

Universidade Estadual de Campinas
Instituto de Química

Tese de Mestrado

Estudo das relações entre a estrutura química e a atividade biológica em compostos anti-úlceras.

Edilson Grünheidt Borges

Orientador: Prof. Dr. Yuji Takahata

Campinas
1998

Estudo das relações entre a estrutura química e a atividade biológica em compostos anti-
úlcera.

Tese para obtenção de grau de MESTRE do aluno Edilson Grünheidt Borges

Banca examinadora

Presidente: Prof. Dr. Yuji Takahata

2º Examinador: Prof. Dr. Fernando Rei Ornellas (IQ-USP)

3º Examinador: Prof. Dr. Nelson Henrique Morgon (IQ-UNICAMP)

Campinas (SP), 07 de agosto de 1998.

Resumo

Investigou-se as relações entre parâmetros calculados por métodos quânticos semi-empíricos e a atividade anti-úlcer de compostos diterpenos naturais e sintéticos, e de 4-fenil 2-guanidinothiazóis com várias substituições. A geometria utilizada para o cálculo das propriedades moleculares foi obtida por análise conformacional, nos mínimos de energia potencial rotacional encontrados para os compostos. O conjunto de dados obtidos foi submetido a uma seleção de variáveis com ferramentas de análise estatística e de reconhecimento de padrões para determinar quais tinham capacidade de modelagem das atividades biológicas *in vivo* ou *in vitro*. Com estas variáveis selecionadas foram realizadas classificações dos grupos mais ou menos ativos com análise hierárquica de agrupamentos (HCA) e por análise por componentes principais (PCA); regressões multivariadas por mínimos quadrados parciais (PLS); e treinamentos de redes neurais para reproduzir os grupos de atividade e os valores de atividade determinados experimentalmente. A técnica de classificação PCA foi a que deu os melhores resultados, com até 93% de acerto do grupo ao qual pertence o composto. As classificações com redes neurais são muito boas, as previsões são razoáveis. As regressões PLS tiveram desempenho bom, tanto as atividades *in vivo* e *in vitro*. As regressões obtidas com redes neurais são complementares, com ótimas classificações e previsões razoáveis. O estudo das relações entre a estrutura química e a atividade biológica revelou que a região do grupo guanidina dos compostos guanidinothiazóis é muito importante para sua atividade biológica. As variações de carga nos átomos, e o volume dos substituintes nesta região determinam a classificação dos compostos nos grupos mais ou menos ativo nos modelos obtidos.

Palavras-chave: *métodos semi-empíricos, regressão multivariada, redes neurais.*

Abstract

It has been investigated that the relationship between calculated semi-empirical parameters and anti-ulcer activity of natural diterpenes and sintetic ones, as well as of a group of 4-fenil 2-guanidinotiazoles. The geometry employed for molecular property calculations was found by conformational analysis, corresponding to calculated minimum rotational potential energy surface of the coumpounds. The data set was submitted to variable selections using statistical analisis tools, in order to determinate which has modeling capability for biological activities *in vivo* and *in vitro*. With these variables we classified high active and low active sets by hierarquical cluster analisis (HCA), and by principal componenets analysis (PCA). Multivariated were done regressions by parcial least squares (PLS). Neural network (NN) was also employed. PCA resulted the best results, with 95% of correct classifications. The NN classifications where very good, but their predictions were reasonable. The PLS regression was also satisfactory, for *in vivo* or *in vitro* activities. The structure activity study of the guanidinotiazoles demonstrated that the guanidine region is most important to its biological activity. The charge variation in atoms near this region, or the Van der Walls volum of the substitutes of this region where important to the correct classification in the high and low activity set in the obtained models.

Keywords: semiempirical methods, multivariate regression, neural network.

Agradecimentos

É prazeroso poder agradecer o apoio dado em qualquer momento da vida.

Eu quero agradecer a todos os colegas de trabalho, professores e amigos que colaboraram comigo na execução deste trabalho. Mencionar todos os nomes é desnecessário, e certamente alguém seria esquecido.

De maneira especial, quero agradecer ao meu orientador, Prof. Yuji Takahata, que com seu trabalho e paciência, ajudou-me a tornar este trabalho muito melhor. O Prof. Dr. Yoshiyuki Hase foi muito compreensivo durante o ano em que o Prof. Takahata ausentou-se. O Prof. Dr. Pedro Vazquez colaborou muito na instalação de sistemas computacionais, e o Julio Trevas ajudou na manutenção. A Profa. Dra. Marcia Miguel de Castro foi muito gentil em colaborar na análise de dados, e com sugestões durante o exame de qualificação. Também nesta ocasião o Prof. Dr. Lauro Euclides Barata sugeriu importantes modificações. O Sergio Teixeira e a Silvana Algueiros sugeriram modificações que ajudaram na clareza e na correção gramatical do texto. A estes também sou especialmente grato.

Também sou grato às instituições que me acolheram: A UNICAMP pela utilização da sua estrutura, à CAPES e a FUNCAMP pelo financiamento da minha pesquisa, e ao CENAPAD-SP pela utilização do seu ambiente computacional.

Abreviações utilizadas

α^{dip}	-Polarizabilidade Molecular calculada por expansão do momento de dipolo.
α^{E4}	-Polarizabilidade Molecular calculada por expansão da energia total.
β^{dip}	-Hiperpolarizabilidade de primeira ordem calculada por expansão do momento de dipolo.
β^{E4}	-Hiperpolarizabilidade de primeira ordem calculada por expansão da energia total
ΔH_f°	-Entalpia-padrão de formação.
ϵ_{HOMO}	-Energia do HOMO.
ϵ_{LUMO}	-Energia do LUMO.
γ^{dip}	-Hiperpolarizabilidade de segunda ordem calculada por expansão do momento de dipolo.
γ^{E4}	-Hiperpolarizabilidade de segunda ordem calculada por expansão da energia total.
μ	-Momento de dipolo total.
μ^{hib}	-Momento de dipolo calculado com os orbitais híbridos.
μ^{car}	-Momento de dipolo calculado com cargas pontuais sobre os átomos.
Å	-Angstrom.
Å ³	-Angstrom cúbico.
ADP	-Difosfato de adenosina.
AM1	- <i>Austin Model 1</i> .
AMP	-Monofosfato de adenosina.
ATP	-Trifosfato de adenosina.
CHELP	- <i>Charge from eletrostatic potential</i> .
CI	- <i>Configuration interaction</i> .
CoMFA	- <i>Comparative molecular field analysis</i> .
CPU	- <i>Central processor unit</i> .
DNA	-Ácido desoxiribonucléico.
F_n^{HOMO}	-Densidade eletrônica em HOMO sobre o n-ésimo átomo.
F_n^{LUMO}	-Densidade eletrônica em LUMO sobre o n-ésimo átomo.
GAMESS	- <i>General atomic and molecular eletronic structure system</i> .
HAM/3	- <i>Hydrogenic atoms in molecules version 3</i> .
HCA	- <i>Hierarchical cluster analisys</i>
HF	-Hartree-Fock.

HOMO -*Highest occupied molecular orbital*

IC₅₀ -Concentração molar de substância capaz de provocar 50% de inibição enzimática.

LUMO -*Lowest unoccupied molecular orbital.*

Mb -*Megabyte*

MHz -*MegaHertz*

MINDO3 -*Modified intermediate neglect of diatomic overlap version 3.*

MM2 -*Molecular mechanics 2.*

MNDO -*Modified neglect of diatomic overlap.*

MOPAC -*Molecular orbital package.*

PC -*Principal component.*

PCA -*Principal component analysis.*

PDB -*Protein data bank.*

PLS -*Partial least square.*

PM3 -*Parameterized method 3.*

PRESS -*Predictive residual sum of squares.*

Q_n^{CHelp} -Carga calculada como derivada do potencial eletrostático no n-ésimo átomo da molécula.

Q_n^{Mull} -Carga de Mulliken no n-ésimo átomo da molécula.

QSAR -*Quantitative structure-activity relationships.*

r -Coeficiente de correlação.

RMS -*Root mean square.*

SCF -*Self-consistent field.*

SDEP -*Standard error of predictions.*

STO -*Slater-type orbital.*

UV -Unidade de variância.

VdWV -Volume de van der Waals.

VL -Variável latente.

Capítulo 1

Introdução

1.1. Motivação do projeto

O custo de desenvolvimento de uma nova droga envolve o investimento de cerca de US\$ 300 milhões e requer 12,4 anos de pesquisa, segundo informação divulgada pelo Centro de Desenvolvimento de Drogas dos EUA. Para o desenvolvimento de um novo agroquímico, cerca de trinta mil novos compostos estão sendo sintetizados e testados. Os custos de desenvolvimento de novos produtos químicos com aplicação comercial têm requerido o investimento crescente de recursos e muitos anos de pesquisa.

O Brasil pode se transformar dentro de pouco tempo num dos maiores produtores de medicamentos do mundo. A indústria farmacêutica deverá investir perto de 1,3 bilhão de dólares na modernização, ampliação e implantação de novas fábricas entre 1997 e o ano 2000. O resultado acumulado do investimento desde 1992 até o ano 2000 terá sido de 2 bilhões de dólares. Ao invés de cortar postos de trabalho, como a grande maioria das indústrias vêm fazendo nos últimos anos, a de medicamentos elevou seu nível de emprego à taxa de 9,4% entre 1992 e 1996.

Diversas multinacionais estão fechando unidades fabris na Europa e mesmo nos EUA para concentrar a produção no Brasil. As empresas montam aqui novas fábricas voltadas basicamente para exportação. Toda essa movimentação indica que o Brasil pode transformar-se num dos maiores exportadores de medicamentos do mundo.

Alguns fatores, entre outros, justificam o atual interesse de investir no Brasil: A explosão do comércio regional desde a formação do Mercosul, o tamanho do mercado interno brasileiro e as boas perspectivas de comércio exterior com a União Européia e com todas as Américas, por intermédio da Alca. Daqui os laboratórios americanos podem colocar seus produtos não só no Mercosul como também na União Européia, sem as barreiras que encontram em seu país de origem por conta da rivalidade dos blocos econômicos.

Os especialistas são unânimes em dizer que o Brasil reúne condições muito especiais, no grupo dos países emergentes, para se tornar um grande centro produtor de medicamentos. Mais do que a China e os tigres do Sudeste asiático, o país oferece – apesar dos problemas

infra-estruturais e administrativos rotulados de Custo Brasil – um conjunto variado de vantagens para os fabricantes no setor. Entre elas estão a capacidade de consumo, a localização geográfica privilegiada para exportação para o Primeiro Mundo, a inserção no bloco do Mercosul, a qualificação da mão-de-obra, a abundância de matérias primas e a existência de um parque manufatureiro razoavelmente moderno e competitivo para o fornecimento de insumos, sem contar o fato de a economia brasileira estar atravessando uma fase de estabilidade monetária como há muitos anos não se viaⁱ.

