncias banidas. Seria necessário então um banco de dados que pudesse ser realizado o filtro de moléculas utilizadas para o uso agrícola.

3.2.1.2. Busca no banco de dados da União Européia

Realizou-se uma pesquisa no banco de dados da União Européia[57], no qual foram encontradas 324 (trezentos e vinte e quatro) substâncias tipificadas como herbicidas e 271

(duzentos e setenta e uma) substâncias tipificadas como inseticidas, incluindo produtos autorizados e não autorizados.

Foi necessário este levantamento para se conhecer a quantidade de moléculas disponíveis que são ou foram utilizadas como inseticidas e herbicidas, para se iniciar o trabalho. Com este levantamento realizado, foi necessário identificar um banco de dados que contivesse informações toxicológicas uma vez que na ANVISA não disponibiliza este tipo de informação e na EU só eram disponibilizados dados toxicológicos de substâncias autorizadas para uso.

3.2.1.3. Busca no banco de dados PPDB

No banco de dados, foram inseridas moléculas autorizadas e não autorizadas, sendo assim foi necessário encontrar um banco de dados com informações confiáveis de toxicidade. Utilizou-se o banco de dados PPDB, que é um banco que disponibiliza resultados de estudos físico-químicos, toxicológicos e ecotoxicológicos e foi desenvolvido pela AERU (Agriculture & Environment Research Unit) na Universidade de Hertfordshire [58]. Através da pesquisa neste banco de dados foram listados 387 (trezentos e oitenta e sete) moléculas da classe de inseticidas e 405 (quatrocentos e cinco) moléculas da classe de herbicidas.

3.2.1.4. Banco Selecionado

Um novo banco de dados foi montado utilizando o programa Instant Jchem, no qual oito campos foram estabelecidos:

- Nome do ativo;
- CAS;
- DL₅₀ dermal em ratos;
- DL₅₀ oral em ratos;
- DL₅₀ inalatória em ratos;
- Irritação ocular;
- Irritação cutânea;
- Referência bibliográfica;

Com os campos determinados estabeleceu-se um filtro para incluir apenas moléculas que apresentassem pelo menos dados de DL₅₀ oral em ratos. O banco de dados foi construindo utilizando-se dados tanto do PPDB como do Pubchem [59]. Com este filtro pode-se criar um

banco de dados com 325 (trezentos e vinte e cinco) moléculas de herbicidas e 292 (duzentos e noventa e duas) moléculas de inseticidas.

3.2.1.5. Outros bancos gerados

A fim de se verificar o comportamento de grupos específicos de compostos, bancos menores foram criados, a partir dos dois bancos maiores de inseticidas e herbicidas. As classes escolhidas foram de acordo com o interesse da empresa e os grupos químicos selecionados foram: Uréia, Triazina, Neonicotinóide, Carbamato e Benzoiluréias.

3.2.2. Estrutura 3D

- 1) O banco de dados gerado no programa Instant Jchem foi exportado em formato sdf.
- 2) Para todos os bancos foram geradas estrutura 3D das moléculas. Para isso utilizou-se o programa SYBYL e a ferramenta CONCORD, que através das estruturas 2D, banco de dados em extensão sdf elaborado pelo programa Instant Jchem, sendo um novo de banco com estruturas 3D obtido.
- 3) As estruturas tridimensionais tiveram suas energias minimizadas usando o campo de força padrão Tripos. As cargas das moléculas foram calculadas através do método Gasteiger-Huckel.

3.2.3. Projeção nos modelos do Volsurf

Os bancos de dados das benzoiluréias, carbamatos, triazinas, uréias e neonicotinóides foram projetadas nos modelos. As sondas utilizadas foram (OH2, DRY e O) e 94 descritores do Volsurf. Os modelos utilizados foram:

- 1) Barreira Hemato-encefálica (BBB)
- 2) Inibidor do canal de potássio (hERG)
- 3) Estabilidade metabólica
- 4) Ligação a proteína

- 5) Permeabilidade em células CaCO₂
- 6) Volume de distribuição

3.2.4. SIMCA aplicado para Reconhecimento de moléculas

Os bancos de dados de herbicidas e inseticidas foram agrupados e constitui-se um único banco de dados. Esse banco de dados foi colocado no programa Sybyl onde se utilizou o Concord para o cálculo da estrutura 3D e em seguida foram calculados os descritores químicos através do Volsurf. Em seguida o peso de Fisher foi utilizado para selecionar o subconjunto de descritores que estão relacionados com a atividade biológica [60-62].

$$W_{12}(i) = \frac{[\bar{X}_i(1) - \bar{X}_i(2)]^2}{S_i^2(1) + S_i^2(2)} \quad (1)$$

$\bar{X}_i(1)$ = média do valor da variável X para os compostos da classe dos herbicidas

$\bar{X}_i(2)$ = média do valor da variável X para os compostos da classe dos inseticidas

$S_i^2(1)$ = variância dos dados para os compostos da classe dos herbicidas

$S_i^2(2)$ = variância dos dados para os compostos da classe dos inseticidas

O próximo passo foi rodar o método SIMCA utilizando uma tabela com o resultado que continha somente os descritores de maiores pesos, utilizando para isso programa SYBYL.

3.2.5. Modelo de PLS

Os bancos de dados das benzoiluréias, carbamatos, triazinas, uréias e neonicotinóides foram utilizados para a elaboração de um modelo preditivo, na qual a DL₅₀ oral foi o endpoint escolhido.

Os bancos de dados foram testados individualmente utilizando o programa Volsurf,

onde os modelos de PLS foram gerados. A validação utilizada foi a de *leave-one-out* (LOO).

3.2.6. Similaridade 3D

Os estudos de similaridade foram realizados para os bancos benzoiluréias, carbamatos, triazinas, uréias e neonicotinóides. A similaridade foi realizada de modo todos contra todos, isso é cada molécula foi comparada com todas as moléculas do banco inclusive ela mesma. A similaridade utilizada foi a de Tanimoto Combo que é a soma da similaridade de forma e cor e a pontuação vai de 0-2. A similaridade de forma é calcula através da equação 2 e a pontuação vai de 0-1, sendo que o resultado igual a 0 nos diz que não existe similaridade entre a molécula alvo e a molécula de entrada e 1 diz que a molécula alvo possui a mesma forma da molécula de entrada. A similaridade por cor apresenta a mesma pontuação e verifica na molécula as regiões de aceitadores ou doadores de H, ânions, cátions, anéis e regiões hidrofóbicas [63].

$$\text{Shape Tanimoto} = \frac{V_c}{V_a + V_b - V_c} \quad (2)$$

V_c = Volume em comum entre as duas moléculas

V_a = Volume total da molécula de entrada

V_b = Volume total da molécula alvo

3.2.6.1. Construção do banco multiconformacional

Para estudos de similaridade 3D é necessário que todos os compostos presentes no banco e o composto de referência apresentem uma biblioteca multiconformacional. A geração das multiconformações foi feita pelo programa OMEGA2. Este programa “quebra” as estruturas em suas ligações e gera um banco de fragmentos que é usado para construir novamente os compostos através da “torção” das ligações quebradas. Esta etapa necessita de uma rigorosa definição de parâmetros. O programa gera e pontua as conformações de acordo com a energia que a mesma apresenta (janela de energia).

O OMEGA2 apresenta uma seleção de parâmetros com valores padrões, no entanto alguns deles foram modificados. O intervalo de energia usado para a janela de energia foi de -

10 kcal mol⁻¹; campo de força usado foi o mmff94; número máximo de conformação por composto foi de 200; o desvio quadrático médio (RMSD) foi de 0,8 Å.

3.2.6.2. Cálculo de similaridade através do ROCS

O programa ROCS necessita dois arquivos de entrada, um que é o banco no qual se deseja realizar o estudo de similaridade e outro que é denominado como estrutura de referência. O ROCS irá alinhar o banco com estrutura de referência e pontuará os compostos de acordo com a similaridade tanto por “forma”, “cor” ou “combinada”. Assim como o OMEGA2, o ROCS também apresenta uma lista de parâmetros que devem ser cuidadosamente avaliados.

3.2.6.3. Estudo de similaridade dos compostos

Foram construídos quatro bancos (benzoiluréias, carbamatos, triazinas, uréias e neonicotinóides) de estruturas multiconformacionais para o estudo de similaridade 3D. O estabelecimento da estrutura de referência foi estabelecido com base em uma estratégia onde todas as estruturas são avaliadas contra todas, gerando então uma matriz de dados. Esse processo foi realizado para todos os bancos e está ilustrado na Figura 10.

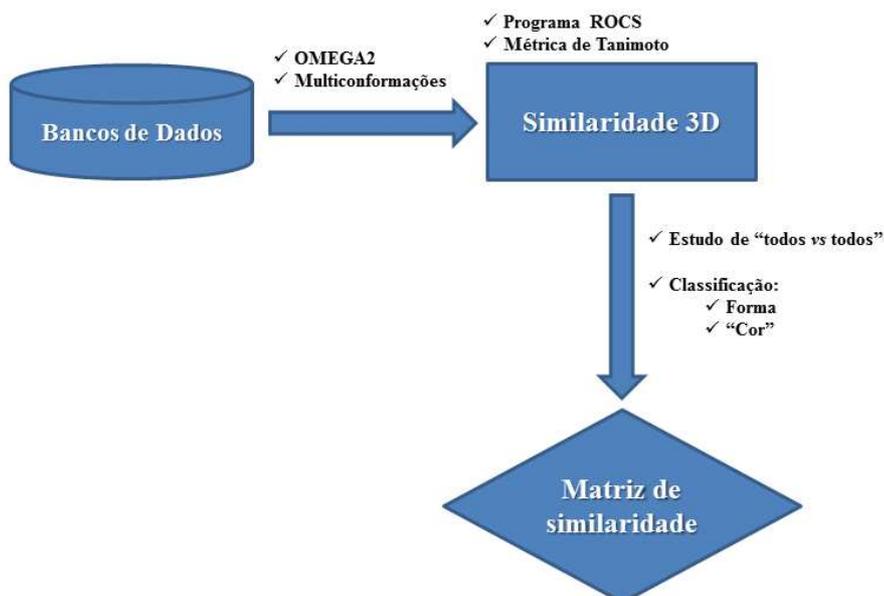


Figura 10. Fluxograma ilustrativo do processo para o cálculo da similaridade 3D dos bancos de dados.

4. RESULTADO E DISCUSSÃO

4.1. Uréias

A Figura 11 representa o gráfico de BBB para os herbicidas das classes das Uréias. O ativo chloroxuron penetra na BBB é o único que se encontra no quarto quadrante. Os ativos dimefuron e ethidimuron estão em uma região do gráfico cuja predição é incerta e possuem poucas substâncias no mesmo ambiente químico, sendo assim, não se pode afirmar as características dessas substâncias para este endpoint. Já os outros herbicidas desta classe penetram na Barreira Hemato-Encefálica e se encontram no primeiro quadrante.

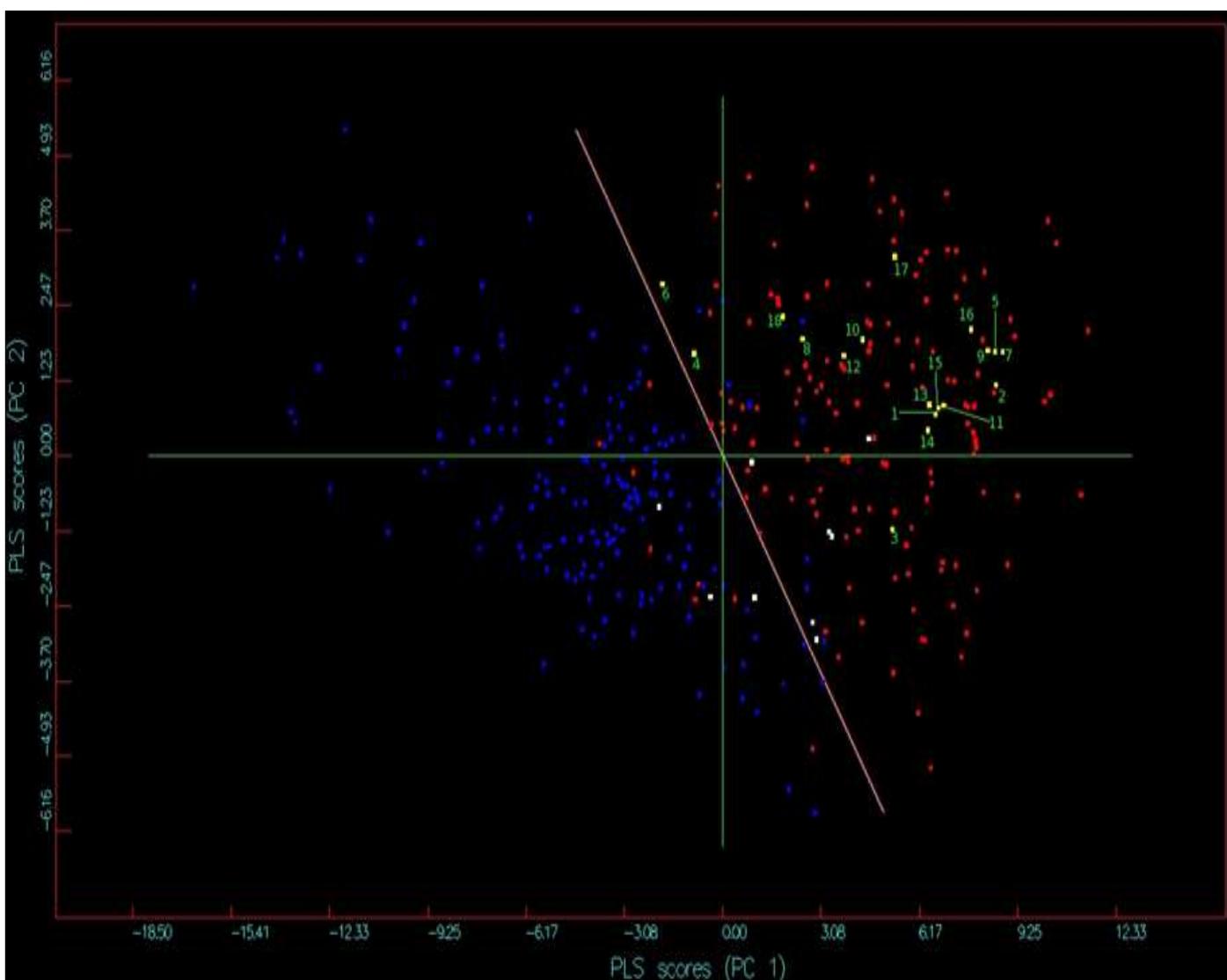


Figura 11. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos herbicidas da classe das Uréias.

A Figura 12 representa o gráfico de hERG para os herbicidas da classe das Uréias. Todos os herbicidas testados desta classe não são bloqueadores do canal de hERG, sendo que a maioria dos compostos se encontram no primeiro quadrante e cinco compostos no segundo quadrante.

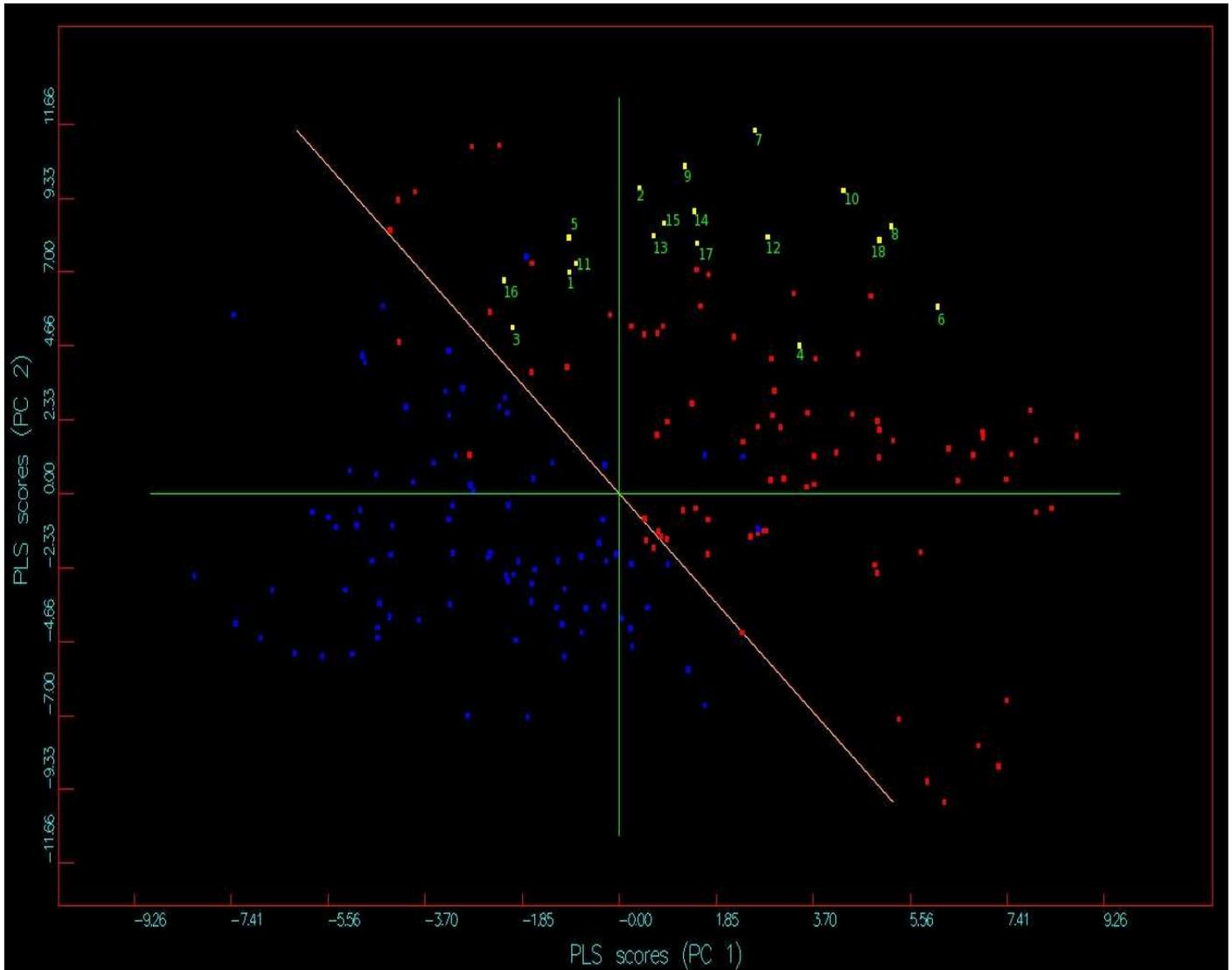


Figura 12. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos herbicidas da classe das Uréias.

A Figura 13 representa o gráfico de estabilidade metabólica para os herbicidas da classe das uréias. Os ativos chloroxuron e o neburon se encontram no terceiro quadrante sendo considerados instáveis metabolicamente. Já a maioria dos compostos testados desta classe é estável metabolicamente e se encontra no quarto quadrante. O dimefuron e ethidimuron se encontram no primeiro quadrante e também são metabolicamente estáveis. Os herbicidas metabolicamente estáveis não são degradados facilmente no organismo, sendo excretados sem que metabólitos tóxicos sejam formados.

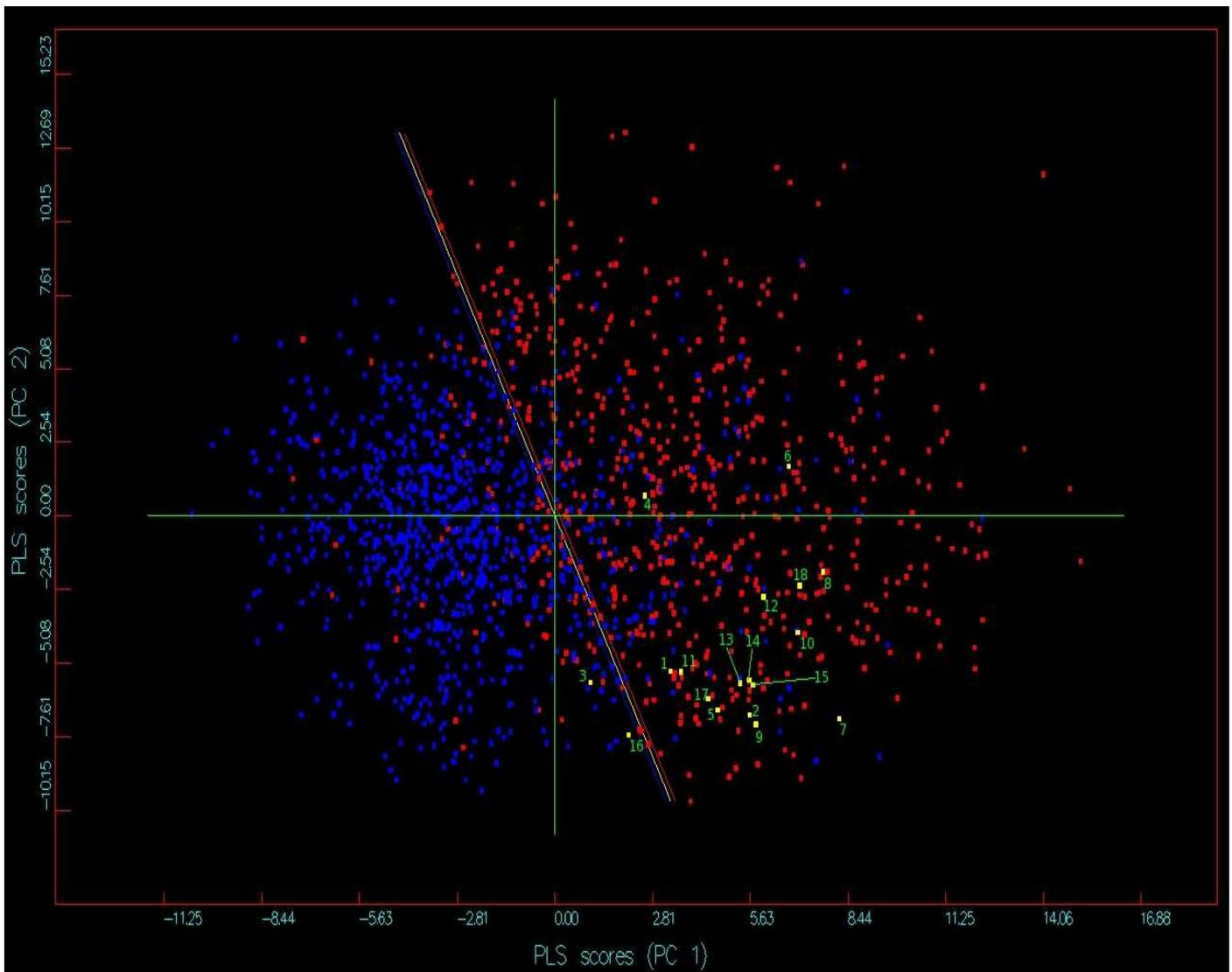


Figura 13. Gráfico de PC1xPC2 da estabilidade metabólica dos herbicidas da classe das Uréias.

A Figura 14 representa o gráfico de ligação à proteína para os herbicidas da classe das uréias. Os ativos chloroxuron, neburon, chlorobromuron, linuron, siduron, diuron, chlorotoluron e isoproturon estão na região do gráfico em que a predição é incerta. Os restantes dos ativos apresentam baixa ligação protéica e estão localizados no segundo e terceiro quadrante.

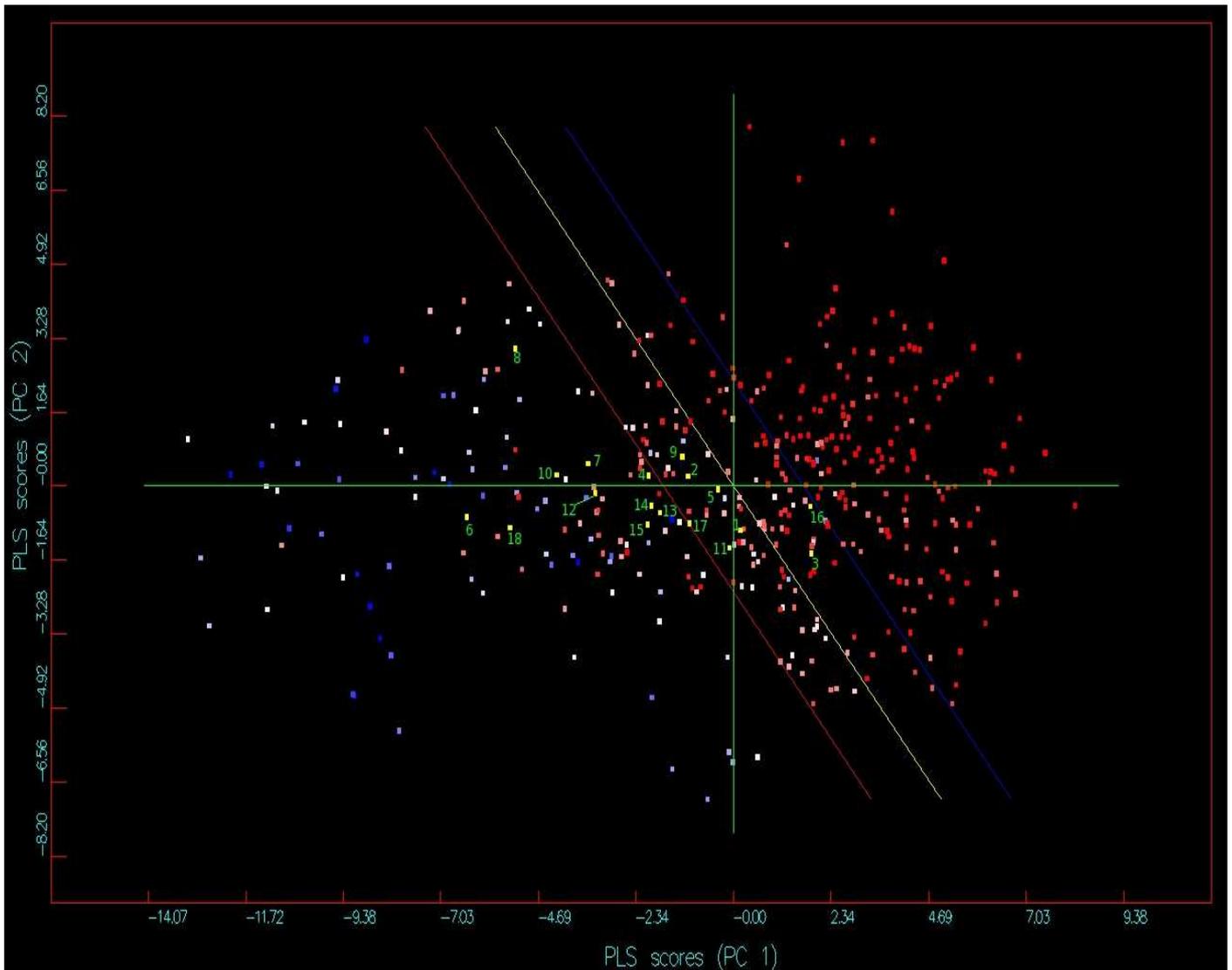


Figura 14. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos herbicidas da classe das Uréias.

A Figura 15 representa o gráfico de CaCO₂ para os herbicidas da classe das uréias. O composto ethidimuron está na região de compostos de baixa permeabilidade. Já os restantes dos compostos estão na região de alta permeabilidade e todos os compostos estão distribuídos entre o primeiro e quarto quadrante.

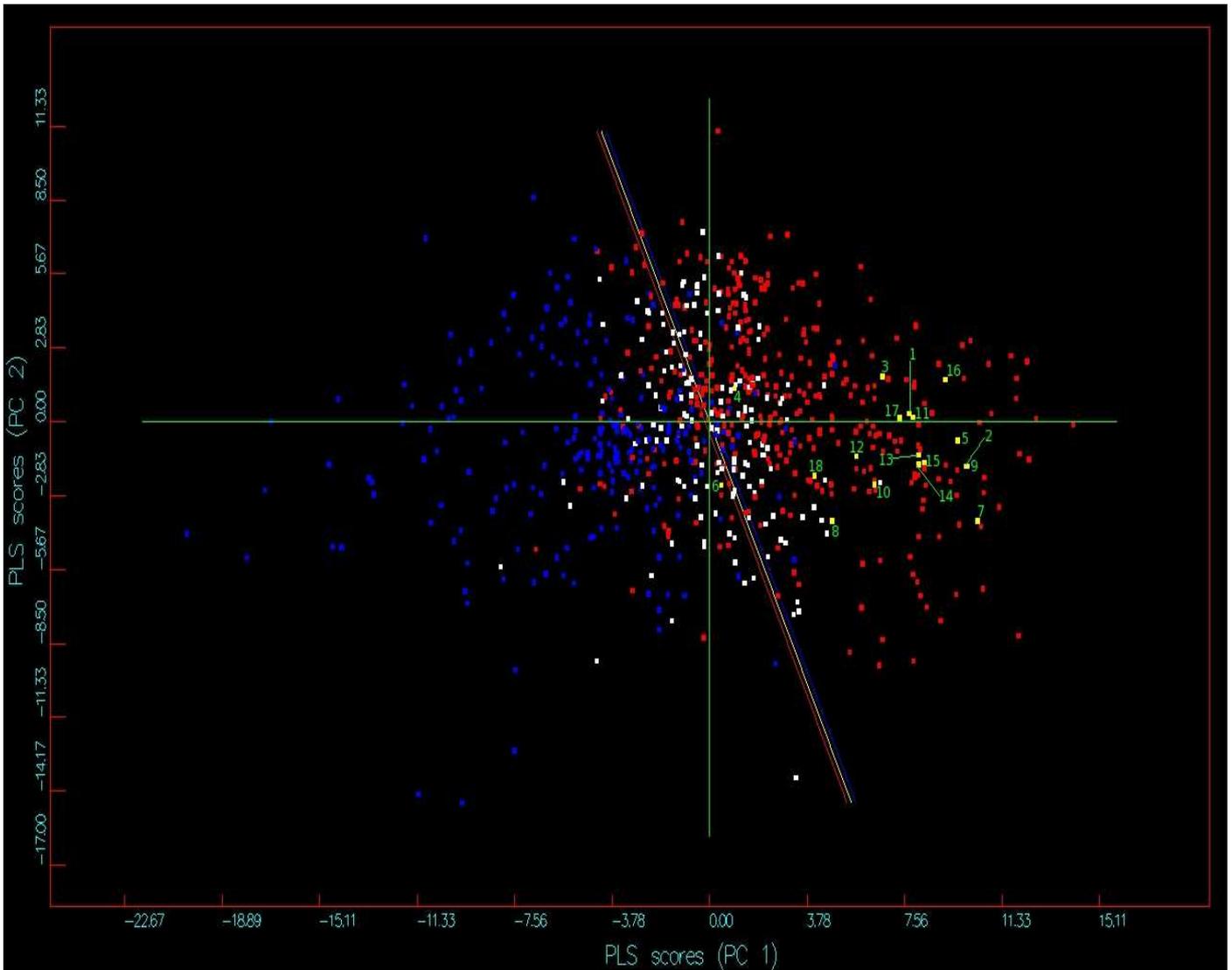


Figura 15. Gráfico PC1xPC2 da permeabilidade em células CaCO₂ dos herbicidas da classe das Uréias.

A Figura 16 representa o gráfico de volume de distribuição (VD) para os herbicidas da classe das uréias. Os compostos neburon e siduron possuem alto volume de distribuição no plasma sanguíneo. O composto chloroxuron se encontra em uma área do gráfico em que a predição é incerta, já os outros compostos estão na região de baixo VD no plasma sanguíneo.

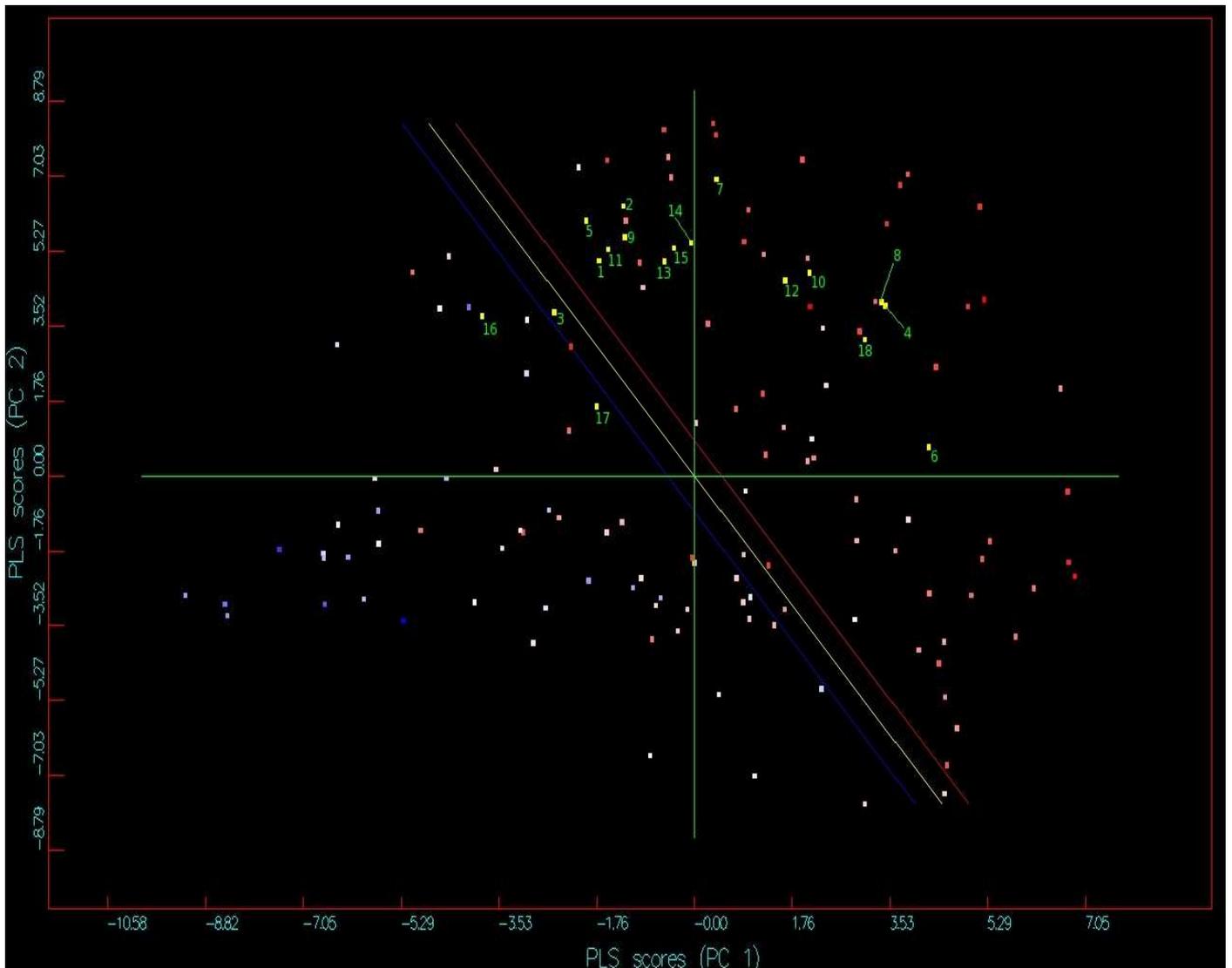


Figura 16. Gráfico PC1xPC2 demonstra o volume de distribuição herbicidas da classe das Uréias.

Os números grifados em verde na Tabela 1 são os ativos que apresentam um nível de similaridade $\geq 70\%$, nas respectivas comparações, para similaridade 3D Tanimoto Combo.

Tabela 1. Similaridade 3D Tanimoto Combo dos herbicidas das classes das Uréias.

Ativos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	2,0																	
2	1,6	2,0																
3	1,1	1,2	2,0															
4	0,9	0,9	1,1	2,0														
5	1,6	2,0	1,3	0,9	2,0													
6	0,9	1,0	0,9	0,9	1,0	2,0												
7	1,6	1,8	0,8	0,8	1,9	0,9	2,0											
8	1,5	1,8	1,2	1,0	1,9	0,9	1,8	2,0										
9	1,8	1,7	1,2	0,8	1,7	1,0	1,6	1,5	2,0									
10	1,4	1,3	1,0	0,9	1,4	1,0	1,3	1,3	1,5	2,0								
11	1,8	2,0	1,1	0,9	2,0	0,9	1,9	1,9	1,6	1,4	2,0							
12	1,0	1,2	1,0	0,9	1,2	1,4	1,0	1,2	1,0	1,2	1,1	2,0						
13	2,0	1,7	0,8	0,9	1,7	0,9	1,7	1,5	1,9	1,4	1,7	1,0	2,0					
14	1,5	1,7	1,5	1,1	1,7	1,0	1,6	1,6	1,5	1,3	1,6	1,1	1,5	2,0				
15	1,7	1,9	0,8	0,9	1,9	0,9	1,9	1,8	1,7	1,4	2,0	1,0	1,8	1,7	2,0			
16	1,5	1,6	0,9	0,9	1,6	0,9	1,6	1,5	1,4	1,2	1,6	0,9	1,5	1,5	1,6	2,0		
17	0,9	1,2	0,8	0,8	0,9	0,9	1,2	0,9	1,0	1,0	0,9	0,9	1,0	1,0	1,2	1,1	2,0	
18	1,1	1,1	0,8	0,8	1,1	1,4	1,0	1,1	1,3	1,4	1,1	1,3	1,1	1,0	1,0	0,9	0,8	2,0

Tabela 2. Legenda para as Figuras 11-16 e Tabela 1 dos herbicidas das classes das Uréias.

Número	Nome	Smile	Fórmula Molecular	DL ₅₀ Oral Aguda em ratos (mg/kg)
1	Chlorobromuron	<chem>CON(C)C(=O)NC1=CC(Cl)=C(Br)C=C1</chem>	C ₉ H ₁₀ BrClN ₂ O ₂	2150,0
2	Chlorotoluron	<chem>CN(C)C(=O)NC1=CC(Cl)=C(C)C=C1</chem>	C ₁₀ H ₁₃ ClN ₂ O	5000,0
3	Chloroxuron	<chem>CN(C)C(=O)NC1=CC=C(OC2=CC=C(Cl)C=C2)C=C1</chem>	C ₁₅ H ₁₅ ClN ₂ O ₂	3700,0
4	Dimefuron	<chem>CN(C)C(=O)NC1=CC(Cl)=C(C=C1)N1N=C(OC1=O)C(C)(C)C</chem>	C ₁₅ H ₁₉ ClN ₄ O ₃	5000,0
5	Diuron	<chem>CN(C)C(=O)NC1=CC(Cl)=C(Cl)C=C1</chem>	C ₉ H ₁₀ Cl ₂ N ₂ O	2000,0
6	Ethidimuron	<chem>CCS(=O)(=O)C1=NN=C(S1)N(C)C(=O)NC</chem>	C ₇ H ₁₂ N ₄ O ₃ S ₂	5000,0
7	Fenuron	<chem>CN(C)C(=O)NC1=CC=CC=C1</chem>	C ₉ H ₁₂ N ₂ O	6400,0
8	Fluometuron	<chem>CN(C)C(=O)NC1=CC=CC(=C1)C(F)(F)F</chem>	C ₁₀ H ₁₁ F ₃ N ₂ O	1450,0
9	Isoproturon	<chem>CC(C)C1=CC=C(NC(=O)N(C)C)C=C1</chem>	C ₁₂ H ₁₈ N ₂ O	1826,0
10	Isouron	<chem>CN(C)C(=O)NC1=NOC(=C1)C(C)(C)C</chem>	C ₁₀ H ₁₇ N ₃ O ₂	630,0
11	Linuron	<chem>CON(C)C(=O)NC1=CC(Cl)=C(Cl)C=C1</chem>	C ₉ H ₁₀ Cl ₂ N ₂ O ₂	1500,0
12	Methabenzthiazuron	<chem>CNC(=O)N(C)C1=NC2=CC=CC=C2S1</chem>	C ₁₀ H ₁₁ N ₃ OS	5000,0
13	Metobromuron	<chem>CON(C)C(=O)NC1=CC=C(Br)C=C1</chem>	C ₉ H ₁₁ BrN ₂ O ₂	2603,0
14	Metoxuron	<chem>COC1=C(Cl)C=C(NC(=O)N(C)C)C=C1</chem>	C ₁₀ H ₁₃ ClN ₂ O ₂	3200,0
15	Monolinuron	<chem>CON(C)C(=O)NC1=CC=C(Cl)C=C1</chem>	C ₉ H ₁₁ ClN ₂ O ₂	1430,0
16	Neburon	<chem>CCCCN(C)C(=O)NC1=CC(Cl)=C(Cl)C=C1</chem>	C ₁₂ H ₁₆ Cl ₂ N ₂ O	11000,0
17	Siduron	<chem>CC1CCCC1NC(=O)NC1=CC=CC=C1</chem>	C ₁₄ H ₂ ON ₂ O	7500,0
18	Tebuthiuron	<chem>CNC(=O)N(C)C1=NN=C(S1)C(C)(C)C</chem>	C ₉ H ₁₆ N ₄ OS	528,0

Analisando-se o grupo de herbicidas da classe das uréias observa-se que os ativos chloroxuron, neburon e siduron podem ser problemáticos à saúde humana, uma vez que os endpoints indicaram que o chloroxuron penetra na BBB e é instável metabolicamente. Já o neburon é instável metabolicamente e possui alto volume de distribuição no plasma sanguíneo. O siduron também apresenta alto valor de distribuição no plasma sanguíneo. Uma característica verificada para todos os compostos da classe da uréia é uma alta permeabilidade para célula CaCO₂, sendo que estas moléculas podem ser absorvidas por células do intestino, apresentando perigo a saúde humana.