Tendo em vista esta situação, a necessidade de aumentar a eficiência de pesquisa e reduzir seus custos é indispensável desenvolver métodos ágeis, com a utilização de novas ferramentas de trabalho. Uma das opções de aumento da eficiência nas pesquisas é a utilização intensiva de informática, ou seja, modelagem molecular com ajuda de computadores, no caso do trabalho químico. O desenvolvimento de modelagem molecular ou Planejamento de Droga com Ajuda de Computador (CADD, *Computer Aided Drug Design*) está chegando ao nível de ser empregado rotineiramente em indústrias químicas e farmacêuticas na criação de novos produtos para um mercado sempre muito competitivoⁱⁱ

O estudo de relações entre a estrutura e atividade, e relações entre a estrutura e propriedade é de grande importância na química e na bioquímica modernas. O desenvolvimento de QSAR (*Quantitative Structure Activity Relationship*) e QSPR (*Quantitative Structure Properties Relationship*) está transformando a busca por compostos com determinadas características usando intuição química e experiência para uma maneira matematicamente quantificada e computadorizada. Uma vez determinada a relação entre uma estrutura ou propriedade com a atividade desejada, qualquer composto, inclusive os ainda não sintetizados, podem ser avaliados no computador para selecionar a estrutura com a propriedade de interesse. Assim é possível escolher os compostos mais promissores para síntese em laboratório.

Pretendemos trabalhar nesta área de modelagem molecular, ou CADD, pois acreditamos haver grandes possibilidades de desenvolvimento no setor, seguindo os passos do enorme salto dado pelo desenvolvimento de *hardware* e *software*, que tem tornado possível a

aplicação prática de métodos teóricos inviáveis num passado próximo em função da enorme quantidade de cálculo matemático envolvido. Ao nosso entender esta é uma área de pesquisa vital para o progresso futuro e que deve ser desenvolvida pelos grupos voltados ao interesse doméstico, sob pena de eterna dependência de políticas externas, ou aumento do atraso no setor.

Para obter boas correlações entre uma propriedade ou estrutura e a sua atividade biológica é fundamental que se utilizem descritores apropriados, indiferentemente se eles vêm de origem teórica, empírica, ou derivados de medidas experimentais facilmente obtidas das estruturas. Muitos dos descritores refletem propriedades moleculares simples, e podem ajudar na compreensão da natureza físico-química da atividade/propriedade sob consideração.

O desenvolvimento de algoritmos eficientes têm ajudado na rotina de cálculo mecânico-quântico. Novos métodos semi-empíricos resultam valores realísticos de grandezas moleculares obtidas por modelos de mecânica quântica em tempos razoáveis de computação. Assim a química quântica é uma nova e atrativa fonte de descritores que pode, a princípio, expressar todas as propriedades eletrônicas e geométricas de uma molécula, e suas interações com o ambiente, fornecendo uma visão mais detalhada e acurada dos efeitos eletrônicos que os métodos empíricosⁱⁱⁱ

1.2 Objetivo do projeto

Este trabalho tem como objetivo final o desenvolvimento de um modelo capaz de prever as alterações na atividade biológicas causadas por modificações na estrutura química de compostos, para sugerir modificações nas estruturas químicas das moléculas estudadas que aumentem sua potência farmacológica. Este objetivo final, bastante ambicioso, foi dividido em objetivos menores a serem cumpridos em cada etapa do projeto. O objetivo inicial foi a familiarização com o ambiente computacional, que usa programas, e sistemas operacionais diferentes, espalhados por diversos computadores. Os dados têm que ser transferidos inúmeras vezes entre as partes do sistema para obter os resultados finais da análise.

O segundo objetivo foi realizar todo o processo de análise conformacional, seleção de variáveis com capacidade de modelagem, e obtenção dos resultados; para um grupo pequeno de compostos, fácil de manipular e compreender o que realmente foi feito. Faz parte desta etapa o desenvolvimento de programas capazes de encontrar os dados

relevantes nos arquivos de saída de cada programa de cálculo quântico e formatar para entrada dos programas estatísticos.

O terceiro objetivo foi estender a metodologia a um grupo maior de compostos, para obtenção de relações com maior significância estatística. A série dos compostos 4-fenil-2-guanidinotiazóis foi escolhida na literatura para objeto do estudo, por ter atividade correlata aos compostos escolhidos como “piloto”. A escolha dos compostos foi orientada em função da atividade, havia disposição em continuar o estudo dos medicamentos anti-úlceras, por ser uma doença realmente muito difundida no mundo todo.

O quarto objetivo é comparar os resultados obtidos com as informações disponíveis na literatura sobre os estudos de SAR (*Structure Activity Relationship*) e QSAR para avaliar os avanços na compreensão das relações entre a estrutura química e a atividade biológica dos compostos. A utilização deste conhecimento na previsão de atividade, ou classe de atividade, em outros compostos a partir de parâmetros calculados com os mesmos métodos quânticos deve ser pensada como consequência dos resultados obtidos, num processo de realimentação entre os experimentalistas, que efetivamente produzem novos compostos e quantificam as atividades no meio biológico, e os teóricos, que podem auxiliar o processo com as conclusões do seu trabalho. O intercâmbio é imprescindível, principalmente do ponto de vista teórico, já que a inclusão de novos compostos, ou a descoberta de diferentes modos de ação pode validar, ou invalidar, relações aparentemente estabelecidas.

1.2. Estudos Preliminares

1.3.1. A úlcera e o tratamento clínico da doença

Aproximadamente cinco milhões de pessoas sofrem de úlcera péptica nos EUA, um grande problema clínico. Associado ao custo do tratamento da doença há o prejuízo na produtividade e o desconforto do doente.

Um quadro da distribuição sociodemográfica da doença foi traçado durante o ‘*National Health Interview Survey*’ em 1989, quando foi distribuído um questionário sobre doenças gastrointestinais para 41.457 pessoas escolhidas ao acaso. Os resultados mostram que 10% dos adultos residentes nos EUA têm relatos que diagnosticam úlcera, e um terço destes reportam caso de úlcera no último ano. Idade avançada, pouca instrução, baixa renda familiar e tabagismo são apontados como fatores de risco independentes. Úlcera gástrica e duodenal afetam indistintamente ambos os sexos. Úlcera do duodeno é mais frequente em indivíduos brancos que em não brancos, ocorrendo o inverso em úlcera gástrica.

O tratamento da doença geralmente é feito com um coquetel de medicamentos^{iv}. Como supressores de secreção gástrica, antagonistas do receptor H₂-histamínico,

usam-se drogas como Cimetidina (Tagamet), Ranitidina (Zantac) e Famotidina (Pepcid). Também atuam na secreção gástrica os supressores da bomba de ácido, o patamar final da produção de ácido no estômago. O FDA (*Federal Drug Administration*), órgão norte americano que controla os medicamentos, aprova o uso de Omeprazol para este fim. Para proteção da mucosa gástrica são utilizados o Sucralfato (Caralfato) e Misoprostol (Citotec), este último uma prostaglandina sintética produzida normalmente pelo organismo. Para este fim também são os medicamentos geralmente usados sem prescrição médica^v, os Antiácidos e Subsalicilato de bismuto, que também é capaz de combater também *H. Pylori*. Para combate efetivo de *H. Pylori*, até a erradicação, devem ser usados diversos antibióticos^{vi}.

1.3.2. A fisiologia celular da secreção gástrica

A fonte da secreção de íons H^+ no organismo são as células parietais, ou oxínticas. O maior passo dado nesta descoberta foi a habilidade de separar uma suspensão relativamente pura de células parietais com a utilização de enzimas que rompiam o contato entre as células para centrifugação diferencial. As evidências de que as células isoladas tinham atividade secretora foram obtidas indiretamente através da taxa de aminopirina marcada nos canalículos em relação ao meio (uma medida da acumulação de H^+ nos canalículos), da aparência morfológica característica das células secretoras vistas ao microscópio eletrônico, e o aumento do consumo de oxigênio em resposta a estimulantes de secreção de H^+ *in vivo* conhecidos.

Os estudos mais abrangentes foram realizados com células parietais caninas. Acetilcolina e histamina claramente estimulam a acumulação de H^+ , consumo de oxigênio, e adoção das características morfológicas de uma célula parietal secretora ativada. Gastrina é um estimulante fraco. Demonstrou-se, todavia, que combinações de histamina e acetilcolina, histamina e gastrina, ou gastrina e acetilcolina resultavam em maior acumulação de aminopirina que as devidas ao efeito aditivo. Posteriormente, a ação de acetilcolina pôde ser bloqueada especificamente por atropina, e a de histamina por cimetidina. Estes resultados permitiram a formulação da hipótese que haviam três receptores nas células parietais: para histamina, acetilcolina e gastrina.

1.3.3. A farmacodinâmica celular da secreção gástrica

Durante séculos a neutralização do ácido gástrico foi o único alívio para a dor causada pela úlcera. Estudos do controle fisiológico da secreção ácida demonstraram que agentes anti-colinérgicos poderiam interromper este processo. O desenvolvimento de antagonistas que agem no receptor H_2 de histamina forneceu uma classe mais específica de inibição da secreção gástrica. Mais recentemente, os benzamidazóis substituídos, inibidores da H^+, K^+ -ATPase ofereceram um meio muito efetivo de bloqueamento seletivo da bomba de prótons que é responsável pela secreção ácida nas células da parede do estômago. Apesar da secreção do ácido gástrico dominar o pensamento sobre o tratamento de úlcera péptica, a observação das maneiras pelas quais a mucosa gástrica normalmente se protege da destruição têm sugerido vários agentes citoprotetores. O reconhecimento do papel da *H. pylori* nas gastrites, contribuindo no processo ulcerativo do estômago e duodeno, teve importância para o desenvolvimento de uma estratégia no combate de úlceras e prevenção da sua recorrência.

A fisiopatologia da doença ácido-péptica pode ser pensada como um desbalanceamento entre os fatores agressivos (ácido, pepsina, infecção por *H. pylori*) e as defesas locais da mucosa (secreção de bicarbonato, muco e prostaglandinas). Apesar do tratamento geralmente se voltar para a redução dos fatores agressivos, pode-se reforçar a defesa do estômago e duodeno pela mucosa com os chamados citoprotetores.

A racionalização do uso de agentes que reduzem a acidez gástrica é melhor entendida em termos da regulação fisiológica da secreção ácida por células parietais. As três principais vias metabólicas que regulam a secreção ácida parietal são (1) estímulo neural por estimulação do nervo vago, (2) estimulação endócrina por gastrina causada por

células G antrais e (3) estimulação paracrina por liberação local de histamina das células tipo enterocromafina (ECL).