4.2. Triazinas

A Figura 17 representa o gráfico de BBB para os herbicidas das classes das Triazinas. Todos os herbicidas desta classe penetram na Barreira Hemato-Encefálica e se encontram entre o primeiro quadrante e o quarto quadrante.

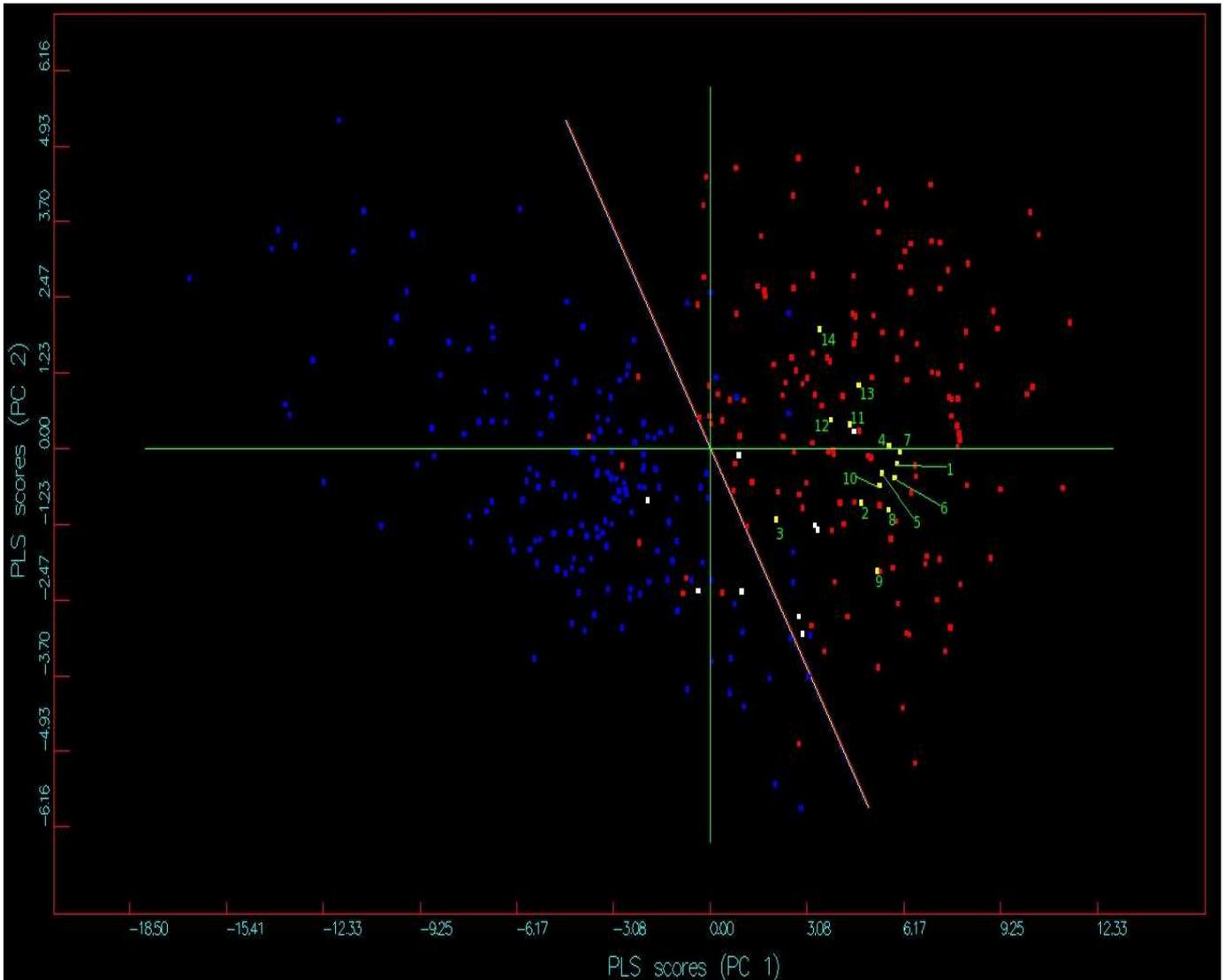


Figura 17. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos herbicidas da classe das triazinas.

A Figura 18 representa o gráfico de hERG para os herbicidas da classe das Triazinas. Todos os herbicidas testados desta classe não são bloqueadores do canal de hERG e se encontram no primeiro quadrante.

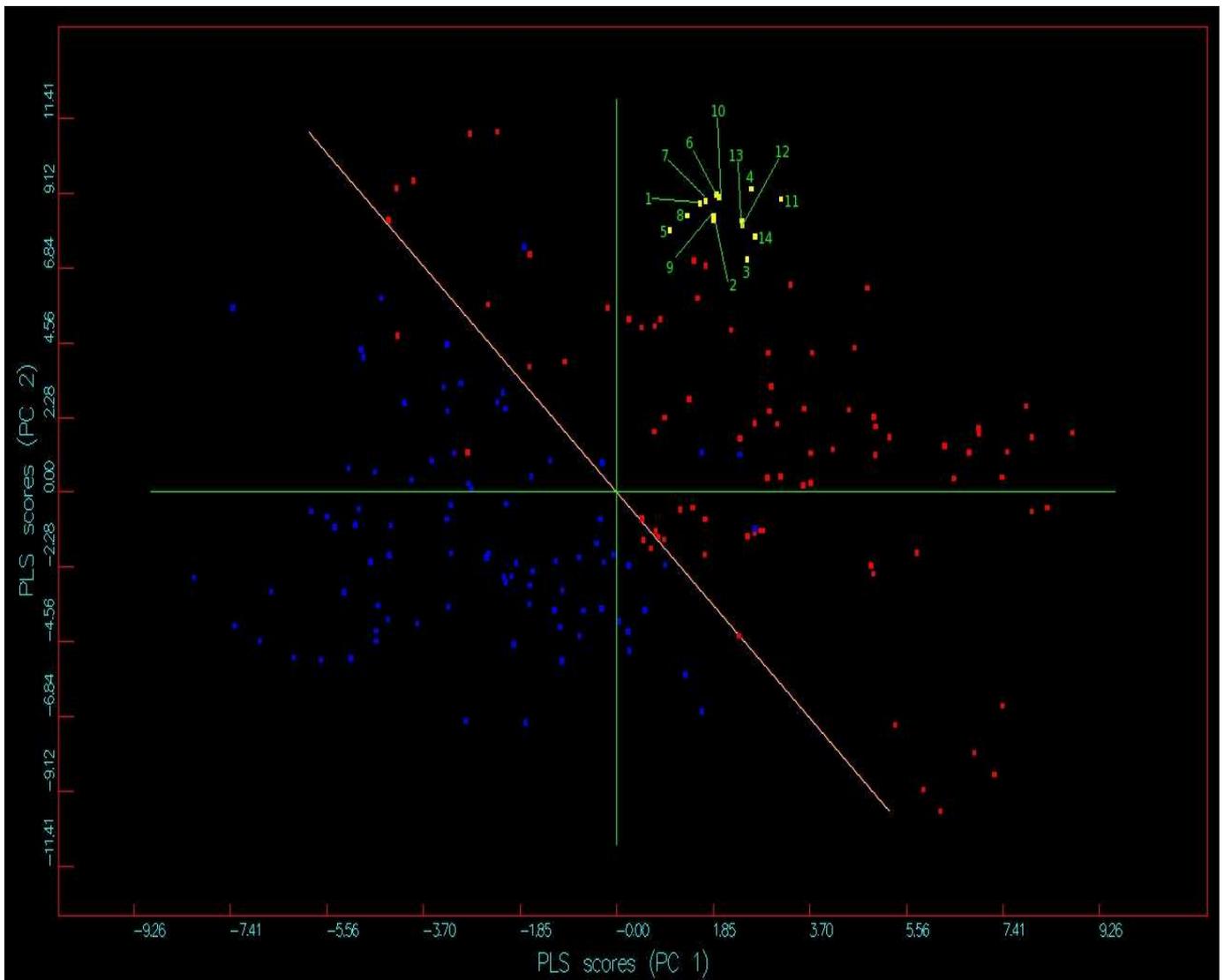


Figura 18. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos herbicidas da classe das triazinas.

A Figura 19 representa o gráfico de estabilidade metabólica para os herbicidas da classe das Triazinas. A maioria dos compostos testados desta classe são estáveis metabolicamente e se encontram no quarto quadrante, exceto a cyanazine, que se encontra na divisa entre o primeiro e o quarto quadrante. Os herbicidas triazínicos são metabolicamente estáveis e não são degradados facilmente no organismo, sendo excretados sem que metabólitos tóxicos sejam formados.

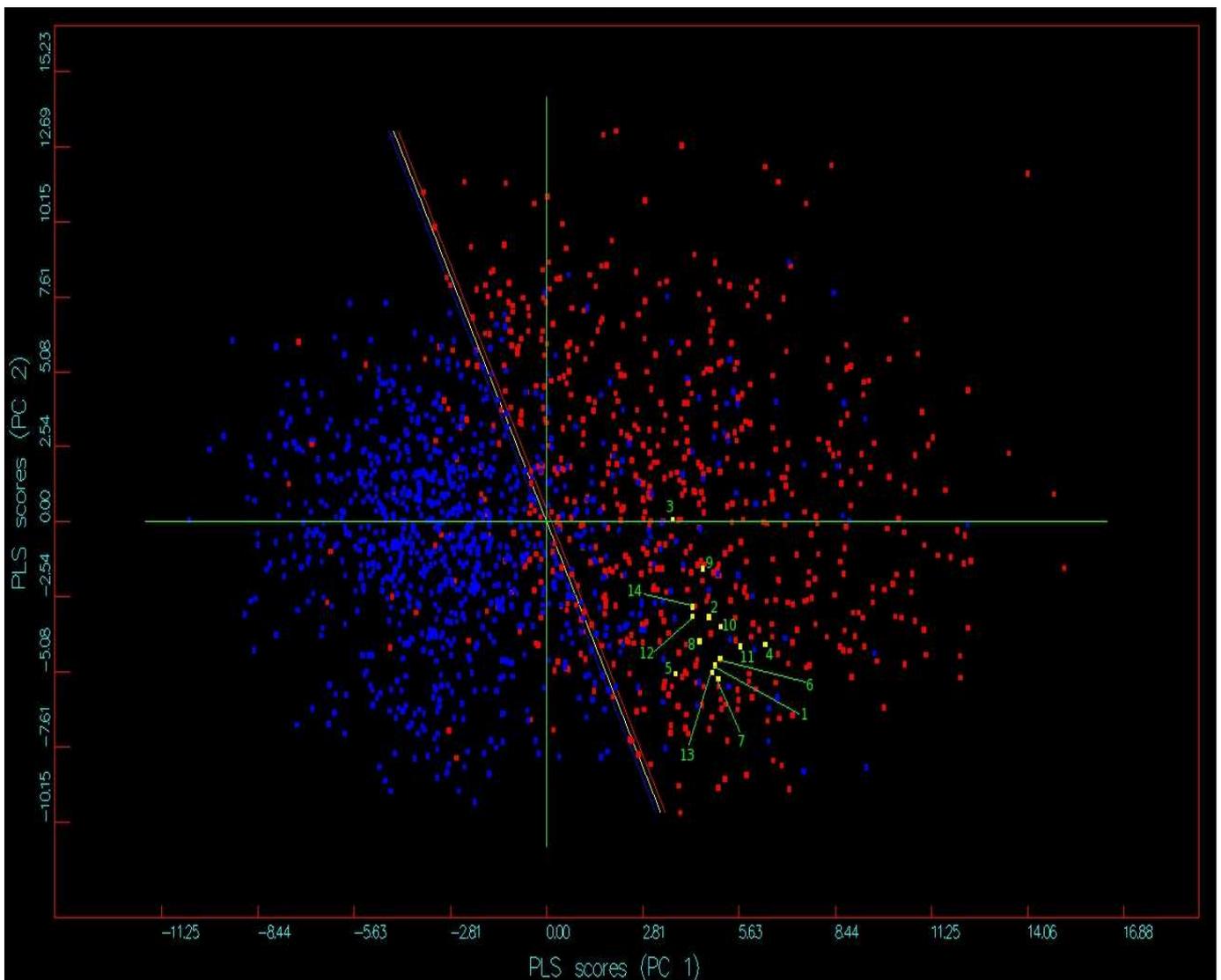


Figura 19. Gráfico de PC1xPC2 da estabilidade metabólica dos herbicidas da classe das triazinas.

A Figura 20 representa o gráfico de ligação à proteína para os herbicidas da classe das Triazinas. Todos os ativos se encontram no terceiro quadrante e na região do gráfico de baixa ligação protéica. Pode-se verificar que no caso de ligação à proteína todos os compostos estão aglomerados, exceto a simazina. Com isso, pode-se dizer que a simazina é uma substância que também apresenta uma baixa ligação protéica, mas o seu comportamento na ligação protéica difere das outras substâncias do grupo, uma vez que essa se encontra em um ambiente químico diferente.

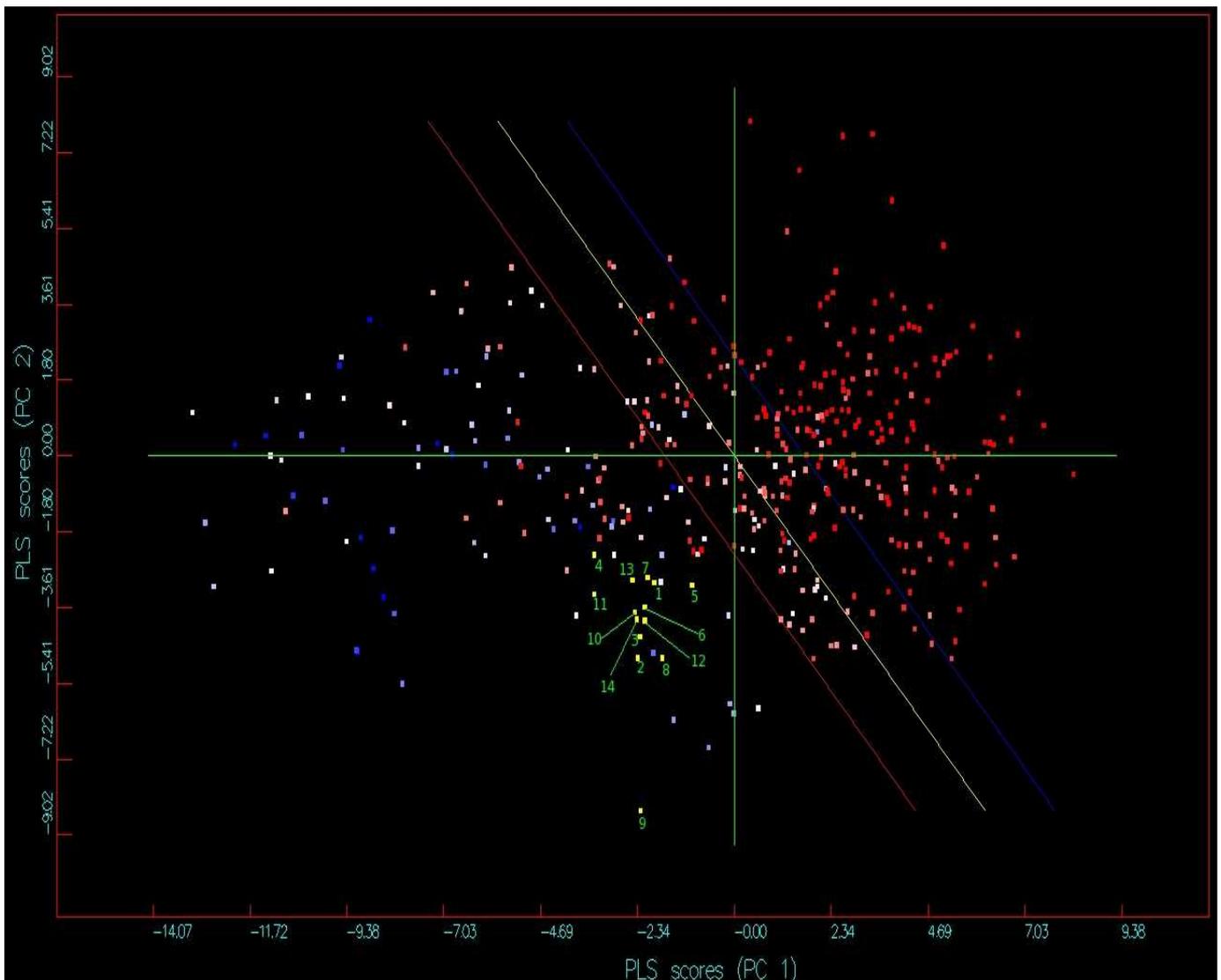


Figura 20. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos herbicidas da classe das triazinas.

A Figura 21 representa o gráfico de CaCO₂ para os herbicidas da classe das Triazinas. Todos os compostos estão no quarto quadrante e são substâncias que apresentam alta permeabilidade para células CaCO₂. Para este endpoint os compostos estão em um mesmo ambiente químico altamente agrupados.

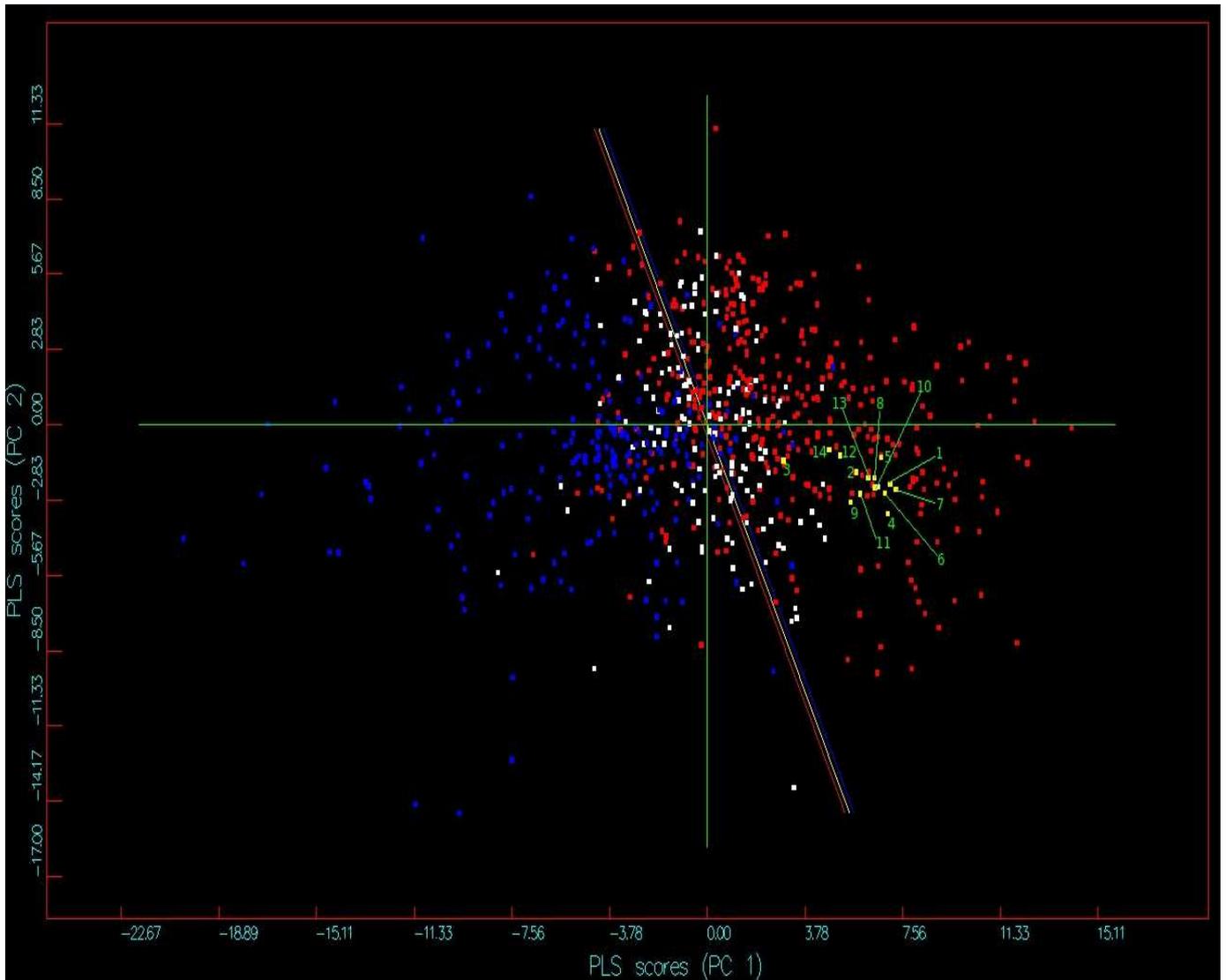


Figura 21. Gráfico PC1xPC2 da permeabilidade em células CaCO₂ dos herbicidas da classe das triazinas.

A Figura 22 representa o gráfico de volume de distribuição (VD) para os herbicidas da classe das triazinas. Todos os compostos possuem baixo volume de distribuição no plasma sanguíneo. Os compostos dimethametryne, propazine e simazine estão localizados no segundo quadrante. Já a prometryne se encontra na divisa do primeiro e segundo quadrante.

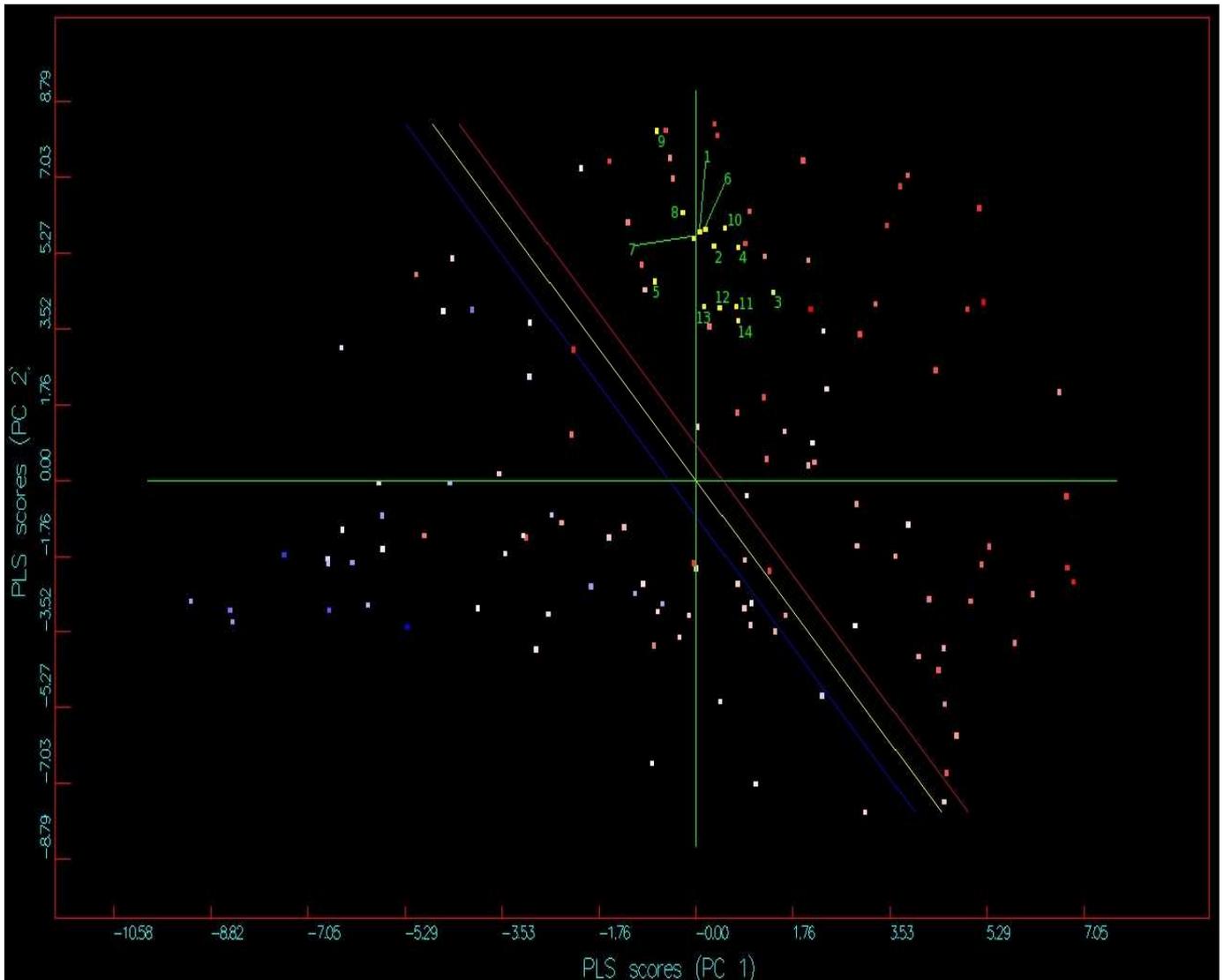


Figura 22. Gráfico PC1xPC2 demonstra o volume de distribuição dos herbicidas da classe das triazinas.

Os números grifados em verde na Tabela 3 são os ativos que apresentam um nível de similaridade $\geq 70\%$ por similaridade 3D Tanimoto Combo.

Tabela 3. Similaridade 3D Tanimoto Combo dos herbicidas das classes das Triazinas.

Ativos	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	2,0													
2	1,8	2,0												
3	1,6	1,8	2,0											
4	1,9	1,8	1,6	2,0										
5	1,7	1,6	1,5	1,7	2,0									
6	1,7	1,7	1,5	1,6	1,4	2,0								
7	1,9	1,7	1,6	1,8	1,6	1,8	2,0							
8	1,7	1,8	1,6	1,7	1,5	1,9	1,9	2,0						
9	1,7	1,8	1,6	1,7	1,6	1,6	1,6	1,7	2,0					
10	1,7	1,7	1,5	1,8	1,7	1,5	1,7	1,6	1,8	2,0				
11	1,7	1,8	1,7	1,7	1,5	1,8	1,6	1,7	1,7	1,6	2,0			
12	1,7	1,9	1,8	1,8	1,5	1,7	1,7	1,8	1,8	1,6	1,9	2,0		
13	1,9	1,8	1,7	1,8	1,6	1,6	1,8	1,7	1,7	1,8	1,7	1,9	2,0	
14	1,5	1,6	1,5	1,5	1,4	1,4	1,4	1,5	1,7	1,6	1,5	1,6	1,5	2,0

Tabela 4. Legenda para as Figuras 17-22 e Tabela 3 dos herbicidas das classes das Triazinas.

Número	Nome	Smile	Fórmula Molecular	DL ₅₀ oral aguda em ratos (mg/kg)
1	Ametryne	CCNC1=NC(NC(C)C)=NC(SC)=N1	C ₉ H ₁₇ N ₅ S	1160,0
2	Atrazine	CCNC1=NC(NC(C)C)=NC(Cl)=N1	C ₈ H ₁₄ ClN ₅	1869,0
3	Cyanazine	CCNC1=NC(NC(C)(C)C#N)=NC(Cl)=N1	C ₉ H ₁₃ ClN ₆	182,0
4	Desmetryn	CNC1=NC(NC(C)C)=NC(SC)=N1	C ₈ H ₁₅ N ₅ S	1390,0
5	dimethametryne	CCNC1=NC(NC(C)C(C)C)=NC(SC)=N1	C ₁₁ H ₂₁ N ₅ S	3000,0
6	prometon	COC1=NC(NC(C)C)=NC(NC(C)C)=N1	C ₁₀ H ₁₉ N ₅ O	2980,0
7	prometryne	CSC1=NC(NC(C)C)=NC(NC(C)C)=N1	C ₁₀ H ₁₉ N ₅ S	2000,0
8	propazine	CC(C)NC1=NC(NC(C)C)=NC(Cl)=N1	C ₉ H ₁₆ ClN ₅	7000,0
9	simazine	CCNC1=NC(Cl)=NC(NCC)=N1	C ₇ H ₁₂ ClN ₅	5000,0
10	simetryne	CCNC1=NC(NCC)=NC(SC)=N1	C ₈ H ₁₅ N ₅ S	750,0
11	terbumeton	CCNC1=NC(NC(C)(C)C)=NC(OC)=N1	C ₁₀ H ₁₉ N ₅ O	651,0
12	terbuthylazine	CCNC1=NC(NC(C)(C)C)=NC(Cl)=N1	C ₉ H ₁₆ ClN ₅	1590,0
13	terbutryne	CCNC1=NC(NC(C)(C)C)=NC(SC)=N1	C ₁₀ H ₁₉ N ₅ S	2045,0
14	trietazine	CCNC1=NC(=NC(Cl)=N1)N(CC)CC	C ₉ H ₁₆ ClN ₅	494,0

Analisando-se o grupo de herbicidas da classe das triazinas observa-se que os compostos dessa classe apresentam os seis endpoints bem definidos, o que permite dizer que os compostos triazínicos penetram na barreira hemato-encefálica, não são bloqueadoras do canal de hERG, são estáveis metabolicamente, apresentam baixa ligação protéica, alta permeabilidade para célula CaCO₂ e baixo volume de distribuição no plasma sanguíneo. Em todos os gráficos os compostos estão agrupados em uma região do gráfico, com isso pode-se dizer que o espaço químico está bem definido para predição de toxicidade de compostos triazínicos utilizando os seis modelos de endpoints (Figura 17-22).

4.3. Neonicotinóides

A Figura 23 representa o gráfico de BBB para os inseticidas da classe dos Neonicotinóides. Todos os inseticidas desta classe podem penetrar na Barreira Hemato-Encefálica, com exceção do composto dinotefuran, que se encontra na divisa do primeiro com o quarto quadrante. Os demais compostos analisados se encontram no quarto quadrante.

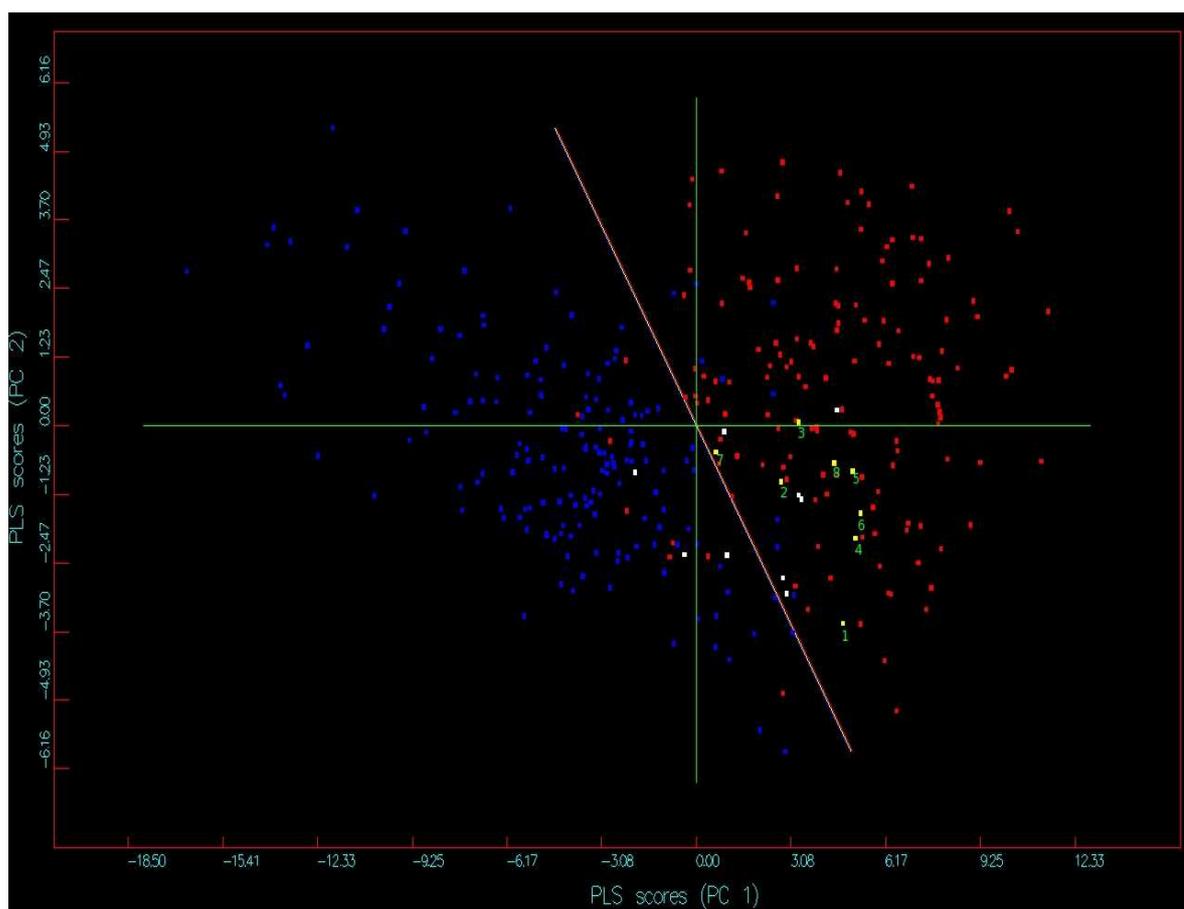


Figura 23. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos inseticidas da classe dos neonicotinóides.

A Figura 24 representa o gráfico de hERG para os inseticidas da classe dos Neonicotinóides. Todos os inseticidas testados desta classe não são bloqueadores do canal de hERG, exceto o composto imidacloprid, que se encontra na divisa do primeiro e o segundo quadrante. Os restantes dos compostos analisados se encontram no primeiro quadrante.

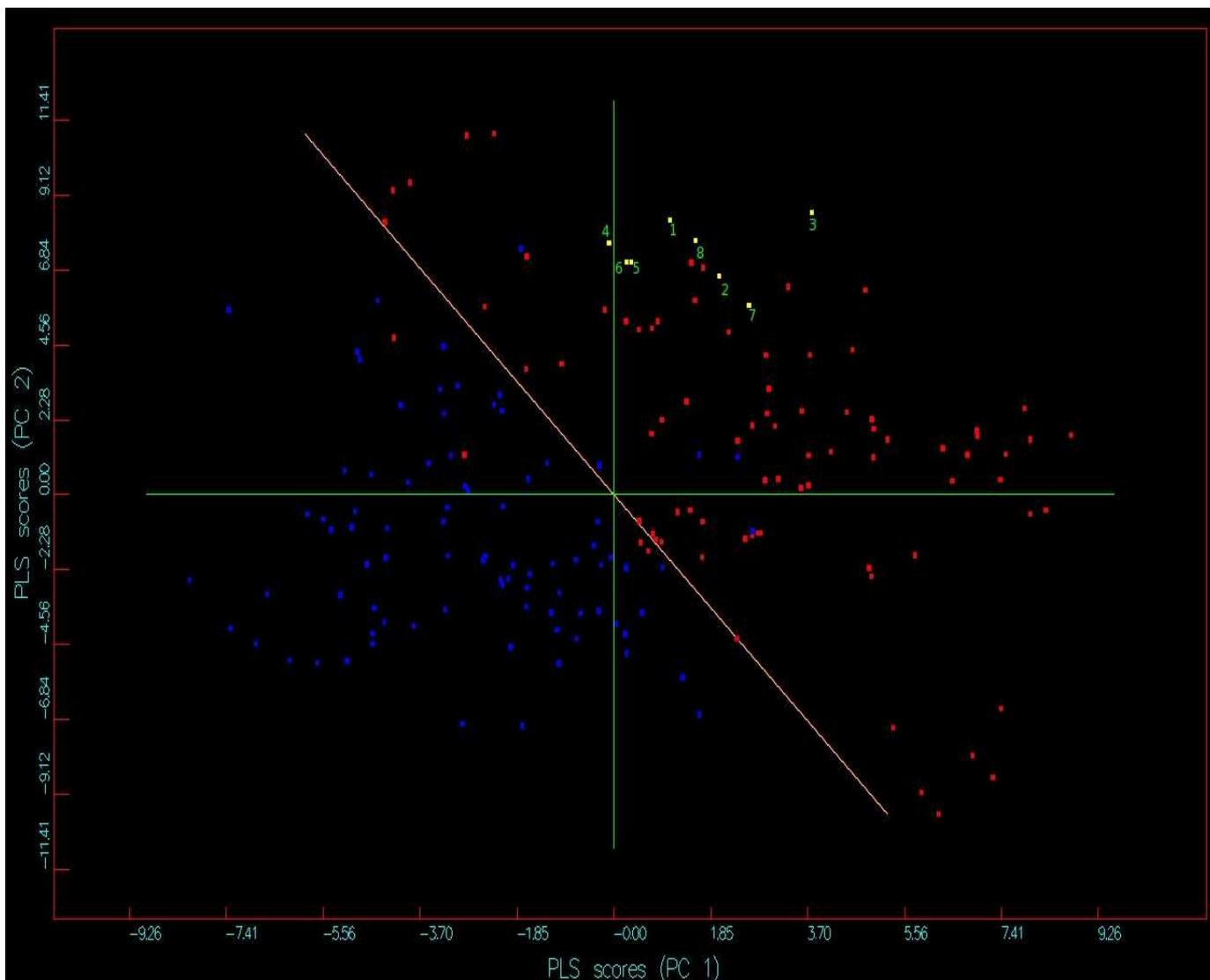


Figura 24. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos inseticidas da classe dos neonicotinóides.

A Figura 25 representa o gráfico de estabilidade metabólica para os inseticidas da classe dos Neonicotinóides. Todos os compostos testados desta classe são estáveis metabolicamente e se encontram no quarto quadrante. Os inseticidas neonicotinóides são metabolicamente estáveis e não são degradados facilmente no organismo, sendo excretados sem que metabólitos tóxicos sejam formados.

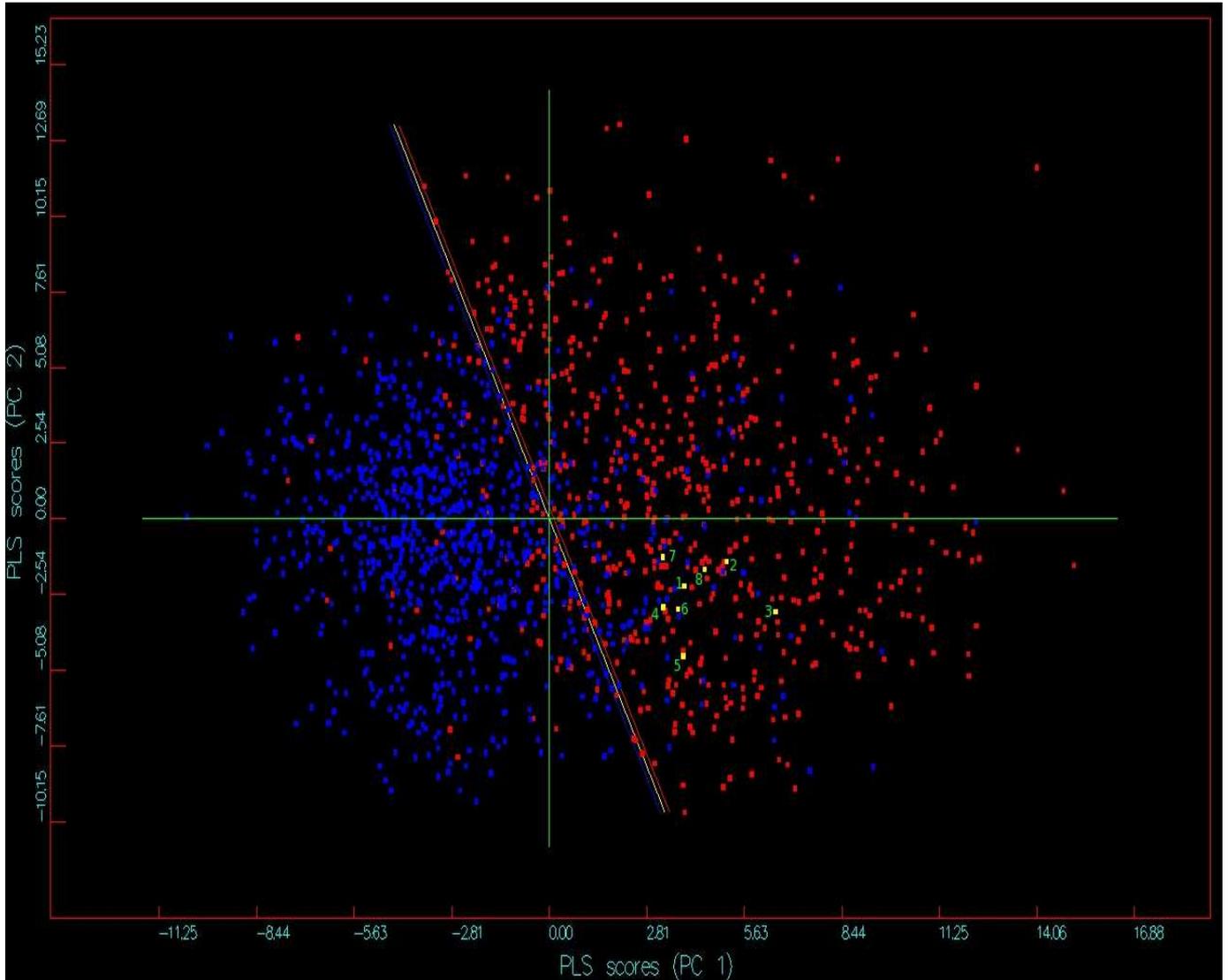


Figura 25. Gráfico de PC1xPC2 da estabilidade metabólica dos inseticidas da classe dos neonicotinóides.

A Figura 26 representa o gráfico de ligação à proteína para os inseticidas da classe dos Neonicotinóides. Os ativos imidacloprid, nitenpyram, thiacloprid e 1H-Imidazol-2-amine-1,6-dichloro-3-pyridinylmethyl-4,5-dihydro-N-nitroso se encontram na região do gráfico em que a predição é incerta, com isso não podemos afirmar se essas substâncias químicas apresentam alta ou baixa ligação protéica. Já os compostos acetamiprido, chlotianidin, dinotefuran e tiametoxam apresentam baixa ligação protéica e se encontram no terceiro quadrante.

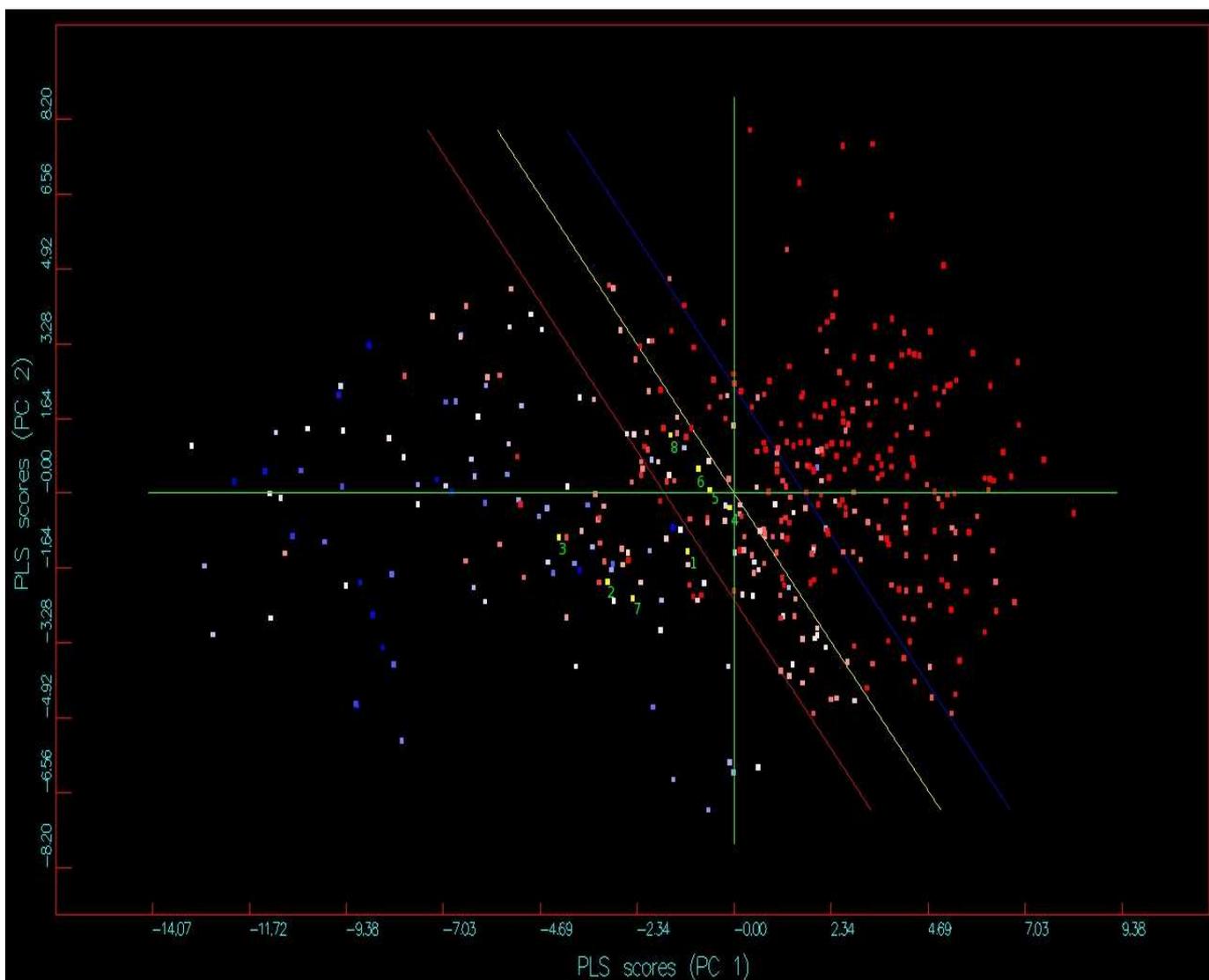


Figura 26. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos inseticidas da classe dos neonicotinóides.

A Figura 27 representa o gráfico de CaCO₂ para os inseticidas da classe dos Neonicotinóides. Todos os compostos estão no quarto quadrante e são substâncias que apresentam alta permeabilidade para células CaCO₂.

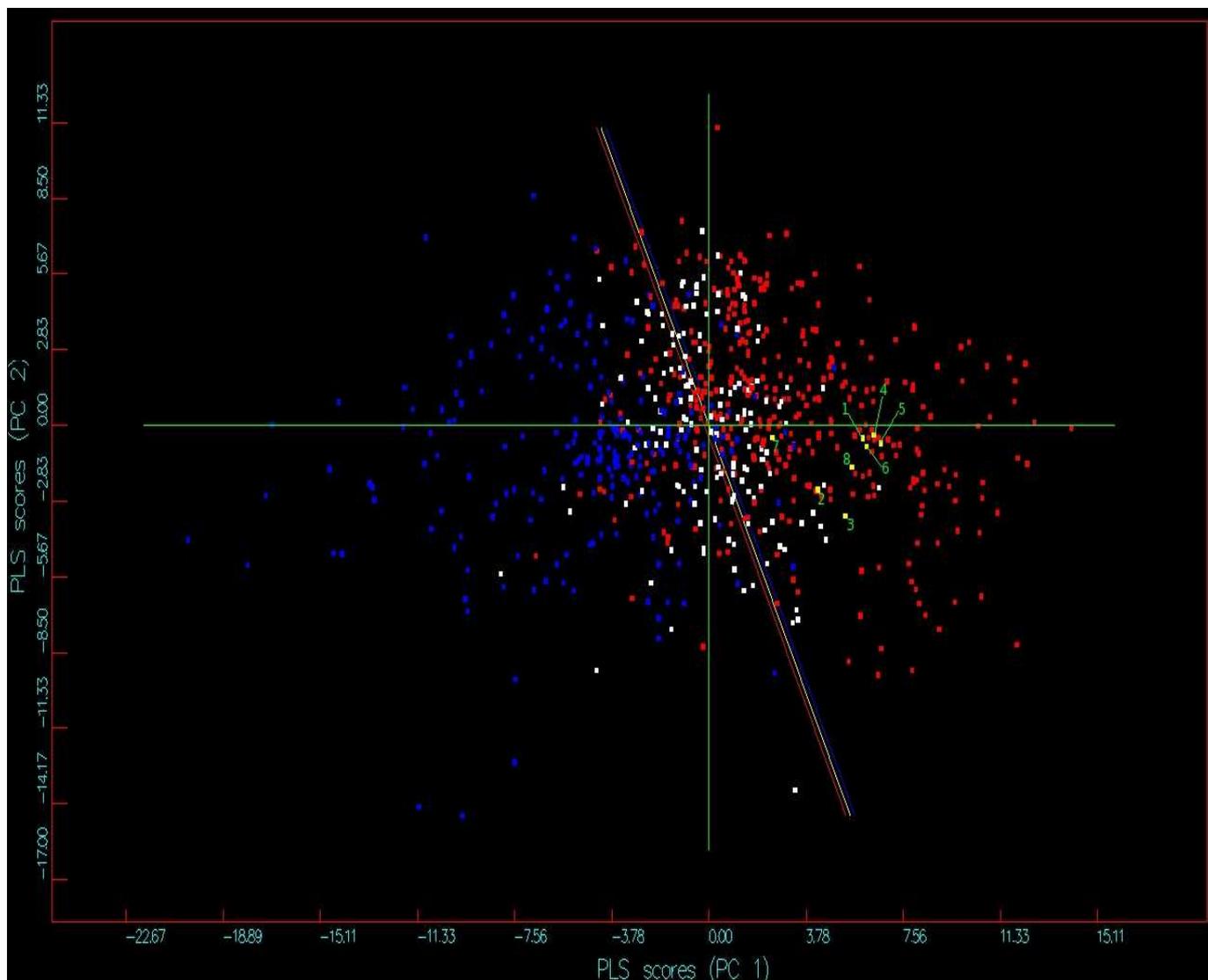


Figura 27. Gráfico PC1xPC2 da permeabilidade em células CaCO₂ dos inseticidas da classe dos neonicotinóides.

A Figura 28 representa o gráfico de volume de distribuição (VD) para os inseticidas da classe dos Neonicotinóides. Todos os compostos possuem baixo volume de distribuição no plasma sanguíneo e estão distribuídos pelo primeiro e segundo quadrante.

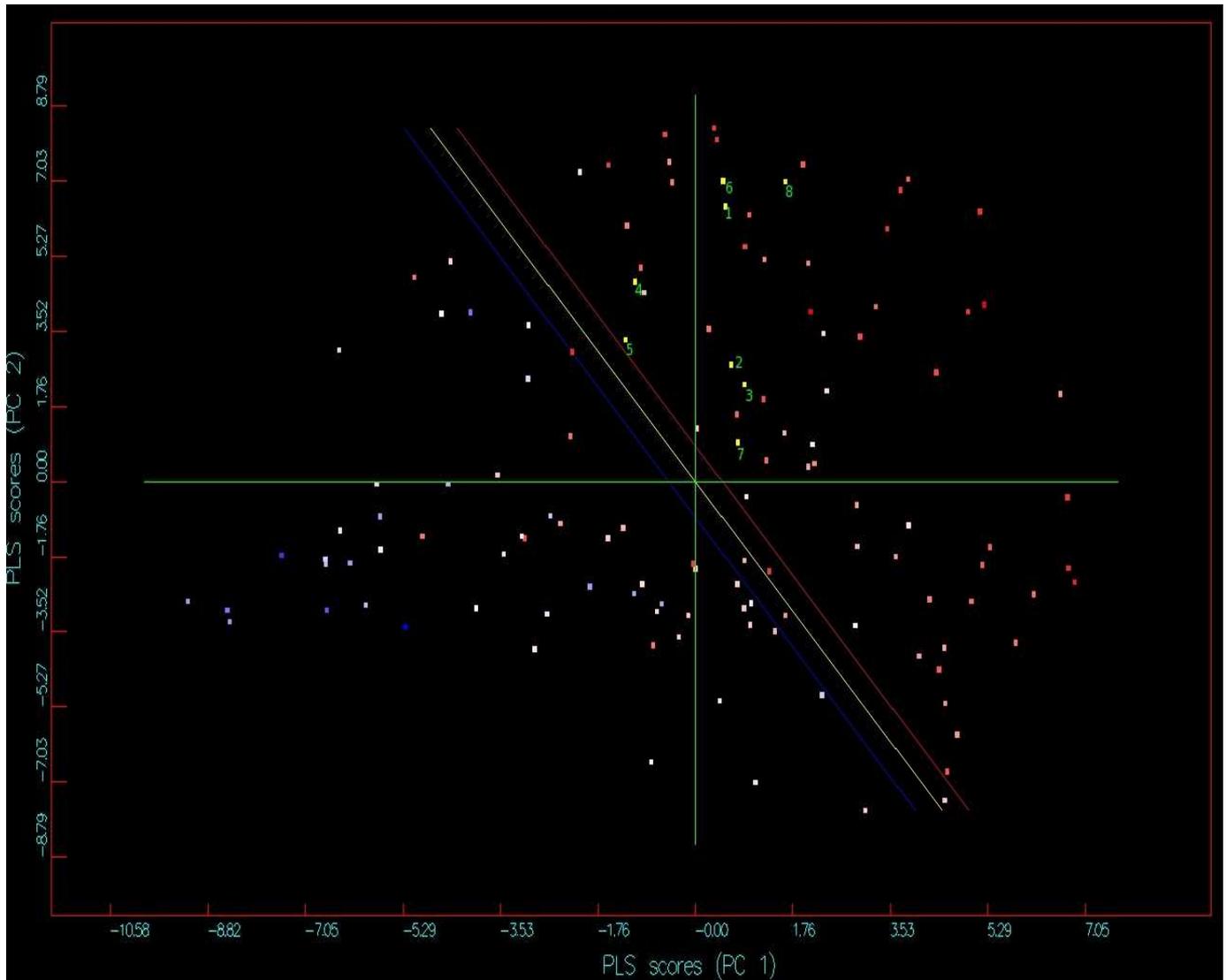


Figura 28. Gráfico PC1xPC2 demonstra o volume de distribuição dos inseticidas da classe dos neonicotinóides.

Os números grifados em verde na Tabela 5 são os ativos que apresentam um nível de similaridade $\geq 70\%$ por similaridade 3D Tanimoto Combo.

Tabela 5. Similaridade 3D Tanimoto Combo dos inseticidas da classe dos neonicotinóides.

Ativos	8	1	2	3	4	5	6	7
8	2,00							
1	1,28	2,00						
2	1,22	1,26	2,00					
3	1,13	1,19	1,78	2,00				
4	1,66	1,18	1,15	1,43	2,00			
5	1,04	1,09	1,15	1,01	1,17	2,00		
6	1,46	1,51	1,17	1,08	1,37	1,13	2,00	
7	1,04	1,02	0,97	1,01	1,07	0,70	0,88	2,00

Tabela 6. Legenda para as Figuras 23-28 e Tabela 5 dos inseticidas da classe dos neonicotinóides.

Número	Nome	Smile	Fórmula Molecular	DL ₅₀ Oral Aguda em ratos (mg/kg)
1	Acetamiprid	<chem>CN(CC1=CN=C(Cl)C=C1)C(C)=NC#N</chem>	C ₁₀ H ₁₁ ClN ₄	217,0
2	Clothianidin	<chem>CN=C(NCC1=CN=C(Cl)S1)N[N+](O-)=O</chem>	C ₆ H ₈ ClN ₅ O ₂ S	5000,0
3	Dinotefuran	<chem>CN=C(NCC1CCOC1)N[N+](O-)=O</chem>	C ₇ H ₁₄ N ₄ O ₃	2804,0
4	Imidacloprid	<chem>[N+](=O)NC1=NCCN1CC1=CN=C(C1)C=C1</chem>	C ₉ H ₁₀ ClN ₅ O ₂	450,0
5	Nitenpyram	<chem>CCN(CC1=CN=C(Cl)C=C1)C(NC)=C[N+](O-)=O</chem>	C ₁₁ H ₁₅ ClN ₄ O ₂	1680,0
6	Thiacloprid	<chem>ClC1=NC=C(CN2CCSC2=NC#N)C=C1</chem>	C ₁₀ H ₉ ClN ₄ S	621,0
7	Thiamethoxam	<chem>CN1COCN(CC2=CN=C(Cl)S2)C1=N[N+](O-)=O</chem>	C ₈ H ₁₀ ClN ₅ O ₃ S	1563,0
8	1H-Imidazol-2-amine-1,6-dichloro-3-pyridinylmethyl-4,5-dihydro-N-nitroso	<chem>ClC1=NC=C(CN2CCN=C2NN=O)C=C1</chem>	C ₉ H ₁₀ ClN ₅ O	2500,0

Analisando-se o grupo dos inseticidas da classe dos neonicotinóides observa-se que os compostos dessa classe apresentam os cinco endpoints bem definidos.

O endpoint de interação à proteína não é totalmente satisfatório uma vez que a

predição não é confiável para metade dos compostos testados, sendo assim esse modelo deve ser melhorado.

Utilizando-se os outros modelos pode-se dizer que esses agrotóxicos penetram na barreira hemato-encefálica, não são bloqueadoras do canal de hERG, são estáveis metabolicamente e possuem alta permeabilidade para célula CaCO2 e baixo volume de distribuição no plasma sanguíneo.

4.4. Carbamatos

A Figura 29 representa o gráfico de BBB para os inseticidas da classe dos Carbamatos. Todos os inseticidas desta classe penetram na Barreira Hemato-Encefálica e estão distribuídos entre o primeiro e o segundo quadrante. Pode-se verificar também que a maioria dos compostos se aglomeram no centro do primeiro quadrante.

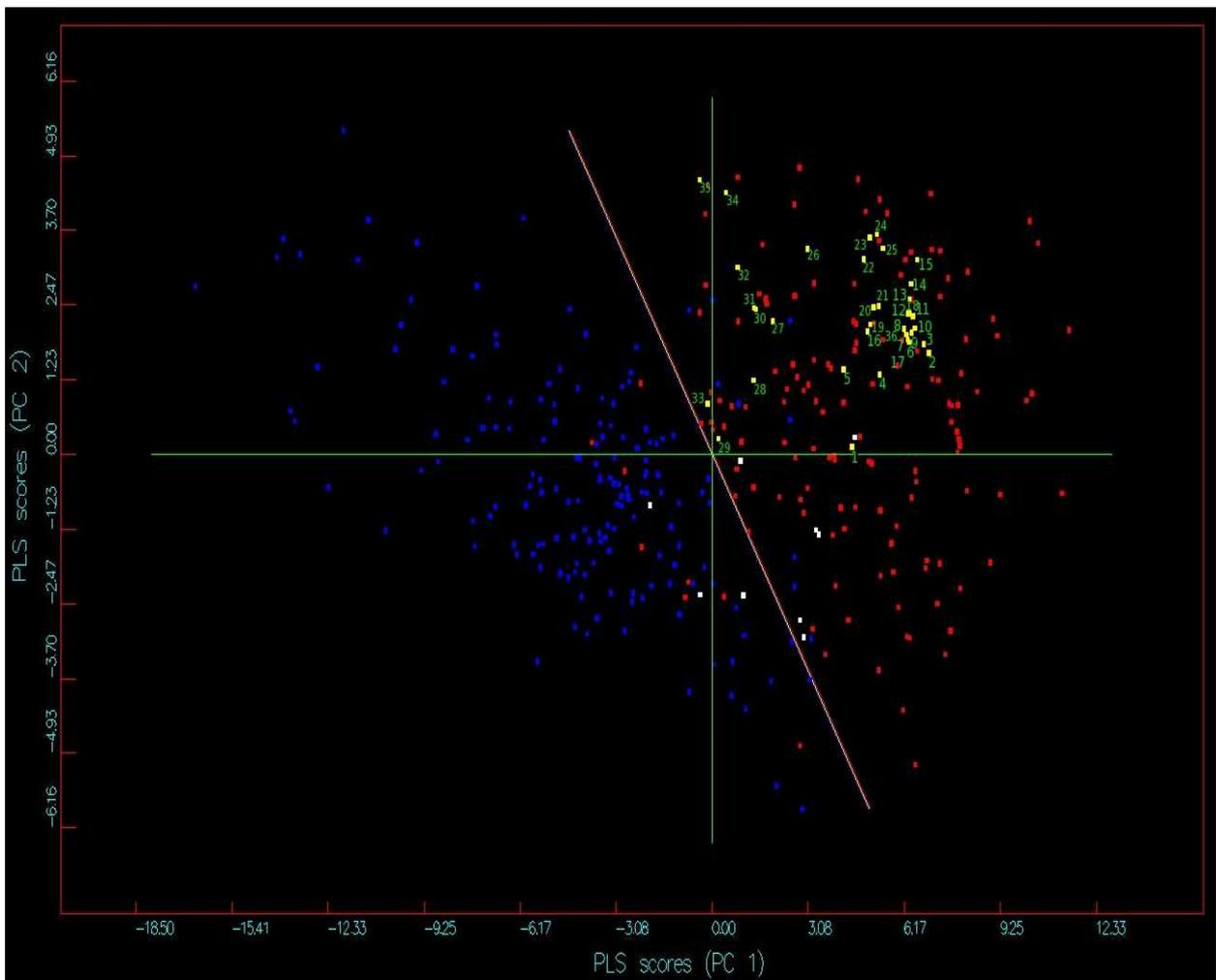


Figura 29. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos inseticidas da classe dos carbamatos.

A Figura 30 representa o gráfico de hERG para os inseticidas da classe dos Carbamatos. Todos os inseticidas testados desta classe não se mostraram bloqueadoras do canal de hERG e se encontram no primeiro quadrante. Pode-se verificar uma aglomeração dos compostos na parte superior do primeiro quadrante.

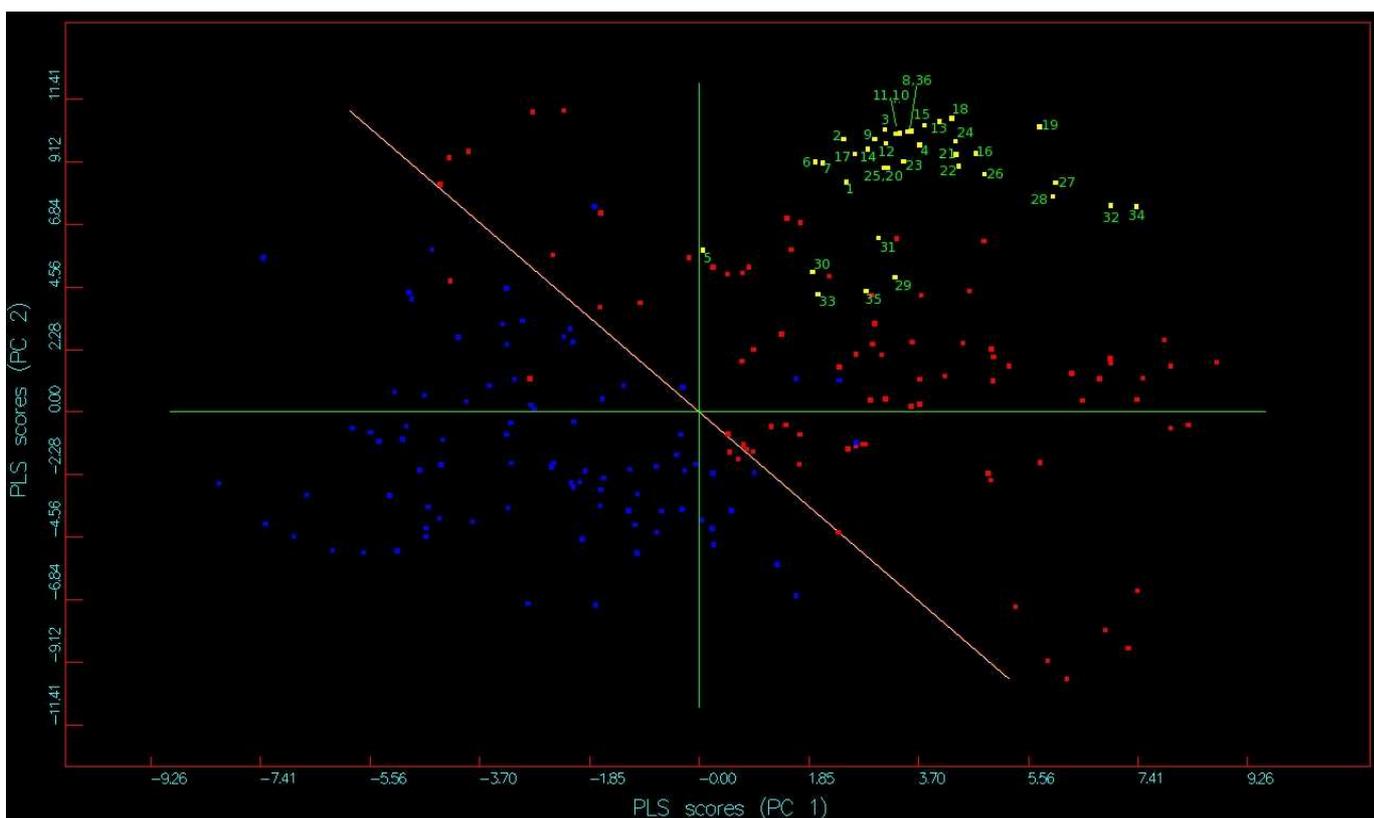


Figura 30. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos inseticidas da classe dos carbamatos.

A Figura 31 representa o gráfico de estabilidade metabólica para os inseticidas da classe dos Carbamatos. Todos os compostos testados desta classe são estáveis metabolicamente e se encontram no primeiro quadrante, sendo que a maioria dos compostos se encontra no centro do primeiro quadrante. Os carbamatos são metabolicamente estáveis e não são degradados facilmente no organismo, sendo excretados.

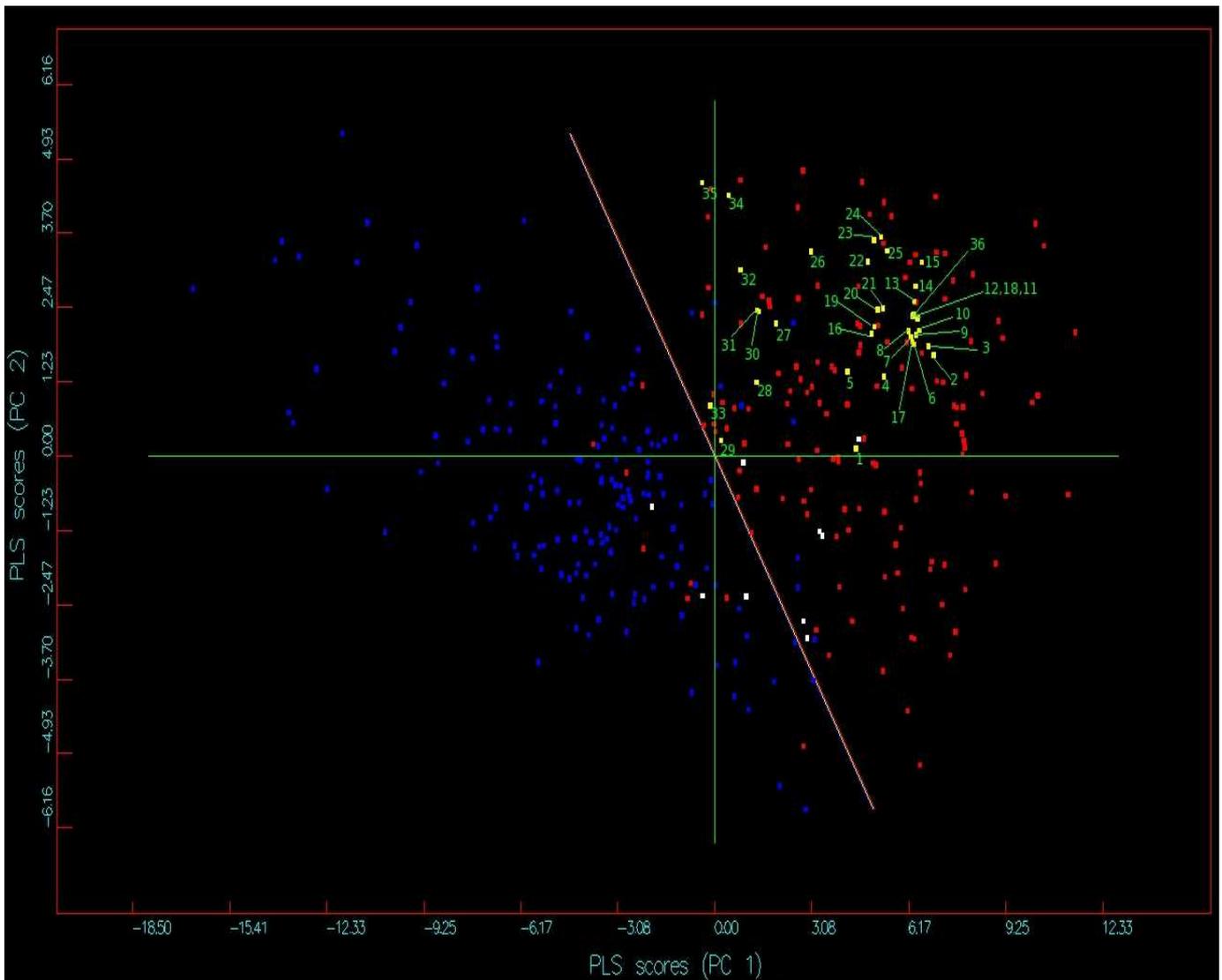


Figura 31. Gráfico de PC1xPC2 da estabilidade metabólica dos inseticidas da classe dos carbamatos.

A Figura 32 representa o gráfico de ligação à proteína para os inseticidas da classe dos Carbamatos. As substâncias carbaryl, Methylbutyl-phenyl-methyl carbamate, Hooker hrs-1422, butacarb, triazamate, benfuracarb, alanycarb, furathiocarb e carbosulfan se encontram na região do gráfico em que a predição é incerta, com isso não se pode afirmar se essas substâncias químicas apresentam alta ou baixa ligação protéica. Já os compostos restantes apresentam baixa ligação protéica e se encontram distribuídos entre o segundo e o terceiro quadrante.

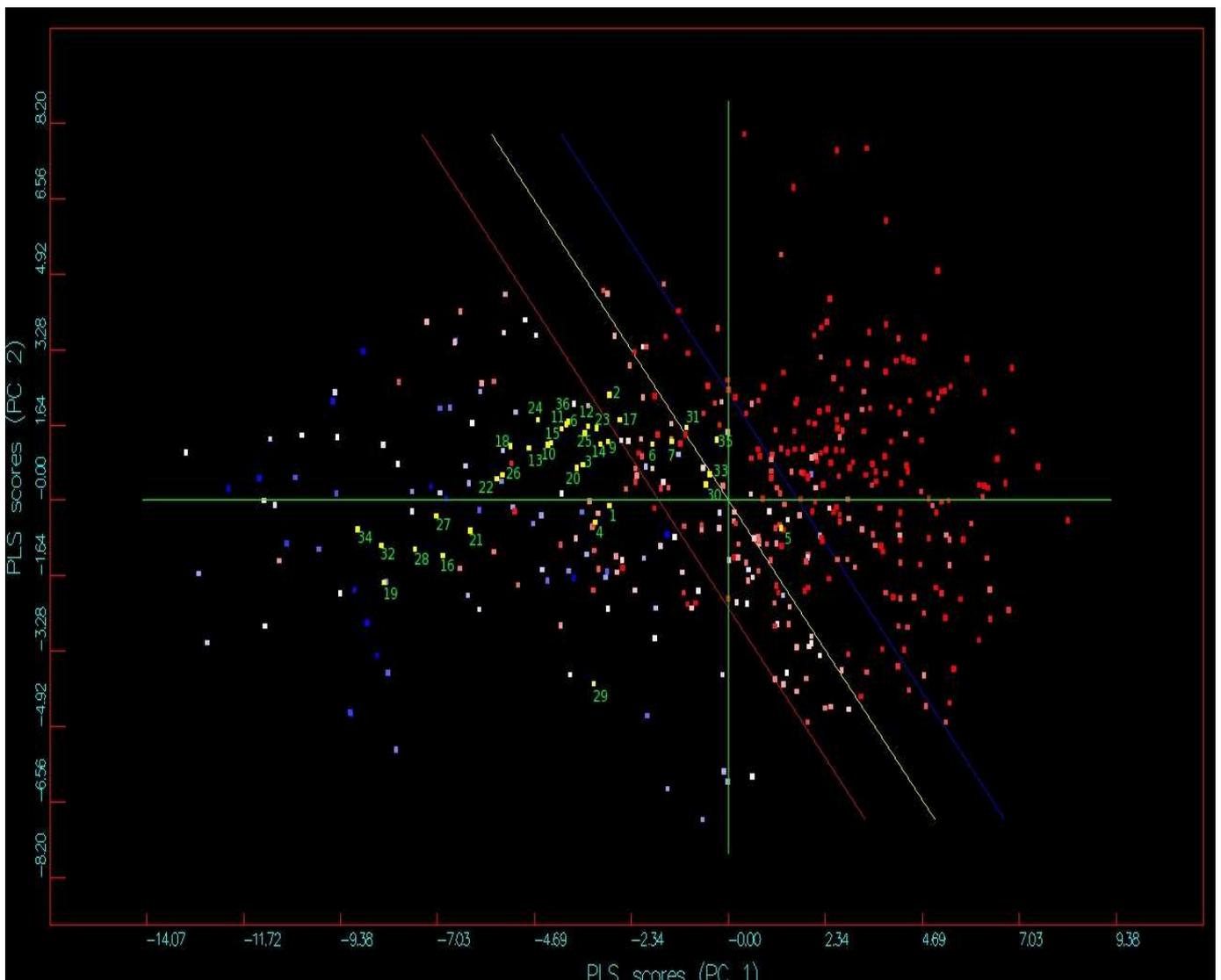


Figura 32. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos inseticidas da classe dos carbamatos.

A Figura 33 representa o gráfico de CaCO₂ para os inseticidas da classe dos Carbamatos. Todos os compostos apresentam alta permeabilidade para células CaCO₂ e estão distribuídas entre o primeiro e quarto quadrante, sendo que há uma concentração dos compostos no quarto quadrante.

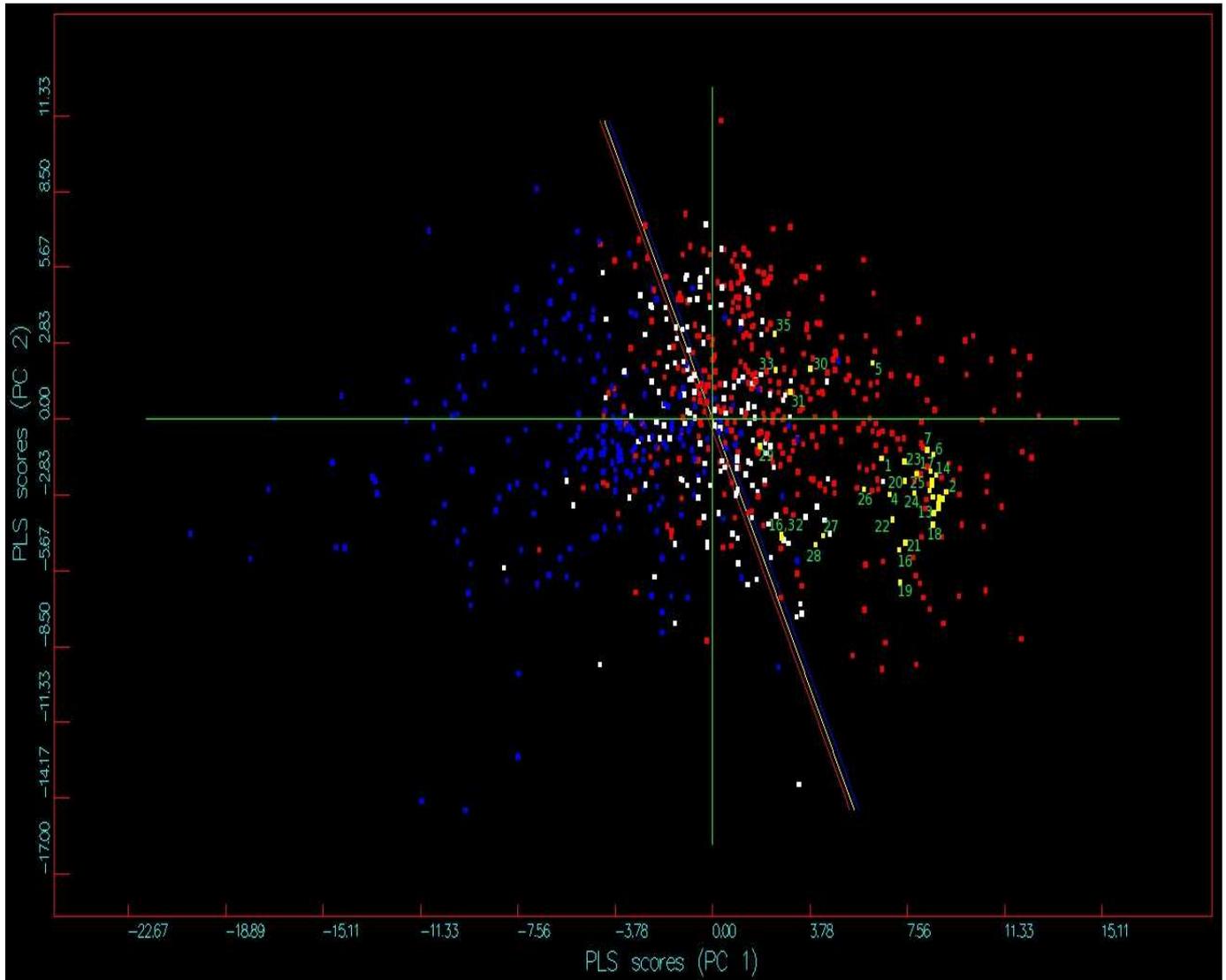


Figura 33. Gráfico PC1xPC2 da permeabilidade em células CaCO₂ dos inseticidas da classe dos carbamatos.

A Figura 34 representa o gráfico de volume de distribuição (VD) para os inseticidas da classe dos Carbamatos. Os compostos butacarb, alanylcarb e benfuracarb se encontram na região do gráfico em que a predição é incerta, com isso não se pode afirmar se essas substâncias químicas apresentam alto ou baixo volume de distribuição. Já o composto carbosulfan apresenta alto volume de distribuição e os outros compostos possuem baixo volume de distribuição no plasma sanguíneo. Os compostos analisados que possuem baixo volume de distribuição se encontram distribuídos no primeiro e segundo quadrante.

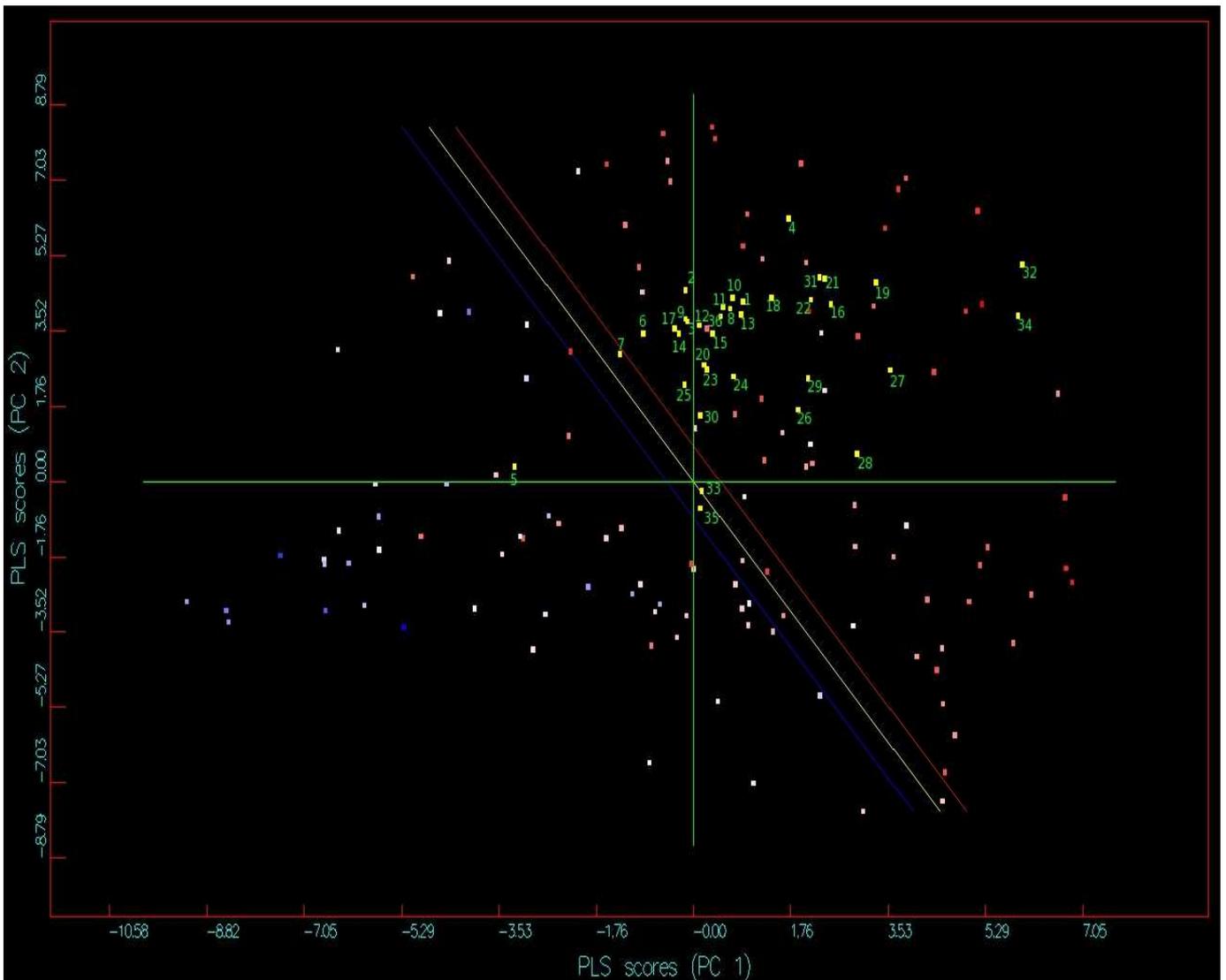


Figura 34. Gráfico PC1xPC2 demonstra o volume de distribuição dos inseticidas da classe dos carbamatos.