A estimulação vagal e a ação da gastrina estimulam a liberação de histamina de células ECL parácrinas. A histamina por sua vez ativa os receptores H₂ das células parietais que estão acoplados a estimulação da adenil-ciclase, causando ativação da via do AMP cíclico. Estímulos por gastrina, e muscarínicos, podem agir diretamente nas células parietais para ativar via sensíveis ao Ca²⁺. O aumento da concentração de Ca²⁺ no citoplasma, e a estimulação das células parietais por agonistas colinérgicos elevam a níveis quase máximos a secreção ácida, sem alteração do AMP cíclico. Os antagonistas do receptor H₂ bloqueiam não só os efeitos da histamina como a resposta brusca a acetilcolina e gastrina, contribuindo portanto para o excepcional efeito clínico destes medicamentos. Estas observações são consistentes com o fato da acetilcolina e a gastrina poderem estimular indiretamente as células parietais causando a liberação de histamina por células parácrinas. A ativação da via AMP-cíclico, ou Ca²⁺ dependente, ou ambas, estimula a ativação da H⁺,K⁺-ATPase nas células parietais, com sua inserção na membrana apical e formação do canal de secreção, com conseqüente taxa de secreção de H⁺ entre 20 e 40 mEq por hora. O resultado final é a acumulação de H⁺ no estômago numa concentração aproximada de 0,1N. Um aumento na permeabilidade da membrana apical para K⁺ e Cl⁻ acompanha a ativação da bomba de prótons.

Esta é a explicação para os resultados modestos de inibição dos medicamentos anti-colinérgicos, da impressionante redução da secreção ácida por antagonistas do receptor H₂, e dos efeitos da neutralização do HCl com anti-ácidos na manutenção do pH estomacal. Inibidores covalentemente ligados à H⁺,K⁺-ATPase, como omeprazol, inibem a secreção ácida, o patamar final e comum entre os mecanismos de secreção gástrica.

Prostaglandinas, por inibirem a atividade estimulante da histamina sobre a adenil-ciclase das células parietais, reduzem a atividade através da via AMP-cíclico histamina dependente citada, e portanto reduz a secreção ácida. Prostaglandinas também estimulam a secreção de muco e bicarbonato por células epiteliais superficiais adjacentes, contribuindo para os efeitos citoprotetores das prostaglandinas endógenas da série E, e dos efeitos protetores de análogos de prostaglandinas E₁ estáveis, como o misoprostol. A importância do papel tonificante das prostaglandinas na citoproteção é manifestada pela ação ulcerativa das drogas anti-inflamatórias não esteroidais (NSAIDs), inibidoras da síntese de prostaglandinas.^{vii}

Aparentemente todas as membranas têm associado pelo menos um mecanismo de transporte acoplado ao ATP. Esta generalização é estritamente verdadeira para membranas plasmáticas de eucariotos e procariotos, e parece valer para a maioria (se não a totalidade) das membranas das organelas dos procariotos. Estas 'bombas de íons' podem ser classificadas em uma de duas categorias, por qualquer um dos critérios comumente utilizados^{viii}..

Até o momento, é conhecido apenas uma classe exemplo de enzimas do Tipo-I com a função de transportar íons, a deslocadora de prótons reversível ATPase da mitocôndria (F₀F₁), dos cloroplastos (CF₀F₁), e de bactérias (B F₀F₁). Na segunda categoria, Tipo II, estão agrupadas a Na⁺,K⁺-ATPase das células plasmáticas animais, a H⁺,K⁺-ATPase da mucosa gástrica, além de outras.

1.3.5. Compostos anti-úlceras inibidores da H⁺,K⁺-ATPase

A úlcera é o resultado do desequilíbrio entre os agressores (ácido, pepsina) e os protetores (bicarbonato, muco) no estômago e no duodeno. A redução da secreção ácida, por antagonistas do receptor H₂, tem mostrado ser útil no tratamento da úlcera, particularmente àquelas da mucosa do duodeno^{ix}. Também foram identificadas drogas que suprimem completamente a secreção ácida por inibição da etapa final de bombeamento de prótons para dentro do estômago. Esta inibição leva a uma profunda e prolongada acloridria, que resulta em taxas de recuperação substancialmente mais

rápidas que aquelas obtidas com antagonistas de H₂ (2-4 semanas contra 6-8 semanas).

A enzima H^+,K^+ -ATPase catalisa o passo final da secreção ácida no estômago, onde a histamina é um dos três mensageiros chave que estimulam a secreção ácida (os outros dois são a acetilcolina e a gastrina). O resultado é que a inibição da H^+,K^+ -ATPase pode causar uma redução da secreção ácida maior que os antagonistas de H_2 . Além do mais, enquanto o receptor H_2 é encontrado em tecidos por todo corpo, a H^+,K^+ -ATPase está localizada predominantemente nas células parietais do estômago. O inibidor protótipo da H^+,K^+ -ATPase é o omeprazol na figura 1.1, que é reconhecidamente efetivo e parece ser específico para a enzima.

O omeprazol contém um grupo sulfinil na ponte que liga o substituinte benzimidazólico ao anel piridínico. É quimicamente estável e desprovido de atividade inibitória em pH neutro. Entretanto a um pH igual ou menor que 5, o composto é protonado, o que desloca o equilíbrio químico no sentido da formação de ácido sulfênico e sulfenamida, que reagem com o grupo sulfidríla da enzima H^+,K^+ -ATPase (figura 1.2). Ocorre inibição completa quando duas moléculas reagentes derivadas do omeprazol se ligam a cada molécula da

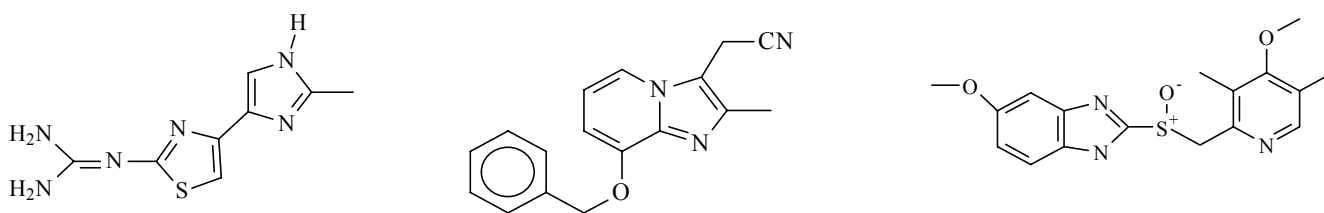


Figura 1.1: Estrutura química dos compostos denominados: Zaltidina à esquerda, SHC-28080 ao centro, e Omeprazol à direita.

enzima por ligações dissulfeto.

A distribuição seletiva da H^+,K^+ -ATPase e a necessidade de um meio ácido para a formação das formas ativas do omeprazol conferem à sua ação um alto grau de especificidade. As espécies ativas são cátions permanentes e estão concentradas no meio fortemente ácido dos canalículos da célula parietal, adjacentes à face luminal da enzima-alvo. A reação com a enzima provoca inibição permanente da atividade enzimática *in vivo*, e a secreção só recomeça após a inserção de novas moléculas de H^+,K^+ -ATPase na membrana luminal.

Porém o resultado dos estudos toxicológicos por longo período, sugerem que a inibição prolongada da secreção gástrica, resultado da inibição irreversível da H^+,K^+ -ATPase por omeprazol, pode levar a sérios efeitos colaterais como carcinoma de estômago, além de elevação do nível plasmático de gastrina. Neste caso um inibidor reversível da H^+,K^+ -ATPase seria melhor, já que poderia resultar num alívio da úlcera mais rápido que os

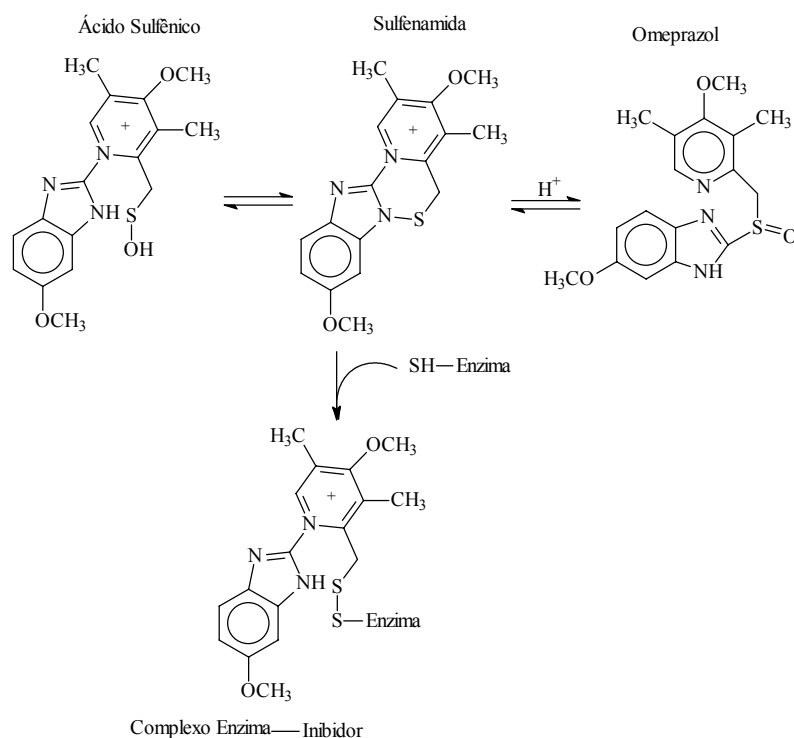


Figura 1.2: Equilíbrio de formação da espécie metabolicamente ativa a partir do Omeprazol.

antagonistas de H₂, livre da acloridria prolongada causada por inibidores irreversíveis. O composto SHC-28080 (figura 1.1) é um inibidor reversível, em fase de teste clínico, e parece causar esta ação no organismo.

Foi relatado o estudo de uma nova série de 4-imidazolil-2-guanidinothiazóis, “liderada” por zaltidina (figura 1.1), já utilizados antagonistas de H₂. Um destes compostos, 4-(2-metil-1H-pirrol-3-il)-2-(guanidino)thiazol, mostrou atividade anti-secretória comparável à da cimetidina, que é um análogo de zaltidina, sem demonstrar nenhuma atividade como antagonista H₂. O mecanismo da ação anti-secretória mostrou que se tratava de um inibidor competitivo, seletivo e reversível da H⁺,K⁺-ATPase, e abriu campo para a pesquisa do potencial dos 4-substituídos 2-guanidinothiazóis como inibidores reversíveis da enzima.