Tabela 7. Similaridade 3D Tanimoto Combo dos inseticidas da classe dos carbamatos.

Ativos	33	16	34	28	26	35	7	21	27	11	12	2	23	5	25	14	3	1	30	6	15	20	19	36	18	32	4	9	24	29	22	31	8	13	10			
33	2,0																																					
16	0,7	2,0																																				
34	0,6	1,6	2,0																																			
28	0,6	1,5	1,8	2,0																																		
26	0,6	1,0	0,9	0,8	2,0																																	
35	0,7	0,7	0,6	0,7	1,2	2,0																																
7	0,7	1,0	0,9	0,9	1,2	0,7	2,0																															
21	0,6	1,5	1,4	1,3	1,0	0,6	1,0	2,0																														
27	0,7	1,3	1,5	1,7	0,9	0,6	0,9	1,3	2,0																													
11	0,6	1,1	1,0	1,0	1,2	0,8	1,3	1,1	1,0	2,0																												
12	0,6	1,1	1,0	1,0	1,3	0,7	1,3	1,1	1,1	1,4	2,0																											
2	0,6	1,0	1,0	0,9	1,5	0,7	1,3	1,0	1,0	1,3	1,6	2,0																										
23	0,6	1,0	1,0	0,9	1,8	1,2	1,2	0,9	0,9	1,3	1,3	1,6	2,0																									
5	0,8	0,7	0,7	0,7	1,2	1,4	0,9	0,6	0,6	0,9	0,8	1,0	1,3	2,0																								
25	0,7	1,0	0,9	0,8	1,2	0,7	1,1	0,9	1,0	1,2	1,6	1,4	1,3	0,9	2,0																							
14	0,6	1,0	1,0	0,9	1,2	0,7	1,5	1,1	1,0	1,5	1,4	1,3	1,3	0,8	1,2	2,0																						
3	0,6	0,9	0,8	0,8	1,1	0,7	1,0	0,9	1,0	1,2	1,4	1,3	1,3	0,9	1,4	1,1	2,0																					
1	0,7	1,0	1,0	1,0	1,1	0,7	1,2	1,0	1,0	1,3	1,5	1,4	1,2	0,7	1,2	1,3	1,2	2,0																				
30	0,8	0,7	0,7	0,7	1,2	1,4	0,8	0,7	0,7	0,8	0,8	1,0	1,3	1,4	0,9	0,8	0,8	0,7	2,0																			
6	0,7	1,0	0,9	0,9	1,2	0,7	1,9	1,0	1,0	1,3	1,4	1,2	1,3	0,8	1,1	1,4	1,1	1,2	0,7	2,0																		
15	0,5	0,9	0,9	0,8	1,3	0,8	1,1	0,9	0,9	1,2	1,6	1,5	1,4	0,9	1,5	1,2	1,6	1,3	0,9	1,2	2,0																	
20	0,6	1,0	1,0	1,0	1,1	0,6	1,3	1,1	1,1	1,3	1,7	1,4	1,2	0,7	1,3	1,4	1,2	1,3	0,7	1,3	1,4	2,0																
19	0,7	1,5	1,4	1,2	1,0	0,6	0,9	1,6	1,3	1,0	1,1	1,0	0,8	0,6	1,0	1,0	1,0	0,9	0,5	1,0	1,0	1,0	2,0															
36	0,6	1,0	1,0	1,0	1,2	0,7	1,3	1,1	1,0	1,5	1,4	1,3	1,3	0,8	1,2	1,5	1,2	1,4	0,8	1,3	1,2	1,3	0,9	2,0														
18	0,6	1,1	1,1	1,0	1,1	0,7	1,3	1,1	1,1	1,5	1,9	1,6	1,2	0,8	1,5	1,5	1,4	1,5	0,8	1,4	1,6	1,6	1,1	1,4	2,0													
32	0,6	1,3	1,3	1,2	1,0	0,6	0,8	1,3	1,3	0,9	1,1	1,0	0,9	0,7	0,9	0,9	1,0	0,9	0,7	0,9	1,1	1,0	1,2	0,9	1,1	2,0												
4	0,6	0,9	0,9	0,9	1,2	0,8	1,0	0,8	0,9	1,1	1,2	1,1	1,1	0,8	1,0	1,1	1,0	1,1	0,8	1,1	1,1	1,1	0,9	1,0	1,2	1,0	2,0											
9	0,7	1,2	1,0	1,0	1,4	0,9	1,6	1,1	1,0	1,5	1,6	1,5	1,4	0,9	1,3	1,6	1,2	1,4	0,9	1,6	1,3	1,5	1,0	1,5	1,6	1,0	1,2	2,0										
24	0,6	1,0	0,8	0,8	1,4	1,0	1,0	0,9	0,9	1,1	1,4	1,3	1,5	0,9	1,4	1,1	1,4	1,1	1,0	1,0	1,5	1,2	0,9	1,1	1,4	1,0	1,0	1,0	1,2	2,0								
29	0,9	0,8	0,8	0,9	0,8	0,8	0,7	0,7	0,9	0,7	0,7	0,7	0,8	0,7	0,8	0,7	0,8	0,8	0,9	0,7	0,8	0,8	0,9	0,8	0,7	1,0	0,9	0,7	0,8	2,0								
22	0,6	1,2	1,1	1,0	1,0	0,6	0,9	1,2	1,2	1,1	1,0	1,0	0,9	0,6	1,0	1,0	1,0	0,8	0,6	1,0	1,0	1,1	1,3	1,0	1,0	1,4	1,1	1,0	1,0	0,6	2,0							
31	0,7	0,8	0,8	0,8	1,1	0,7	1,1	0,8	0,8	0,9	0,9	0,9	1,1	0,8	0,8	1,0	0,8	0,9	0,8	1,1	0,9	0,9	0,9	0,9	0,9	0,9	0,9	1,4	1,1	1,0	0,7	0,9	2,0					
8	0,6	1,1	1,0	0,9	1,2	0,8	1,4	1,0	1,1	1,4	1,9	1,6	1,3	0,8	1,6	1,5	1,4	1,5	0,8	1,5	1,6	1,6	1,0	1,4	1,9	1,0	1,2	1,7	1,4	0,6	0,9	0,8	2,0					
13	0,6	1,1	1,1	1,0	1,1	0,7	1,4	1,1	1,1	1,5	1,9	1,6	1,2	0,7	1,5	1,5	1,4	1,5	0,8	1,5	1,6	1,7	1,1	1,4	2,0	1,0	1,2	1,7	1,4	0,6	1,0	0,8	2,0	2,0				
10	0,6	1,1	1,1	1,0	1,1	0,7	1,3	1,1	1,1	1,5	2,0	1,6	1,2	0,8	1,5	1,5	1,4	1,5	0,8	1,4	1,6	1,7	1,1	1,4	2,0	1,1	1,2	1,6	1,4	0,6	1,0	0,7	1,9	1,9	2,0			

Os números grifados em verde na Tabela 7 são os ativos que apresentam um nível de similaridade $\geq 70\%$, nas respectivas comparações.

Tabela 8. Legenda para as Figuras 29-34 e Tabela 7 dos inseticidas da classe dos carbamatos.

Número	Nome	Smile	Fórmula Molecular	DL ₅₀ Oral Aguda em ratos (mg/kg)
1	Formetanate	<chem>CNC(=O)OC1=CC=CC(=C1)N=CN(C)C</chem>	C ₁₁ H ₁₅ N ₃ O ₂	14,8
2	Carbaryl	<chem>CNC(=O)OC1=CC=CC2=CC=CC=C12</chem>	C ₁₂ H ₁₁ NO ₂	264,0
3	Fenobucarb	<chem>CCC(C)C1=CC=CC=C1OC(=O)NC</chem>	C ₁₂ H ₁₇ NO ₂	524,0
4	Pirimicarb	<chem>CN(C)C(=O)OC1=NC(=NC(C)=C1C)N(C)C</chem>	C ₁₁ H ₁₈ N ₄ O ₂	142,0
5	Carbosulfan	<chem>CCCCN(CCCC)SN(C)C(=O)OC1=CC=CC2=C1OC(C)(C)C2</chem>	C ₂₀ H ₃₂ N ₂ O ₃ S	250,0
6	Hooker hrs-1422	<chem>CNC(=O)OC1=CC(=CC(=C1)C(C)C)C(C)C</chem>	C ₁₄ H ₂₁ NO ₂	200,0
7	Butacarb	<chem>CNC(=O)OC1=CC(=CC(=C1)C(C)(C)C)C(C)C</chem>	C ₁₆ H ₂₅ NO ₂	1800,0
8	Trimethacarb	<chem>CNC(=O)OC1=CC(C)=CC(C)=C1C</chem>	C ₁₁ H ₁₅ NO ₂	130,0
8	Trimethacarb	<chem>CNC(=O)OC1=CC(C)=C(C)C(C)=C1</chem>	C ₁₁ H ₁₅ NO ₂	178,0
9	Promecarb	<chem>CNC(=O)OC1=CC(C)=CC(=C1)C(C)C</chem>	C ₁₂ H ₁₇ NO ₂	35,0
10	Xylylcarb	<chem>CNC(=O)OC1=CC(C)=C(C)C=C1</chem>	C ₁₀ H ₁₃ NO ₂	290,0
11	Carbamic_acid_methyl_2,3_4_trimethylphenyl_ester	<chem>CNC(=O)OC1=C(C)C(C)=C(C)C=C1</chem>	C ₁₁ H ₁₅ NO ₂	318,0
12	Carbamic acid_methyl_m_sec_butyphenyl_ester	<chem>CCC(C)C1=CC(OC(=O)NC)=CC=C1</chem>	C ₁₂ H ₁₇ NO ₂	10,0
13	XMC	<chem>CNC(=O)Oc1cc(C)cc(C)c1</chem>	C ₁₀ H ₁₃ NO ₂	542,0
14	Fenethacarb	<chem>CCC1=CC(OC(=O)NC)=CC(C)C=C1</chem>	C ₁₂ H ₁₇ NO ₂	1600,0
15	Isoprocarb	<chem>CNC(=O)OC1=CC=CC=C1C(C)C</chem>	C ₁₁ H ₁₅ NO ₂	188,0
16	Aldicarb	<chem>CNC(=O)ON=CC(C)(C)SC</chem>	C ₇ H ₁₄ N ₂ O ₂ S	0,9
17	m_1_Methylbutylphenyl_methylcarbamate	<chem>CCCC(C)C1=CC(OC(=O)NC)=CC=C1</chem>	C ₁₃ H ₁₉ NO ₂	50,0
18	Metolcarb	<chem>CNC(=O)OC1=CC=CC(C)=C1</chem>	C ₉ H ₁₁ NO ₂	268,0
19	Methomyl	<chem>CNC(=O)ON=C(C)SC</chem>	C ₅ H ₁₀ N ₂ O ₂ S	34,0
20	Methiocarb	<chem>CNC(=O)OC1=CC(C)=C(SC)C1</chem>	C ₁₁ H ₁₅ NO ₂ S	33,0

C)=C1				
21	Butocarboxim	<chem>CNC(=O)ON=C(C)C(C)SC</chem>	<chem>C7H14N2O2S</chem>	153,0
22	Thiofanox	<chem>CNC(=O)ON=C(CSC)C(C)(C)C</chem>	<chem>C9H18N2O2S</chem>	8,5
23	Carbofuran	<chem>CNC(=O)OC1=CC=CC2=C1OC(C)(C)C2</chem>	<chem>C12H15NO3</chem>	8,0
24	Propoxur	<chem>CNC(=O)OC1=CC=CC=C1OC(C)C</chem>	<chem>C11H15NO3</chem>	50,0
25	Ethiofencarb	<chem>CCSCC1=CC=CC=C1OC(=O)NC</chem>	<chem>C11H15NO2S</chem>	200,0
26	Bendiocarb	<chem>CNC(=O)OC1=CC=CC2=C1OC(C)(C)O2</chem>	<chem>C11H13NO4</chem>	25,0
27	Butoxycarboxim	<chem>CNC(=O)ON=C(C)C(C)S(C)(=O)=O</chem>	<chem>C7H14N2O4S</chem>	458,0
28	Aldoxycarb	<chem>CNC(=O)ON=CC(C)(C)S(C)(=O)=O</chem>	<chem>C7H14N2O4S</chem>	20,0
29	Thiodicarb	<chem>CSC(C)=NOC(=O)N(C)SN(C)C(=O)ON=C(C)SC</chem>	<chem>C10H18N4O4S3</chem>	66,0
30	Furathiocarb	<chem>CCCCOC(=O)N(C)SN(C)C(=O)OC1=CC=CC2=C1OC(C)(C)C2</chem>	<chem>C18H26N2O5S</chem>	53,0
31	Triazamate	<chem>CCOC(=O)CSC1=NC(=NN1C(=O)N(C)C)C(C)(C)C</chem>	<chem>C13H22N4O3S</chem>	100,0
32	Oxamyl	<chem>CNC(=O)ON=C(SC)C(=O)N(C)C</chem>	<chem>C7H13N3O3S</chem>	3,1
33	Alanylcarb	<chem>CCOC(=O)CCN(CC1=CC=CC=C1)SN(C)C(=O)ON=C(C)SC</chem>	<chem>C17H25N3O4S2</chem>	440,0
34	Aldicarb sulfoxide	<chem>CNC(=O)ON=CC(C)(C)S(C)=O</chem>	<chem>C7H14N2O3S</chem>	0,5
35	Benfuracarb	<chem>CCOC(=O)CCN(SN(C)C(=O)OC1=CC=CC2=C1OC(C)(C)C2)C(C)C</chem>	<chem>C20H30N2O5S</chem>	222,6
36	Phenol, 3-(1-methylethyl)-, methylcarbamate	<chem>CNC(=O)OC1=CC=CC(=C1)C(C)C</chem>	<chem>C11H15NO2</chem>	16,0

Os carbamatos penetram na barreira hemato-encefálica, não são bloqueadoras do canal de hERG, são estáveis metabolicamente e possuem baixa ligação protéica, alta permeabilidade para célula CaCO2 e baixo volume de distribuição no plasma sanguíneo, exceto o carbosulfan, que apresenta alto volume de distribuição.

4.5. Benzoiluréias

A Figura 35 representa o gráfico de BBB para os inseticidas da classe das Benzoiluréias. O modelo utilizado mostra que o ativo diflubenzurom penetra na barreira hemato-encefálica e o triflumuron se encontra em uma região do gráfico em que a predição é incerta, com isso não se pode afirmar se essa substância penetra ou não na BBB. Já os outros inseticidas analisados não penetram na BBB. Verificou-se que a maioria dos compostos estão distribuídos entre o segundo e o terceiro quadrante.

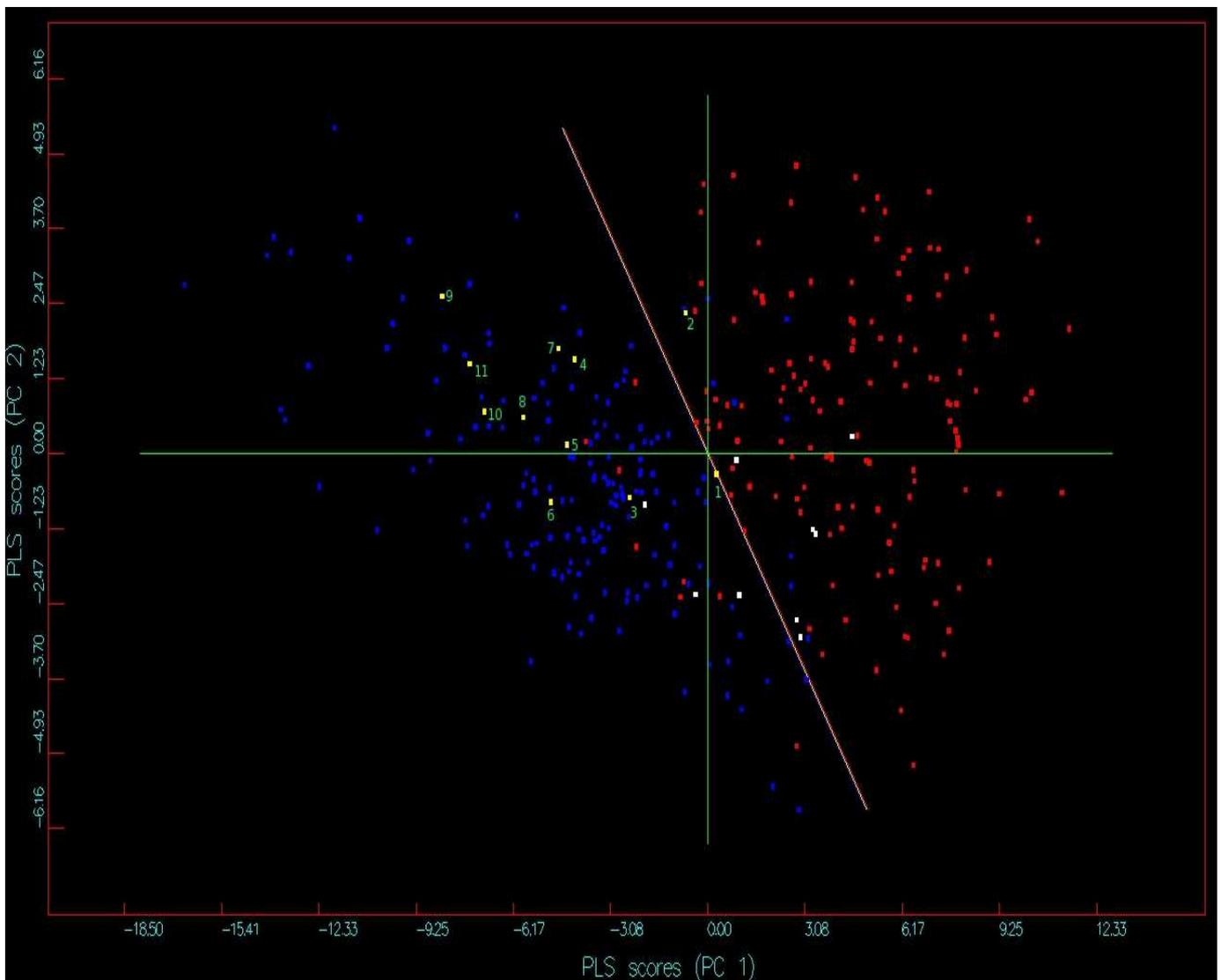


Figura 35. Gráfico de PC1xPC2 para Barreira Hemato-Encefálica (BBB) dos inseticidas da classe das benzoiluréias.

A Figura 36 representa o gráfico de hERG para os inseticidas da classe das benzoiluréias. Todos os inseticidas testados desta classe não são bloqueadores do canal de hERG e a maioria dos compostos se encontram no primeiro quadrante.

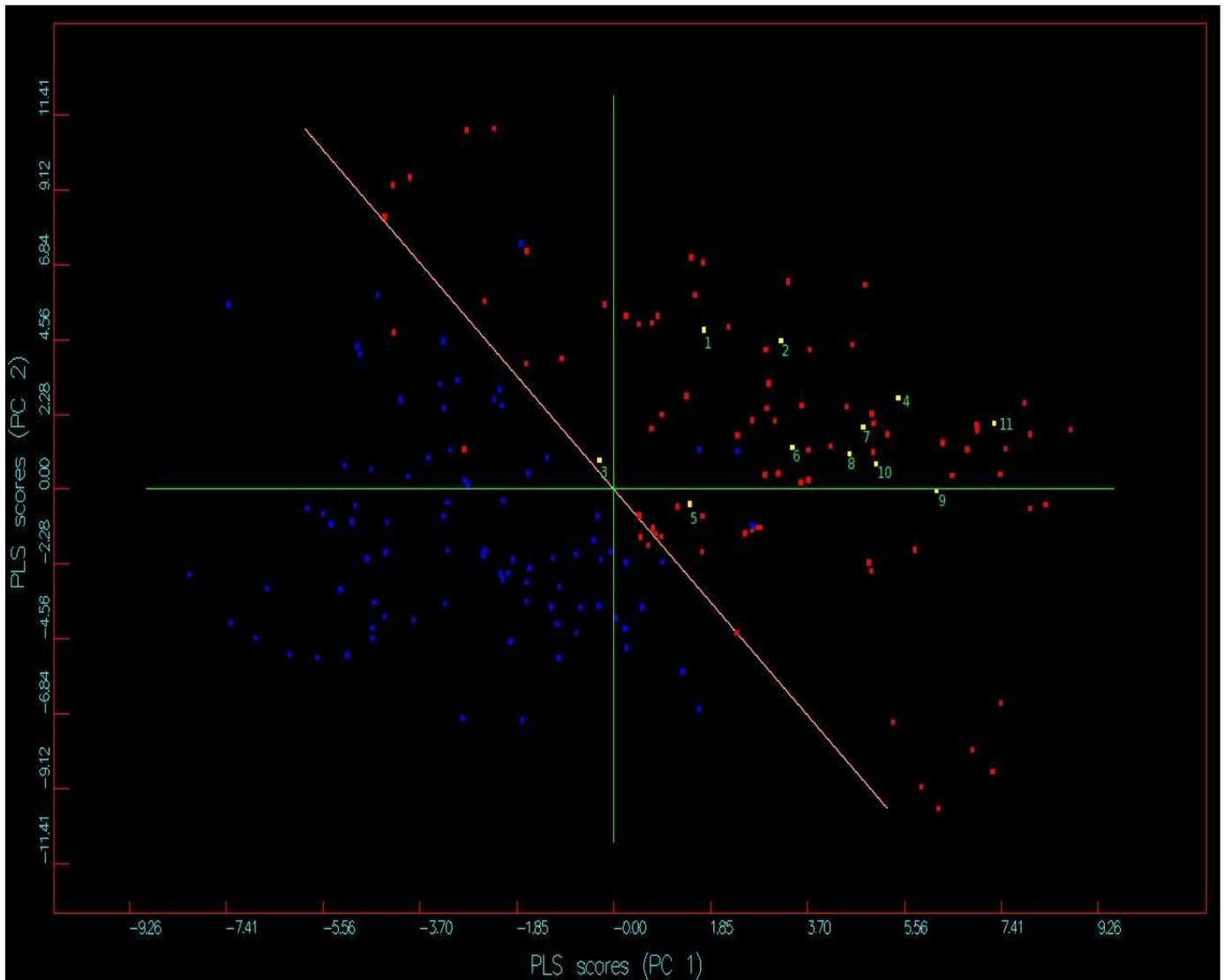


Figura 36. Gráfico PC1xPC2 do inibidor do canal de potássio (hERG) dos inseticidas da classe das benzoiluréias.

A Figura 37 representa o gráfico de estabilidade metabólica para os inseticidas da classe das Benzoiluréias. O composto chlorfluazuron se encontra em uma região do gráfico em que a predição é incerta, com isso não se pode afirmar se essa substância é estável ou não metabolicamente. Já o composto flucycloxuron se encontra no segundo quadrante e é instável metabolicamente. Os outros compostos testados desta classe foram estáveis metabolicamente e se encontram no primeiro quadrante, sendo que a maioria dos compostos se encontra no centro do primeiro quadrante. As benzoiluréias, com exceção do flucycloxuron, são metabolicamente estáveis e não são degradados facilmente no organismo, sendo excretados sem que metabólitos tóxicos sejam formados.

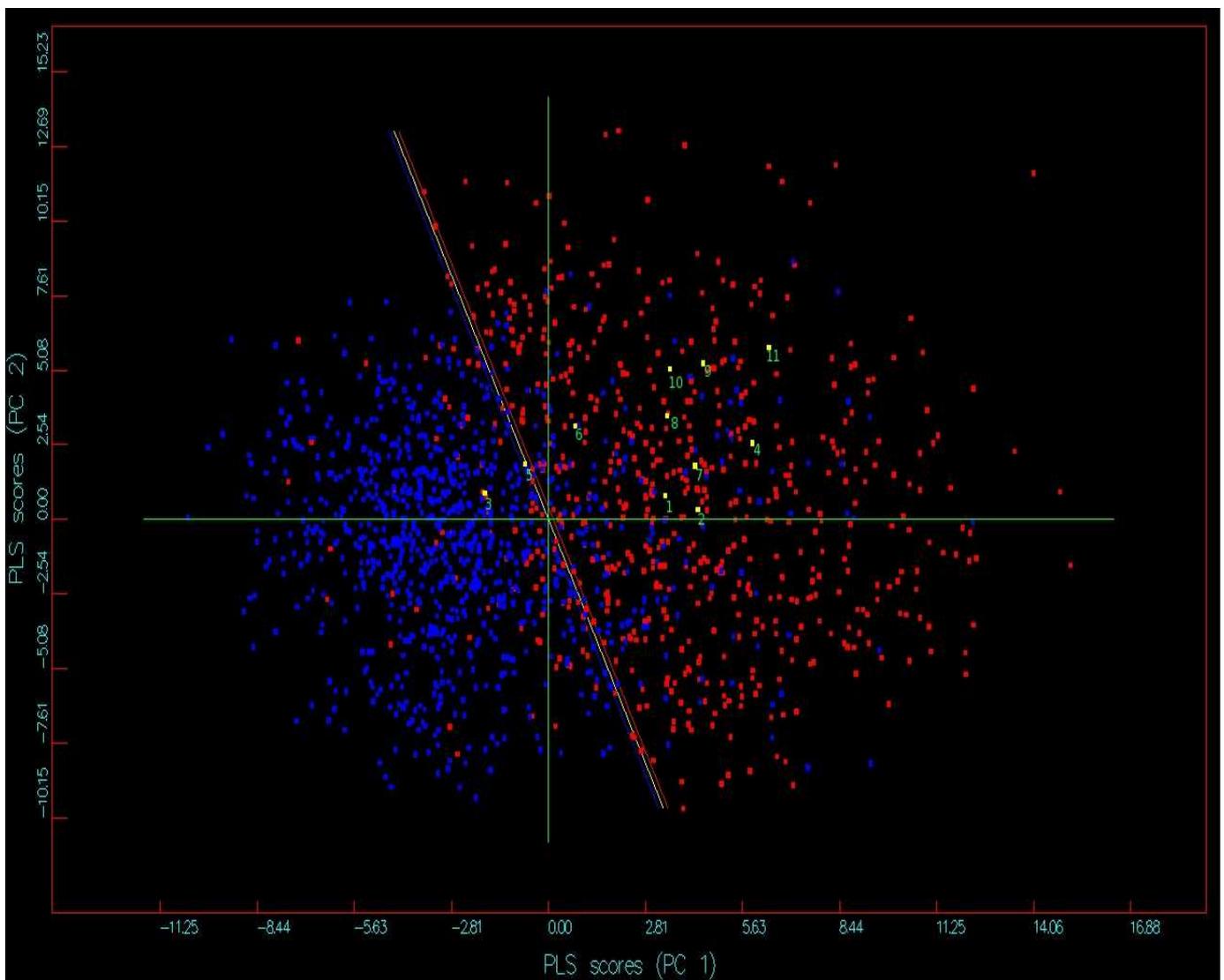


Figura 37. Gráfico de PC1xPC2 da estabilidade metabólica dos inseticidas da classe das benzoiluréias.

A Figura 38 representa o gráfico de ligação à proteína para os inseticidas da classe das Benzoiluréias. As substâncias diflubenzuron, flufenoxuron e chlorfluazuron se encontram na região do gráfico em que a predição é incerta, com isso não se pode afirmar se essas substâncias químicas apresentam alta ou baixa ligação protéica. Os compostos triflumuron e flucycloxuron apresentam alta ligação protéica. Já os compostos restantes se encontram, em sua maioria, no segundo quadrante e apresentam baixa ligação protéica.

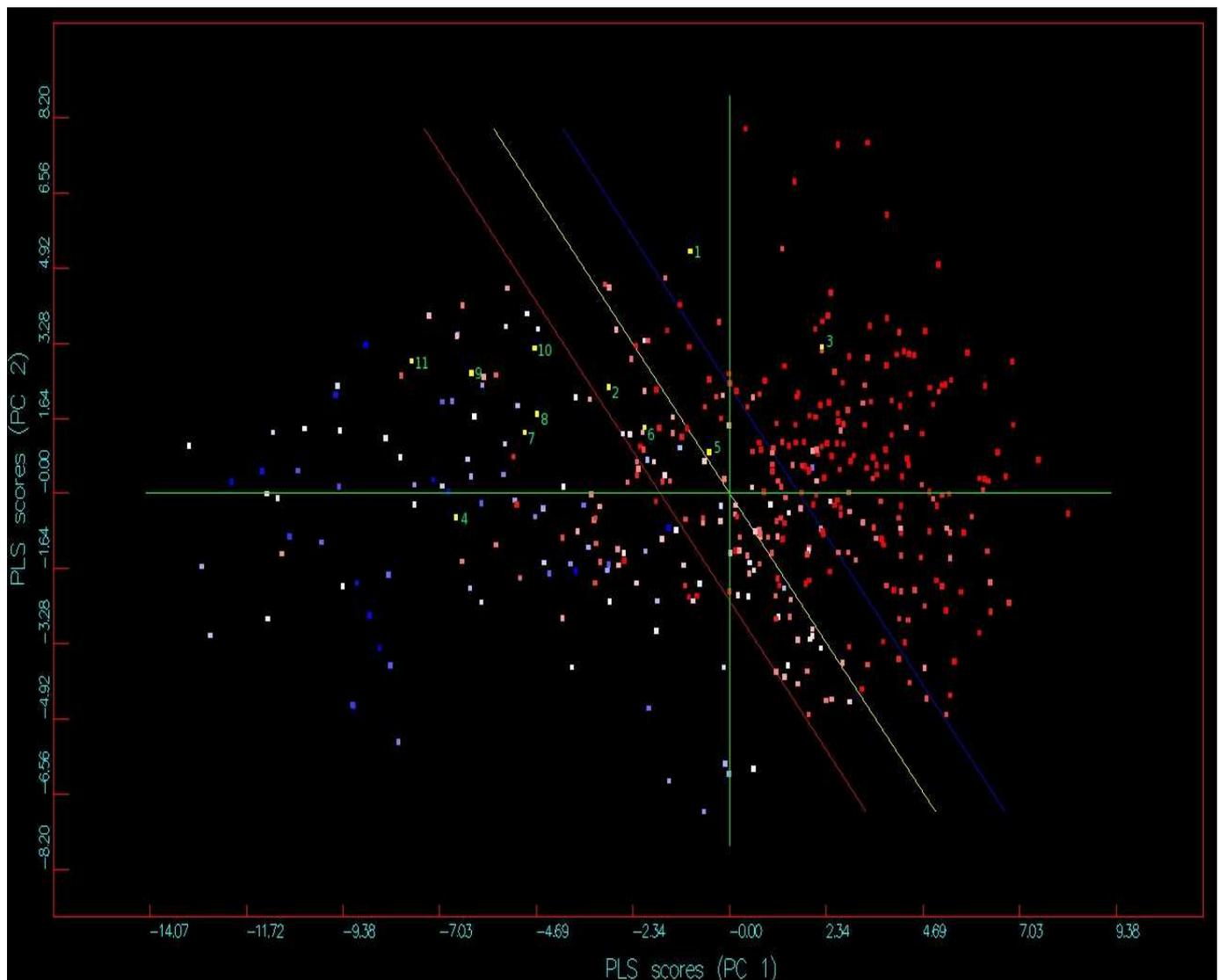


Figura 38. Gráfico PC1xPC2 da ligação a proteína Albumina Sérica de Humanos dos inseticidas da classe das benzoiluréias.

A Figura 39 representa o gráfico de CaCO₂ para os inseticidas das Benzoiluréias. Os compostos triflumuron, diflubenzuron e flucycloxuron apresentam alta permeabilidade para células CaCO₂. Os compostos que apresentam baixa permeabilidade às células CaCO₂ se encontram no terceiro quadrante (-,-).

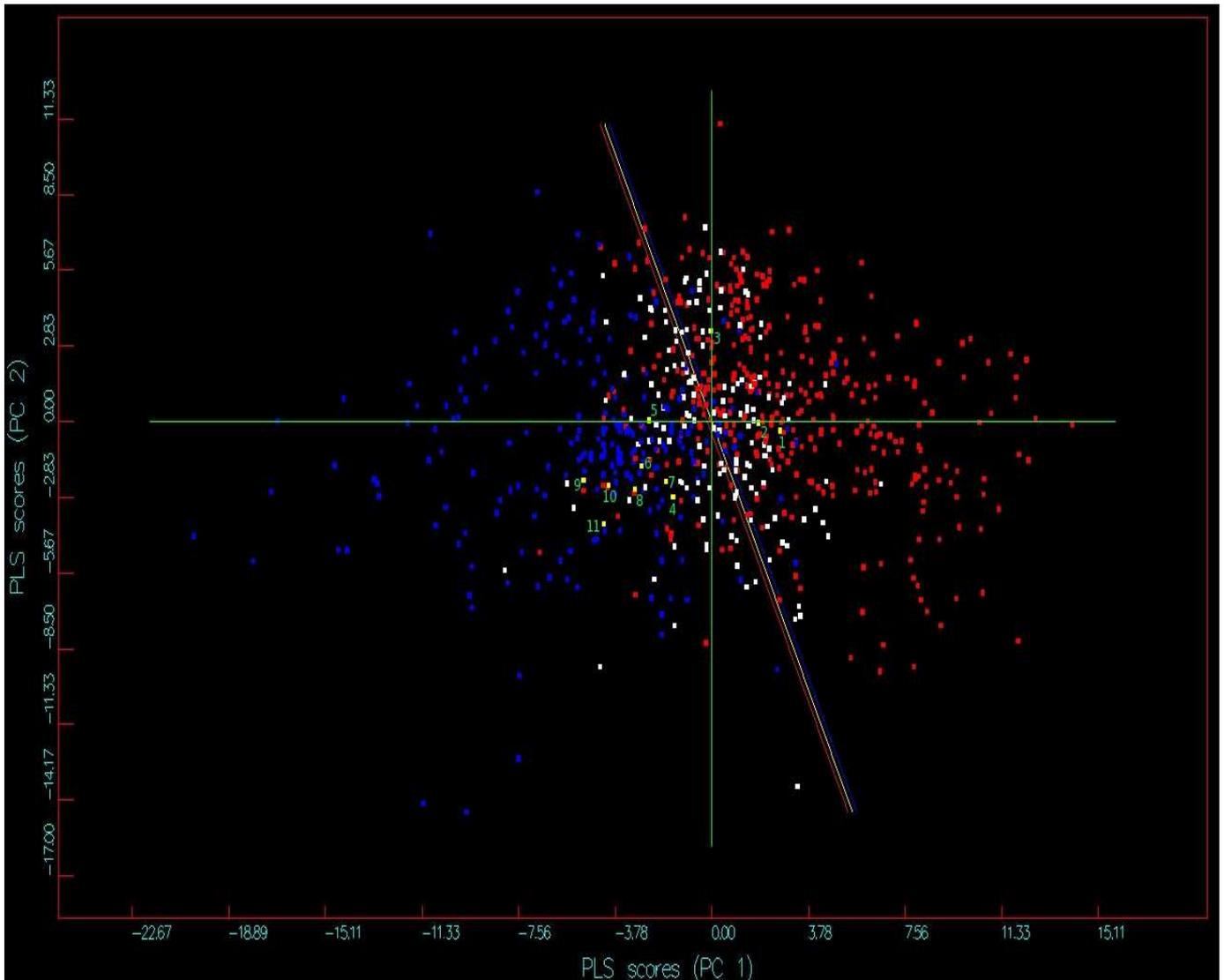


Figura 39. Gráfico PC1xPC2 da permeabilidade em células CaCO₂ dos inseticidas da classe das benzoiluréias.

A Figura 40 representa o gráfico de volume de distribuição (VD) para os inseticidas da classe das Benzoiluréias. Os compostos flucycloxuron e hexaflumuron se encontram na região do gráfico em que a predição é incerta, com isso não se pode afirmar se essas substâncias químicas apresentam alto ou baixo volume de distribuição. O composto chlorfluazuron apresenta alto volume de distribuição e os outros compostos possuem baixo volume de distribuição no plasma sanguíneo. Os compostos analisados que possuem baixo volume de distribuição se encontram distribuídos no primeiro e o quarto quadrante.

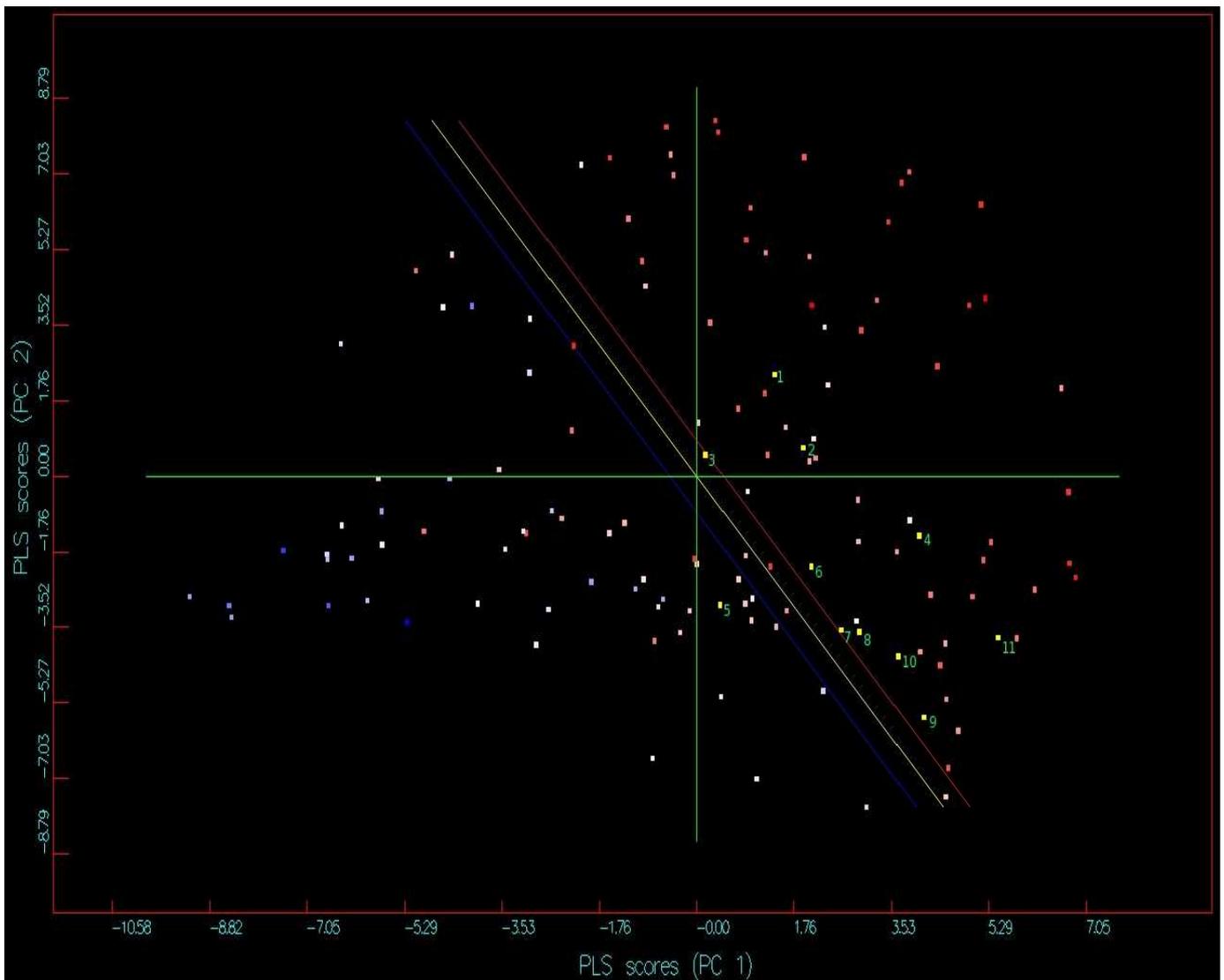


Figura 40. Gráfico PC1xPC2 demonstra o volume de distribuição dos inseticidas da classe das benzoiluréias.

Os números grifados em verde na Tabela 9 são os ativos que apresentam um nível de similaridade $\geq 70\%$ por similaridade 3D Tanimoto Combo.

Tabela 9. Similaridade 3D Tanimoto Combo dos inseticidas da classe das benzoiluréias.

Ativos	11	5	2	3	6	7	8	10	9	4	1
11	2,00										
5	0,98	2,00									
2	1,75	0,65	2,00								
3	1,29	1,27	1,39	2,00							
6	1,28	1,78	1,42	1,34	2,00						
7	1,58	1,56	0,99	0,77	1,60	2,00					
8	1,53	1,48	0,79	1,40	1,63	1,89	2,00				
10	1,39	1,44	1,30	1,42	1,57	1,76	1,78	2,00			
9	1,54	1,53	0,78	1,42	1,60	1,93	1,90	1,78	2,00		
4	1,88	0,76	1,94	0,82	0,82	1,66	0,90	0,83	0,91	2,00	
1	1,48	1,44	1,65	0,90	1,46	1,83	1,78	1,47	1,76	1,62	2,00

Tabela 10. Legenda para as Figuras 35-40 e Tabela 9 dos inseticidas da classe das benzoiluréias.