1.4. Os compostos estudados.

1.4.1. Compostos derivados de guanidinothiazóis

Os compostos analisados neste estudo têm a estrutura básica mostrada na figura 1.3, chamados de 4-fenil 2-guanidinothiazóis e 4-pirrol 2-R₁ thiazol. As atividades biológicas reportadas para esta série de compostos são mostradas na tabela 1.1, tanto para atividade *in vivo* como para atividades *in vitro*. Estes valores foram utilizados para definir os patamares em duas classes, muito ou pouco ativo.

Estes compostos já eram conhecidos como antagonistas de histamina no sítio H₂, e durante o estudo dos derivados sintéticos foi observado que a substituição do anel imidazol por pirrol causava uma grande mudança no mecanismo de ação. A partir daí iniciou-se o estudo do potencial desta classe de compostos como inibidores reversíveis da H⁺,K⁺-ATPase.

Após a descoberta inicial, tentou-se estabelecer a natureza cinética da inibição enzimática. Experimentos com lixiviação mostraram que se tratava de inibição reversível. Por meio de um gráfico de Lineweaver-Burk ficou demonstrado que estes compostos agiam competitivamente no sítio do potássio da H⁺,K⁺-ATPase.

Estudos de SAR para esta os 4-fenil 2-guanidinothiazóis foram realizados por LaMattina e colaboradores, responsáveis pela síntese dos compostos. Foram consideradas três partes da molécula para efeitos de estudo. A guanidina, o grupo substituinte na posição C(4) do thiazol, e o grupo thiazol.

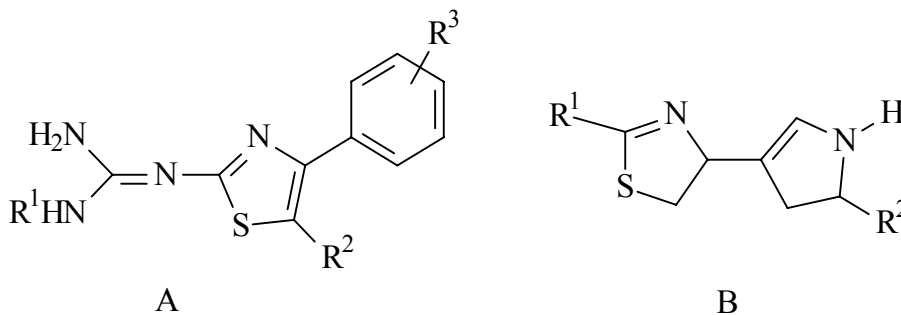


Figura 1.3: Esqueleto fundamental da estrutura dos compostos guanidinothiazóis: ‘A’ para 48-91, ‘B’ para 1-6.

A substituição do grupo guanidina por

um grupo metil ou amino causa uma completa perda da atividade (os compostos em que esta substituição foi feita não fizeram parte do estudo realizado durante meu trabalho de mestrado). Com diversos substituintes na guanidina analisados, verificou-se que a monossustituição leva a um aumento da potência, e que a substituição adicional no mesmo nitrogênio, ou no outro nitrogênio, resultaram compostos sem nenhum aumento da potência em relação aos monossustituídos.

A segunda parte examinada da molécula foi o substituinte na posição C(4) do thiazol. A única indicação encontrada é que este substituinte não deve ser básico. O estudo de SAR mostrou que a capacidade de formar ponte de hidrogênio do grupo NH do anel pirrol é relativamente pouco importante para atividade inibitória. Foram testados uma grande

variedade de substituintes no anel pirrol, que resultaram em modestos aumentos na atividade. A substituição no nitrogênio do anel pirrol com um grupo fenilsulfonil melhora a atividade inibitória em todos os compostos da série de 4-pirrol-2-guanidino-2-tiazóis. A combinação de benzilação da guanidina com N-fenilsulfonação resulta em compostos que são menos ativos que seus familiares. Portanto estes melhoramentos não são aditivos na atividade do composto. A segunda série sintetizada, os 4-fenil-2-guanidino-2-tiazóis foi o objeto de pesquisa do meu projeto de mestrado. Estes compostos não permitiram bons resultados em estudos de SAR, com a metodologia aplicada então pelo grupo responsável por sua síntese. Nenhum padrão foi conseguido em relação à natureza estérica ou eletrônica dos substituintes. A única conclusão obtida foi que os compostos substituídos com *orto*-fenil derivados são mais ativos que os derivados *meta*- e *para*-. Esta foi uma grande motivação para que os estudos fossem iniciados com esta série de compostos.

A terceira porção da molécula que foi investigada é o anel tiazol. Imaginaram que servia como mero espaçador entre a guanidina e o 4-*aril* substituinte. Todavia a sua substituição por um grupo etil fez com que houvesse quase nenhuma atividade do produto.

Assumindo que o grupo espaçador devesse ser planar, foi tentada a substituição por fenil, que resultou também na perda quase total da atividade. A conclusão foi a de que o nitrogênio do grupo tiazol deve ter um papel importante na ligação com o sítio ativo, ou que a variação do pK_a do sistema causada pela modificação seja importante em algum momento.

Foram testados análogos com maior semelhança estrutural. Com a utilização de um oxazol no lugar de tiazol, nenhuma atividade foi observada. A constatação que o grupo tiazol não é essencial para atividade vem do fato que a substituição deste grupo por cloropirimidina resulta num composto até mais ativo que o correspondente tiazol.

Infelizmente a generalização da troca por cloropirimidina não é geralmente bem sucedida, e a preferência recai sobre os tiazóis.

Tabela 1.1: Atividades anti-úlceras dos 4-fenil-2-guanidino-2-tiazóis.

Atividade Inibitória na H^+, K^+ -ATPase gástrica para os compostos com base da figura 1.3-A						Atividade Inibitória na H^+, K^+ -ATPase gástrica para os compostos com base da figura 1.3-A					
no.	R1	R2	R3	% inibição à 50 μ M	IC ₅₀ , μ M	no.	R1	R2	R3	% inibição à 50 μ M	IC ₅₀ , μ M
48	H	H	H	52	40	70	H	H	4-NH ₂	63	32
49	Bn	H	H	-1		71	H	H	4-NMe ₂	75	14
50	H	H	2-Cl	71	18	72	Bn	H	4-NH ₂	77	5,2
51	H	H	3-Cl	48	53	73	H	H	2-OMe	70	27
52	H	H	4-Cl	48		74	H	H	3-OMe	48	60
53	H	H	4-F	61	33	75	H	H	4-OMe	54	35
54	Bn	H	2-Cl	99	6	76	Bn	H	2-OMe	99	2,3
55	Bn	H	3-Cl	80		77	Bn	H	3-OMe	85	10
56	Bn	H	4-Cl	38		78	pentil	H	3-OMe	25	
57	Bn	H	4-F	35		79	Bn	H	4-OMe	24	
58	H	H	2-Me	78	20	80	H	H	3,4-(OMe) ₂	41	
59	H	H	2-CH ₂ NMe ₂	7		81	Bn	H	3,4-(OMe) ₂	25	
60	H	H	3-Me	60	32	82	H	H	2-OH	26	
61	H	H	3-CH ₂ NMe	56	37	83	H	H	3-OH	69	23
62	H	H	4-Me	61		84	H	H	4-OH	75	16
63	H	H	4-Ph	25		85	Bn	H	1-OH	85	11
64	Bn	H	2-Me	111	10	86	Bn	H	3-OH	8	
65	Bn	H	3-Me	53		87	Bn	H	4-OH	49	
66	Bn	H	4-Me	16		88	H	H	3,4-(OH) ₂	100	1,5
67	Bn	H	4-Ph	5		89	Bn	H	3,4-(OH) ₂	111	0,3
68	H	H	3-NH ₂	45		90	H	Me	3,1-(OH) ₂	97	1,6

69	H	H	3-NMe ₂	52	91	Bn	Me	3,4-(OH) ₂	103	1,3
Atividade Inibitória na H ⁺ ,K ⁺ -ATPase gástrica para os compostos com base da figura 1.3-B.					Atividade Inibitória na H ⁺ ,K ⁺ -ATPase gástrica para os compostos com base na figura 1.3-B.					
no.	R1		R2	% inibição, à 50 uM	no.	R1		R2	% inibição, à 50 uM	
1	(H ₂ N) ₂ C=N		Me	70	2	Me		Me	0	
3	NH ₂		Me	17	4	(H ₂ N) ₂ C=N		CHO	44	
5	Me		CHO	5	6	NH ₂		CHO	9	

Os inibidores da H⁺K⁺-ATPase mais potentes foram testados *in vivo* em camundongos. A correlação entre as atividades *in vivo* e *in vitro* não foi observada. Alguns dos compostos com a maior atividade *in vitro* na série do 4-fenil-2-guanidinothiazóis exibiram pouca atividade *in vivo*. Especula-se que pouca absorção, rápida eliminação, ou metabolização sejam os fatores responsáveis por isso.

Também foram encontrados publicados resultados em estudos de SAR para estes compostos com uso de índices de conectividade e topológicos^{xi}. Os autores utilizam o índice topológico de Wiener que é uma medida do grau de compactação da molécula. O valor 'W' é obtido como a soma dos elementos da matriz de distâncias nucleares, $W(G) = \frac{1}{2} \sum_{i,j} D_{ij}(G)$, onde $D_{ij}(G)$ representa os elementos fora da diagonal da matriz de distâncias nucleares. Também usam o índice de conectividade de Randic, que é definido $\chi = \sum_{ij} (V_i V_j)^{-0.5}$, onde V_i e V_j são os ângulos nos vértices em todas as arestas ij para todos os átomos das moléculas. Os resultados de classificação *in vitro* para os 44 compostos 4-fenil-2-guanidinothiazóis não classificam corretamente 17 deles. Para os resultados *in vivo* de 24 compostos, apenas 3 são erradamente classificados. Este tipo de descritor não fornece nenhuma indicação sobre qual estrutura é responsável pela variação no índice.

1.4.2. Compostos anti-úlceras diterpenos

O estudo começou com uma série de compostos homólogos provenientes do extrato fracionado de diversas espécies vegetais do grupo *Euphorbiaceae* que têm demonstrado atividade biológica em estudos *in vivo* e *in vitro*.