Número	Nome	Smile	Fórmula Molecular	DL ₅₀ oral Aguda em ratos (mg/kg)
1	Triflumuron	<chem>FC(F)(F)OC1=CC=C(NC(=O)NC(=O)C2=CC=CC=C2Cl)C=C1</chem>	C ₁₅ H ₁₀ ClF ₃ N ₂ O ₃	5000,0
2	Diflubenzuron	<chem>FC1=CC=CC(F)=C1C(=O)NC(=O)NC1=CC=C(Cl)C=C1</chem>	C ₁₄ H ₉ ClF ₂ N ₂ O ₂	4640,0
3	Flucycloxuron	<chem>FC1=CC=CC(F)=C1C(=O)NC(=O)NC1=CC=C(CON=C(C2CC2)C2=CC=C(Cl)C=C2)C=C1</chem>	C ₂₅ H ₂₀ ClF ₂ N ₃ O ₃	5000,0
4	Teflubenzuron	<chem>FC1=CC=CC(F)=C1C(=O)NC(=O)NC1=CC(Cl)=C(F)C(Cl)=C1F</chem>	C ₁₄ H ₆ Cl ₂ F ₄ N ₂ O ₂	5000,0
5	Chlorfluazuron	<chem>FC1=CC=CC(F)=C1C(=O)NC(=O)NC1=CC(Cl)=C(OC2=C(Cl)C=C(C=N2)C(F)(F)F)C(Cl)=C1</chem>	C ₂₀ H ₉ Cl ₃ F ₅ N ₃ O ₃	8500,0
6	Flufenoxuron	<chem>FC1=CC=CC(F)=C1C(=O)NC(=O)NC1=C(F)C=C(OC2=C(Cl)C=C(C=C2)C(F)(F)F)C=C1</chem>	C ₂₁ H ₁₁ ClF ₆ N ₂ O ₃	3000,0
7	Hexaflumuron	<chem>FC(F)C(F)(F)OC1=C(Cl)C=C(NC(=O)NC(=O)C2=C(F)C=CC=C2F)C=C1Cl</chem>	C ₁₆ H ₈ Cl ₂ F ₆ N ₂ O ₃	5000,0
8	Lufenuron	<chem>FC(C(F)(F)F)C(F)(F)OC1=C(Cl)C=C(NC(=O)NC(=O)C2=C(F)C=CC=C2F)C(Cl)=C1</chem>	C ₁₇ H ₈ Cl ₂ F ₈ N ₂ O ₃	2000,0
9	Noviflumuron	<chem>FC(C(F)(F)F)C(F)(F)OC1=C(Cl)C=C(NC(=O)NC(=O)C2=C(F)C=CC=C2F)C(F)=C1Cl</chem>	C ₁₇ H ₇ Cl ₂ F ₉ N ₂ O ₃	5000,0
10	Novaluron	<chem>FC(OC(F)(F)F)C(F)(F)OC1=C(Cl)C=C(NC(=O)NC(=O)C2=C(F)C=CC=C2F)C=C1</chem>	C ₁₇ H ₉ ClF ₈ N ₂ O ₄	5000,0

11	Bistrifluron	<chem>FC1=CC=CC(F)=C1C(=O)NC(=O)NC1=CC(=CC(=C1Cl)C(F)(F)F)C(F)(F)F</chem>	$C_{16}H_7ClF_8N_2O_2$	5000,0
----	--------------	---	------------------------	--------

Analisando-se o grupo das benzoiluréias observa-se que os compostos dessa classe não penetram na barreira hemato-encefálica, não são bloqueadoras do canal de hERG, são estáveis metabolicamente, com exceção do flucycloxuron, a maioria apresenta baixa ligação protéica, exceto os triflumuron e flucycloxuron, baixa permeabilidade para célula CaCO₂ para a maioria dos compostos testados, exceto o triflumuron, diflubenzuron e flucycloxuron e todos os compostos apresentam baixo volume de distribuição no plasma sanguíneo, exceto o chlorfluazuron.

4.5.1. Modelo PLS para benzoiluréias

O modelo foi desenvolvido utilizando três sondas químicas: DRY, H₂O e O descritores químicos do programa VOLSURF e os dados de DL₅₀ oral aguda, que foram transformados em PDL₅₀ (potência).

Os valores de DL₅₀ oral aguda em ratos maiores que 2000 mg/kg foram harmonizados para este valor, uma vez que o protocolo de estudo do EPA utiliza 2000 mg/kg como a maior dose testada. A Figura 41 mostra o gráfico de valores preditos e os experimentais obtidos.

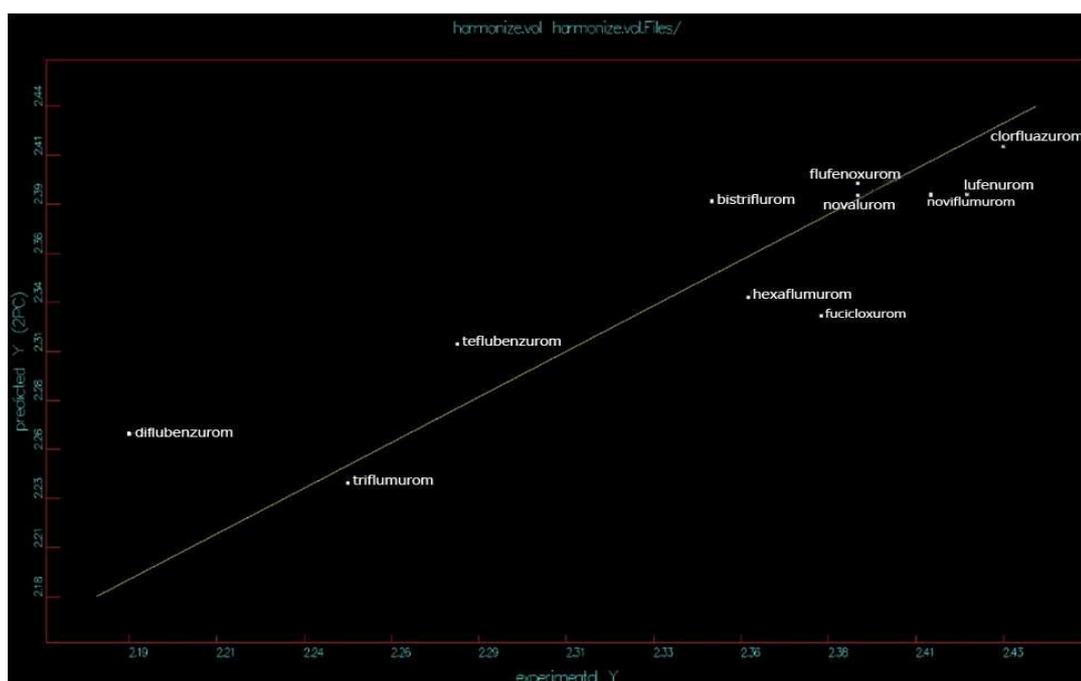


Figura 41. Gráfico de valores preditos x experimentais para DL₅₀ oral aguda em ratos para inseticidas da classe das benzoiluréias.

O modelo de predição apresentou um coeficiente angular (r^2) de 0,99 utilizando-se três componentes principais (PC). A validação live one out (LOO) apresentou um capacidade de predição (q^2) = 0,78. O modelo apresenta um resultado satisfatório, apesar de ser um modelo pouco robusto, uma vez que a faixa de predição é pequena levando em consideração que a diferença logarítmica entre o maior e o menor valor predito de PDL₅₀ é de apenas 0,11. No entanto, o modelo é satisfatório para predição de compostos dessa classe, visto que os DL50 orais são altos para todos os compostos e, quando harmonizados, apresentam baixa variação nos valores de PDL₅₀.

A Tabela 11 apresenta os resultados para os inseticidas da Figura 41, com dados experimentais, os calculados e os preditos.

Tabela 11. Dados para os inseticida da classe das benzoiluréias.

Ativos	pDL50 Calculado	pDL50 Experimental	pDL50Predito
diflubenzurom	1,83	1,82	1,87
triflumurom	1,84	1,85	1,86
teflubenzurom	1,89	1,88	1,93
bistrifluron	1,96	1,95	1,98
hexaflumuron	1,95	1,96	1,94
flucycloخورon	1,99	1,98	1,94
novaluron	1,99	1,99	1,99
noviflumuron	2,01	2,02	1,99
flufenoxuron	2,02	2,02	2,00
chlorfluazuron	2,04	2,03	2,03

4.6. Método SIMCA no reconhecimento de herbicidas e inseticidas em um banco de dados compostos por 617 moléculas

O método SIMCA se mostrou eficaz para separar inseticidas de herbicidas, em um banco de dados com 617 moléculas de agrotóxicos, contendo 325 herbicidas e 292 inseticidas, com uma ampla diversidade química. Essas moléculas foram ou são utilizadas como agrotóxicos na agricultura brasileira.

Utilizando-se o CONCORD para a geração das estruturas 3D, as sondas e descritores do programa VOLSURF para o cálculo dos descritores químicos e o programa SYBYL para rodar o método SIMCA, não se obteve um resultado satisfatório, sendo que o modelo gerado não foi robusto o suficiente para separar herbicidas de inseticidas.

Na etapa seguinte, por meio do método de seleção de variáveis de FISHER, foram selecionados 22 descritores de maior peso. Os resultados podem ser visualizados no quadro abaixo:

	Herbicida	Inseticida	Sem resposta
Herbicida	282	39	4
Inseticida	161	128	3
Total predito corretamente	410 compostos		
Total predito erroneamente	200 compostos		
Total sem predição	7 compostos		
Total de moléculas preditas corretamente	66,45 %		
% Herbicidas preditos corretamente	86,80%		
% Inseticidas preditos corretamente	43,80%		

O resultado foi um modelo que, de um total de 617 moléculas, conseguiu obter uma predição correta para 443 compostos, sendo um percentual de 66,45 %. O modelo, de forma geral, é considerado satisfatório devido à grande diversidade química do banco de dados inicial. A predição para os inseticidas não foi satisfatória, uma vez que somente 128 compostos de 292 possíveis foram preditos corretamente, um percentual de acerto de 43,80 %. Já o modelo se mostra robusto para a predição de herbicidas, uma vez que entre 325 moléculas foram preditas corretamente 282, um percentual de 86,80 %.

4.7. Aplicação de Redes Neurais para o banco de dados de inseticidas e herbicidas

Trezentos e vinte e cinco herbicidas e duzentos e noventa e dois inseticidas foram usados. A rede foi treinada de forma aleatorizada com conjuntos de validações e testes. Todos os cálculos foram realizados usando a versão 7.8 do MatLab. O algoritmo usado foi o de *trainrp Resilient Backpropagation*, com função de desempenho avaliada pelo erro quadrático médio com divisão aleatória dos dados e 25 neurônios. Os treinamentos encerraram-se após 5 e 35 interações quando chegaram aos erros quadráticos médios de 0,067 e 0,064 para os herbicidas e inseticidas, respectivamente. A Figura 42 mostra o acompanhamento do erro em cada época do treinamento.

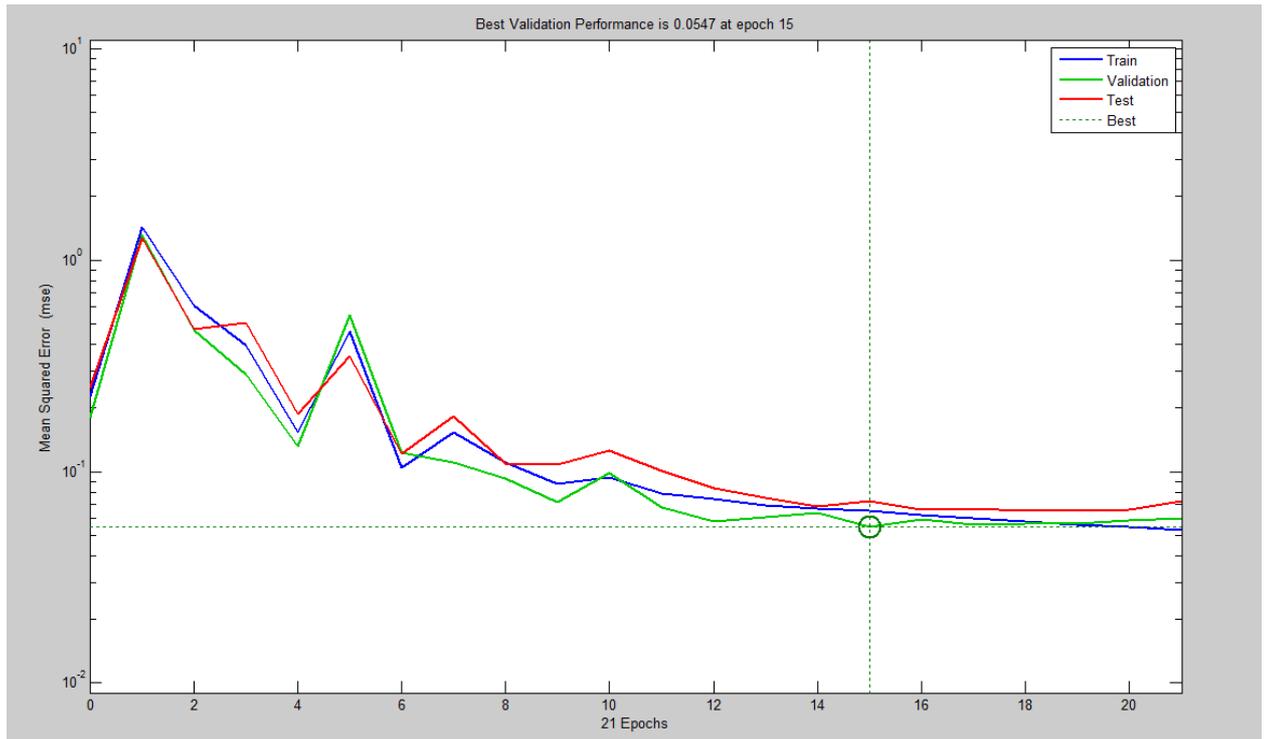


Figura 42. Erro quadrático médio da série de treinamento, teste e validação de 325 herbicidas.

O menor erro ocorreu durante a 15^a. época do treinamento, obtendo-se o valor de 0,055. Pode-se observar a inexistência de um *overfitting* sobre o conjunto de treinamento, visto que o erro sobre o conjunto de validação e teste não é essencialmente reduzido após a 15^a. época.

A Figura 43 mostra o ajuste linear para a série de treinamento e para a série completa (treinamento + validação + teste). Pode-se observar que a capacidade de predição de 61 % pode ser considerada boa tendo-se em vista a elevada complexidade do espaço químico descrito pelos herbicidas.

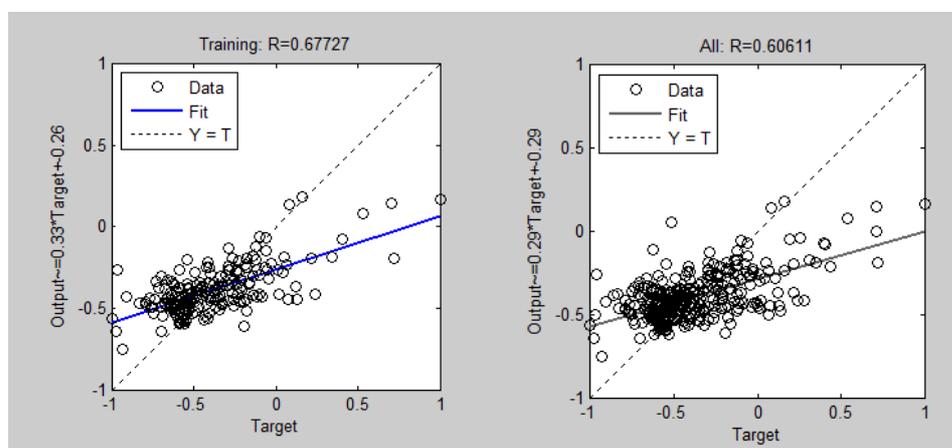


Figura 43. Ajuste linear para a série completa de herbicidas (treinamento + validação + teste).

Para a série de inseticidas, os resultados são mostrados nas Figuras 44 e 45.

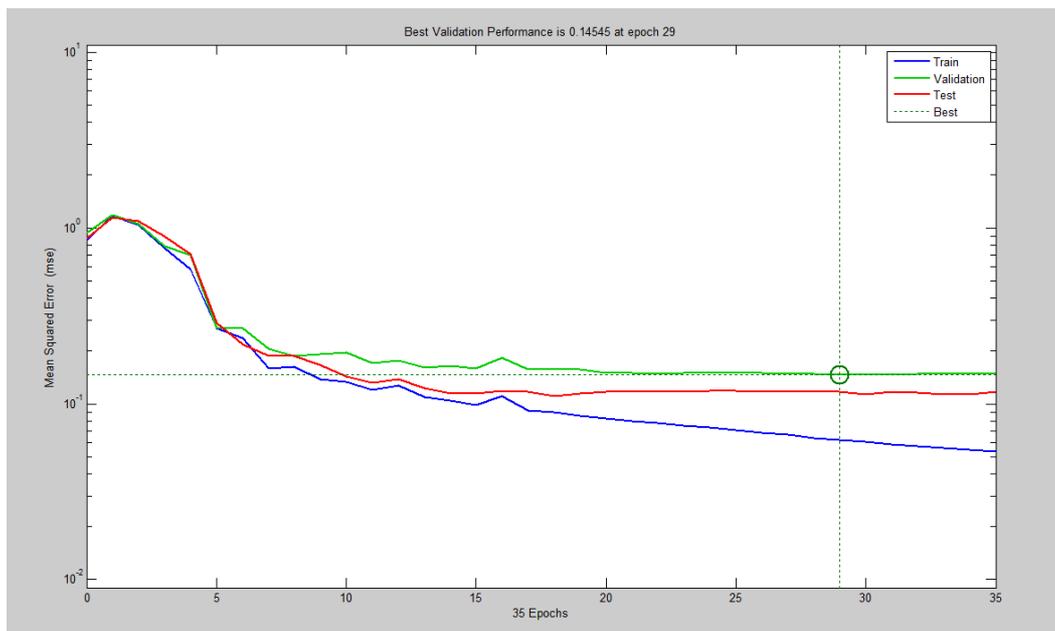


Figura 44. Erro quadrático médio da série de treinamento, teste e validação de 292 inseticidas.

O menor erro ocorreu durante a 29ª. época do treinamento, obtendo-se o valor de 0,145. Pode-se observar a inexistência de um *overfitting* sobre o conjunto de treinamento, visto que o erro sobre o conjunto de validação e teste não é essencialmente reduzido após a 29ª. época.

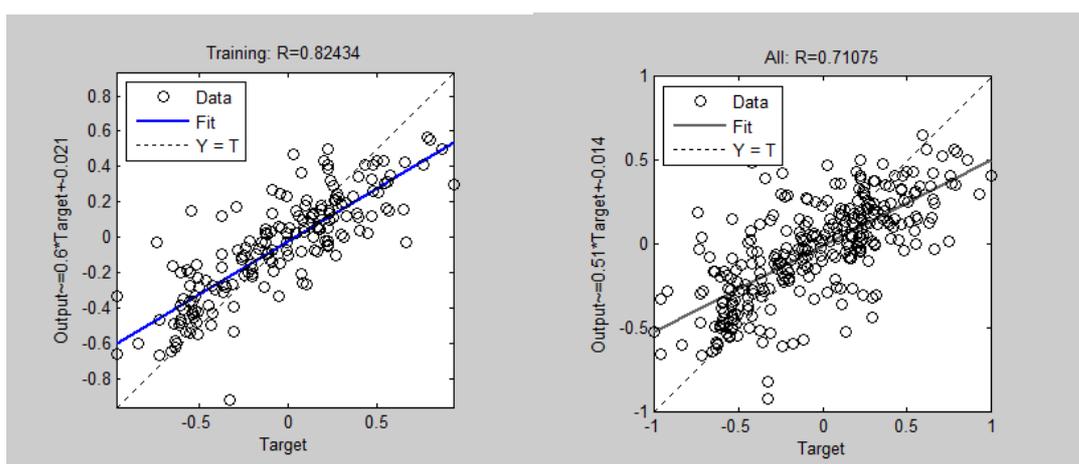


Figura 45. Ajuste linear para a série completa de inseticidas (treinamento + validação + teste).

A Figura 45 mostra o ajuste linear para a série de treinamento e para a série completa (treinamento + validação + teste). Pode-se observar que a capacidade de predição de 71 % pode ser considerada muito boa tendo-se em vista a elevada complexidade do espaço químico descrito pelos herbicidas.

5. ESTUDO DE CASOS

A aplicação direta dos métodos empregados neste trabalho pode ser diretamente observada nos exemplos de caso que se seguem. O principal objetivo é demonstrar que a aplicabilidade, as capacidades de ajuste/calibração e validação do método são suficientemente apropriadas para estabelecer e definir espaços químicos capazes de englobar espaços biológicos. Os exemplos selecionados são de origem real para produtos técnicos em uso corrente no mercado brasileiro (e também internacional) e servem apenas como demonstração da viabilidade e validação geral desta proposta de trabalho.

5.1. Caso 1: Determinação das Relações entre Estrutura e Toxicidade (STR) do Glifosato e suas Impurezas

Resumo

A toxicidade do glifosato e suas impurezas foi avaliada por métodos de relações estrutura-toxicidade (Q)STR. Os campos moleculares de interação (MIF) foram calculados para o ingrediente ativo glifosato e todas as impurezas desse produto técnico. Os estudos de similaridade química tridimensional foram realizados através de métodos de reconhecimento de padrões, PCA, PLS e PLS-D. O perfil toxicológico de todos os componentes da mistura do glifosato técnico e suas impurezas estão definido em espaço químico-toxicológico similar àquele exibido pelo próprio glifosato e vários de seus similares. Para validar os resultados obtidos, uma série de 616 herbicidas extraída do banco de dados PAN e do Manual de Pesticidas foi usada como padrão de comparação do perfil toxicológico do glifosato, seus análogos e impurezas, bem como seus similares. Os resultados obtidos demonstram, claramente, que eles estão classificados dentro das expectativas toxicológicas esperadas e aceitáveis dentro do padrão de toxicidez definido. A predição dos valores putativos de DL_{50} para as impurezas do glifosato técnico demonstra que todas elas apresentam valores superiores a 500 mg kg^{-1} , obtidos dos ensaios por via oral em ratos.

Introdução

O presente estudo de caso propõe o emprego de métodos em relações qualitativas e quantitativas entre estrutura, propriedade e toxicidade (Q)SAR/SPR/STR do glifosato como

produto técnico e suas impurezas. O objetivo é demonstrar que através desses estudos é possível identificar o produto técnico e suas impurezas dentro do mesmo perfil de toxicidade de outros herbicidas empregados na agricultura universal.

O glifosato

O glifosato é o ingrediente ativo de várias formulações usadas como herbicidas. Ele é usado mundialmente, inclusive no Brasil, no controle efetivo, não-seletivo de plantas daninhas através de um modo de ação que não está presente em insetos, pássaros, peixes, mamíferos e humanos. Embora o glifosato seja considerado pouco tóxico, há evidências de efeitos deletérios ao ambiente e também devido à resistência adquirida por algumas espécies de plantas daninhas, fungos e bactérias [64].

A toxicidade aguda do glifosato é considerada baixa. De acordo com a Organização Mundial da Saúde (WHO), o valor da DL_{50} oral do glifosato puro em ratos é de 5.600 mg kg^{-1} , enquanto que a do produto técnico é maior que 5.000 mg kg^{-1} para ratos fêmeas.

Considerando-se seu uso largo na agricultura, o glifosato é controlado por agências governamentais reguladoras que inclui também a necessidade de avaliação de seu principal metabólito e também de suas eventuais impurezas.

Neste estudo, há demonstração de seu perfil de toxicidade similar a vários herbicidas empregados universalmente, bem como de suas impurezas.

O glifosato e suas impurezas

O ingrediente ativo objeto deste estudo é o glifosato e suas impurezas. A estrutura e dados toxicológicos do glifosato estão representados na Figura 46

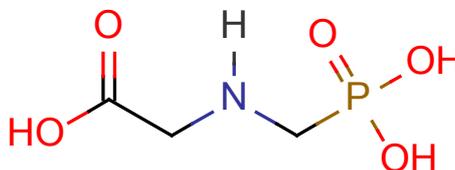
a) Glifosato**Nome comum:** Glifosato**Nome químico (IUPAC):** N-(phosphomethyl)glycine**Estrutura molecular:****Número CAS:** 1071-83-6**Formula empírica:** C₃H₈NO₅P**Lote:** 80814**Código Plantec:** G01/10**Pureza:** 96,0 %**Marca:** Dr. Ehrenstorfer**Validade:** 09/2012**Dados toxicológicos (Pesticide Manual):****DL₅₀ oral Agudo:** ratos = 5600 mg kg⁻¹
camundongos = 11300 mg kg⁻¹**DL₅₀ Cutâneo:** coelhos = > 5000 mg kg⁻¹**CL₅₀ Inalação (4h):** ratos = 0,81 mg L⁻¹ de ar**Dados Ecotoxicológicos (Pesticide Manual):****DL₅₀ oral Agudo:** pássaro (bobwhite quail) = > 3850 mg kg⁻¹**EC₅₀ (48 h):** Daphnia = > 780 mg L⁻¹**CL₅₀ (peixe):** Truta = 120 mg L⁻¹Bluegill sunfish = 120 mg L⁻¹**DL₅₀ contato:** 0,39 mg/abelha

Figura 46. Dados de literatura do Glifosato Técnico.

Métodos e Fundamentos

A forma molecular e a distribuição eletrostática desempenham um papel crucial no processo de reconhecimento molecular de padrões necessário para a descrição da similaridade molecular [65]. Dois objetos geométricos são chamados similares, se ambos tiverem a mesma forma. Mais precisamente, um é congruente com o resultado de um dimensionamento uniforme (aumentando ou diminuindo) do outro. O uso de abordagens baseadas em similaridade através dos campos moleculares para a obtenção de alinhamentos moleculares de qualidade e para a identificação de padrões em moléculas bioativas/tóxicas é necessário para a

descrição das propriedades tridimensionais de moléculas. Os campos moleculares de interação (MIF) [66], são usados para a identificação de regiões moleculares que possuem características semelhantes. Os MIF são determinados através do uso de sondas químicas virtuais descritivas da superfície molecular cujas projeções de sobreposição resultam em regiões espaciais de estabelecida similaridade. Três sondas foram usadas neste estudo: água (OH₂), hidrofóbica – lipofilia (DRY) e oxigênio carbonílico (O). Através do alinhamento das substâncias em grades cartesianas faz-se uma varredura espacial das interações atrativas e de repulsões intermoleculares e identificam-se as regiões hidrofílicas, hidrofóbicas e aceitadoras de ligação hidrogênio, respectivamente, com complementaridade de forma e energia.

O conceito fundamental empregado é que estruturas similares apresentam propriedades similares. A similaridade é descrita através de *fingerprints* (“impressão digital”) [67] e índices de similaridade [68]. No presente estudo, a similaridade de Tanimoto foi avaliada através da equação $c/(a + b - c)$, onde a é o número de *bits* da molécula A, b da molécula B e c é o número comum de *bits* entre A e B. Duas moléculas serão iguais quando o índice de Tanimoto for igual a 1, mas se nada em comum for encontrado o índice de Tanimoto será igual a 0. Por exemplo, o índice de Tanimoto do glifosato é 1,00 em relação à sua forma molecular e 1,00 em relação aos heteroátomos contidos na sua estrutura (“cor do átomo”), que combinados resultam no valor máximo igual a 2,00 (veja a Tabela 1). Esta é uma ferramenta poderosa para o reconhecimento de padrões e, portanto, muito importante para a descrição da toxicidade de pesticidas.

Análises quimiométricas baseadas em métodos de mínimos quadrados parciais (PLS) [69, 70] foram realizadas para a obtenção de componentes principais latentes em que regiões próximas no espaço bidimensional que representam padrões similares de propriedades moleculares e perfil de toxicidade.

Os parâmetros estatísticos foram definidos como segue:

Análise de PLS-D: 42 compostos. Sondas químicas: H₂O, DRY, O, representativas de campos moleculares de interação hidrofílico, hidrofóbico e aceitador de ligações hidrogênio; 41 descritores por substância, totalizando 1.722 dados matriciais para análise. O coeficiente de ajuste foi $R^2 = 0,72$, com capacidade preditiva $Q^2 = 0,63$. A variância total medida foi de PC1: 32,8 %, PC2: 23,9 % e PC1XPC2: 56,7 %.

Para as análises de PCA, foram usados 616 herbicidas e geradas três componentes principais cuja variância está assim distribuída: PC1: 37,3 %, PC2: 29,2 % e PC1XPC2: 66,5 %. Doze herbicidas com similaridade superior a 1 com pelo menos uma das impurezas ou do próprio glifosato foram usadas para a determinação do nível de similaridade comparativa. As

mesmas sondas H₂O, DRY e O foram empregadas e foram calculados 96 descritores por composto, totalizando 60.288 dados matriciais para análise.

Nas análises de reconhecimento de padrões através do método de similaridade por analogia (SIMCA), os campos moleculares de interação do tipo forma, hidrofílico, eletrostático e hidrofóbico foram usados. Os dados *a priori* obtidos das análises de similaridade química e PCA foram usados como parâmetros de entrada para a supervisão matemática do método.

Estudos de Relação Estrutura-Toxicidade: (Q)SAR, (Q)SPR, (Q)STR

Perfil toxicológico comparativo a fármacos empregados na quimioterapia humana

Com o objetivo de estabelecer, em nível molecular, o perfil de toxicidade das impurezas do glifosato em alvos de interesse para o ser humano, três análises distintas foram inicialmente realizadas: capacidade de interação do glifosato e suas impurezas com proteínas [71], toxicidade em canal iônico hERG (gene relacionado ao ether-à-go-go) [72] e perfil metabólico[73]. Através do emprego de métodos *in silico* (computacionais) robustos [74], foram usados 408 compostos nos estudos de interação com proteínas cujos campos moleculares de interação intermolecular foram obtidos a partir das sondas químicas OH₂, hidrofóbica (DRY) e O, para um total de 94 variáveis empregando o método dos mínimos quadrados (PLS). Todas as substâncias foram agrupadas em região desejada de baixa ligação proteica.

O gene hERG codifica uma proteína conhecida como canal iônico de potássio. Este canal iônico é mais conhecido por sua contribuição para a atividade elétrica do coração, que coordena seu batimento (ou seja, o canal hERG media a corrente repolarizante de ação cardíaca potencial). Quando este canal perde a capacidade de conduzir corrente elétrica através da membrana celular, uma doença potencialmente fatal chamada síndrome de QT longo pode resultar em morte súbita. Portanto, uma potencial inibição do hERG deve ser evitada durante o desenvolvimento e/ou uso de medicamentos ou substâncias tóxicas. A importância de inibição do hERG é de grande dimensão e por isso o mesmo é conhecido como anti-alvo[75, 76].

O perfil de toxicidade para o glifosato em canal iônico foi aceitável para os padrões de similaridade estabelecidos por todas as impurezas e pelo próprio glifosato. O número de compostos usados foi de 167, cujos campos moleculares foram obtidos usando as mesmas sondas químicas: OH₂, DRY, O e a mesma avaliação estatística. A resposta foi observada em

termos da classe de inibição do canal iônico hERG e a discriminação para os componentes da mistura foi feita classificando-se as substâncias pertencentes a áreas de bloqueadores do canal hERG, intermediária de bloqueadores/não-bloqueadores do canal hERG, intermediária de não-bloqueadores/bloqueadores de canal hERG e de não-bloqueadores do canal hERG. Todos os componentes da mistura, incluindo o glifosato, foram classificados como não bloqueadores.

A estabilidade metabólica é uma propriedade importante dos xenobióticos, uma vez que afeta parâmetros como a depuração, meia-vida e biodisponibilidade oral em diferentes espécies. Conseqüentemente, o conhecimento da estabilidade metabólica é essencial na descrição do perfil farmacocinético e farmacodinâmico do xenobiótico. Ela é determinada com o objetivo de avaliar o potencial tóxico do xenobiótico ou como metabolito farmacologicamente inativo devido à fase I do metabolismo ou ainda devido ao seu acúmulo no corpo devido à falta ou degradação metabólica insignificante. A determinação da estabilidade metabólica é, portanto, uma medida para descrever o destino metabólico[77].

O glifosato e suas impurezas demonstraram um perfil metabólico com indício de estabilidade metabólica aceitável para os padrões toxicofóricos exibidos. Ou seja, através do controle de depuração, tempo de meia-vida e biodisponibilidade é possível avaliar o nível de segurança metabólica dos mesmos.

A estabilidade metabólica foi acessada através de banco de dados indicativos do metabolismo pela isozima CYP3A4 – o grupo mais proeminente de enzimas metabolizadoras de xenobióticos da super família de citocromo P450s [35]. Foram usados 1507 compostos com campos determinados usando as mesmas sondas já mencionadas.

Perfil toxicológico comparativo a herbicidas

A toxicidade aguda do glifosato é considerada baixa [78] (a DL_{50} oral do glifosato puro em ratos é de 4.230 mg kg^{-1}) tendo-se em vista que sua rota bioquímica de ação nas plantas – a rota do ácido chiquimico, não existe em animais. Embora o glifosato possa impedir a ação de funções enzimáticas nos animais, os resultados acima (para o glifosato e suas impurezas) demonstram e corroboram claramente que seu correto uso e manuseio pelo ser humano, como estabelecido pelas agências reguladoras no mundo, está definido. Estes resultados estão de acordo com a classificação toxicológica IV - pouco tóxico para o ser humano, mas para o impacto ambiental (classe II - moderadamente perigoso). Veja o Quadro 1 abaixo. Neste caso, é necessário realizar estudos comparativos para assegurar a validade de utilização dentro dos conceitos preconizados pela órgãos reguladores e outras agências reguladoras como a US EPA. Dessa forma, apesar da toxicidade relativamente baixa do

glifosato, é necessário demonstrar que ele e suas impurezas estão em padrões toxicológicos comparáveis aceitáveis a outros herbicidas em uso corrente.

O glifosato é classificado segundo a US-EPA (Agência de Proteção Ambiental dos Estados Unidos da América), como um composto moderadamente tóxico, pertencente à classe II de toxicidade, conforme mostrado na Tabela 12.

Tabela 12. Classificação de toxicidade segundo a EPA.

Classe	Toxicidade	Indicação
I	Altamente tóxico	Perigo-Veneno
II	Moderadamente tóxico	Perigo
III	Ligeiramente tóxico	Cuidado
IV	Praticamente não tóxico	Sem perigo

Os parâmetros de toxicidade da classe II (moderadamente tóxico), onde se inclui o glifosato ($\sim 5.000 \text{ mg kg}^{-1}$) são:

- **DL₅₀ oral: > 50-500 mg kg⁻¹**
- **DL₅₀ dérmica: > 200-2.000 mg kg⁻¹**
- **DL₅₀ inalação: > 0,2-2,0 mg kg⁻¹**
- **Irritação na pele: moderada**

Ainda segundo a US EPA e também a Canadá Agricultura (E.U. EPA 1993; Canadá Agricultura IPCS 1994, 1991), a dose aguda ou concentração de glifosato que irá causar a morte são os seguintes:

- Oral DL₅₀ (rato) > 4.320 mg kg⁻¹**
- Dermal CL₅₀ (coelho) > 2.000 mg kg⁻¹**

Esses valores são equivalentes a outros mostrados acima.

A Organização Mundial da Saúde também apresenta quatro classes de toxicidade conforme segue:

Classe 1 - a: extremamente perigosos;

Classe 1 - b: altamente perigosos;

Classe 2: moderadamente perigosas;

Classe 3: pouco perigoso.

O sistema é baseado na determinação da DL_{50} em ratos. Assim, um agente oral sólido, com uma DL_{50} de 5 mg ou menos por kg de peso corporal é Classe Ia; em 5-50 $mg\ kg^{-1}$ de classe Ib; em 50-500 $mg\ kg^{-1}$ de Classe II, e em mais de 500 $mg\ kg^{-1}$ Classe III para concentrações de 95-100 % do ingrediente ativo na formulação). Os valores podem variar quando líquidos orais são empregados ou para agentes dérmicos.

Modelo de QSAR para classificação

A Tabela 13 mostra os índices de similaridade para o glifosato, todas as impurezas encontradas no produto técnico e doze herbicidas comerciais (Tabela 14). As estruturas similares exibirão comportamentos similares. Assim, se o interesse estiver no estabelecimento do grau de toxicidade das impurezas o índice de similaridade das impurezas deverá ser calculado. Se, entretanto, um elevado índice de similaridade não resultar na toxicidade esperada, estudos de relações estrutura-atividade servirão para a identificação das características estruturais e propriedades físico-químicas necessárias para tornar a informação útil.

Para o cálculo da similaridade, são necessários dois parâmetros: o descritor e o índice de similaridade. Nestes estudos, os descritores usados foram os tridimensionais de forma e átomo-combinados (combo), a partir dos quais o índice de Tanimoto foi calculado. O valor máximo obtido é igual a 2 e indica similaridade combinada total entre a forma e os átomos que compõem a fórmula estrutural. Os dados apresentados na Tabela 13 revelam que todas as impurezas apresentam valores de similaridade superior a 1 com o componente ativo da formulação (Glifosato, CAS 1071-83-6), o que sugere serem suas estruturas de similaridade suficientemente elevada para permitir inferir que estes devem exibir perfis toxicológicos semelhantes. Além disso, também pode-se observar altos graus de similaridade dos contaminantes com herbicidas atualmente disponíveis no mercado, que possuem toxicidade de

no mínimo 650 mg Kg^{-1} . Deste modo, pode-se seguramente atribuir perfis semelhantes de toxicidade para os contaminantes da formulação em estudo.

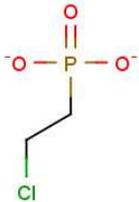
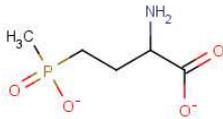
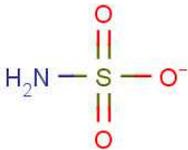
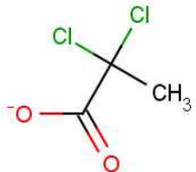
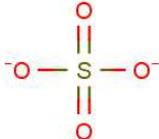
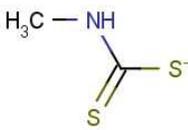
Com base nestas análises é possível também estabelecer uma serie de compostos como controle positivo para o desenvolvimento dos modelos matemáticos de STR propostos neste estudo.

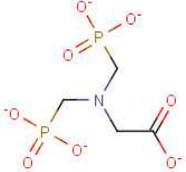
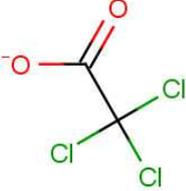
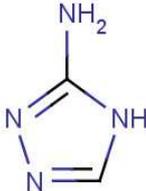
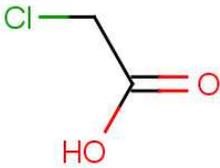
Considerando-se os índices de similaridade aceitáveis, uma série de estudos de relações estrutura-toxicidade foi realizada. O primeiro deles consistiu de uma análise de componentes principais (PCA), empregando os campos moleculares de interação entre pequenas sondas químicas virtuais e todas as impurezas e também o glifosato. Com o objetivo de estabelecer o espaço químico e demonstrar que o glifosato, suas impurezas e os herbicidas em uso comum podem ser descritos pelos mesmos parâmetros, mesmo que haja grande diversidade estrutural, uma série de outros 616 herbicidas foi incluída no banco de dados.

Tabela 13. Similaridade 3D combinada (Tanimoto/Combo) entre os pares de substâncias contidas no conjunto de dados: Glifosato (CAS: 1071-83-6, em azul); Impurezas em amarelo; Outros herbicidas (laranja). Similaridades (S) $1 \leq S < 1,4$ (verde claro); $S \geq 1,4$ (verde escuro).

	1071-83-6	IMPUREZA 1	IMPUREZA 2	IMPUREZA 3	IMPUREZA 4	IMPUREZA 5	IMPUREZA 6	IMPUREZA 7
1071-83-6	2,00	1,27	1,72	1,74	1,63	1,10	1,62	1,10
IMPUREZA 1	1,27	2,00	1,08	1,17	1,13	1,69	1,40	1,61
IMPUREZA 2	1,72	1,08	2,00	1,53	1,55	0,94	1,43	1,20
IMPUREZA 3	1,74	1,17	1,53	2,00	1,59	1,02	1,59	0,95
IMPUREZA 4	1,63	1,13	1,55	1,59	2,00	0,98	1,34	0,96
IMPUREZA 5	1,10	1,69	0,94	1,02	0,98	2,00	1,17	1,38
IMPUREZA 6	1,62	1,40	1,43	1,59	1,34	1,17	2,00	1,12
IMPUREZA 7	1,10	1,61	1,20	0,95	0,96	1,38	1,12	2,00
Amitrole	0,70	1,04	0,72	0,72	0,61	1,17	0,71	1,12
Chloroacetate	0,88	1,33	1,00	1,00	0,92	1,31	0,87	1,57
Dalapon	0,87	1,24	1,01	1,01	0,99	1,24	0,90	1,45
Daminozide	1,05	1,09	1,09	1,00	0,98	1,09	1,11	1,16
Ethephon	1,18	1,57	1,08	1,10	1,23	1,58	1,22	1,21
Fosamine	0,92	1,21	0,96	0,86	0,96	1,20	0,91	1,10
Glufosinate	1,09	1,20	1,18	1,10	1,09	0,98	1,11	1,51
Glyphosine	1,32	0,84	1,18	1,22	1,48	0,76	1,08	0,72
Metam	0,72	1,16	0,76	0,67	0,73	1,17	0,72	1,27
Sulfamate	1,03	1,65	0,94	0,96	1,01	1,66	1,11	1,48
Sulfuric acid	1,07	1,68	0,91	0,99	1,08	1,70	1,16	1,41
TCA	0,86	1,23	1,00	0,78	0,99	1,23	0,89	1,44

Tabela 14. Estruturas moleculares e dados toxicológicos dos doze herbicidas utilizados nas análises de similaridade 3D.

CAS	Estrutura molecular	Nome comum	DL ₅₀ oral agudo Ratos
19767-45-4		Ethephon	3400
53369-07-6		Glufosinate	2030
5329-14-6		Sulfamate	3160
75-99-0		Dalapon	9330
7664-93-9		Sulfuric acid	2140
144-54-7		Metam	1700

59682-52-9		Fosamine	10125
2439-99-8		Glyphosine	3925
650-51-1		TCA	3200
1596-84-5		Daminozide	5000
61-82-5		Amitrole	1100
79-11-8		Chloroacetate	650

A Figura 47 mostra o gráfico das duas primeiras componentes principais descritas nos níveis de variância de 66,5 %.

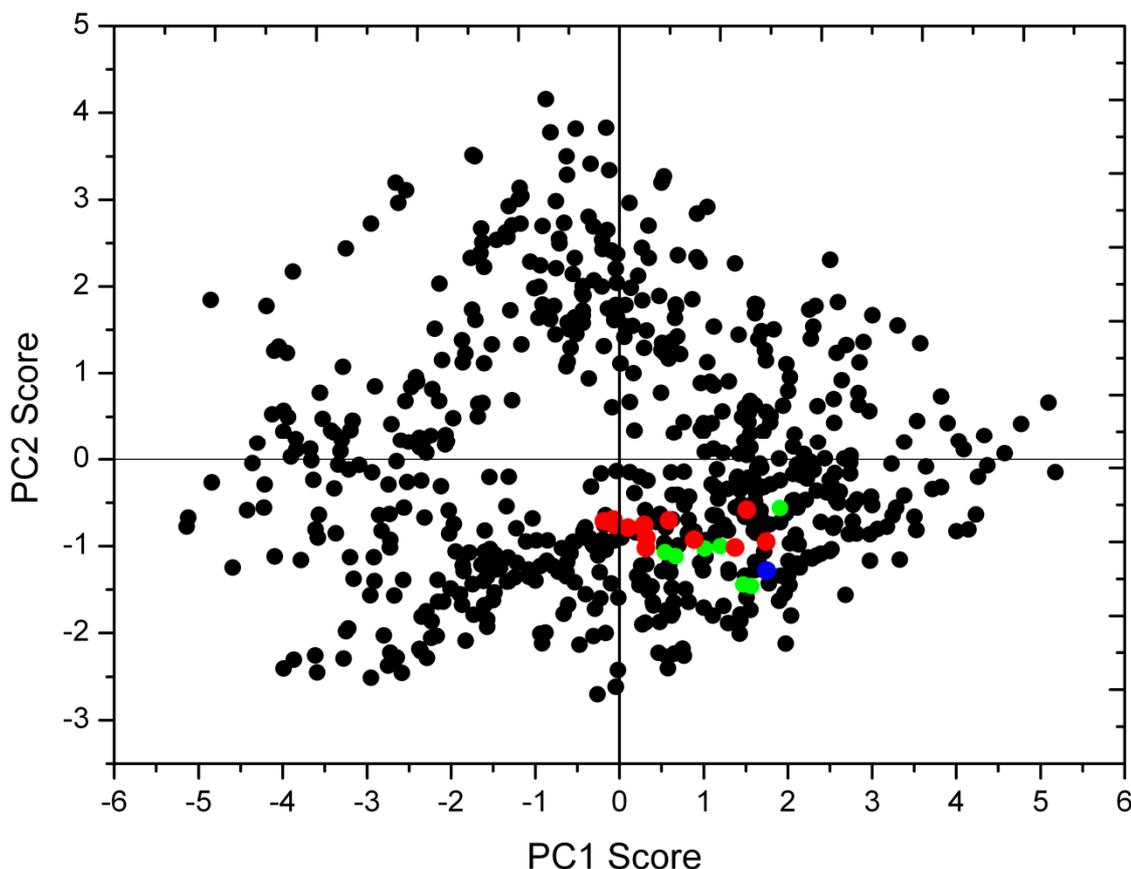


Figura 47. Projeção do Glifosato (azul), 7 Impurezas (verde) e 12 Herbicidas diversos (vermelho) sobre o modelo de PCA gerado a partir de 614 herbicidas descritos no banco de dados PAN (preto).

Todos os herbicidas em vermelho possuem uma similaridade Tanimoto/Combo (Forma+Cor) de pelo menos uma unidade com uma ou mais impurezas. Além disso, suas toxicidades (DL_{50} oral em ratos) são superiores a 400 mg kg^{-1} .

A análise de componentes principais (PCA)[79], mostrada na Figura 47 descreve o conteúdo de informação matricial referente aos objetos em estudo (herbicidas). A grande utilidade dessa análise dá-se pelo entendimento da distribuição e diferença entre os objetos que estão representados por cada ponto dentro do espaço molecular descrito, de tal forma que informações similares ou completamente independentes podem ser observadas. A informação é mais bem apresentada na forma do gráfico de escores bidimensionais que representam as posições relativas dos objetos no espaço. A Figura 47 mostra o gráfico de escores de PC1 versus PC2 com o glifosato (ponto em azul), suas impurezas (verde) e herbicidas comercializados similares (vermelho) agrupados na região do quadrante (-,+), com poucos

dos herbicidas estendendo-se para o quadrante (-,-). Estes resultados demonstram o agrupamento similar entre os herbicidas e são particularmente interessantes porque não consideram os valores de DL_{50} explicitamente. Mais importante ainda é o fato de que a PCA foi capaz de reconhecer todos os similares na mesma região com base apenas em suas estruturas e propriedades moleculares. Estes dados estão corroborados pela análise de similaridade química mostrada na Tabela 13, em que o glifosato tem elevado nível de similaridade com as impurezas contidas no produto técnico. Dessa forma, as propriedades biológicas/toxicológicas esperadas são similares. Este resultado está demonstrado através dos dados da Figura 47 e da Tabela 13. Esses dados podem ser extrapolados para as impurezas para as quais não há dados de DL_{50} , já que possuem perfil de similaridade elevado e estão agrupadas na mesma região do espaço definido na Figura 47.

Todos os pontos observados no gráfico da Figura 47 definem o espaço químico dos herbicidas em estudo, juntamente com as impurezas do produto técnico glifosato. O glifosato, suas impurezas e os herbicidas similares, estão contidos nesse grande espaço químico e agrupados por similaridade como descrito acima. Portanto, a validade da diversidade molecular estabelece-se pelo reconhecimento de padrões necessários para o correto agrupamento dos similares.

A análise de PCA tem como característica fundamental descrever o espaço químico de acordo com as propriedades estruturais de forma não supervisionada. Ou seja, nenhuma informação sobre o efeito biológico/toxicológico é usado para o estudo de agrupamento. Como extensão natural dessa análise, a análise de similaridade por analogia estrutural (SIMCA) pode ser usada para a classificação dos herbicidas similares em grupos distintos, através do reconhecimento de padrões químicos. Embora seja legítimo descrever superposições de substâncias químicas estruturalmente diversas na mesma classe, o método oferece a oportunidade de calcular a distância entre as eventuais classes de acordo com a similaridade intrínseca de cada série de moléculas. A modelagem independente de classes usada no método SIMCA tem a vantagem de avaliar cada membro de uma determinada classe como se ele tivesse uma característica única não superposta que pode ser incluída no modelo independentemente do “modelo” ou classe previamente existente. Isto porque cada grupo é modelado independentemente através de sua componente principal, dependendo apenas da estrutura espacial de hiperplanos. A robustez do método consiste, então, na classificação inequívoca (quando pertinente) através da identificação separada de cada um dos hiperplanos. A distância entre as classes é calculada como a distância geométrica das componentes principais em que a substância desconhecida é classificada na região do espaço à qual

pertencerá. Isso significa que a mesma substância será projetada sobre as classes na tentativa de identificar a distância que melhor refletir a similaridade com os membros daquela classe. O resultado obtido está mostrado na Figura 48, para o gráfico de Cooman que é obtido a partir das coordenadas residuais que maximiza a separação entre as classes ao invés de explicar a variação máxima dos dados. Além disso, é possível usar o modelo para predição da classe de substâncias independentes com base em uma série de treinamento pertencente a uma classe conhecida, como é demonstrado no próximo item.

O método SIMCA usa as características das componentes principais para criar limites significativos das classes previamente especificadas para os escores das substâncias e a direção residual. O gráfico de Cooman mostrado na Figura 48 representa o mapa de visualização de duas classes em que os dois modelos independentes de componentes principais são obtidos quando o mapeamento das substâncias de “cada classe” for identificado para cada substância que seja classificada e pertencente a uma das duas classes, a ambas as classes ou a nenhuma das classes. Como pode ser observado na Figura 48, há classificação dos herbicidas em estudo em duas classes. Em vermelho estão os herbicidas similares ao glifosato. As impurezas do produto técnico também estão mapeadas nesta classe, bem como os herbicidas similares ao glifosato e às suas impurezas. Os pontos em preto representam todos os demais herbicidas e estão classificados na outra classe. Estes resultados corroboram a análise não-supervisionada previamente mostrada na Figura 47 e ratificam estatisticamente a natureza da classificação. Para os herbicidas similares ao glifosato o percentual de acerto na classificação é de 95,4 % para a categoria 1. Para os demais herbicidas do banco de dados, o percentual de acerto é 92,3 %. Esses valores são excepcionalmente elevados considerando-se a diversidade molecular do banco de dados, a natureza das propriedades descritas e as posições espaciais relativas da Figura 47.

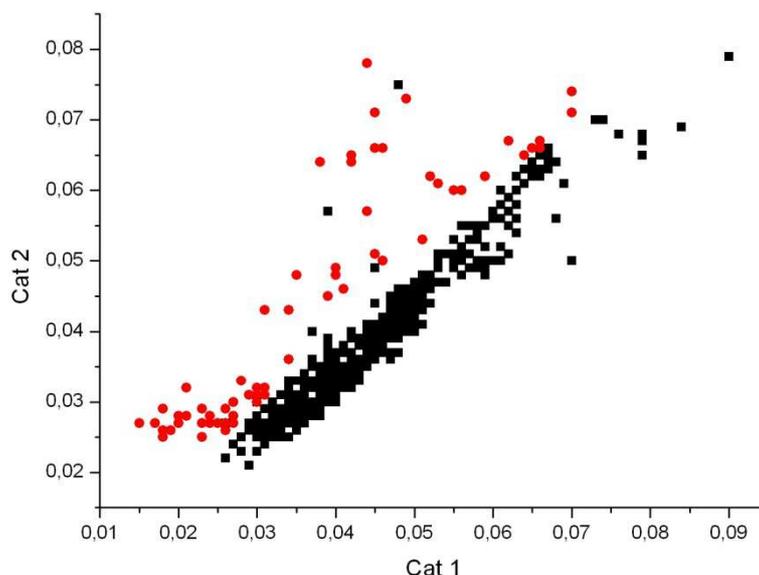


Figura 48. Gráfico de Cooman. Glifosato e Impurezas (vermelho); Herbicidas (preto): 95,4 % e 92,3 % de classificação correta, respectivamente. Veja o texto para a explicação.

Modelo de QSAR para predição

1. Análise de Similaridade Química por Analogia (SIMCA)

Os métodos quimiométricos não-supervisionados como o da Figura 48, funcionam muito bem quando o número de classes é bem definido. Os resultados de PCA mostrados na Figura 47 e já amplamente discutidos, demonstrando claramente a existência de agrupamentos definidos que também foram apropriadamente classificados pela análise supervisionada SIMCA. Além disso, o método SIMCA possibilita ainda a predição da substância em uma determinada classe. Dessa forma, com o objetivo de validar o modelo de classificação, uma nova análise SIMCA foi realizada.

O procedimento quimiométrico supervisionado empregado – aquele em que incorpora conhecimento *a priori* da identidade da classe pode ser usado para prever independentemente a classe de herbicidas com base em uma série de treinamento cuja classe seja conhecida. Tipicamente, neste caso, o banco de dados é dividido em duas séries distintas: uma série de treinamento e uma série de validação e os modelos preditivos são construídos com base na série de treinamento. No presente estudo, o banco de dados contendo todos os herbicidas, inclusive as impurezas do glifosato técnico, foi separado em uma série de treinamento contendo 70 % do total de herbicidas e 30 % para a série de validação. O processo de escolha foi realizado de forma aleatória com três análises distintas: cada série de treinamento foi modificada para conter herbicidas distintos. A validação do modelo foi então

realizada para a predição dos herbicidas das respectivas séries testes independentes. A importância desse resultado é de particular relevância porque no modelo de treinamento gerado os herbicidas da série de validação não são incluídos na análise e, dessa forma, se classificados corretamente permitem qualificar a natureza do perfil toxicológico dos herbicidas em estudo, em uma determinada classe química.

O processo de classificação do glifosato e suas impurezas em comparação classificatória com os herbicidas similares permitiu determinar um acerto de 100%. Ou seja, o nível de similaridade entre as classes está definido assim como está definida dentro das respectivas classes. As sete impurezas do glifosato técnico pertencem à mesma classe do glifosato. Quando o restante do banco de herbicidas é incluído na análise, as 7 impurezas continuam com 100 % de classificação na mesma categoria. Para os herbicidas similares ao glifosato técnico e às impurezas, o acerto é de 92,3 % e para os demais herbicidas do banco, 95,2 %.

Os percentuais de classificação correta das séries de treinamento são de 100 % para quando o banco inclui o glifosato, suas impurezas e similares e 95,2 % para todos os demais herbicidas do banco. Este resultado que se repete ao longo das análises demonstra uma característica já observada anteriormente no gráfico da Figura 46 em que outros herbicidas do conjunto total têm similaridade elevada ao glifosato e suas impurezas e similares.

Considerando-se a robustez da série de treinamento em corretamente classificar os herbicidas em suas respectivas categorias, o modelo foi então usado para a predição da série de validação constituída dos 30 % de substâncias excluídas da análise inicial. Em média, os resultados favoráveis à correta predição foram: 87,5 % para o glifosato, suas impurezas e similares; 98,9 % para os demais herbicidas. Tipicamente, o número total dos demais herbicidas para cada análise individual foi de 180, com exceção de uma análise que continha ca. de 80 estruturas de herbicidas. Naturalmente, o glifosato técnico, suas impurezas e similares fizeram parte tanto de uma quanto de outra série no processo de predição para permitir a identificação positiva de classificação na classe correta.

2. Análise de PLS discriminante (PLS-D)

A extensão da PCA por regressão resulta na análise dos mínimos quadrados parciais discriminantes (PLS-D). O modelo gerado descreve a separação máxima entre as classes definidas *a priori*, que está mais focalizada na variação discriminativa da classe de dados. A correta interpretação da regressão provê informação quantitativa baseada nas diferenças entre as classes em que os coeficientes estão relacionados a determinada região associada à uma

estrutura molecular específica. Dessa forma, a predição é realizada por projeção do herbicida “desconhecido” sobre as classes previamente identificadas. Portanto, a predição resultará na identificação da classe do herbicida teste (“desconhecido”) ou, eventualmente, em nenhuma das classes previamente definidas que, naturalmente, também constitui informação valiosa. No presente estudo, a PLS-D foi usada para distinguir entre as classes esperadas para serem tóxicas ou não, de acordo com a classificação toxicológica da WHO. Dessa forma, um novo banco de dados constituído de 23 substâncias (herbicidas de diferentes classes) com valores de DL_{50} inferiores a 500 mg kg^{-1} e 19 outros com valores de DL_{50} superiores a 500 mg kg^{-1} foi usado como treinamento da PLS-D. A diferença fundamentalmente importante em relação ao método SIMCA, vem do fato de que nesta análise a definição prévia das classes é realizada apenas com base nos valores de DL_{50} e não nas características estruturais dos herbicidas. Portanto, se o modelo for capaz de quantificar as substâncias “desconhecidas” de forma correta, a classificação na respectiva classe resultará em valiosíssima informação para a tipificação do perfil toxicológico das substâncias de interesse. A Figura 48 mostra o gráfico de escores obtido da PLS-D que classifica o banco de dados de acordo com os limites de DL_{50} previamente definidos. Os pontos correspondentes aos valores de DL_{50} inferiores a 500 mg kg^{-1} estão representados em vermelho e em preto os valores de DL_{50} superiores a 500 mg kg^{-1} .

Como pode ser observado na Figura 48, a análise de PLS-D produz uma clara discriminação entre as substâncias em cada um dos limites definidos, estabelecendo assim duas classes distintas com base apenas em seus valores de DL_{50} . Esta análise, baseada na magnitude dos valores de DL_{50} estabelece as diferenças entre os perfis toxicológicos de acordo com os efeitos da dose que são então correlacionados às respectivas estruturas através de suas propriedades moleculares. No presente caso, as interações intermoleculares com as respectivas sondas químicas usadas para o cálculo dos campos moleculares de interação definem adequadamente as regiões classificatórias dos valores de DL_{50} .

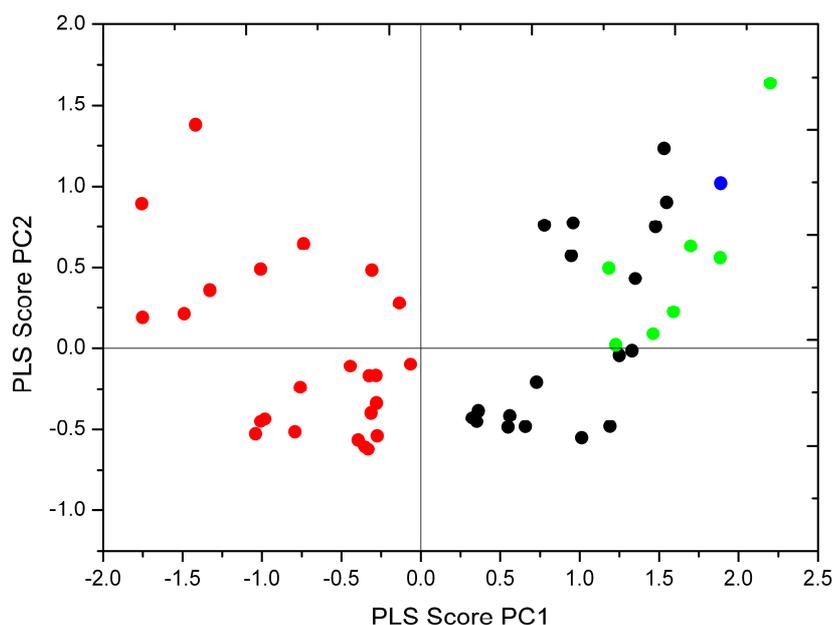


Figura 49. Gráfico de escores referente à análise de regressão por PLS Discriminante obtido a partir de um conjunto contendo 23 substâncias com valores de DL_{50} inferiores a 500 mg kg^{-1} (vermelho), 19 substâncias com DL_{50} superior a 500 mg kg^{-1} .

A projeção do glifosato técnico e suas impurezas está representada por pontos em azul e verde, respectivamente.

A análise de regressão discriminante classificou o glifosato técnico (em azul) e todas as suas impurezas (em verde) no espaço constituído de herbicidas com valores de DL_{50} superiores a 500 mg kg^{-1} , Figura 49. Este resultado é extremamente útil e valioso para a correta classificação do perfil toxicológico das impurezas. Este resultado é confirmado experimentalmente para três das impurezas do glifosato técnico, para os quais os valores de DL_{50} são conhecidos. Como pode ser claramente observado, esses valores corroboram os resultados mostrados nas Figuras 47, 48 e 49 e demonstram, por validação explícita a capacidade preditiva do modelo. Dessa forma, e em conjunto com todas as demais análises realizadas neste estudo, pode-se claramente inferir que o perfil toxicológico esperado para as demais impurezas do produto técnico esteja definido.

Justificativa para o uso de Relações Estrutura-Toxicidade (Q)STR, (Q)SAR, (Q)SPR

⇒ Maximizar a utilização dos dados existentes, juntamente com dados cientificamente adequados, relacionados aos produtos químicos, a fim de minimizar os ensaios com animais.

As relações estrutura-atividade (SAR), estrutura-propriedade (SPR) ou estrutura-toxicidade (STR) podem ser usadas para reduzir o número de ensaios:

1. Os princípios de SAR podem ser aplicados a um único produto químico que esteja intimamente relacionado a um ou mais produtos químicos caracterizados como "análogos" ou "derivados". Os dados aqui analisados levam ao fornecimento de dados toxicológicos virtuais, suficientemente adequados para demonstrar que os análogos/impurezas do glifosato são similares e estruturalmente aceitáveis, dentro do mesmo nível de toxicidade identificada e que, portanto, os ensaios em animais são dispensáveis nesse estágio.
2. As estruturas químicas das impurezas e do glifosato são similares, porque eles estão estreitamente relacionados em estrutura, propriedades químicas e propriedades do ambiente.
3. Numerosos estudos ambiental, toxicológico e ecotoxicológico têm sido realizados com o glifosato e formulações de glifosato. Revisões extensivas e avaliações desses estudos concluíram que o glifosato não é genotóxico e não cancerígeno.
4. Testes de toxicidade de algumas das impurezas indicam semelhança ao glifosato na produção resultados indicativos de baixa toxicidade. As impurezas, na verdade, estão em um nível baixo nas especificações técnicas do grau de pureza do glifosato.
5. Com base na similaridade de estrutura e propriedades e as propriedades toxicológicas conhecidas do glifosato e suas impurezas, pode-se afirmar com certeza razoável que o pequeno potencial de exposição não resultaria em danos excessivos aos seres humanos ou ao meio ambiente, além daqueles já conhecidos e relatados.
6. Os dados apresentados são considerados suficientes para satisfazer os requisitos da dos órgãos reguladores sem a necessidade de realizar quaisquer novos ou adicionais ensaios.

Conclusões do caso 1

O avanço das tecnologias quimiométricas tem permitido a criação de grandes e complexos bancos de dados. A capacidade de acumular dados e a necessidade enfatizada de recuperar dados conduzem ao aprimoramento dos métodos de mineração de dados e análises. Um dos grandes desafios da análise quimiométrica é a extração de dados relevantes biológica e toxicologicamente capazes de utilidade prática que resulte na redução do número de experimentos e principalmente no uso reduzido de animais sem que a informação desejada seja perdida. Existem várias estratégias para análises relacionadas. Entretanto, aqueles métodos capazes de tratar bancos de dados matricialmente megavariados são poderosos no processo de correlação entre estrutura-propriedade e parâmetros toxicológicos. Neste estudo foram gerados para as mais de 600 estruturas moleculares, mais de 60.000 descritores. Através dos métodos de PCA, SIMCA e PLS-D, esses dados foram corretamente balanceados em grupos identificativos das respectivas classes de substâncias químicas com acerto total

para o glifosato técnico e suas impurezas e excessivamente robusto para o banco de dados. Portanto, estes métodos oferecem a possibilidade de explorar as relações entre dados tipicamente distintos tais como os gerados por dados toxicológicos. Um dos principais usos desta tecnologia vem da capacidade de predição dos efeitos toxicológicos que ofereça informações robustas sobre as estruturas e suas propriedades. Estes sistemas inteligentes são capazes, portanto, de prever a toxicidade das substâncias em estudo, identificar respostas idiossincráticas e avaliar as diferenças entre as classes que as compoem.

Neste estudo de relações estrutura-toxicidade (STR), ficou demonstrado que, por similaridade química, as impurezas do glifosato técnico apresentam toxicidade parecida com o produto principal – o próprio glifosato e também outros herbicidas similares que estão classificados dentro do mesmo perfil toxicofórico. Ou seja, os campos moleculares de interação (MIF) do glifosato e suas impurezas, avaliados por métodos quimiométricos robustos, associados aos índices de similaridade tridimensional demonstram que o perfil toxicofórico é mantido em níveis semelhantes.

Inicialmente, três avaliações distintas foram realizadas: interação a proteínas, bloqueio do canal iônico hERG e estabilidade metabólica. Os resultados são coerentes e condizentes com a toxicidade do glifosato postulada para os seres humanos. As impurezas apresentaram perfis semelhantes ao do glifosato.

Modelos robustos e significativos foram desenvolvidos para estabelecer as relações entre as estruturas de herbicidas e os valores de DL_{50} . Os métodos de PCA, PCA-SIMCA, SIMCA-preditivo e PLS-D mostraram-se úteis para a descrição do espaço químico-biológico baseado em informações toxicológicas, bem como na predição do perfil toxicológico de interesse. Das relações estabelecidas depreende-se claramente que o nível de toxicidade das impurezas do glifosato técnico apresenta perfil similar, suficientemente amparado no mesmo espaço químico-toxicológico definido pelos campos moleculares de interação e relacionados com as magnitudes dos valores de DL_{50} . O banco de dados construído e selecionado para a análise contém informações da diversidade química de herbicidas em uso comercial com aprovação de agências reguladoras, cujas informações estruturais e propriedades coadunam com o perfil conhecido e descrito para várias das impurezas. Para aquelas impurezas encontradas em concentrações diminutas no glifosato técnico, as magnitudes foram estimadas e os resultados demonstraram que o espaço químico toxicológico a que pertencem – juntamente com as demais impurezas, são suficientes para inferir o perfil toxicológico postulado (veja a Figura 49).

Os métodos empregados juntamente com os descritores estruturais obtidos para uma

série diversa de substâncias herbicidas mostraram-se adequados para a geração de bons modelos quantitativos e qualitativos confiáveis dentro do espaço estabelecido.

Os testes podem consistir de ensaios *in vivo* ou *in vitro* e também podem ser realizados pelo uso de modelagem via QSAR. Neste caso, os testes *in silico*, mais apropriados na relação custo/tempo e na diminuição do número de animais produziram resultados cujas informações preliminares estão consubstanciadas no mesmo espaço químico-toxicológico de inúmeros herbicidas em uso comercial. Os resultados são típicos para priorizar as informações necessárias para sua descrição nesse espaço e dessa forma estabelecer os níveis de exigência da toxicidade química que precisa ou não ser avaliada em maior detalhe. As relações matemáticas entre os aspectos quantificados da estrutura química e as suas propriedades químicas ou toxicológicas foram desenvolvidas usando uma série adequada de herbicidas. A capacidade de demonstrar apropriadamente o espaço ocupado pelo glifosato técnico e suas impurezas como similar provê informação suficiente para o estabelecimento das relações entre várias propriedades da estrutura molecular e toxicológica. Essas relações permitiram, então, a predição das propriedades químicas e toxicológicas das demais substâncias em estudo, com estruturas similares. A presunção de que moléculas ou estruturas químicas similares tenham propriedades químicas e potencial de toxicidade similares foi demonstrada.

Considerando-se que os objetivos da (Q)SAR são apropriados para facultar a realização eventual de outros protocolos experimentais é importante observar que a estratégia estabelecida neste estudo foi a de complementar cada resultado com uma análise distinta e ortogonal mas que conduziu ao mesmo resultado. Neste caso, a similaridade química e a magnitude dos valores de DL_{50} foram modelados e produziram resultados que permitem avaliar positivamente os estudos de (Q)SAR, como indicativos da comparabilidade das estruturas e suas relações de toxicidade.

O fato de que os modelos de QSAR são reducionistas e por isso estão associados a limitações, foi testado neste estudo por validações externas de um grande e representativo banco de dados contendo herbicidas comerciais e muitos aprovados para uso por agências reguladoras.

Em conclusão, foi demonstrado que o glifosato técnico e seus produtos ocupam o mesmo espaço químico e de toxicidade estabelecido para outros herbicidas que é também ocupado pelas impurezas cujos níveis estão definidos dentro de limites inferiores a 5 %, conforme as especificações técnicas do produto.

Sendo assim, foram conduzidos estudos toxicológicos agudos para o glifosato com as impurezas estudadas nesse caso e o resultado corroboram os modelos utilizados nesse estudo.

5.2. Caso 2: Determinação das Relações entre Estrutura e Toxicidade (STR) da Cipermetrina

Resumo

A toxicidade da cipermetrina e um subproduto foram avaliados por métodos de relações estrutura-toxicidade (STR). Os campos moleculares de interação (MIF) foram calculados para o ingrediente ativo Cipermetrina Técnico e a impureza do produto técnico. Os estudos da similaridade química tridimensional foram realizados através de métodos de reconhecimento de padrões PLS. O perfil toxicológico da amida da cipermetrina técnica está definido em espaço químico-biológico parecido àquele exibido pela própria cipermetrina.

Piretróides

Os piretróides são compostos químicos sintéticos, similares às piretrinas, que são compostos químicos naturais produzidos por flores do piretro (*Chrysanthemum cinerariaefolium* e *C. coccineum*)[80]. As piretrinas são constituídas de um par de compostos orgânicos naturais que têm atividade inseticida potente [81]. Os piretróides constituem uma parte importante do mercado de inseticidas sintéticos e são comuns em produtos comerciais, tais como inseticidas domésticos. Nas concentrações utilizadas nos produtos, eles também podem ter propriedades repelentes de insetos e são geralmente inofensivos para os seres humanos em doses baixas.

As piretrinas são usadas em muitas variedades de inseticidas; portanto, são tóxicas. Cuidados devem ser tomados ao usar essa classe de substâncias quando humanos e animais fizerem parte do ambiente ao quais os inseticidas forem aplicados. Administração involuntária em animais domésticos e em super-dosagem podem resultar em uma variedade de sintomas, especialmente em animais de estimação, incluindo sialorréia, letargia, tremores musculares, vômitos, convulsões e morte. Os sintomas de toxicidade em humanos incluem respiração asmática, espirros, congestão nasal, dor de cabeça, náusea, falta de coordenação, tremores, convulsões, rubor facial e edema e sensação de ardor e prurido. Embora seja possível o controle da dose para minimizar esses problemas tóxicos, as piretrinas são pouco tóxicas para certas espécies de aves, como os patos. Esses compostos também são tóxicos para as abelhas [82].

A cipermetrina

A cipermetrina é um piretróide sintético utilizado como inseticida em grande escala em aplicações comerciais agrícolas, bem como em produtos de consumo para fins domésticos. Ele atua sobre neurotoxinas com ação rápida sobre insetos. É facilmente degradada no solo e plantas. A exposição à luz solar, água e oxigênio irá acelerar sua decomposição.

As impurezas e dados de Toxicidade

A Figura 50 mostra a representação da estrutura química da Cipermetrina Técnica.

O índice de similaridade 3D entre a cipermetrina e a Impureza 1 é 1,7, que indica elevado grau de similaridade.

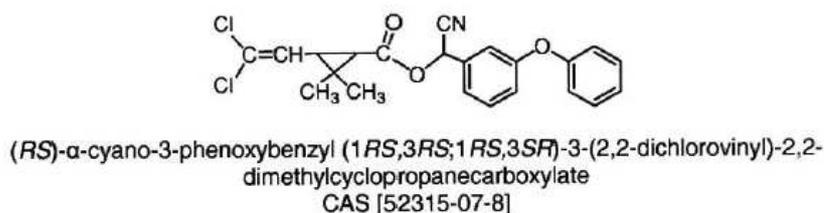


Figura 50. Estrutura química da Cipermetrina Técnica.

Métodos e Fundamentos

A forma molecular e a distribuição eletrostática desempenham um papel crucial no processo de reconhecimento molecular de padrões necessário para a descrição da similaridade molecular [65]. Dois objetos geométricos são chamados similares, se ambos tiverem a mesma forma. Mais precisamente, um é congruente com o resultado de um dimensionamento uniforme (aumentando ou diminuindo) do outro. O uso de abordagens baseadas em similaridade através dos campos moleculares para a obtenção de alinhamentos moleculares de qualidade e para a identificação de padrões em moléculas bioativas é necessário para a descrição das propriedades tridimensionais de moléculas. Os campos moleculares de interação (MIF) [66] são usados para a identificação de regiões moleculares que possuem características semelhantes. Os MIF são determinados através do uso de sondas químicas virtuais descritivas da superfície molecular cujas projeções de sobreposição resultam em regiões espaciais de estabelecida similaridade. Três sondas foram usadas neste estudo: água (OH₂), hidrofóbica – lipofilia (DRY) e oxigênio (O). Através do alinhamento das substâncias em grades cartesianas faz-se uma varredura espacial das interações e repulsões intermoleculares e identificam-se as

regiões hidrofílicas, hidrofóbicas e aceitadoras de ligação hidrogênio, respectivamente, com complementaridade de forma e energia.

O conceito fundamental empregado é que estruturas similares apresentam propriedades similares. A similaridade é descrita através de *fingerprints* (“impressão digital”)[67] e índices de similaridade[68]. No presente estudo, a similaridade de Tanimoto foi avaliada através da equação $c/(a + b - c)$, onde a é o número de bits da molécula A, b da molécula B e c é o número comum de bits entre A e B. Duas moléculas serão iguais quando o índice de Tanimoto for igual a 1, mas se nada em comum for encontrado o índice de Tanimoto será igual a 0. Por exemplo, o índice de Tanimoto da cipermetrina é 1 quando em relação a ela mesma – índice de forma. Se acrescentarmos seus átomos o índice torna-se 2 (soma da forma com a combinação de átomos de sua estrutura molecular). Esta é uma ferramenta poderosa para o reconhecimento de padrões e, portanto, é muito importante para a descrição da toxicidade de piretróides.

Análises quimiométricas baseadas em métodos de mínimos quadrados parciais (PLS)[69, 70] foram realizadas para a obtenção de componentes principais latentes em que regiões próximas no espaço bidimensional representam padrões similares de propriedades moleculares e perfil de toxicidade.

Estudos de Relação Estrutura-Toxicidade

Com o objetivo de estabelecer, em nível molecular, o perfil de toxicidade do subproduto da cipermetrina, três análises distintas foram realizadas: capacidade de interação de todas as substâncias com proteínas [71], toxicidade em canal iônico hERG (gene relacionado ao ether-à-go-go)[72] e perfil metabólico [73].

Através do emprego de métodos *in silico* (computacionais) robustos [74], pode-se concluir que, por similaridade estrutural à cipermetrina, os estudos de relação estrutura-toxicidade demonstram um perfil de toxicidade semelhante ao mesmo.

A capacidade de interação de pequenas substâncias químicas com proteínas plasmáticas é de fundamental importância porque apenas a porção livre de interação é que exibe um determinado efeito farmacológico ou toxicológico. É essa porção que pode também ser metabolizada e/ou excretada do organismo.

A Figura 51 mostra o resultado obtido para a interação com proteínas (PB). Foram usados 408 compostos cujos campos moleculares de interação intermolecular foram obtidos a partir das sondas químicas OH2, hidrofóbica (DRY) e O, para um total de 94 variáveis empregando o método dos mínimos quadrados (PLS). Duas componentes principais foram

definidas para valores de resposta no intervalo de $0\% < PB < 99\%$. A importância de se avaliar a interação com albumina sérica de humanos, como principal carreador de substâncias bioativas no organismo, está relacionada à facilidade de distribuição aos respectivos tecidos alvos através do sistema circulatório.

O modelo 2D de pontuação PLS oferece uma discriminação entre os compostos com valores altos de ligação à proteína (entre 90% e 100%) e valores baixos de ligação à proteína (de 10% para 50%). A cor do espectro indica que os pontos vermelhos referem-se a pontos de elevada ligação à proteína e azul para baixa ligação proteica.

Como pode ser observado na Figura 51, tanto a cipermetrina como seu derivado amida estão na região de elevada ligação proteica.

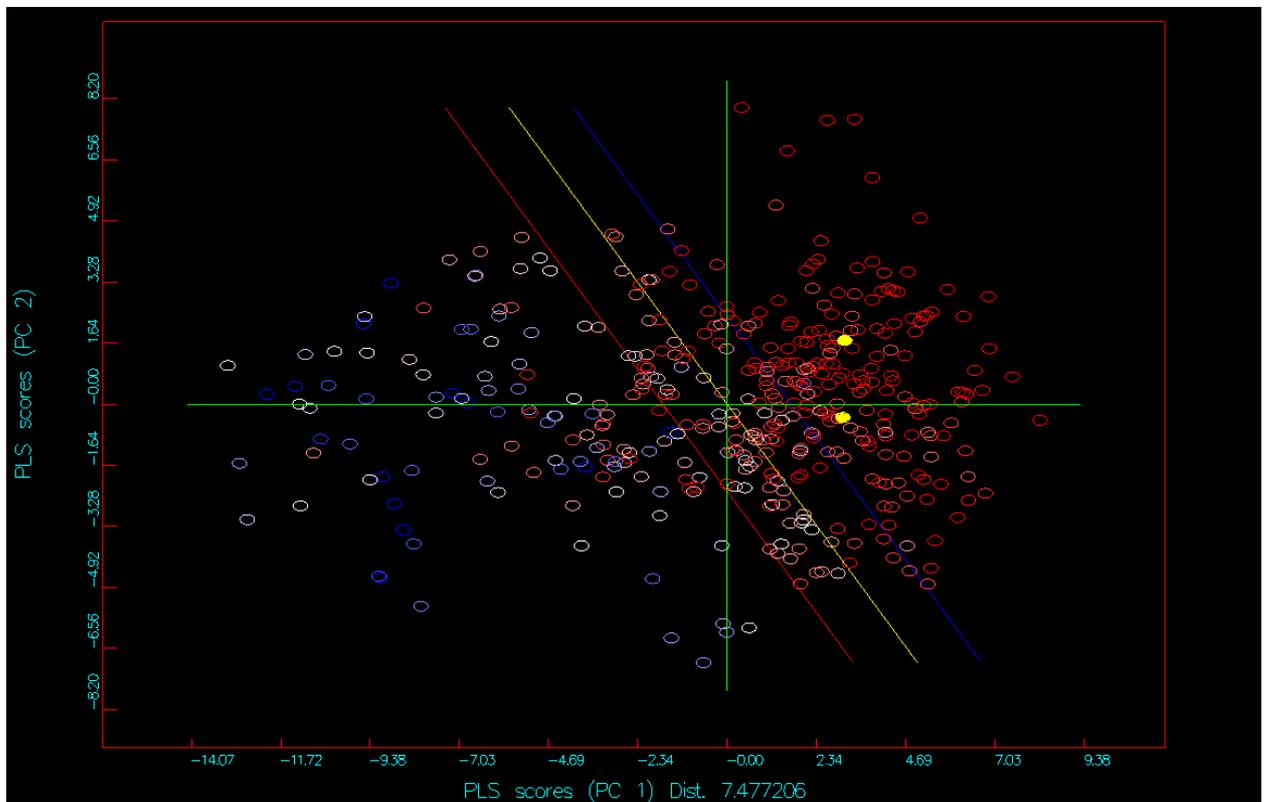


Figura 51. Interação a proteína. Em amarelo, análise qualitativa que implica serem a cipermetrina e sua impureza 1 de elevada ligação proteica.

O gene *hERG* codifica uma proteína conhecida como canal iônico de potássio. Este canal iônico é mais conhecido por sua contribuição para a atividade elétrica do coração, que coordena seu batimento (ou seja, o canal *hERG* media a corrente repolarizante de ação cardíaca potencial). Quando este canal perde a capacidade de conduzir corrente elétrica através da membrana celular, uma doença potencialmente fatal chamada síndrome de QT longo pode resultar em morte súbita. Portanto, uma potencial inibição do *hERG* deve ser evitada durante o desenvolvimento e/ou uso de medicamentos ou substâncias tóxicas. A

importância de inibição do hERG é de grande dimensão e por isso o mesmo é conhecido como anti-alvo[75].

Na Figura 52 o perfil de toxicidade em canal iônico é aceitável para os padrões de similaridade estabelecidos entre a cipermetrina e a sua impureza 1. O número de compostos usados foi de 167, cujos campos moleculares foram obtidos usando as mesmas sondas químicas: OH2, DRY, O e a mesma avaliação estatística. A resposta é observada em termos da classe de inibição do canal iônico hERG e a discriminação para os componentes da mistura é feita dentro dos seguintes parâmetros: $\text{hERG} < -0,5$ ---> área de bloqueadores do canal hERG; $-0,5 < \text{hERG} < 0$ ---> área intermediária de bloqueadores/não-bloqueadores do canal hERG; $0 < \text{hERG} < 0,5$ ---> área intermediária de não-bloqueadores/bloqueadores de canal hERG e $\text{hERG} > 0,5$ ---> para a área de não-bloqueadores do canal hERG. Portanto, tanto a cipermetrina quando seu derivado amida podem ser classificados como bloqueadores de interface aos não bloqueadores.