O extrato do caule de *Croton sublyratus* Kurz foi identificado como inibidor de úlcera induzida por reserpina em camundongos e úlcera Shay em ratos. O fracionamento do extrato guiado pela atividade anti-úlceras levou a identificação do 18-hidroxigeranilgeraniol como o princípio anti-úlceras induzida por reserpina (figura 1.4) em camundongos e uma série de furanoditerpenóides como os princípios ativos anti-úlceras Shay em ratos^{xii}

A nomenclatura sistemática para estes compostos é apresentada a seguir: (Plaunol A: 11aH-3a,9-Ethano-9H,10H-difuro[2,3-c:3',4'-i][2]benzopyran-7(3bH)-one,2-(3-furanyl)-2,3,4,5-tetrahydro-10-hydroxy-13-methylene-, [2S-(2.alpha.,3a.alpha.,3b.alpha.,9.alpha.,9aR*,10.alpha.,11a.alpha.)]- [69748-99-8](figura 3.5) ; Plaunol B: Spiro[furan-3(2H),6'-[6H]naphtho[1,8-bc]furan]-2,2'(4'H)-dione, 5-(3-furanyl)-4,5,5',5'a,7',8',8'a,8'b-octahydro-8'b-(hydroxymethyl)-7'-methylene-, [5'aS-[5'a.alpha.,6'.beta.(R*),8'a.beta.,8'b.beta.]]- [69749-00-4](figura 3.5); Plaunol C: Spiro[furan-3(2H),6'-[6H]naphtho[1,8-bc]furan]-2,2'(4'H)-dione, 5-(3-furanyl)-4,5,5',5'a,7',8',8'a,8'b-octahydro-4'-hydroxy-8'b-(hydroxymethyl)-7'-methylene-, [4'S-[4'.alpha.,5'a.alpha.,6'.beta.(R*),8'a.beta.,8'b.beta.]]- [69749-01-5](figura 3.5); Plaunol D: 1,7-Ethano-1H,10H-furo[3,4-i][2]benzopyran-3(5H)-one, 7-[2-(3-furanyl)-2-hydroxyethyl]-6,6a,7,8-tetrahydro-5,8-dihydroxy-11-methylene-, [1R-[1.alpha.,5.alpha.,6a.alpha.,7.alpha.,7(S*),8.beta.,10aR*]]- [66302-50-9](figura 3.3); Plaunol E: 1,7-Ethano-1H,10H-furo[3,4-i][2]benzopyran-3(5H)-one, 7-[2-(acetyloxy)-2-(3-furanyl)ethyl]-6,6a,7,8-tetrahydro-5,8-dihydroxy-11-methylene-, [1R-[1.alpha.,5.alpha.,6a.alpha.,7.alpha.,7(S*),8.beta.,10aR*]]- [69749-02-6])(figura 3.2).

Estudos em andamento no laboratório de bioquímica da UNICAMP conduzidos por Patrícia S. Melo, aluna de doutorado da Prof.^a Dra. Alba Regina S. Brito, têm demonstrado que compostos semelhantes aos encontrados no caule de *Croton sublyratus* Kurz, extraídos de *Croton cajucara* apresentam atividade anti-úlceras Shay da mesma ordem de grandeza, através da mesma metodologia de quantificação ^{xiii}, como pode ser observado na tabela 1.2. Testes farmacológicos indicam que a atividade é causada pela depressão de secreção gástrica nos ratos.

Em busca de aumento da atividade biológica dos compostos obtidos, derivados da t-crotonina (t-Crotonin: Spiro[furan-3(2H),1'(7'H)-naphthalene]-2,7'-dione, 5-(3-furanyl)decahydro-2',5'-dimethyl-, [1'R[1'.alpha.(R*),2'.alpha.,4'a.alpha.,5'.alpha.,8'a.beta.]]- [123357-12-0]) têm sido preparados. O composto t-desidrocrotonina (figura 3.5) é um 19-nor-clerodano diterpeno, já testado e apresentando atividade anti-úlceras, sendo previsto para breve o estudo da atividade dos derivados sintéticos.

Os resultados obtidos no seguinte trabalho poderão ser usados como auxiliar na escolha de novos derivados a serem preparados e testados em sua atividade biológica por P.S.

Melo, já havendo contato entre os grupos neste sentido.

Croton cajucara Benth, vulgarmente conhecida como Sacaca, é uma árvore de 4 a 6 metros de altura, nativa da Amazônia, com centro de dispersão no estado do Pará. É muito conhecida pelos índios e seu nome indígena significa feitiço. É usado na medicina popular na forma de um chá hipoglicemiante contra diarreia e diabetes ^{xiv}.

Há relatos na literatura sobre a atividade dos diterpenos no

crescimento de insetos em experimentos realizados com *Pectinophora gossypiella* ^{xv}. Também foram encontrados relatos sobre a atividade antiinflamatória em experimentos realizados com indução de inflamação por teleocidina em orelha de camundongo ^{xvi}. Além destas foi encontrado relato da atividade antiinflamatória de "Sangre de Grado", o látex proveniente de *C. draconoides* ou *C. lechlesi*, originárias da Amazônia peruana, cujo princípio ativo isolado, conhecido como Taspine, tem estrutura química com alguma semelhança aos furanoditerpenóides extraídos de *C. sublyratus* Kurz ^{xvii}.

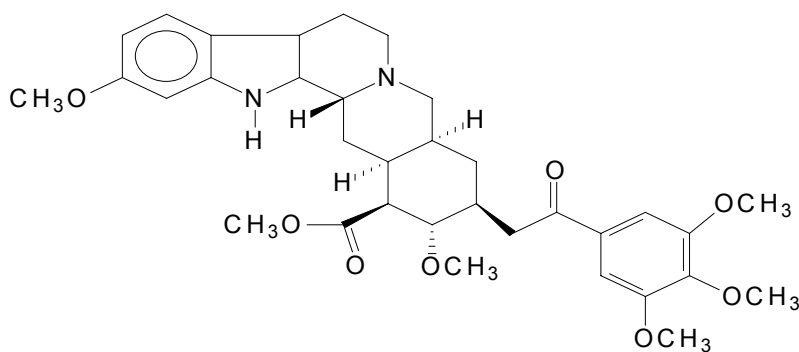


Figura 1.4: Estrutura química do composto denominado Reserpina.

A t-crotonina é presente em grande quantidade na casca de *C. cajucara*, que é uma espécie abundante na Amazônia. Por estes motivos sua obtenção por extração do produto natural é simples e barata, constituindo portanto um bom produto de partida para síntese de derivados.

A atividade depressora da secreção gástrica é uma característica muito desejada em um composto químico, tendo em vista o grande número de medicamentos para este fim disponíveis no mercado e a popularidade da sua aplicação. De fato, doenças no trato gástrico em todos os níveis de gravidade afetam diariamente o bem estar de milhares de pessoas no mundo todo. Apenas a pequena parcela dos casos mais graves recorre a tratamento médico nestas moléstias. A grande parcela dos casos sem gravidade e que causam apenas desconforto do paciente preferem a auto medicação, com todos os riscos inerentes à prática.

Tabela 1.2: Atividade biológica medida experimentalmente para a série de furanoditerpenóides.

COMPOSTO	PERCENTUAL DE INIBIÇÃO	
	Dose 10mg/Kg	Dose 3mg/Kg
Desidrocrotonina	52	Não determinada.
Plaunol A	0	0
Plaunol B	85	55
Plaunol C	88	36
Plaunol D	61	44
Plaunol E	82	52

Capítulo 2

Procedimentos e métodos adotados

2.1. Busca da conformação de mínima energia

A primeira etapa do planejamento de drogas, com ajuda de computador, é a determinação da geometria de menor energia de um dos compostos de interesse, que será usada em substituição à geometria biologicamente ativa, uma vez que não se tem informação suficiente sobre qual é o mecanismo de ação da droga. Os métodos semi-empíricos foram utilizados na tarefa, tendo em vista a experiência anterior do grupo com estes métodos. A busca de geometria feita com os pacotes Spartan4.1 , disponível no ambiente do CENAPAD-SP , e MOPAC6 , disponível nos computadores do Instituto de Química da UNICAMP .

As geometrias iniciais das moléculas foram criados com a ajuda do módulo gráfico do Spartan4.1. Um composto “A” submetido a otimização inicial de sua geometria com AM1 foi usado como esqueleto padrão para a construção de toda a série. A partir de “A”, elimina-se os fragmentos não existentes no composto “B”, que se deseja obter a geometria, seguido da adição em “B” dos fragmentos e átomos para completar a estrutura. O composto “A” deve ser o mesmo para a série estudada, e a seqüência de adição dos átomos ou fragmentos deve ser mantida a mesma, de modo a facilitar a consulta do valor das cargas e demais propriedades moleculares calculadas nas posições dos átomos.

Após a otimização inicial o procedimento foi diferente para os grupos de compostos estudados. Para os compostos diterpenos, um grupo menos numeroso, foi realizada a busca pelas conformações de menor calor de formação pela metodologia sistemática , que analisa todas as conformações criadas pela rotação das ligações. O programa foi orientado para aceitar ângulos diedro envolvendo hidrogênio.

O pacote MOPAC6 compilado para 35 átomos pesados e 40 hidrogênios foi usado com o intuito de mapear com maior precisão a energia das moléculas em função da rotação dos ângulos diedro para Plaunol A, Plaunol B e Desidrocrotonina. Os cálculos foram realizados em computadores Pentium de 100 e 160Mhz com sistema operacional FreeBSD 2.2.5. A geometria inicial foi obtida nos mínimos locais encontrados com o Spartan4.1. As opções usadas no arquivo de entrada especificam o uso do Hamiltoniano AM1, a não impressão da matriz de distâncias nucleares, e o uso de incremento num fator de 100 no critério de convergência. Os ângulos diedro foram marcados um de cada vez, e o ângulo correspondente ao mínimo calor de formação foi deixado livre na rotação da ligação seguinte.

Para os compostos criados com o esqueleto inicial de 2-guanidinothiazol os compostos tiveram os ângulos diedro das ligações livres para rotação variados sistematicamente com o MOPAC-6, para alguns dos compostos da série, e os resultados foram aplicados nas coordenadas internas dos demais, a partir da visualização das coordenadas internas dos compostos com o programa MOLDEN. As novas coordenadas geradas foram submetidas a otimização das suas energias com o ângulo diedro substituído pelo valor do ângulo correspondente no mínimo de energia do composto selecionado para estudo detalhado, feito com variação sistemática de 5 graus (até 0,5 graus em alguns pontos) no ângulo diedro.

2.2. A obtenção das propriedades moleculares

O MOPAC6 foi usado para calcular as propriedades, com a mudança do Hamiltoniano para AM1, PM3, MNDO, ou MINDO3. Os resultados dos cálculos forneceram os

valores das energias de HOMO e LUMO, os índices de orbitais de fronteira, calores de formação, cargas de Mulliken, momentos de dipolo total; e suas componentes calculadas com cargas pontuais sobre os átomos, e nos centros de carga dos orbitais híbridos; e as polarizabilidades e hiperpolarizabilidades de primeira e segunda ordem. A carga derivada do potencial eletrostático foi calculada com o Spartan4.1, nas geometrias obtidas com o MOPAC, exportadas para a interface gráfica do programa em coordenadas cartesianas.