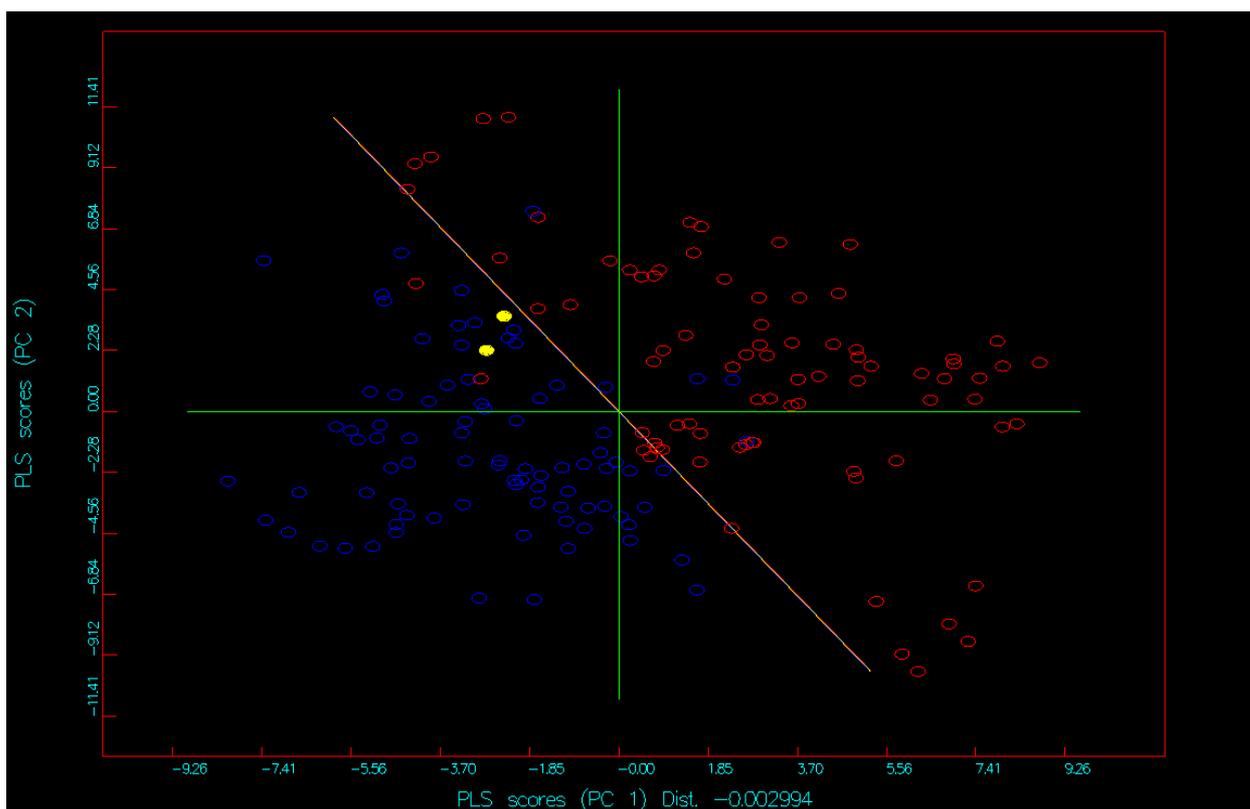


Figura 52. Modelo hERG. Os círculos em vermelho definem as regiões em que os compostos não são bloqueadores do canal de potássio hERG. Em amarelo, observam-se a cipermetrina e a impureza 1 em estudo.

A estabilidade metabólica é uma propriedade importante dos xenobióticos, uma vez que afeta parâmetros como a depuração, meia-vida e biodisponibilidade oral em diferentes espécies. Consequentemente, o conhecimento da estabilidade metabólica é essencial tanto na

descrição do perfil farmacocinético e farmacodinâmico do xenobiótico. Ela é determinada com o objetivo de avaliar o potencial tóxico do xenobiótico ou como metabolito farmacologicamente inativo devido à fase I do metabolismo ou ainda devido ao seu acúmulo no corpo devido à falta ou degradação metabólica insignificante. A determinação da estabilidade metabólica é, portanto, uma medida para descrever o destino metabólico [77].

A Figura 53 demonstra claramente que o perfil metabólico dos componentes da mistura pode ser indício de baixa estabilidade metabólica. Ou seja, através do controle de depuração, tempo de meia-vida e biodisponibilidade é possível avaliar o nível de segurança metabólica.

A estabilidade metabólica foi acessada através de banco de dados indicativos do metabolismo pela isozima CYP3A4 – o grupo mais proeminente de enzimas metabolizadoras de xenobióticos da super família de citocromo P450s [35]. Foram usados 1507 compostos com campos determinados usando as mesmas sondas já mencionadas. Os valores das respostas de estabilidade metabólica (MS) são: 1 para $MS > 40\%$ e -1 para $MS < 40\%$.

Considerando-se, que a impureza 1 da cipermetrina não foi previamente descrita quanto à sua toxicidade, realizou-se uma busca na literatura com o objetivo de identificar substâncias químicas altamente similares à impureza 1 para as quais informações toxicológicas estivessem disponíveis. Uma busca por similaridade química em níveis superiores a 80% no banco de dados SciFinder, resultou na identificação de 711 substâncias para as quais há relato de efeitos toxicológicos para apenas 11 (no. CAS: 83860-32-6, Alfametrina (83860-31-5: 72204-44-5: 72204-43-4: 67375-30-8: 66841-24-5: 65732-07-2: 65731-84-2), 66290-21-9, Cipermetrina (65731-83-1: 52315-07-8)).

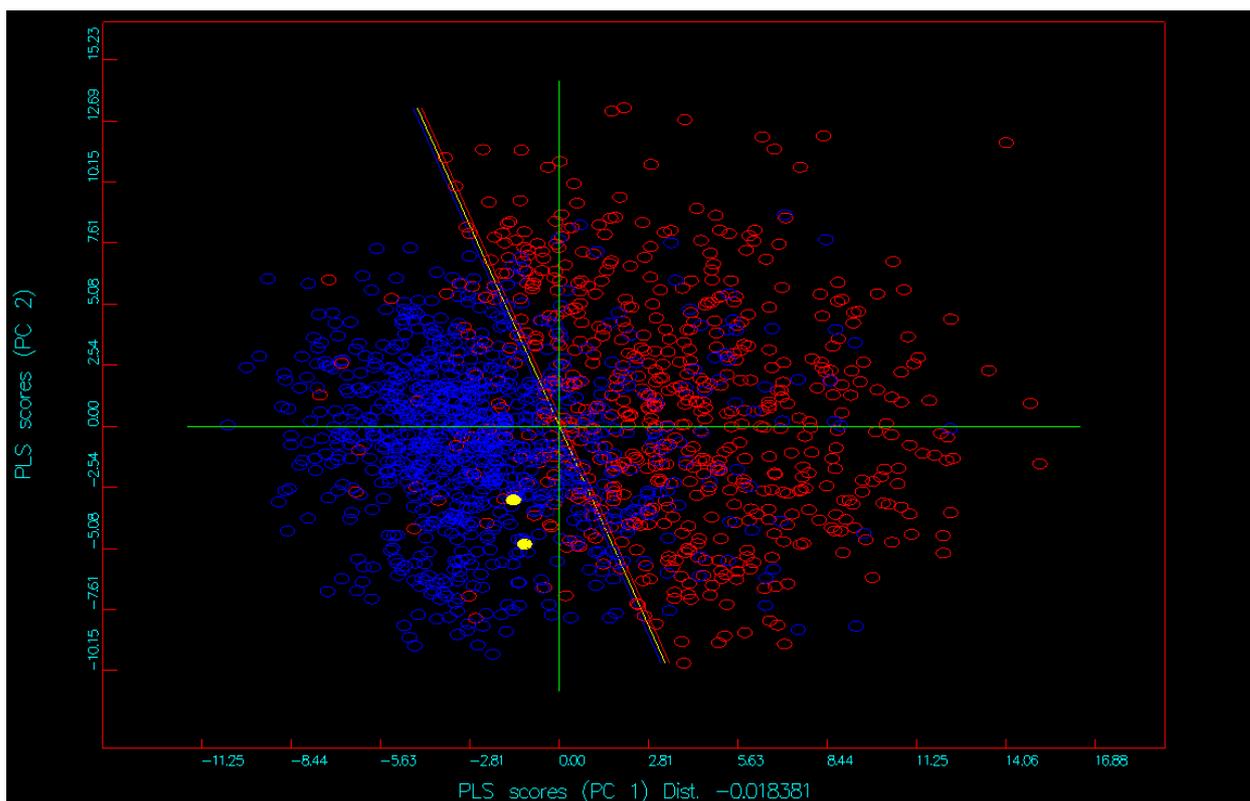


Figura 53. Estabilidade metabólica. Pontos em vermelho representam compostos estáveis; os pontos azuis indicam os compostos instáveis metabolicamente.

A projeção de piretróides sobre os espaços químicos definidos nas Figuras 54-56, identificados no respectivo quadro sob cada Figura, demonstram similaridade de agrupamento em conformidade com o mesmo perfil observado para a cipermetrina e a impureza 1 (Figuras 51-53).

Conclusões do caso 2

Neste estudo de relações estrutura-toxicidade (STR), ficou demonstrado que, por similaridade química, a cipermetrina e a impureza 1 apresentam toxicidade parecida com o produto principal – a própria cipermetrina. Os campos moleculares de interação (MIF) da cipermetrina e seu a impureza 1, avaliados por métodos quimiométricos, bem como os índices de similaridade de Tanimoto demonstram que o perfil toxicofórico é mantido em níveis semelhantes.

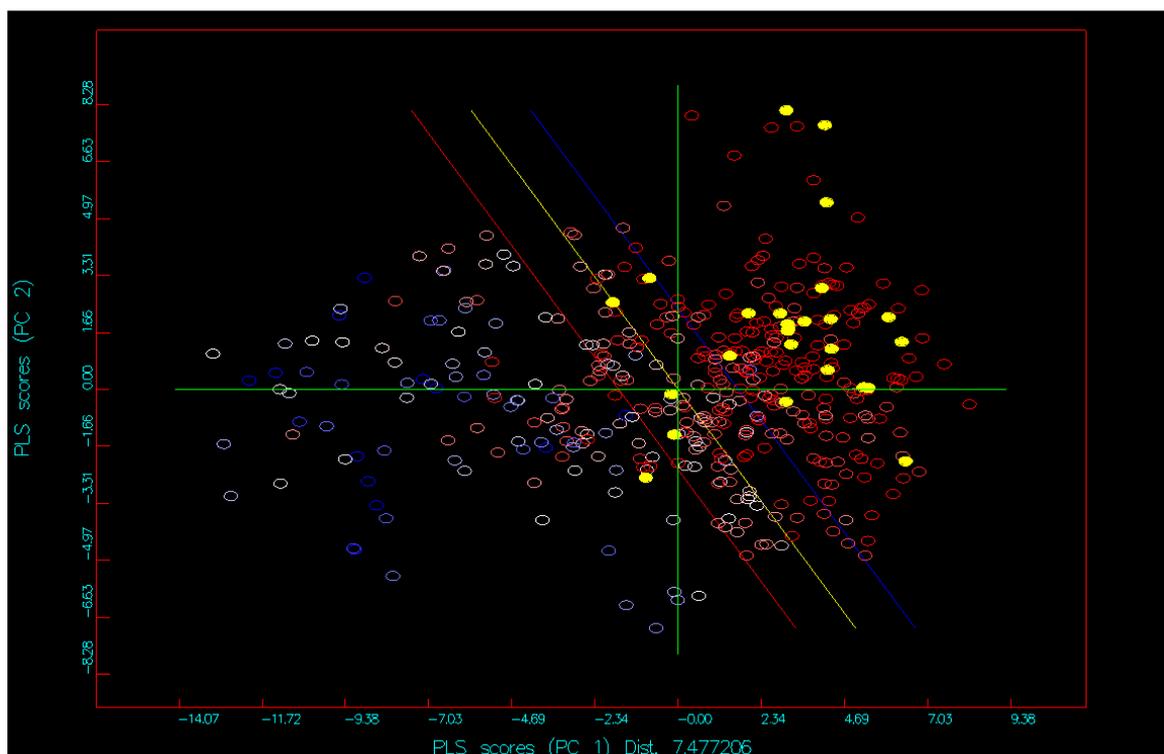


Figura 54. Interação a proteína (Albumina Sérica Humana, HSA). Substâncias do SciFinder similares à impureza 1 em níveis de agrupamento similar. Todas anotadas em amarelo.

Impureza 1	Y-LV2
Acrinathrin	117.64
Allethrin	70.04
Alpha_cypermethrin	94.14
104926	96.52
Bifenthrin	114.22
Bioallethrin	70.04
Bioresmethrin	93.05
Cyhalomethrin	94.50
Cypermethrin	93.90
Cyphenothrin	96.87
Deltamethrin	98.67
Empenthrin__1R_isomers_	84.34
Esfenvalerate	98.12
Fenpropathrin	90.61
Fenvalerate	95.25
Flucythrinate	102.15
Lambda_cyhalothrin	110.68
Permethrin	100.38
Phenothrin	94.58
Prallethrin	73.48
Resmethrin	97.70
tau_Fluvalinate	107.78
Tefluthrin	74.67
Tetramethrin	62.49
Tralomethrin	107.14
Transfluthrin	81.49
Zeta_cypermethrin	93.90

Piretróides inseticidas usados para a validação do espaço químico demonstrado pela cipermetrina e a sua impureza 1. Todos os inseticidas apresentam percentuais semelhantes de interação protéica (Y-LV2) que confirmam o comportamento similar de todos os inseticidas em estudo.

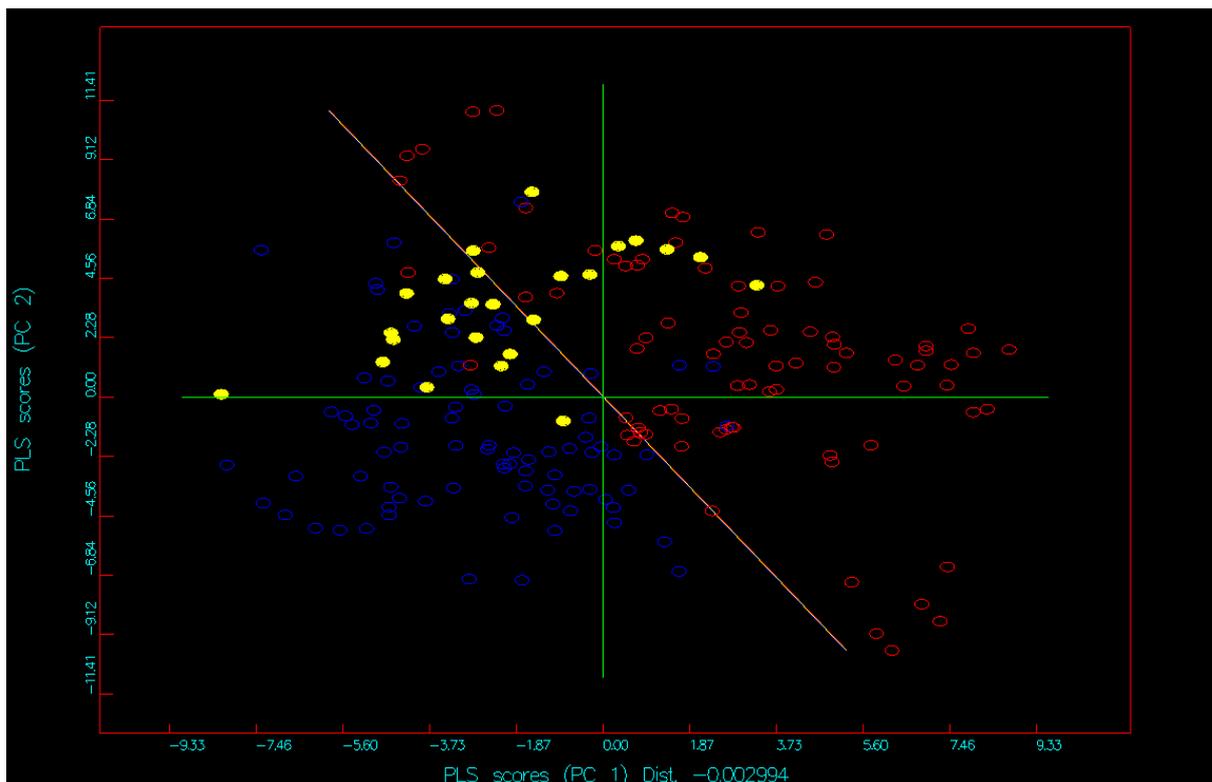


Figura 55. Modelo hERG de interação com a identificação em amarelo das substâncias químicas similares à impureza 1 da cipermetrina para as quais há relatos na literatura corrente sobre efeitos toxicológicos regulamentados por agências reguladoras.

Impureza 1	Y-IV2
Acrinathrin	-0.30
Allethrin	-0.27
Alpha_cypermethrin	0.67
104926	-0.11
Bifenthrin	-1.49
Bioallethrin	0.37
Bioresmethrin	0.67
Cyhalomethrin	0.01
Cypermethrin	-0.01
Cyphenothrin	-0.11
Deltamethrin	-0.19
Emperthrin__1R_isomers_	-0.61
Esfenvalerate	0.45
Fenpropathrin	-0.62
Fenvalerate	0.25
Flucythrinate	-0.34
Lambda_cyhalothrin	-0.23
Permethrin	-0.31
Phenothrin	-0.40
Prallethrin	-0.05
Resmethrin	0.58
tau_Fluvalinate	-0.20
Tefluthrin	-0.67
Tetramethrin	0.97
Tralomethrin	0.86
Transfluthrin	-0.75
Zeta_cypermethrin	0.76
	-0.11

Validação do espaço químico frente a vários inseticidas cujo espaço químico é semelhante àquele demonstrado pela cipermetrina e a impureza 1. A cipermetrina e sua impureza 1 estão classificadas dentro do intervalo $-0,5 < \text{hERG} < 0$ cuja área intermediária de bloqueadores/não-bloqueadores do canal hERG está dentro do esperado para essa classe de inseticidas.

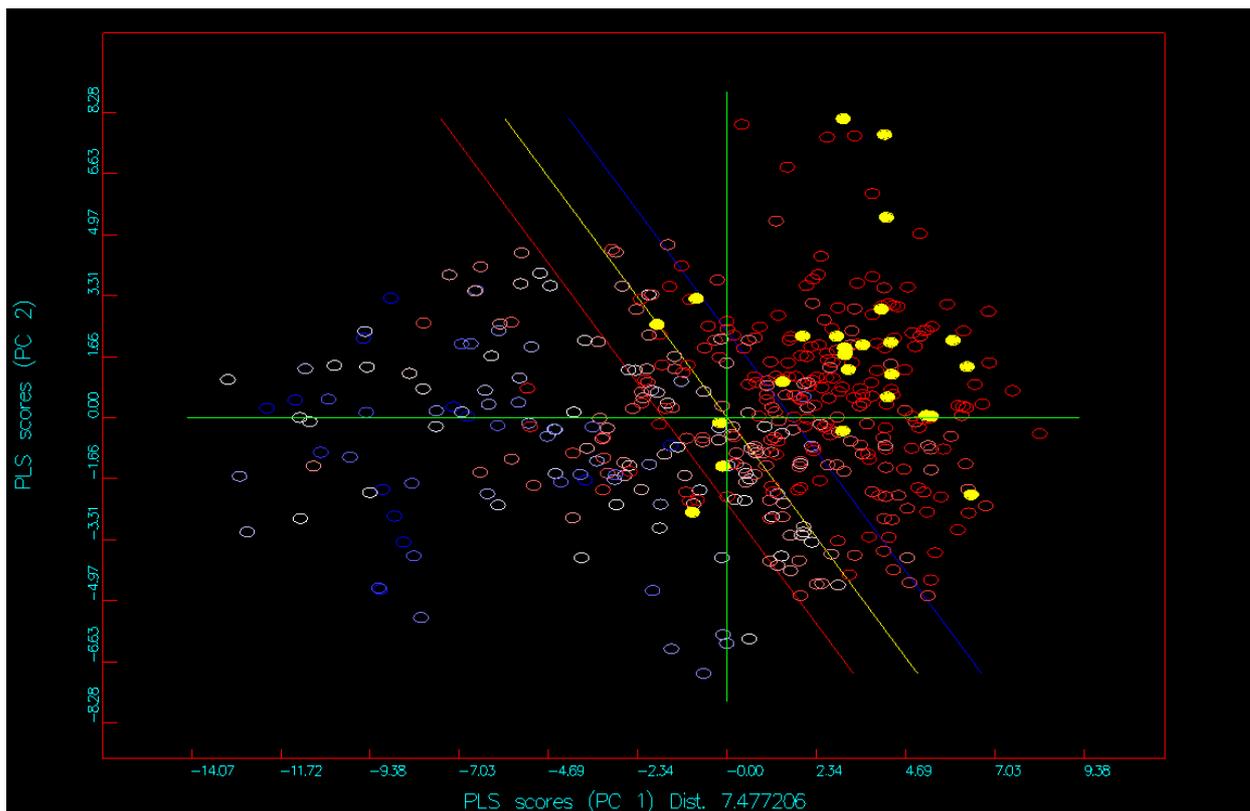


Figura 56. Perfil de estabilidade metabólica que demonstra similaridade química da cipermetrina e sua impureza 1 com agrupamento a substâncias que apresentam perfil toxicológico conhecido. Todas estão anotadas em amarelo.

Impureza 1	Y-LV2
Acrinathrin	-0.52
Allethrin	-0.63
Alpha_cypermethrin	-0.21
104926	-0.56
Bifenthrin	-0.57
Bioallethrin	-0.35
Bioresmethrin	-0.21
Cybalomethrin	-0.53
Cypermethrin	-0.63
Cyphenothrin	-0.56
Deltamethrin	-0.57
Emperthrin__1R__isomers_	-0.87
Esfenvalerate	-0.24
Fenpropathrin	-0.89
Fenvalerate	-0.35
Flucythrinate	-0.75
Lambda_cyhalothrin	-0.69
Permethrin	-0.62
Phenothrin	-0.64
Prallethrin	-0.48
Resmethrin	-0.14
tau_Fluvalinate	-0.69
Tefluthrin	-0.91
Tetramethrin	-0.03
Tralomethrin	-0.12
Transfluthrin	-1.03
Zeta_cypermethrin	-0.02
	-0.56

Validação do espaço químico de estabilidade metabólica que demonstra serem a cipermetrina e sua impureza 1 classificadas dentro do espaço descrito por inseticidas similares.

Três avaliações distintas foram realizadas: interação a proteínas, bloqueio do canal iônico hERG e estabilidade metabólica. Os resultados são coerentes e condizentes com a toxicidade da cipermetrina e vários outros piretróides que estão em uso corrente. Embora a toxicidade da impureza 1 da cipermetrina não seja conhecida observa-se, por similaridade do campo espacial que ela também está classificada na região de toxicidade conhecida.

A elevada probabilidade de ligação a proteínas e a estabilidade metabólica moderada da cipermetrina e de sua impureza 1 caracterizam seus níveis de toxicidade dentro de parâmetros aceitáveis para o nível de toxicidade da própria cipermetrina e de outros piretróides em uso corrente.

Um dos mais conhecidos e mais significativos anti-alvos é o hERG que em sendo bloqueado pode resultar em problemas cardíacos sérios ou simplesmente morte súbita. Por isso, é fundamental demonstrar que o canal de potássio hERG não é bloqueado pelos agentes químicos componentes da mistura que contém a cipermetrina. Os resultados obtidos neste estudo demonstram claramente que o nível de toxicidade da impureza 1 da cipermetrina é similar ao de várias outras substâncias para cujos efeitos toxicológicos estão descritos dentro do mesmo espaço de toxicidade demonstrado nestes estudos.

Deve ser considerada ainda a fração máxima de impureza 1 da cipermetrina que não é suficientemente elevada para classificar o produto Cipermetrina Técnico em níveis de toxidez inferiores a 500 mg kg^{-1} . Caso a dose oral em ratos (DL_{50}) fosse inferior a 500 mg kg^{-1} o nível extra de toxicidade seria observado. Como os valores de toxidez (DL_{50}) da cipermetrina estão na faixa de 250 a 4000 mg kg^{-1} em diferentes sistemas pode-se inferir que a impureza 1 da cipermetrina não é responsável por qualquer aumento de toxicidade. Sua classificação toxicológica pode, portanto, ser considerada dentro do mesmo nível de classe II, como estabelecido pela Organização Mundial de Saúde ou classe Ib para a zeta-cipermetrina. Além disso, das 711 substâncias químicas similares à impureza 1 da cipermetrina, em níveis superiores a 80%, avaliadas no SciFinder [83], somente os piretróides classificados similarmente nas Figuras 54-56 apresentaram o mesmo tipo de descrição toxicológica.

A impureza 2 por não apresentar toxicidade aguda ($DL_{50} \text{ rato/oral} = 25.000 \text{ mg kg}^{-1}$) e não estar registrado na US EPA e nem mesmo listado no PAN, e ainda por não ter uma estrutura química similar ao banco de dados empregados nestes estudos não deve ser considerado perigoso.

5.3. Caso 3: Determinação das Relações entre Estrutura e Toxicidade (STR) do Carbosulfano e suas Impurezas

Resumo

A toxicidade do Carbosulfano e suas impurezas foram avaliadas por métodos de Relações Estrutura-Toxicidade (STR) e Relações Quantitativas Estrutura-Toxicidade (QSTR). Os campos moleculares de interação (MIF) foram calculados para o ingrediente ativo Carbosulfano e as impurezas do produto técnico. Os estudos da similaridade química tridimensional por análise do índice de Tanimoto combinado demonstraram que o Carbosulfano e as impurezas são similares em níveis superiores a 1,3. Vários carbamatos com DL_{50} (oral em ratos) conhecidos, apresentam similaridades $> 1,2$. Através de métodos de reconhecimento de padrões usando PLS discriminante, o perfil toxicológico das impurezas do Carbosulfano Técnico está definido em espaço químico-biológico parecido com o perfil do próprio Carbosulfano. Os valores de DL_{50} para o Carbosulfano foram calculados via *fingerprints* moleculares e o modelo quantitativo desenvolvido é capaz de prever o valor do DL_{50} do Carbosulfano. Dessa forma, os valores calculados para as impurezas estão dentro de limites aceitáveis para a predição dos respectivos valores de D_{50} .

O Carbosulfano e as impurezas

A toxicidade aguda do Carbosulfano é considerada moderada (Classe II da WHO), mas ele pode atacar o sistema nervoso por ser inibidor da colinesterase (a DL_{50} oral do Carbosulfano em ratos é 250 mg kg^{-1}) [84]. Por isso, deve ser manuseado de acordo com as normas estabelecidas pelos Órgão reguladores.

Na Tabela 11 são apresentados os dados do *Pesticide Manual* para o produto técnico Carbosulfano.

O sistema de classificação da Organização Mundial da Saúde (WHO) é baseado na determinação da DL_{50} em ratos. Assim, um agente oral sólido, com uma DL_{50} de 5 mg ou menos por kg de peso corporal é pertencente à Classe Ia; entre $5-50 \text{ mg kg}^{-1}$ à classe Ib; entre $50-500 \text{ mg kg}^{-1}$ à Classe II, e em mais de 500 mg kg^{-1} Classe III para concentrações de 95-100 % do ingrediente ativo na formulação). Esses valores podem variar quando são empregados líquidos orais ou agentes dérmicos. Conforme dito anteriormente o valor de DL_{50} para o Carbosulfano é 250 mg kg^{-1} e, segundo a WHO pertence à Classe II.

O nível de similaridade das impurezas e os análogos do Carbosulfano técnico foram

determinados. A classificação por similaridade foi realizada com base na estrutura molecular do Carbosulfano. Os resultados são apresentados na Tabela 16.

Para o cálculo da similaridade, são necessários dois parâmetros: o descritor e o índice de similaridade. Nestes estudos, os descritores usados foram os tridimensionais de forma e átomo-combinados (combo), a partir dos descritores o índice de Tanimoto foi calculado. O valor máximo obtido é igual a 2 e indica similaridade combinada total entre a forma e os átomos que compõem a fórmula estrutural. Os dados apresentados na Tabela 16 revelam que todas as impurezas apresentam valores de similaridade superiores a 1,2 em relação à estrutura molecular do Carbosulfano (Carbosulfano, CAS 55285-14-8), o que sugere serem suas estruturas de similaridade suficientemente elevada para permitir inferir que as impurezas devem apresentar perfis toxicológicos semelhantes. Além disso, pode-se observar elevados graus de similaridade das impurezas com inseticidas e outros agrotóxicos da mesma classe. Deste modo, podem-se atribuir perfis semelhantes de toxicidade para as impurezas do produto em estudo. Os estudos de similaridade bidimensional (Tabela 17) mostram que todas as estruturas selecionadas apresentam elevados índices de similaridade. Neste caso, o valor máximo adotado para a similaridade total é 1.

Tabela 15. Dados Toxicológicos e Ecotoxicológicos do Pesticida Manual.

Toxidade Oral (ratos)	DL₅₀ dermal aguda	Irritação Ocular	Tox. Inalatória (ratos)	Tox. Abelhas	<i>Daphia Magna</i>
DL ₅₀ oral machos = 250 mg kg ⁻¹	DL ₅₀ > 2000 mg kg ⁻¹	levemente irritante	CL ₅₀ machos = 1,53 mg L ⁻¹	Tóxico	EC ₅₀ (48h) = 1,5 ug L ⁻¹
DL ₅₀ oral fêmeas = 185 mg kg ⁻¹			CL ₅₀ fêmeas = 0,61 mg L ⁻¹		

Tabela 16. Similaridade 3D combinada (Tanimoto/Combo) entre os pares de substâncias contidas no conjunto de dados: Carbosulfano (azul escuro); Impurezas (azul claro); carbamatos com DL₅₀ (oral em ratos) conhecidos (laranja), Similaridades > 1,2 (verde); ≤1,2 (amarelo).

	Carbosulfano	impureza E	impureza K	impureza L	impureza M	impureza N	82560-29-0	82560-32-5	82560-66-5	82560-71-2
Carbosulfano	2,0	1,7	1,7	1,6	1,4	1,3	1,5	1,7	1,5	1,5
impureza_E	1,7	2,0	1,6	1,5	1,4	1,3	1,5	1,6	1,3	1,7
impureza_K	1,7	1,6	2,0	1,5	1,3	1,2	1,5	1,6	1,4	1,5
impureza_L	1,6	1,5	1,5	2,0	1,5	1,3	1,5	1,5	1,3	1,5
impureza_M	1,4	1,4	1,3	1,4	2,0	1,6	1,2	1,3	1,2	1,3
impureza_N	1,3	1,3	1,2	1,3	1,6	2,0	1,2	1,3	1,2	1,3
82560-29-0	1,5	1,5	1,5	1,5	1,2	1,2	2,0	1,5	1,3	1,7
82560-32-5	1,7	1,7	1,6	1,5	1,3	1,3	1,5	2,0	1,4	1,6
82560-66-5	1,5	1,4	1,4	1,4	1,2	1,2	1,3	1,4	2,0	1,3
82560-71-2	1,5	1,7	1,5	1,5	1,3	1,3	1,7	1,6	1,3	2,0
82560-73-4	1,7	1,7	1,6	1,5	1,3	1,3	1,7	1,7	1,6	1,6
82560-75-6	1,3	1,3	1,3	1,4	1,2	1,2	1,5	1,4	1,6	1,4
86627-63-6	1,5	1,7	1,4	1,5	1,3	1,3	1,5	1,5	1,3	1,7
86627-65-8	1,5	1,7	1,4	1,5	1,4	1,3	1,4	1,4	1,5	1,6
86627-68-1	1,3	1,4	1,2	1,4	1,2	1,2	1,3	1,3	1,2	1,4

	82560-73-4	82560-75-6	86627-63-6	86627-65-8	86627-68-1
Carbosulfano	1,7	1,3	1,5	1,5	1,4
impureza_E	1,6	1,3	1,7	1,7	1,4
impureza_K	1,6	1,3	1,4	1,4	1,2
impureza_L	1,4	1,4	1,5	1,6	1,3
impureza_M	1,3	1,2	1,3	1,4	1,2
impureza_N	1,3	1,2	1,3	1,3	1,2
82560-29-0	1,7	1,5	1,5	1,4	1,3
82560-32-5	1,7	1,4	1,5	1,4	1,3
82560-66-5	1,6	1,6	1,3	1,5	1,2
82560-71-2	1,6	1,4	1,7	1,6	1,4
82560-73-4	2,0	1,5	1,5	1,4	1,3
82560-75-6	1,5	2,0	1,3	1,4	1,3
86627-63-6	1,5	1,3	2,0	1,6	1,5
86627-65-8	1,4	1,4	1,6	2,0	1,4
86627-68-1	1,3	1,3	1,5	1,4	2,0

A Tabela 17 mostra os índices de similaridade de Tanimoto em duas dimensões. Os valores próximos da unidade representam elevado grau de similaridade entre o Carbosulfano, suas impurezas e outros carbamatos usados para o desenvolvimento do modelo.

Tabela 17. Similaridade 2D combinada (Tanimoto) entre os pares de substâncias contidas no conjunto de dados: Carbosulfano (azul escuro); Impurezas (azul claro); carbamatos com DL₅₀ (oral em ratos) conhecidos (laranja), Similaridades > 0,6 (verde).

	Carbosulfano	Impureza E	Impureza K	Impureza L	Impureza M	Impureza N	82560-71-2	86627-68-1	86627-65-8	86627-63-6
Carbosulfano	1,00	0,94	0,97	0,99	0,77	0,74	0,95	0,76	0,76	0,78
Impureza E	0,94	1,00	0,93	0,93	0,73	0,70	0,93	0,77	0,77	0,79
Impureza K	0,97	0,93	1,00	0,96	0,75	0,74	0,94	0,73	0,73	0,75
Impureza L	0,99	0,93	0,96	1,00	0,76	0,74	0,94	0,76	0,76	0,77
Impureza M	0,77	0,73	0,75	0,76	1,00	0,92	0,74	0,69	0,69	0,71
Impureza N	0,74	0,70	0,74	0,74	0,92	1,00	0,71	0,67	0,67	0,68
82560-71-2	0,95	0,93	0,94	0,94	0,74	0,71	1,00	0,75	0,75	0,77
86627-68-1	0,76	0,77	0,73	0,76	0,69	0,67	0,75	1,00	1,00	0,97
86627-65-8	0,76	0,77	0,73	0,76	0,69	0,67	0,75	1,00	1,00	0,97
86627-63-6	0,78	0,79	0,75	0,77	0,71	0,68	0,77	0,97	0,97	1,00
82560-32-5	0,96	0,91	0,95	0,95	0,75	0,73	0,95	0,73	0,73	0,75
82560-75-6	0,96	0,90	0,94	0,97	0,75	0,72	0,97	0,74	0,74	0,75
82560-73-4	0,97	0,91	0,96	0,96	0,76	0,73	0,98	0,74	0,74	0,76
82560-29-0	0,96	0,91	0,94	0,95	0,75	0,73	0,98	0,73	0,73	0,75
82560-66-5	0,88	0,83	0,86	0,89	0,72	0,71	0,85	0,71	0,71	0,71

	82560-32-5	82560-75-6	82560-73-4	82560-29-0	82560-66-5
Carbosulfano	0,96	0,96	0,97	0,96	0,88
Impureza E	0,91	0,90	0,91	0,91	0,83
Impureza K	0,95	0,94	0,96	0,94	0,86
Impureza L	0,95	0,97	0,96	0,95	0,89
Impureza M	0,75	0,75	0,76	0,75	0,72
Impureza N	0,73	0,72	0,73	0,73	0,71
82560-71-2	0,95	0,97	0,98	0,98	0,85
86627-68-1	0,73	0,74	0,74	0,73	0,71
86627-65-8	0,73	0,74	0,74	0,73	0,71
86627-63-6	0,75	0,75	0,76	0,75	0,71
82560-32-5	1,00	0,96	0,97	0,96	0,86
82560-75-6	0,96	1,00	0,99	0,97	0,88
82560-73-4	0,97	0,99	1,00	0,99	0,87
82560-29-0	0,96	0,97	0,99	1,00	0,86
82560-66-5	0,86	0,88	0,87	0,86	1,00

Métodos e Fundamentos

A forma molecular e a distribuição eletrostática desempenham um papel crucial no processo de reconhecimento molecular de padrões necessário para a descrição da similaridade molecular [65]. Dois objetos geométricos são chamados similares, se ambos tiverem a mesma

forma. Mais precisamente, um é congruente com o resultado de um dimensionamento uniforme (aumentando ou diminuindo) do outro. O uso de abordagens baseadas em similaridade através dos campos moleculares para a obtenção de alinhamentos moleculares de qualidade e para a identificação de padrões em moléculas bioativas é necessário para a descrição das propriedades tridimensionais de moléculas. Os campos moleculares de interação (MIF) [66] são usados para a identificação de regiões moleculares que possuem características semelhantes. Os MIF são determinados através do uso de sondas químicas virtuais descritivas da superfície molecular cujas projeções de sobreposição resultam em regiões espaciais de estabelecida similaridade. Três sondas foram usadas neste estudo: água (OH₂), hidrofóbica – lipofilia (DRY) e oxigênio carbonílico (O). Através do alinhamento das substâncias em grades cartesianas faz-se uma varredura espacial das interações e repulsões intermoleculares e identificam-se as regiões hidrofílicas, hidrofóbicas e aceitadoras de ligação hidrogênio, respectivamente, com complementaridade de forma e energia.

O conceito fundamental empregado é que estruturas similares apresentam propriedades similares. A similaridade é descrita através de *fingerprints* (“impressão digital”) e índices de similaridade [68]. No presente estudo, a similaridade de Tanimoto foi avaliada através da equação $c/(a + b - c)$, onde a é o número de *bits* da molécula A, b da molécula B e c é o número comum de bits entre A e B. Duas moléculas serão iguais quando o índice de Tanimoto for igual a 1, mas se nada em comum for encontrado o índice de Tanimoto será igual a 0. Por exemplo, o índice de Tanimoto do Carbosulfano é 1,00 em relação à sua forma molecular e 1,00 em relação aos heteroátomos contidos na sua estrutura (“cor do átomo”), que combinados resultam no valor máximo igual a 2,00 (veja a Tabela 16). Esta é uma ferramenta poderosa para o reconhecimento de padrões de similaridade e, portanto, é muito importante para a descrição da toxicidade de inseticidas.

Análises quimiométricas baseadas em métodos de mínimos quadrados parciais (PLS) [69, 70] foram realizadas para a obtenção de componentes principais latentes em que regiões próximas no espaço bidimensional representam padrões similares de propriedades moleculares e perfil de toxicidade.

Foram construídos também, modelos quantitativos de estrutura atividade utilizando hologramas moleculares (HQSAR) para a predição da DL₅₀ teórica das impurezas. Desta forma, estabelece-se uma relação quantitativa dos compostos que compõem o Carbosulfano Técnico.

O holograma QSAR (HQSAR – Hologram Quantitative Structure-Activity Relationships) é um método de QSAR bidimensional que converte a estrutura química em

notações lineares (sequências numéricas) que codificam a estrutura bidimensional das moléculas contidas nos conjuntos de dados. Baseado no conceito fundamental da QSAR de que a toxicidade é função da estrutura molecular, é possível estabelecer relações quantitativas entre a toxicidade de uma série de moléculas a partir das informações codificadas nos descritores bidimensionais. O método é baseado no conceito de impressão digital molecular (*fingerprints* – semelhantes àqueles usados para o cálculo da similaridade química entre pares). Os hologramas moleculares são originários da fragmentação molecular bidimensional. Informações adicionais como hibridação e quiralidade também são codificadas nos hologramas moleculares. Os modelos de QSAR são desenvolvidos empregando-se o método de PLS e técnicas de validação cruzada [85].

Os parâmetros importantes para o desenvolvimento do modelo são: (i) comprimento do holograma, (ii) tamanho dos fragmentos e (iii) distinção dos fragmentos. O comprimento do holograma define o número de posições no holograma disponível para que diferentes fragmentos sejam incorporados. O tamanho de fragmentos é definido pelo número máximo e mínimo de átomos que é imposto aos fragmentos a serem gerados, enquanto a distinção de fragmentos define as características moleculares que distinguem um fragmento do outro. Os parâmetros utilizados para distinguir os fragmentos são variáveis ajustáveis e define o tipo de informação bidimensional que será incluída na criação dos descritores moleculares, estes são definidos por: (i) Átomos (A), Ligações (B), Conectividade (C), Hidrogênio (H), Quiralidade (Ch) e Doador e Aceitador (DA); A combinação destes parâmetros é utilizada para determinar a melhor correlação entre as estruturas e atividade.

A geração dos hologramas moleculares foi realizada utilizando-se várias combinações de distinção de fragmentos: AB, ABC, ABCH, ABCHCh, ABCHChDA, ABH, ABCCh, ABDA, ABCDA, ABHDA, ABCHDA, e ABHChDA. O tamanho padrão de fragmentos (4-7) foi empregado para a obtenção dos modelos iniciais; os melhores modelos foram então submetidos à variação do tamanho dos fragmentos (2-5, 3-6, 5-8, 6-9 e 7-10). Os padrões de comprimento dos hologramas foram 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 e 401 [86].

Estudos de Relação Estrutura-Toxicidade (STR/QSTR)

Com o objetivo de estabelecer, em nível molecular, o perfil de toxicidade do Carbosulfano Técnico e dos subprodutos do Carbosulfano, quatro análises distintas foram realizadas: capacidade de interação de todas as substâncias com proteínas [71], toxicidade em

canal iônico hERG (gene relacionado ao *ether-à-go-go*) [72] e perfil metabólico [73] e cálculo dos valores de DL_{50} .

Através do emprego de métodos *in silico* (computacionais) robustos [74], pode-se concluir que, por similaridade estrutural ao Carbosulfano, os estudos de relação estrutura-toxicidade demonstram um perfil de toxicidade semelhante ao mesmo.

A Figura 57 mostra o resultado obtido para a interação do Carbosulfano e impurezas com proteínas (PB). Foram usados 408 compostos cujos campos moleculares de interação intermolecular foram obtidos a partir das sondas químicas OH2, hidrofóbica (DRY) e O, para um total de 94 variáveis empregando o método dos mínimos quadrados (PLS). Duas componentes principais foram definidas para valores de resposta no intervalo de $0\% < PB < 99\%$. A importância de se avaliar a interação com albumina sérica de humanos, como principal carreador de substâncias bioativas no organismo, está relacionada à facilidade de distribuição aos respectivos tecidos alvos através do sistema circulatório.

O modelo 2D de pontuação PLS oferece uma discriminação entre os compostos com valores elevados de ligação à proteína (entre 90% e 100%) e valores baixos de ligação à proteína (de 10% para 50%). A cor do espectro indica que os pontos vermelhos referem-se a pontos de elevada ligação à proteína e azul para baixa ligação proteica.

Como pode ser observado, todos os componentes da mistura estão na região de moderada interação proteica.

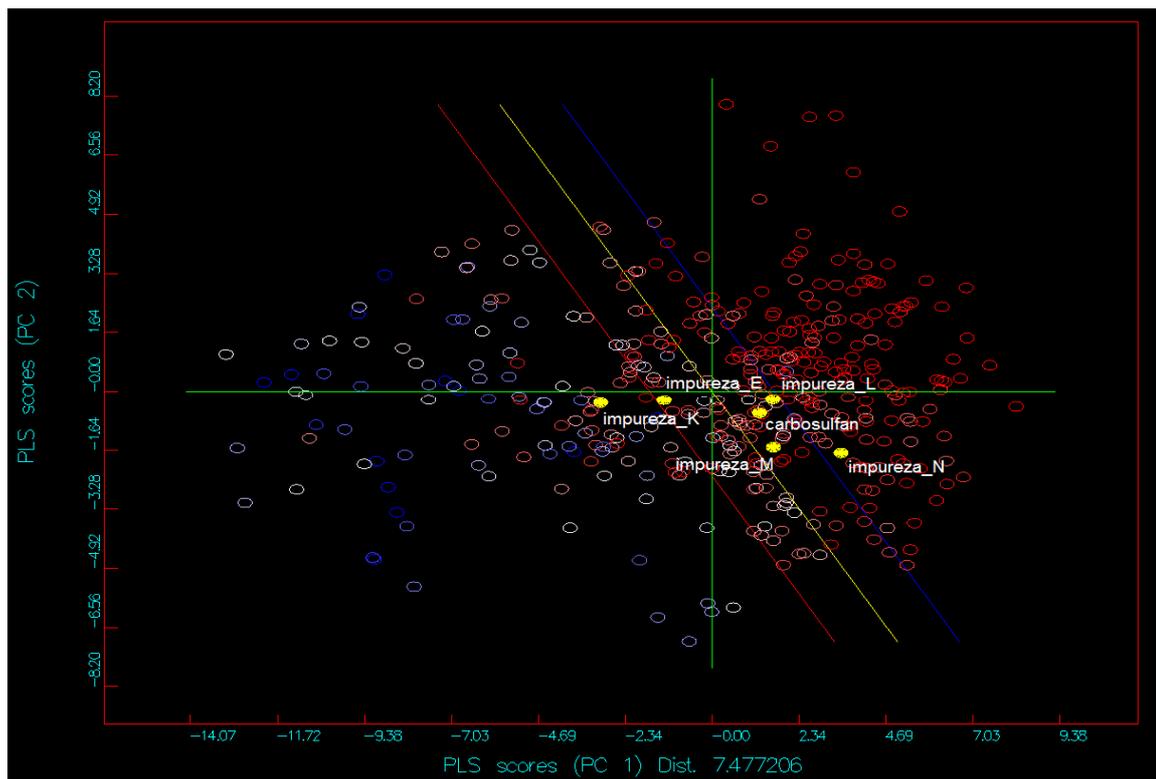


Figura 57. Análise qualitativa das interações de moléculas com a proteína. Em amarelo, substâncias com moderada ligação proteica; em vermelhos, elevada ligação proteica e em azuis substâncias que apresentam baixa ligação proteica.

O gene *hERG* codifica uma proteína conhecida como canal iônico de potássio. Este canal iônico é mais conhecido por sua contribuição para a atividade elétrica do coração, que coordena seu batimento (ou seja, o canal *hERG* media a corrente repolarizante de ação cardíaca potencial). Quando este canal perde a capacidade de conduzir corrente elétrica através da membrana celular, uma doença potencialmente fatal chamada síndrome de QT longo pode resultar em morte súbita. Portanto, uma potencial inibição do *hERG* deve ser evitada durante o desenvolvimento e/ou uso de medicamentos ou substâncias tóxicas. A importância de inibição do *hERG* é de grande dimensão e por isso o mesmo é conhecido como *anti-alvo* [75]. A Figura 58 mostra que o perfil de toxicidade em canal iônico é aceitável para os padrões de similaridade estabelecidos por todos as impurezas e pelo Carbosulfano. O número de compostos usados foi de 167, cujos campos moleculares foram obtidos usando as mesmas sondas químicas: OH2, DRY, O e a mesma avaliação estatística. A resposta é observada em termos da classe de inibição do canal iônico *hERG* e a discriminação para os componentes da mistura é feita dentro dos seguintes parâmetros: $hERG < -0.5$ ---> área de bloqueadores do canal *hERG*; $-0.5 < hERG < 0$ ---> área intermediária de bloqueadores/não-bloqueadores do canal *hERG*; $0 < hERG < 0.5$ ---> área intermediária de não-bloqueadores/bloqueadores de canal *hERG* e $hERG > 0.5$ ---> para a área de não-

bloqueadores do canal hERG. Portanto, os componentes da mistura podem ser classificados como não bloqueadores. A impureza N está na região de bloqueio potencial do canal hERG.

A estabilidade metabólica é uma propriedade importante dos xenobióticos, uma vez que afeta parâmetros como a depuração, meia-vida e biodisponibilidade oral em diferentes espécies. Conseqüentemente, o conhecimento da estabilidade metabólica é essencial tanto na descrição do perfil farmacocinético e farmacodinâmico do xenobiótico. Ela é determinada com o objetivo de avaliar o potencial tóxico do xenobiótico ou como metabolito farmacologicamente inativo devido à fase I do metabolismo ou ainda devido ao seu acúmulo no corpo devido à falta ou degradação metabólica insignificante. A determinação da estabilidade metabólica é, portanto, uma medida para descrever o destino metabólico do xenobiótico [77].

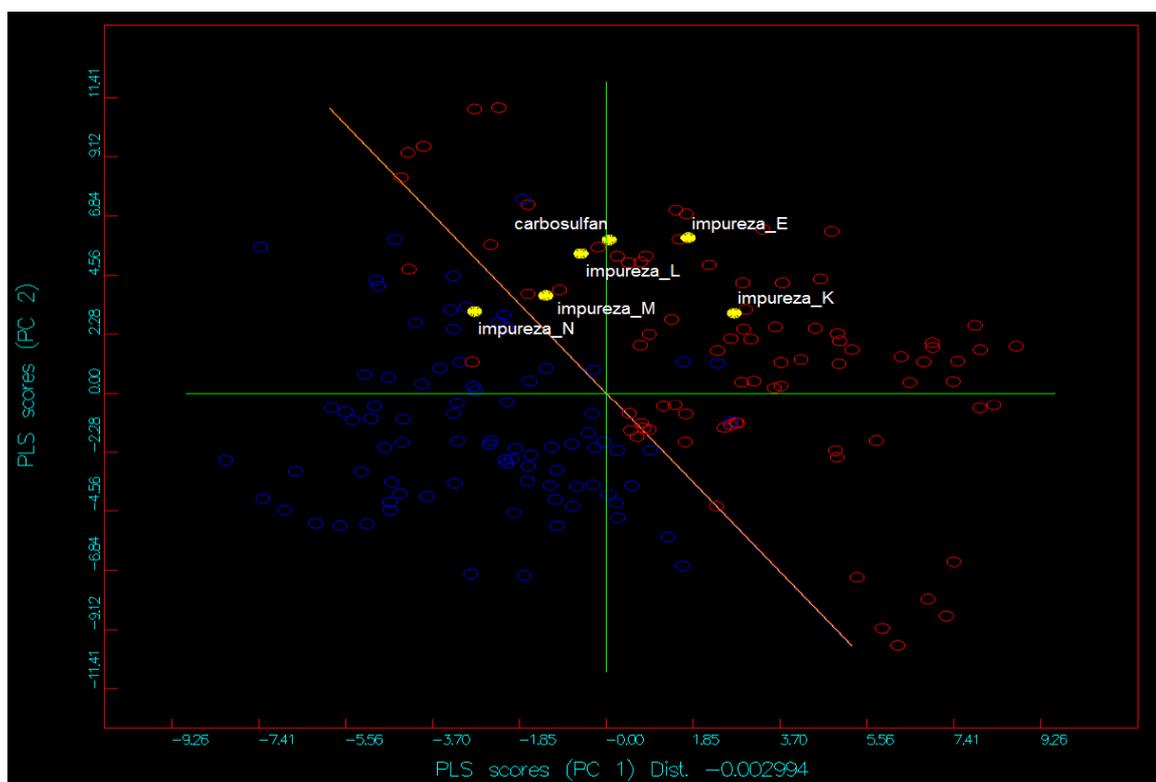


Figura 58. Modelo hERG. Os círculos em vermelho definem as regiões em que os compostos não são bloqueadores do canal de potássio hERG. Em amarelo, observam-se os componentes da mistura em estudo.

Composto	Predito (Y- LV2)	SDEP (Y- LV2)
Carbosulfano	0,55	0,03
Impureza_E	0,86	0,33
Impureza_K	0,75	0,55
Impureza_L	0,39	0,66
Impureza_M	0,10	0,57
Impureza_N	-0,23	0,50

Valores superiores a 0,1 não são bloqueadores do canal de potássio hERG. Somente a impureza N da mistura é potencialmente bloqueadora do canal de potássio hERG (círculos amarelos).

A Figura 59 demonstra claramente que o perfil metabólico das impurezas L, M e N apresentam estabilidade metabólica similar àquela demonstrada pelo Carbosulfano. As demais impurezas E e K são estáveis metabolicamente no modelo apresentado. A estabilidade metabólica foi acessada através de banco de dados indicativos do metabolismo pela isozima CYP3A4 – o grupo mais proeminente de enzimas metabolizadoras de xenobióticos da super família de citocromo P450s [35]. Foram usados 1507 compostos com campos determinados usando as mesmas sondas já mencionada. Os valores das respostas de estabilidade metabólica (MS) são: 1 para MS > 40% e -1 para MS < 40%. Portanto, os componentes da mistura são considerados dentro do perfil de elevada estabilidade metabólica que pode resultar no conceito já estabelecido para a rápida excreção do organismo sem biotransformação apreciável para eventuais metabólitos tóxicos.

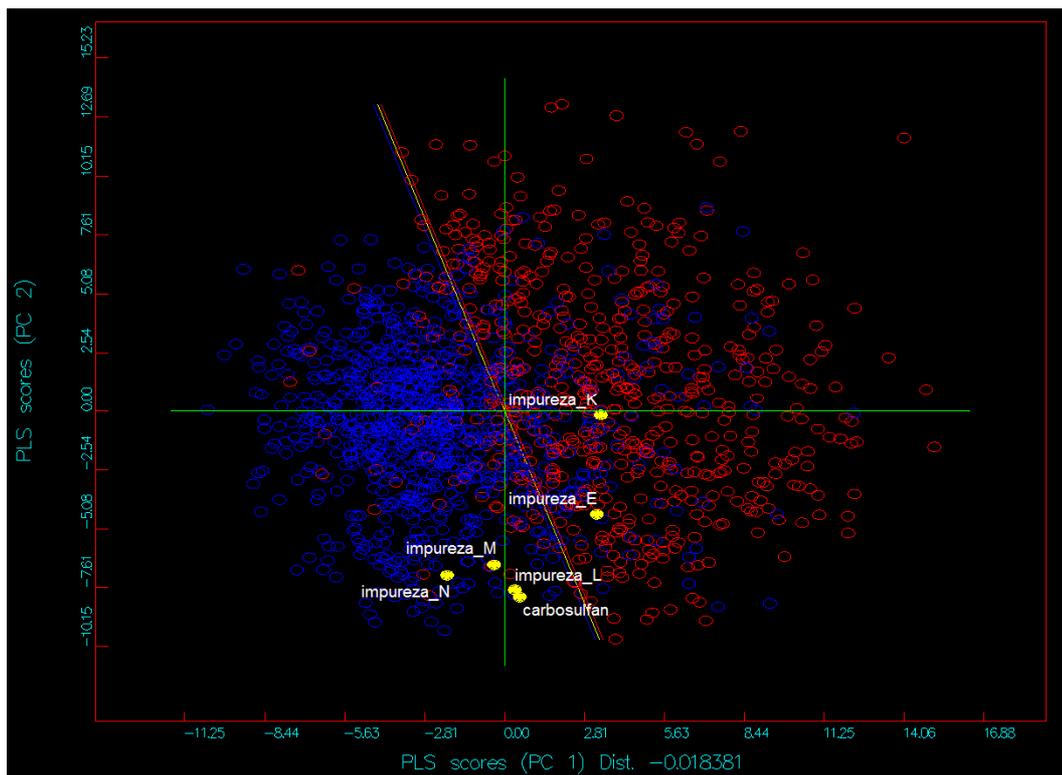


Figura 59. Estabilidade metabólica. Em amarelo estão codificadas as substâncias em estudo. As impurezas L, M e N apresentam estabilidade metabólica similar àquela demonstrada pelo Carbosulfano, enquanto que as impurezas E e K são estáveis metabolicamente.

Composto	Predito (Y- LV2)	SDEP (Y- LV2)
Carbosulfano	-0,48	0,18
Impureza_E	0,05	0,27
Impureza_K	0,27	0,49
Impureza_L	-0,49	0,68
Impureza_M	-0,54	0,66
Impureza_N	-0,79	0,63

O escore de pontuação das duas componentes principais mostra os compostos codificados por cores de acordo com a sua estabilidade metabólica (pontos vermelhos representam compostos estáveis, os pontos azuis indicam os compostos instáveis). Valores acima de zero representam substâncias estáveis metabolicamente.

Algumas das impurezas do Carbosulfano Técnico apresentam características detrimenais similares àquela já observada para o próprio Carbosulfano. Portanto, com o objetivo de melhor caracterizar suas estruturas químicas e relacioná-las com a estrutura do Carbosulfano em termos do perfil toxicológico, um estudo quantitativo necessário para o cálculo dos valores de DL_{50} foi realizado. O conjunto de substâncias empregadas para a construção dos modelos de HQSAR com os respectivos valores calculados e preditos pode ser

encontrado na Tabela 18. O resultado estatístico para o melhor modelo gerado pelos agrotóxicos é mostrado na Tabela 19.

Os resultados corroboram os cálculos de similaridade e demonstram, claramente, que os valores de DL_{50} calculados e preditos são comparáveis aos observados experimentalmente. Dessa forma, pode-se assumir que as impurezas apresentarão o mesmo nível de toxicidade do Carbosulfano.

Tabela 18. Valores preditos a partir do melhor modelo de HQSAR. Carbosulfano (azul); serie treinamento (laranja escuro); serie teste (laranja claro); valores preditos de pDL50 (verde escuro).

ID	DL_{50} experimental (mg kg ⁻¹)	Massa molecular (g mol ⁻¹)	pDL50	pDL ₅₀ predito	Desvio
Carbofuran	250	380,5	3,18	3,29	0,11
Impureza E	-	324,4	-	3,90	
Impureza K	-	381,5	-	3,68	
Impureza L	-	380,5	-	3,67	
Impureza M	-	384,6	-	2,65	
Impureza N	-	416,6	-	2,58	
82560-71-2	65	335,4	3,71	3,92	0,21
86627-68-1	340	437,6	3,11	3,66	0,55
86627-65-8	37	353,5	3,98	3,89	-0,09
86627-63-6	43	297,4	3,84	3,78	-0,06
82560-32-5	103	363,5	3,55	3,53	-0,02
82560-75-6	105	433,6	3,62	3,66	0,04
82560-73-4	110	377,5	3,54	3,70	0,16
82560-29-0	75	374,5	3,70	3,88	0,18
82560-66-5	135	452,6	3,53	3,73	0,20
Alanycarb	330	399,5	3,08	2,88	-0,20
Aldicarb	1	190,3	5,31	5,45	0,14
Bendiocarb	55	223,2	3,61	3,45	-0,16
Benfuracarb	205	410,5	3,30	3,41	0,11
Butocarboxim	158	190,3	3,08	3,50	0,42
Butoxycarboxim	288	222,3	2,89	2,73	-0,16
Carbaryl	300	201,2	2,83	2,75	-0,08
Carbetamide	10000	236,3	1,37	1,11	-0,26
Carbofuran	8	221,3	4,44	4,33	-0,12
Carbosulfano	250	380,5	3,18	3,29	0,11
Chlorpropham	5000	213,7	1,63	1,90	0,26
Ethiofencarb	200	225,3	3,05	3,20	0,14
Fenobucarb	620	207,3	2,52	2,41	-0,12
Fenothiocarb	1150	253,4	2,34	2,37	0,03
Fenoxycarb	10000	301,3	1,48	1,74	0,26
Formetanate	21	221,3	4,02	4,00	-0,02
Furathiocarb	42	382,5	3,96	3,91	-0,05
Isoprocarb	403	193,2	2,68	2,88	0,20
Methiocarb	20	225,3	4,05	3,74	-0,31
Methomyl	17	162,2	3,98	3,76	-0,22
Metolcarb	268	165,2	2,79	3,02	0,23
Oxamyl	6	219,3	4,56	4,33	-0,23
Pirimicarb	147	238,3	3,21	3,15	-0,06
Propoxur	95	209,2	3,34	2,97	-0,37
Thiodicarb	66	354,5	3,73	4,18	0,45
Thiofanox	8	218,3	4,44	4,29	-0,15
XMC	542	179,2	2,52	2,85	0,33
Xylylcarb	380	179,2	2,67	2,52	-0,16

Tabela 19. Dados estatísticos do modelo.

Distinção dos Fragmentos	HL	N	Tamanho dos Fragmentos	r^2	q^2	SEE	SEP
A/B	61	4	4-7	0,85	0,52	0,40	0,71

Conclusões do caso 3

Neste estudo de relações estrutura-toxicidade (STR/QSTR), ficou demonstrado que, por similaridade química, as impurezas do Carbosulfano apresentam toxicidade similar ao produto principal – o próprio Carbosulfano. Como o Carbosulfano está classificado em nível II de toxicidade moderada pela WHO, espera-se que suas impurezas apresentem o mesmo perfil de toxicidade e que dessa forma o Carbosulfano Técnico seja usado de acordo com as recomendações já estabelecidas para o emprego do Carbosulfano.

Quatro avaliações distintas foram realizadas: interação a proteínas, bloqueio do canal iônico hERG, estabilidade metabólica e cálculo dos valores de DL_{50} . Os resultados são coerentes e condizentes com a toxicidade do Carbosulfano.

Um dos mais conhecidos e mais significativos anti-alvos é o hERG que em sendo bloqueado pode resultar em problemas cardíacos sérios ou simplesmente morte súbita. Por isso, é fundamental demonstrar que o canal de potássio hERG não é bloqueado pelos agentes químicos componentes da mistura que contém o Carbosulfano. Os resultados obtidos neste estudo demonstram claramente que eles não são potencialmente bloqueadores do canal de potássio hERG, com provável exceção da impureza N. Não obstante, seu DL_{50} predito é o mais favorável dentre as impurezas do Carbosulfano Técnico. Apesar de sua baixa concentração (% m/m), é recomendado – como esperado para qualquer inseticida, que o Carbosulfano Técnico seja usado com o mesmo cuidado e limites de aplicação preconizados pelas agências reguladoras.

6. CONCLUSÃO

Pode-se verificar que as benzoiluréias são o grupo menos tóxico à saúde humana, uma vez que nenhum dos anti-alvos testados foram inibidos, conforme tabela 20. Analisando-se o DL50 oral como um dado toxicológico relevante, nota-se que essa classe apresenta a maior média de DL50 oral, portando menos tóxica.

Os grupos carbamatos, neonicotinóides, carbamatos e triazinas apresentaram o mesmo perfil, quando comparados os seis endpoints testados, Tabela 20.

Tabela 20. Resultados para os seis endpoints testados dos bancos: Benzoiluréias, Carbamatos, Neonicotinóides, Triazinas e Uréias.

	BBB	hERG	Estabilidade metabólica	Ligação à proteína	Permeabilidade em célula CaCO2	Volume de distribuição no plasma
Benzoiluréias						
Carbamatos						
Neonicotinóides						
Triazinas						
Uréias						

Legenda Tabela 20.

	Anti-alvos, que não foram inibidos
	Anti-alvos que foram inibidos
	Modelo insatisfatório

O método PLS foi aplicado para os bancos de dados benzoiluréias, carbamatos, neonicotinóides, triazinas e uréias e somente com a classe das benzoiluréias foi possível a obtenção de um modelo linear. Esse modelo apresenta uma pequena variação de 0,11 na escala logarítmica, sendo esse valor a diferença entre o maior o menor valor de pDL50, quando o ideal seria uma diferença de 2 na escala logarítmica. No entanto, esse modelo apresentou uma predição de pDL50 oral aguda com uma predição de 78 %. Visto que os compostos dessa classe apresentam pequena variação no seu DL50 oral, pode-se considerar que o modelo é satisfatório para a predição de pDL50 oral para compostos dessa classe e similares.

O método SIMCA usado para a separação de compostos inseticidas de herbicidas apresentou 66,4 % de capacidade de predição, que pode ser considerada satisfatória, uma vez que se tem uma grande diversidade química no conjunto de dados. A predição correta de

herbicidas ao nível de 86 % estabelece uma capacidade de acerto de oito para cada dez compostos encontrados no banco de dados.

As redes neurais artificiais apresentaram bons resultados tanto para o banco de dados de inseticidas quanto para o de herbicidas. De forma comparativa, pode-se observar que o método SIMCA é mais robusto para a predição de herbicidas, mas esse método não foi suficientemente apropriado para a descrição do espaço químico constituído pelos inseticidas.

Os resultados constantes deste trabalho sugerem a possibilidade de identificação de novos agrotóxicos com características biológicas similares àqueles dos bancos de dados ora em estudo. Entretanto, a racionalização do espaço químico-biológico permite ampliar a base de conhecimento necessária para a proposição de novos agentes com melhores características estruturais, físico-químicas e biológicas. Isto porque, a identificação do perfil toxicofórico encontrado ao longo dos modelos ora propostos, permitirá a seleção daqueles com menor toxicidade e, portanto, mais adequados para o desenvolvimento de novas substâncias com melhores perfis gerais. Ou seja, mais seguros e eficazes.

Novos modelos preditivos da potência/afinidade, estabilidade, mutagenicidade, carcinogenicidade e disruptor-endócrino dentre outros, deverão ser estabelecidos complementarmente aos presentes resultados para a descrição completa do perfil químico e biológico de um novo candidato a agrotóxico.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Disponível em: http://www.bbc.co.uk/portuguese/noticias/2009/03/090331_brasil_agricultura_ac.shtml
Acesso em: 05/11/2010.
- [2] CAMPACCI, M. H. O mercado de defensivo agrícola. *Agroanalysis FGV*; 2010.
- [3] FAO. The State of Food Insecurity in the World. In: NATION, U., ed.; 2010.
- [4] FISCHER, V. Agroquímicos. *Crop World South America*. São Paulo; 2010.
- [5] Disponível em: <http://www.andef.com.br/noticias/noticia.asp?cod=154> Acesso em: 24/10/2010.
- [6] COMMUNITY, E. Directive 86/609EC - The protection of animals used for experimental and other scientific purposes. 1986.
- [7] UNESCO. Declaração Universal dos Direitos do Animais janeiro de 1978.
- [8] GUIMARÃES, A. C. *Congresso Crop World South America*. São Paulo; 2010.
- [9] FRANK K, B. Chapter 35. Chemoinformatics: What is it and How does it Impact Drug Discovery. In: JAMES, A. B., ed. *Annual Reports in Medicinal Chemistry, volume Volume 33*: Academic Press; 1998: 375.
- [10] JOHANN, G. "The central role of chemoinformatics". *Chemometrics and Intelligent Laboratory Systems* **82**: 200; 2006.
- [11] GILLET, V. J.; LEACH, A. R. Chemoinformatics. In: EDITORS-IN-CHIEF: JOHN, B. T.; DAVID, J. T., eds. *Comprehensive Medicinal Chemistry II*. Oxford: Elsevier; 2007: 235.
- [12] WALTERS, W. P.; STAHL, M. T.; MURCKO, M. A. "Virtual screening-an overview". *Drug Discovery Today* **3**: 160; 1998.
- [13] EGLÉN, R. M.; SCHNEIDER, G.; BOHM, H.-J. High-Throughput Screening and Virtual Screening: Entry Points to Drug Discovery. *Virtual Screening for Bioactive Molecules*: Wiley-VCH Verlag GmbH; 2008: 1.
- [14] FREITAS, R. F.; OPREA, T. I.; MONTANARI, C. A. "2D QSAR and similarity studies on cruzain inhibitors aimed at improving selectivity over cathepsin L". *Bioorganic & Medicinal Chemistry* **16**: 838; 2008.
- [15] RUSH, T. S.; GRANT, J. A.; MOSYAK, L.; NICHOLLS, A. "A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein-Protein Interaction". *Journal of Medicinal Chemistry* **48**: 1489; 2005.
- [16] SUTHERLAND, J. J.; O'BRIEN, L. A.; WEAVER, D. F. "A Comparison of Methods for Modeling Quantitative Structure-Activity Relationships". *Journal of Medicinal Chemistry*

47: 5541; 2004.

[17] ERIC D, C. "Beyond physical properties—Application of Abraham descriptors and LFER analysis in agrochemical research". *Bioorganic & Medicinal Chemistry* **17**: 4153; 2009.

[18] ARNAUD, L.; TAILLANDIER, G.; KAOUADJI, M.; RAVANEL, P.; TISSUT, M. "Photosynthesis Inhibition by Phenylureas: A QSAR Approach". *Ecotoxicology and Environmental Safety* **28**: 121; 1994.

[19] NAKAGAWA, Y.; KITAHARA, K.; NISHIOKA, T.; IWAMURA, H.; FUJITA, T. "Quantitative structure-activity studies of benzoylphenylurea larvicides: I. Effect of substituents at aniline moiety against *Chilo suppressalis* walker". *Pesticide Biochemistry and Physiology* **21**: 309; 1984.

[20] SPARKS, T.; CROUSE, G.; DRIPPS, J.; ANZEVENO, P.; MARTYNOW, J.; DEAMICIS, C.; GIFFORD, J. "Neural network-based QSAR and insecticide discovery: spinetoram". *Journal of Computer-Aided Molecular Design* **22**: 393; 2008.

[21] KWAK, Y.; COPPOLA, G.; FORSTER, C. J.; GILMORE, T. A.; GONG, Y.; KANTER, A.; NEUBERT, A.; STROUP, B.; SZKLENNIK, P.; GLOWIENKE, S.; STADELMANN, P.; BELL, L.; BICKFORD, S.; GANGL, E.; GUNDUZ, M.; JAIN, M.; ZHAN, J.; SERRANO-WU, M. H. "A new structural alert for benzimidazoles: 2,6-Dimethylphenyl substituents increase mutagenic potential and time-dependent CYP3A4 inhibition risk". *Bioorganic & Medicinal Chemistry Letters* **21**: 1422; 2011.

[22] SPECK-PLANCHE, A.; KLEANDROVA, V. V.; SCOTTI, M. T. "Fragment-based approach for the in silico discovery of multi-target insecticides". *Chemometrics and Intelligent Laboratory Systems* **111**: 39; 2012.

[23] CLARK, R. D. "A perspective on the role of QSAR and QSPR in herbicide discovery". *Pest Management Science*: n/a; 2011.

[24] CRUCIANI, G.; PASTOR, M.; GUBA, W. "VolSurf: a new tool for the pharmacokinetic optimization of lead compounds". *European Journal of Pharmaceutical Sciences* **11, Supplement 2**: S29; 2000.

[25] GOODFORD, P. J. "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules". *Journal of Medicinal Chemistry* **28**: 849; 1985.

[26] WADE, R. C.; CLARK, K. J.; GOODFORD, P. J. "Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds". *Journal of Medicinal Chemistry* **36**: 140; 1993.

[27] BOOBYER, D. N. A.; GOODFORD, P. J.; MCWHINNIE, P. M.; WADE, R. C. "New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure". *Journal of Medicinal Chemistry* **32**: 1083; 1989.

- [28] CRUCIANI, G.; WATSON, K. A. "Comparative Molecular Field Analysis Using GRID Force-Field and GOLPE Variable Selection Methods in a Study of Inhibitors of Glycogen Phosphorylase b". *Journal of Medicinal Chemistry* **37**: 2589; 1994.
- [29] CRUCIANI, G.; CRIVORI, P.; CARRUPT, P. A.; TESTA, B. "Molecular fields in quantitative structure–permeation relationships: the VolSurf approach". *Journal of Molecular Structure: THEOCHEM* **503**: 17; 2000.
- [30] GARG, P.; VERMA, J. "In Silico Prediction of Blood Brain Barrier Permeability: An Artificial Neural Network Model". *Journal of Chemical Information and Modeling* **46**: 289; 2005.
- [31] PEARLSTEIN, R.; VAZ, R.; RAMPE, D. "Understanding the Structure–Activity Relationship of the Human Ether-a-go-go-Related Gene Cardiac K⁺ Channel. A Model for Bad Behavior". *Journal of Medicinal Chemistry* **46**: 2017; 2003.
- [32] Disponível em: www.moldiscovery.com Acesso em: 13/05/2011.
- [33] FUJIWARA, S.-I.; YAMASHITA, F.; HASHIDA, M. "Prediction of Caco-2 cell permeability using a combination of MO-calculation and neural network". *International Journal of Pharmaceutics* **237**: 95; 2002.
- [34] GONÇALVES, J. E. Padronização das condições para cultura de células Caco-2 visando à obtenção de membranas viáveis ao estudo da permeabilidade in vitro da rifampicina. *volume Doutorado*: Faculdade de Ciências Farmacêuticas Universidade de São Paulo; 2010.
- [35] CRIVORI, P.; ZAMORA, I.; SPEED, B.; ORRENIUS, C.; POGGESI, I. "Model based on GRID-derived descriptors for estimating CYP3A4 enzyme stability of potential drug candidates". *Journal of Computer-Aided Molecular Design* **18**: 155; 2004.
- [36] LOMBARDO, F.; OBACH, R. S.; SHALAEVA, M. Y.; GAO, F. "Prediction of Volume of Distribution Values in Humans for Neutral and Basic Drugs Using Physicochemical Measurements and Plasma Protein Binding Data". *Journal of Medicinal Chemistry* **45**: 2867; 2002.
- [37] DAOLIO, C. Aplicação de técnicas espectrométricas, cromatográficas e quimiométricas na avaliação da autenticidade de plantas utilizadas como fitoterápicos. *volume Doutorado em Química*. São Carlos: Universidade Federal de São Carlos; 2006.
- [38] LIRA, T. O. Perfil Metabólico de duas variedades transgênicas de cana-de-açúcar modificadas com os genes inibidores de proteinase Bowman-Birk e Kunitz. *volume Doutorado em Química*. São Carlos: Universidade de São Paulo; 2010.
- [39] MAS, S.; DE JUAN, A.; TAULER, R.; OLIVIERI, A. C.; ESCANDAR, G. M. "Application of chemometric methods to environmental analysis of organic pollutants: A review". *Talanta* **80**: 1052; 2010.
- [40] QIN, Y.; DENG, H.; YAN, H.; ZHONG, R. "An accurate nonlinear QSAR model for the antitumor activities of chloroethylnitrosoureas using neural networks". *Journal of Molecular Graphics and Modelling* **29**: 826; 2011.

- [41] LARS, M. "Analysis of air pollution plant exposure data: The soft independent modelling of class analogy (SIMCA) and partial least squares modelling with latent variable (PLS) approaches". *Environmental Pollution* **53**: 313; 1988.
- [42] GUPTA, V. K.; KHANI, H.; AHMADI-ROUDI, B.; MIRAKHORLI, S.; FERREYDUNI, E.; AGARWAL, S. "Prediction of capillary gas chromatographic retention times of fatty acid methyl esters in human blood using MLR, PLS and back-propagation artificial neural networks". *Talanta* **83**: 1014; 2011.
- [43] CASALE, M.; OLIVERI, P.; CASOLINO, C.; SINELLI, N.; ZUNIN, P.; ARMANINO, C.; FORINA, M.; LANTERI, S. "Characterisation of PDO olive oil Chianti Classico by non-selective (UV-visible, NIR and MIR spectroscopy) and selective (fatty acid composition) analytical techniques". *Analytica Chimica Acta* **712**: 56; 2012.
- [44] GÓMEZ-DE-ANDA, F.; DORANTES-ÁLVAREZ, L.; GALLARDO-VELÁZQUEZ, T.; OSORIO-REVILLA, G.; CALDERÓN-DOMÍNGUEZ, G.; MARTÍNEZ LABAT, P.; DE-LA-ROSA-ARANA, J.-L. "Determination of *Trichinella spiralis* in pig muscles using Mid-Fourier Transform Infrared Spectroscopy (MID-FTIR) with Attenuated Total Reflectance (ATR) and Soft Independent Modeling of Class Analogy (SIMCA)". *Meat Science*.
- [45] FLÅTEN, G. R.; GRUNG, B.; KVALHEIM, O. M. "A method for validation of reference sets in SIMCA modelling". *Chemometrics and Intelligent Laboratory Systems* **72**: 101; 2004.
- [46] A. M. JOHNSON, G. M. M. *Concepts and Applications of Molecular Similarity*. New York; 1990.
- [47] GLEN, R. C.; ADAMS, S. E. "Similarity Metrics and Descriptor Spaces – Which Combinations to Choose?". *QSAR & Combinatorial Science* **25**: 1133; 2006.
- [48] *OpenEye Scientific Software. ROCS. [Programa de computador]*. Santa Fe, New Mexico.
- [49] GRANT, J. A.; GALLARDO, M. A.; PICKUP, B. T. "A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape". *Journal of Computational Chemistry* **17**: 1653; 1996.
- [50] AZEVEDO, F. A. C., ET AL. "Equal numbers of neuronal and non-neuronal cells make the human brain an isometrically scaled-up primate brain.". *Journal of Comparative Neurology* **513**: 532; 2009.
- [51] MCCULLOCH, W. S., PITTS, W., . "A Logical Calculus of the Ideas Immanent in Nervous Activity". In: *Bulletin of Mathematical Biophysics volume 5*; 1943: 115.
- [52] RUMELHART, D. E., MCCLELLAND, J. L. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* MIT Press. 1986.
- [53] ChemAxon's products. Instant J Chem. [Programa de Computador].

- [54] *OpenEye Scientific Software. OMEGA2. [Programa de computador].* Santa Fe, New Mexico.
- [55] *Tripos Company. SYBYL. [Programa de computador].*
- [56] Disponível em: www.anvisa.gov.br Acesso em: 20/11/2010.
- [57] Disponível em:
http://ec.europa.eu/sanco_pesticides/public/index.cfm?event=activesubstance.selection
Acesso em: 20/10/2010.
- [58] Disponível em: <http://sitem.herts.ac.uk/aeru/footprint/en/> Acesso em: 10/11/2010.
- [59] Disponível em: <http://pubchem.ncbi.nlm.nih.gov/search/search.cgi> Acesso em: 15/12/2010.
- [60] MOTA, S. G. R. B., TÂNIA F.; CASTILHO, MARCELO S. "In vitro screening and chemometrics analysis on a series of azole derivatives with fungicide activity against *moniliophthora perniciosa*". *J. Braz. Chem. Soc* **21**; 2010.
- [61] CAVALCANTI, A. R. O.; SOARES LEITE, E.; NETO, B. B.; FERREIRA, R. "On the Classes of Aminoacyl-tRNA Synthetases, Amino Acids and the Genetic Code". *Origins of Life and Evolution of Biospheres* **34**: 407; 2004.
- [62] MOREDA-PIÑEIRO, A.; FISHER, A.; HILL, S. J. "The classification of tea according to region of origin using pattern recognition techniques and trace metal data". *Journal of Food Composition and Analysis* **16**: 195; 2003.
- [63] TRESADERN, G.; BEMPORAD, D.; HOWE, T. "A comparison of ligand based virtual screening methods and application to corticotropin releasing factor 1 receptor". *Journal of Molecular Graphics and Modelling* **27**: 860; 2009.
- [64] Disponível em: <http://www.epa.gov/iris/subst/0057.htm>. Acesso em: 21/06/2010.
- [65] ECKERT, H. B., J. . "Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches". *Drug Discov. Today* **12**: 225; 2007.
- [66] CROSS, S., CRUCIANI, G. . "Molecular fields in drug discovery: getting old or reaching maturity?". *Drug Discov. Today* **15**; 2010.
- [67] HERT, J. W., P.; WILTON, D.J.; ACKLIN, P.; AZZAOU, K.; JACOBY, E.; SCHUFFENHAUER, A. . "Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures". *J. Chem. Info. Comp. Sci.* **44**: 1177; 2004.
- [68] MARTIN, Y. C. K., J.L.; TRAPHAGEN, L.M. . "Do structurally similar molecules have similar biological activity?". *J. Med. Chem.* **45**: 4350; 2002.
- [69] WOLD, S. H., M.; MARTENS, H.; TRYGG, J.; WESTAD, F.; MACGREGOR, J.; WISE, B.M. . "The PLS model space revisited". *J. Chemom* **23**; 2009.

- [70] WOLD, S. E., K.; GELADI, P. "Principal component analysis". *Chemom. Intel. Lab. Syst* **2**; 1987.
- [71] COHEN, S. D. P., N.R.; KHAIRALLAH, E.A.; BOEKELHEIDE, K.; POHL, L.R.; AMOUZADEH, H.R.; HINSON, J.A. "Selective protein covalent binding and target organ toxicity". *Toxicol. Ap. Pharmacol*; 1997.
- [72] VISKIN, S. "Long QT syndromes and torsade de pointes". *Lancet* **354**: 1925; 1999.
- [73] SHEN, M. X., Y.; GOLBRAIKH, A.; GOMBAR, V.K.; TROPSHA, A. "Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates". *J.Med. Chem.* **46**: 3013; 2003.
- [74] MERLOT, C. "Computational toxicology-a tool for early safety evaluation". *Drug Discov. Today*; 2010.
- [75] SANGUINETTI, M. C. T.-F., M. . "hERG potassium channels and cardiac arrhythmia". *Nature* **440**; 2006.
- [76] PRICE, D. B., J.; JONES, L.; GREENE, N.; WAGER, T. . "Physicochemical drug properties associated with in vivo toxicological outcomes: a review". *Ex. Op. Drug Metabol. Toxicol.* **5**: 921; 2009.
- [77] BARANCZEWSKI, P. S., A.; SUNDBERG, K.; SVENSSON, R.; WALLIN, A.; JANSSON, J.;GARBERG, P.; POSTLIND, H. . "Introduction to in vitro estimation of metabolic stability and drug interactions of new chemical entities in drug discovery and development". *Pharmacol. Rep* **58**: 453; 2006.
- [78] *The Pesticide Manual: Incorporating the Agrochemicals Handbook*. 1994.
- [79] *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Richard G. Brereton. Copyright ©. John Wiley & Sons, . 2003.
- [80] RAMACHANDRA RAO, S.; RAVISHANKAR, G. A. "Plant cell cultures: Chemical factories of secondary metabolites". *Biotechnology Advances* **20**: 101; 2002.
- [81] PÉREZ-FERNÁNDEZ, V.; GARCÍA, M. Á.; MARINA, M. L. "Characteristics and enantiomeric analysis of chiral pyrethroids". *Journal of Chromatography A* **1217**: 968; 2010.
- [82] Disponível em: http://www.pesticideinfo.org/Detail_Chemical.jsp?Rec_Id=PC34291
Acesso em: 19/04/2010.
- [83] Disponível em: <https://scifinder.cas.org> Acesso em: 19/04/2010.
- [84] *The WHO recommended classification of pesticides by hazard and guidelines to classification*. 2009.
- [85] KONG, D. Z., W.; WU, D.; SHEN, X.; JIANG, H. . *J. Theor. Comp. Chem.*; 2007.
- [86] Manual do Programa SYBYL8.0- HQSAR™Manual SYBYL® 8.0. 2007.