O volume molecular de van der Waals, a área molecular de van der Waals, o volume acessível ao solvente, e a área acessível ao solvente foram calculados para todas as moléculas utilizando o programa SURF.

Para os diterpenos também foram determinadas propriedades eletrônicas através do programa HAM/3 para os confôrmeros de menor energia encontrados pela busca sistemática de geometria. O HAM/3^{xviii} também foi usado com base na experiência anterior do grupo. As determinações destes parâmetros com o HAM/3 nos guanidinotiazóis não foi feita porque não há parametrização do elemento enxofre disponível para este programa. Os parâmetros calculados utilizados para correlação entre a estrutura e a atividade foram as energias de HOMO e LUMO, a energia de excitação para os estados singlete e triplete, o potencial de ionização, a afinidade eletrônica, o calor de formação, o momento de dipolo total e suas componentes com cargas pontuais sobre os átomos e com cargas pontuais nos orbitais híbridos. As cargas de Mulliken em algumas posições também foram utilizadas.

2.2.1. Métodos quânticos para cálculo de parâmetros moleculares

Visto que as moléculas podem ser consideradas unidades básicas que preservam as propriedades dos sistemas macroscópicos, nada mais natural que procurar determinar as propriedades dos sistemas moleculares, e das interações entre moléculas e o meio nas quais estão envolvidas.

Um outro aspecto relevante é tentar justificar determinadas propriedades em termos de contribuições atômicas. Aqui o uso do conceito de cargas parciais sobre os átomos de uma molécula torna-se um instrumento valioso, na interpretação do comportamento químico de uma molécula e de tendências nas propriedades moleculares, tais como polarizabilidade, momento de dipolo elétrico, reatividade química^{xix}.

De forma geral a reatividade das moléculas é controlada por dois caminhos, e há duas espécies de reações que resultam deste controle: (1) Reações controladas por orbitais de fronteira, que são representados pela densidade eletrônica das camadas mais externas da molécula e (2) Reações com o controle de cargas, estas são controladas pela densidade líquida de elétrons ao redor de cada átomo^{xx}.

2.2.1.1. Os índices de orbitais de fronteira

As energias de HOMO e LUMO são descritores quânticos muito utilizados. Já foi demonstrado^{xxi} que estes orbitais têm papel fundamental na reatividade dos compostos em muitos tipos de reação química, e na determinação dos 'gaps' das bandas eletrônicas nos sólidos. Estes orbitais também são responsáveis pela formação de muitos complexos de transferência de carga^{xxii}. De acordo com a teoria dos orbitais de fronteira em reações químicas, a formação do estado de transição se deve a interação entre os orbitais de fronteira (HOMO e LUMO) das espécies reagentes. Portanto o tratamento diferenciado para estes orbitais moleculares é baseado em princípios gerais que governam a natureza das reações químicas.

A energia de HOMO é diretamente relacionada ao Potencial de Ionização, e caracteriza a suscetibilidade da molécula ao ataque de eletrófilos. A energia de LUMO é relacionada à afinidade eletrônica, e caracteriza a suscetibilidade ao ataque por nucleófilos. As energia de HOMO e LUMO são importantes no caso de reações radicalares. O conceito de dureza e moleza para nucleófilos e eletrófilos também está diretamente ligado à energia relativa de HOMO/LUMO. Nucleófilos duros têm a energia para HOMO baixa, e nucleófilos moles têm a energia para LUMO alta. Eletrófilos duros têm alta energia para LUMO, e eletrófilos moles têm baixa energia para LUMO.

O 'gap' entre HOMO e LUMO, ou seja, a diferença de energia entre eles, é um importante índice de estabilidade química. Uma grande diferença entre HOMO e LUMO significa que a molécula tem alta estabilidade, ou baixa reatividade em reações químicas. A diferença entre HOMO e LUMO também é usada para aproximar a energia da primeira excitação eletrônica da molécula. Porém esta idéia não leva em conta que um novo estado eletrônico é formado no momento da excitação, e pode levar a resultados conceitualmente equivocados.

O conceito de dureza de ativação também é derivado da diferença de energia entre HOMO e LUMO ^{xxiii}. A dureza de ativação distingue a reatividade da molécula em diferentes sítios de reação, e portanto é útil na previsão do efeito da orientação dos reagentes.

A definição qualitativa da dureza é relacionada à polarizabilidade, já que geralmente a diminuição na diferença de energia entre HOMO e LUMO facilita a polarização da molécula submetida a um campo eletromagnético orientado.

As densidades eletrônicas nos orbitais de fronteira sobre os átomos são uma maneira de caracterização detalhada das possíveis interações entre doadores e aceptores de elétrons. De acordo com a teoria dos orbitais de fronteira para a reatividade, a maioria das reações químicas acontecem na posição em que HOMO e LUMO dos respectivos reagentes podem ter a maior sobreposição. No caso de uma molécula doadora de elétrons, a densidade de HOMO é crítica para a transferência de carga. O descritor utilizado para este carácter da molécula no átomo “a” é f_a^{HOMO} , definido da seguinte maneira:

$$f_a^{HOMO} = \sum_a (C_{HOMO}^a)^2 \quad (1)$$

onde $C_{HOMO,n}$ são os coeficientes das “n” orbitais atômicas da camada de valência do átomo “a” no HOMO.

Analogamente para um receptor de elétrons, a densidade em LUMO é importante. Para LUMO o descritor usado é f_a^{LUMO} , definido como:

$$f_a^{LUMO} = \sum_n (C_{LUMO}^a)^2 \quad (2)$$

onde $C_{LUMO,n}$ são os coeficientes das “n” orbitais atômicas da camada de valência do átomo “a” no LUMO.

Os descritores citados devem ser usados para descrever a reatividade em diferentes átomos da mesma molécula. Para comparar as reatividades dos átomos em diferentes moléculas os índices devem ser normalizados pelo valor da energia do orbital de fronteira correspondente:

$$F_a^{HOMO} = f_a^{HOMO} / \epsilon_{HOMO} \quad (3)$$

$$F_a^{LUMO} = f_a^{LUMO} / \epsilon_{LUMO} . \quad (4)$$

2.2.1.2. Polarizabilidade Molecular

A polarização de uma molécula por um campo eletromagnético externo é dada em termos dos tensores de suscetibilidade magnética de ordem “n”. O termo de primeira ordem é chamado de polarizabilidade (α), o termo de segunda ordem é conhecido como hiperpolarizabilidade (β), e o termo de terceira ordem, segunda hiperpolarizabilidade (γ). A propriedade mais significativa da polarizabilidade molecular é a sua relação com o volume, ou tamanho, da molécula. Foi demonstrado que os valores de polarizabilidade têm relação com a hidrofobicidade dos compostos. Este é um fator associado aos fenômenos de transporte passivo das moléculas que se acredita importante na ação dos fármacos em geral. O tensor de polarizabilidade de primeira ordem contém também informações sobre possíveis interações indutivas da molécula, e a anisotropia total da polarizabilidade (segunda ordem) caracteriza a propriedade acetona de elétrons de uma molécula. Em estudos recentes foi demonstrado que moléculas orgânicas com conjugação eletrônica, e capacidade doadora ou aceptora de elétrons têm um grande valor de β^{xxiv} .

A polarização (P) é induzida num meio qualquer por um campo eletromagnético externo (**F**) segundo as equações de Maxwell^{xxv}. De maneira sucinta, pode se escrever:

$$P = P_0 + \chi^1 \bullet F + \chi^2 \bullet F \bullet F + \chi^3 \bullet F \bullet F \bullet F + \dots \quad (5)$$

Onde χ^n é o tensor de suscetibilidade de ordem n do meio, e o campo eletromagnético $\mathbf{F}$ é a matriz com as componentes de intensidade nas direções x,y, e z. A suscetibilidade do meio pode ser expressa em termos do momento de dipolo molecular. O momento de dipolo de uma molécula interagindo com um campo eletromagnético pode ser escrito como:

$$\mu_i = \mu_i^0 + \alpha_{ij} F_j + (1/2)\beta_{ijk} F_j F_k + (1/6)\gamma_{ijkl} F_j F_k F_l + \dots \quad (6)$$

Onde μ_i^0 é o momento de dipolo permanente, α_{ij} , β_{ijk} , γ_{ijkl} , são os tensores da polarizabilidade linear, e as hiperpolarizabilidades de primeira e segunda ordem, respectivamente. O termo de segunda ordem, β_{ijk} , dá origem ao “batimento” na mistura de frequências (geração de harmônicos com a soma e a diferença das frequências misturadas, incluindo o segundo harmônico), e a retificação óptica. O termo de terceira ordem, γ_{ijkl} , é responsável pela geração do terceiro harmônico e ressonância entre dois fótons.

A obtenção dos valores para a polarizabilidade e hiperpolarizabilidades é feita tanto por expansão da energia quanto do momento de dipolo no MOPAC. A acuidade numérica das equações utilizadas é sensível à precisão do cálculo realizado da energia, ou do momento de dipolo. Por exemplo, para um campo eletromagnético aplicado de 0,001 au, um erro da ordem de 10^{-15} nas energias produz um erro de 10^{-3} em γ_{ijkl} . Para valores similares no erro do momento de dipolo resultam erros menores em γ_{ijkl} por causa da menor potência em $\mathbf{F}$ na equação. A tentativa de minimizar o erro com o aumento da intensidade do campo aplicado, pode levar a mudanças na configuração eletrônica, ou aumento da significação dos termos de ordem superior nas equações de campo finito usadas na derivação.

As grandezas obtidas são transformadas em valores observáveis, então. A segunda hiperpolarizabilidade é usada como a média:

$$\gamma = 1/5 \{ \gamma_{xxxx} + \gamma_{yyyy} + \gamma_{zzzz} + 2[\gamma_{xxyy} + \gamma_{xxzz} + \gamma_{yyzz}] \} \quad (7)$$

A componente ao longo do eixo do momento de dipolo é a grandeza comparada na hiperpolarizabilidade:

$$\beta_{\mu} = (3/5) \{ \beta_x \mu_x + \beta_y \mu_y [\beta_{iii} + \beta_{ijj} + \beta_{ikk}] + \beta_z \mu_z \} / \|\mu\| \quad (8)$$

2.2.1.3. A determinação das cargas numa molécula

O conceito de carga atômica é muito usual para se compreender as propriedades químicas e físicas de uma molécula. Infelizmente, carga atômica não é o que se pode obter diretamente determinado por uma função de onda Hartree-Fock.

Alguns procedimentos fazem uso de funções de bases, o mais antigo e também mais utilizado método de cálculo de carga atômica, é a análise populacional de Mulliken^{xxvi}, que faz uso da matriz densidade, assumindo só os elementos ao longo da sua diagonal. Os elementos que estão fora da diagonal da matriz densidade são desprezados. Ainda que amplamente usado sabe-se que o procedimento de Mulliken tem forte dependência do conjunto de funções de bases utilizados (o conjunto de funções, geralmente Gaussianas, são fixadas nos átomos)^{xxvii}. A Análise Populacional Natural (NPA) desenvolvida por Winhold-Reed^{xxviii} visa eliminar alguns dos problemas encontrados na análise populacional de Mulliken^{xxix}.

Algumas outras aproximações disponíveis incluem cargas derivadas de ajustes do potencial eletrostático. Esta é outra quantidade observável, e representa a energia de uma carga teste em uma dada posição^{xxx,xxxi}. Nesta classe estão os métodos CHELP e CHELPG.

A maioria dos métodos semi-empíricos são parametrizados para reproduzir os calores de formação dos compostos, e portanto as cargas obtidas pelos diferentes métodos geralmente são bem diferentes entre si. Porém servem para se ter uma idéia da distribuição das cargas nas moléculas.

2.2.1.3.1. A análise populacional de Mulliken

A análise populacional de Mulliken é sem dúvida o método mais utilizado entre os químicos. Este método foi derivado por Mulliken e baseia-se na teoria de orbitais moleculares, onde um conjunto de orbitais moleculares Ψ_i é definida por um conjunto de combinações lineares de k orbitais, também chamado de funções de bases, χ_i , cujos coeficientes são obtidos pelo método de Hartree-Fock. A equação que representa esta combinação é:

$$\psi_i = \sum_{\mu=1}^k (c_{i\mu} \chi_{\mu}) \quad (9)$$

e assim, em termos de combinações lineares a equação pode ser escrita como,

$$\Psi_i = c_{i1}\chi_1 + c_{i2}\chi_2 + \dots + c_{ik}\chi_k \quad (10)$$

Para maior simplicidade, $C_{i\mu}$ e χ_{μ} são considerados reais, e existe apenas uma função de base sobre cada átomo na molécula. Desta forma a probabilidade de densidade associado com um elétron em Ψ é;

$$|\Psi_i|^2 = c_{i1}^2 \chi_1^2 + c_{i2}^2 \chi_2^2 + \dots + 2c_{i1}c_{i2}\chi_1\chi_2 + 2c_{i1}c_{i3}\chi_1\chi_3 + 2c_{i2}c_{i3}\chi_2\chi_3 + \dots \quad (11)$$

Integrando esta equação sobre todo o espaço tridimensional e usando o fato que Ψ e χ são normalizados, temos:

$$1 = c_{i1}^2 + c_{i2}^2 + \dots + 2c_{i1}c_{i2}S_{12} + 2c_{i2}c_{i3}S_{23} + \dots \quad (12)$$

Onde S_{12} são integrais de recobrimento, definidas como:

$$S_{\mu\nu} = \int \chi_{\mu} \chi_{\nu} d\sigma_{\mu} d\sigma_{\nu} \quad (13)$$

A proposta de Mulliken, em termos da equação 11 pode ter a seguinte interpretação:

Um elétron na orbital molecular Ψ_i contribui em c_{i1}^2 para a população líquida em χ_1 e c_{i2}^2 para χ_2 , e então $2c_{i1}c_{i2}S_{12}$ para a população de recobrimento entre χ_1 χ_2 , e $2c_{i2}c_{i3}S_{23}$ para o recobrimento entre χ_2 χ_3 .

Definindo $q_{i\mu}^i$ como a população líquida do elétron na orbital molecular Ψ_i pela contribuição da orbital atômica χ_{μ} , e $p_{\mu\nu}^i$ como a população eletrônica de recobrimento entre as orbitais atômicas χ_{μ} e χ_{ν} , teremos as seguintes expressões.

$$q_{i\mu}^i = c_{i\mu}^2 \quad (14)$$

$$p_{\mu\nu}^i = 2c_{i\mu}c_{i\nu}S_{\mu\nu} \quad (15)$$

Então, somando as contribuições $q_{i\mu}^i$ e $p_{i\mu}^i$ chegaremos à Equação 19, obtendo desta forma a população líquida de Mulliken.

A população de Mulliken é muito útil para determinar a localização de cargas eletrônicas em moléculas e a natureza de suas orbitais. No caso de $p_{\mu\nu}^i(1) < 0$, temos uma função orbital antiligante. Se $p_{\mu\nu}^i(1) > 0$ temos uma orbital ligante, e se $p_{\mu\nu}^i(1) = 0$ temos uma orbital não ligante.

Mulliken chamou o resultado de população bruta das orbitais atômicas. Ele usou $N_{\mu}(i)$ para simbolizar a população bruta em χ_{μ} . Onde χ_{μ} pode ser uma orbital atômica (isto é, a soma de todas as populações brutas de um átomo) assim,

$$N_{\mu}(i) = q_{i\mu}^i + 1/2 \sum_{\mu \neq \nu} p_{\mu\nu}^i \quad (16)$$

Nos casos dos métodos computacionais semi-empíricos, AM1, PM3, CNDO entre outros^{xxxii}, que consideram a integral diferencial de recobrimento como sendo nula, as

orbitais atômicas são considerados como ortogonais entre si, isto é $S_{\mu\nu}=0$ para $\mu \neq \nu$. Neste caso específico a equação 16 torna-se,

$$N_{\mu}(i)=(C_{i\mu})^2 \quad (17)$$

Desta forma a carga bruta de Mulliken torna-se dependente dos coeficientes que são otimizados num processo de auto consistência^{xxxiii}

Os motivos para a ampla difusão do método populacional de Mulliken foram a intensa aplicação dos métodos que empregam a teoria do orbital molecular e a facilidade com que pode ser calculada, e o fato da análise populacional ser obtida de forma direta, sem custo computacional adicional.

Apesar de ser um método muito utilizado pela comunidade química, ele possui algumas arbitrariedades como a divisão das populações de recobrimento em contribuições atômicas, dificultando a separação de cargas entre dois átomos de diferentes eletronegatividades. Este fator faz com o que não haja reprodução do momento de dipolo, μ , calculado por auto-consistência utilizando as cargas de Mulliken.

No método populacional de Mulliken há uma forte dependência dos conjuntos de bases escolhidos^{xxxiv}, uma consequência da função densidade ter sido projetada sobre as funções de bases utilizadas.

Outros pesquisadores propuseram modificações no método original de Mulliken, entretanto nenhuma atingiu grande resultado. Alguns pesquisadores como Christofferson e Baker^{xxxv}, Pollak e Rein^{xxxvi} Löwdin^{xxxvii} e Davison^{xxxviii} tentaram resolver o problema da separação da população de recobrimento.

Yanes^{xxxix}, Ferreira^{xl}, Politzer e Harris^{xli}, usaram outras estratégias para melhorar o método original de Mulliken, mas o que atingiu a maior popularidade foi a análise populacional

natural de Weinhold e Reed, onde este utiliza um procedimento mais complexo na partição de densidade eletrônica.

2.2.1.3.2. Métodos de determinação de cargas derivadas do potencial eletrostático

Nestes métodos estão envolvidos os ajustes do potencial eletrostático. Ele tem reproduções do momento dipolar e do potencial eletrostático só satisfatório. O seu custo computacional e dependência de bases são baixos. Estes métodos surgiram da necessidade de se corrigir algumas imperfeições na distribuições de cargas que a análise populacional de Mulliken possui, desta forma alguns métodos foram introduzidos, tais como: o método **CHELP** (*Charge from Eletrostatic Potencial*), o método **CHELPG** (*Charge from Eletrostatic Potencial Grid based*) e o método **RESP** (*Restrained Eletrostatic Potencial*)^{xliii}.

Método CHELP

O método **CHELP** proposto por Chirlian e Franci^{xxx}. baseia-se no ajuste do potencial eletrostático. Neste método o potencial eletrostático é determinado selecionando um determinado número de pontos ao redor da molécula. Tipicamente é selecionado 1 ponto por Å² em quatro camadas situadas a distâncias de 1,4; 1,6; 1,8; e 2,0 vezes o raio de van der Waals da molécula. Os pontos que caem dentro do raio de van der Waals da molécula são desprezados, devido as grandes distorções da proximidade do núcleo. Desta forma são considerados no final um número total de pontos que variam de 100 a 300, para moléculas pequenas, tais como por exemplo: ácido acético, metanol, formamida, acetonitrila, ácido fluorídrico, etc. O melhor ajuste para as cargas obtidas pelo potencial eletrostático é encontrando os mínimos quadrados de y ,

$$y(q_1, q_2, \dots, q_n) = \sum_{i=1}^m \left[V_i - E_i(q_1, q_2, \dots, q_n) \right]^2 \quad (18)$$

onde V_i é dado por:

$$V(r) = \sum_A \frac{Z_A}{|r - R_A|} - \sum_{\mu\nu} P_{\mu\nu} \int \frac{\phi_\mu \phi_\nu}{|r - r'|} dr' \quad (19)$$

onde Z_A é a carga nuclear no átomo A centrado em R_A , $P_{\mu\nu}$ é a matriz densidade determinada do procedimento Hartree-Fock, e ϕ_μ e ϕ_ν são funções de base usadas.

E_i é o potencial eletrostático dado por:

$$E_i = \sum_{j=1}^n \left(q_j / r_{ij} \right) \quad (20)$$

n = número de átomos na molécula.

O mínimo de y pode ser obtido encontrando os pontos estacionários da função de

Lagrange Z

$$z(q_1, q_2, \dots, q_n) = y(q_1, q_2, \dots, q_n) + \lambda g(q_1, q_2, \dots, q_n) \quad (21)$$

Onde g é imposto para forçar o ajuste e λ é o multiplicador associado a ele. Este ajuste é necessário para reproduzir a carga molecular total. A correspondente função é:

$$g(q_1, q_2, \dots, q_n) = \left(\sum q_i \right) - q_{\text{tot}} = 0 \quad (22)$$

A solução de z é encontrada fazendo $\partial z / \partial \lambda = 0$ e $\partial z / \partial q_k = 0$. A solução deste conjunto equações gera um conjunto de n+1 equações para n+1 átomos. A solução da correspondente equação, gera a matriz de cargas^{xxx}.

Método CHELPG

O método CHELPG proposto por Breneman e Wiberg^{xxxi} também é derivado do potencial eletrostático. Tal qual foi discutido para o método CHELP, o CHELPG também usa a

técnica do multiplicador de Lagrange, técnica esta que tem a vantagem de ser rápida e não iterativa.

A principal diferença entre os dois métodos, CHELP e CHELPG, é que o segundo emprega um algoritmo que seleciona os pontos baseados na regularidade dos espaçamentos do mesmos. O algoritmo usado para selecionar os pontos ao redor do núcleo é semelhante ao que foi anteriormente utilizado por Cox e Williams^{xliii}. Este algoritmo, ao contrário do que foi empregado no CHELP, obtém-se o potencial eletrostático calculando os pontos que se encontram sobre a superfície da molécula sob a forma de um cubo com dimensões de 0,3-0,8 Å³. Estes pontos são estendidos até a distância de 2,8 Å de cada aresta do cubo. De forma semelhante ao CHELP, no CHELPG todos os pontos que caem dentro do raio de van der Waals da molécula e aqueles que caem fora do raio máximo de 2,8 Å são eliminados do procedimento de ajuste. Após o procedimento de seleção de pontos, o potencial eletrostático de cada amostra é calculado analiticamente. Estes dados são então usados como arquivo de entrada para o multiplicador de Lagrange na rotina de mínimos quadrados, que força um ajuste exato na carga da molécula.

Como um exemplo de comparação entre ambos os métodos, para determinar a carga eletrônica ao redor da formamida seriam necessários o cálculo de 300 a 400 pontos pelo método CHELP e 5000-6000 pontos pelo método CHELPG^{xliiv}.

2.2.1.4. A determinação dos parâmetros estéricos das moléculas

Os parâmetros estéricos moleculares calculados foram o volume molecular, o volume acessível ao solvente, área da superfície molecular, e área acessível ao solvente. Para calcular estes parâmetros foi utilizado o programa SURF, desenvolvido por Gaudio, e adaptado por mim. Este programa cria paralelogramos de tamanho determinado (quanto menor o paralelogramo maior a precisão), completa o volume de cada molécula com os paralelogramos até o raio de van der Waals, conta quantos paralelogramos couberam, e calcula o volume. Os raios utilizados para cada átomo^{xliv} são mostrados na tabela 2.1. O volume e a área acessíveis ao solvente são calculados adicionando-se 1,5Å, o raio médio de uma molécula de água, ao raio de van der Waals de cada átomo.

Tabela 2.1: Raios atômicos utilizados pelo programa SURF para calcular as áreas e volumes moleculares.

Tipo do átomo	Símbolo	Raio vdW	Tipo do átomo	Símbolo	Raio vdW (a)
1	H	1,20	8	S	1,85
2	C aliph	1,60	9	Cl	1,80
3	C arom	1,70 (b)	10	As	2,00
4	N	1,50	11	Se	2,00
5	O	1,40	12	Br	1,95
6	F	1,35	13	Sb	2,20
7	P	1,90	14	Te	2,20
15	I	2,15			

2.2.1.5. Inclusão do efeito de solvatação no modelo AM1

Segundo Dixon e colaboradores^{xlv}, a energia livre de solvatação de uma molécula de soluto, 'S', é considerada a diferença entre a energia da molécula em solução e em fase gasosa.

Tabela 2.2: Parâmetros utilizados para simulação do efeito do solvente.

Elemento	ξ^C	ξ^E	σ
H	-0,0651	-0,1164	0,0 (assumido)
C	-0,0569	0,0380	0,256
N	0,0206	-0,0697	-9,509
O	-0,1760	0,1361	-14,025
F	-0,0672	0,0307	0,867
S	0,1089	0,0581	-1,664
Cl	0,0039	0,0109	-0,248
λ_D			-6,501

2.3. Análise Exploratória dos Dados

2.3.1. Regressão em componentes principais e análise hierárquica de agrupamentos

A correlação entre conjuntos de valores obtidos e a atividade biológica foi feita com modelos de reconhecimento de padrão utilizando o programa PIRUETTE^{xlvii}. O conjunto inicial de dados foi autoescalado aplicando-se a equação (36).

$$x_{ij(as)} = \frac{x_{ij} - \bar{x}_j}{s_j} \Rightarrow \bar{x}_j = \frac{1}{n} \sum_1^n x_{ij} \Rightarrow s_j = \sqrt{\frac{1}{n-1} \sum_1^n (x_{ij} - \bar{x}_j)^2} \quad (36)$$

Num grande conjunto de dados, nem sempre é fácil determinar o que pode ser relacionado ao problema em questão. A análise exploratória dos dados foi feita com técnicas de decomposição em componentes principais (PCA) e análise hierárquica de agrupamentos (HCA). O objetivo é identificar e enfatizar o que é relevante, e desconsiderar o restante, de modo que as informações mais importantes tornem-se óbvias.

Decomposição em componentes principais é uma forma de definir novas variáveis, combinações lineares das iniciais em um novo sistema de eixos ortogonais especialmente construído. Cada eixo é definido de maneira a acomodar a maior variância possível, e sua contribuição é subtraída nos dados para a construção do eixo seguinte. O novo sistema tem os novos eixos denominados componentes principais, em ordem decrescente na variância do sistema original descrita. A representação em outra base não afeta as relações entre as variáveis originais.

Matematicamente o processo consiste na decomposição em valores singulares de uma matriz **X**, não quadrada, em que cada amostra é representada numa linha, e as variáveis

são as colunas. A matriz passa a ser representada pelo produto de três outras matrizes, a matriz **V** denominada matriz de 'loadings', e o produto **U S**, matriz de 'scores'.

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (38)$$

O gráfico das colunas umas pelas outras permite identificar os autovetores capazes de promover separação entre grupos que se sabe serem diferentes. Assim pode ser possível selecionar as variáveis mais importantes e criar um modelo com alguma capacidade de prever a posição de amostras cujo grupo é desconhecido.

Os 'scores' são as novas coordenadas das amostras, e revelam características de colinearidade entre as variáveis dos dados iniciais. É um produto entre **U**, os autovetores das colunas de **X**, e **S**, a matriz diagonal dos valores singulares.

2.3.2. A regressão por mínimos quadrados parciais

O método de Mínimos Quadrados Parciais (PLS) é uma técnica de regressão considerada robusta, no sentido que o resultado não é grandemente afetado por exclusão de uma ou outra amostra no conjunto de treinamento para a criação do modelo. Permite a correlação entre um número maior de variáveis que de amostras, por meio da extração de variáveis latentes, que são combinações lineares das variáveis originais, encontradas por método iterativo. Tem como objetivo maximizar o ajuste da variância das variáveis dependentes, conjuntamente com o melhor ajuste possível às variáveis independentes.

O tratamento matemático é feito pelo algoritmo NIPALS, sigla inglesa para Mínimos Quadrados Parciais Iterativo Não Linear. Consiste na seguinte sequência de passos:

$$\mathbf{u}_h = \mathbf{Y} \quad (39)$$

Inicialização das variáveis latentes da matriz de respostas, $\mathbf{u}_h$ igual a matriz de respostas.

Onde h refere-se à dimensão ($h=1...a$).

$$\mathbf{w}_h = \mathbf{u}_h' \mathbf{X} \quad (40)$$

Os pesos $\mathbf{w}_h$ para **X** são calculados, o normalizados para unidade.

$$\mathbf{t}_h = \mathbf{X} \mathbf{w}_h' \quad (41)$$

As variáveis latentes $\mathbf{t}_h$ são calculadas.

$$\mathbf{u}_h = \mathbf{v}_h \mathbf{t}_h + d \quad (42)$$

$$\mathbf{v}_h = \mathbf{t}_h' \mathbf{u}_h / |\mathbf{t}_h|^2 \quad (43)$$

O coeficiente $\mathbf{v}_h$ é determinado. Este coeficiente relaciona as variáveis latentes de $\mathbf{X}$ com as de $\mathbf{Y}$.

$$\mathbf{b}_{hx} = \mathbf{t}_h' \mathbf{X} \quad (44)$$

O vetor de 'loadings' para $\mathbf{X}$, dado por $\mathbf{b}_{hx}$ é calculado, e normalizado para unidade.

$$\mathbf{E}_x = \mathbf{X} - \mathbf{t}_h \mathbf{b}_{hx} \quad (45)$$

$\mathbf{E}_y = \mathbf{Y} - \mathbf{v}_h \mathbf{t}_h$: os resíduos $\mathbf{E}_x$ e $\mathbf{E}_y$ são calculados. Para as próximas dimensões $\mathbf{E}_x$ e $\mathbf{E}_y$ são usados em lugar de $\mathbf{X}$ e $\mathbf{Y}$, e o procedimento retorna para o primeiro passo.

O processo é interrompido na dimensão ($h=a$), quando o teste de validação cruzada indicar que ($h=a+1$) não é mais significativo. Este resultado deve convergir para a solução por mínimos quadrados após um número de iterações igual ao número de variáveis. Para usar o modelo em predição, os dados são escalados como na criação do modelo. A primeira iteração começa com $h=0$, com estimativa inicial $y=y_{(média)}$, a média dos valores do conjunto de instrução. No passo seguinte $h=h+1$, $\mathbf{t}_h$ é calculado:

$$\mathbf{t}_h = \mathbf{X} \mathbf{w}_h' \quad (46)$$

E em seguida:

$$\mathbf{Y} = \mathbf{Y} + \mathbf{v}_h \mathbf{t}_h \mathbf{b}_{hy} \quad (47)$$

$$\mathbf{X} = \mathbf{X} - \mathbf{t}_h \mathbf{b}_{hx} \quad (48)$$

Estes passos são repetidos até $h>a$. O valor predito para y é escalado, e deve ser revertido^{xlviii}.

O ajuste de um vetor de dados testado ao modelo pode ser obtido pelo desvio padrão do resíduo, $\mathbf{e}_i$.

$$\mathbf{e}_i = \mathbf{x}_i - \mathbf{x}_i \mathbf{b}' \mathbf{b} \quad (49)$$

$s^2 = |\mathbf{e}|^2 / (\mathbf{p} - \mathbf{a})$, onde $(\mathbf{p} - \mathbf{a})$ indica os graus de liberdade. Um critério de ajuste aceitável pode ser $s^2 < s_a^2(F)$, onde s_a^2 é o desvio padrão do resíduo do conjunto de treinamento com número (a) de dimensões, e F o valor estatístico apropriado.

Para medir a capacidade do modelo fazer previsões em relação a amostras não incluídas na sua construção, pode ser feita uma validação cruzada. A forma mais popular deste método é deixar uma amostra de fora, construir o modelo, e determinar as características da amostra com o modelo construído. Isso deve ser feito para cada uma das amostras, e então pode ser determinado o valor de PRESS (*predictive residual sum of squares*)^{xlix}.

$$PRESS = \sum (Y_i - \hat{Y}_i)^2 \quad (50)$$

O erro padrão das previsões é calculado de acordo com a seguinte equação:

$$SDEP = \sqrt{\frac{PRESS}{n}} \quad (51)$$

PRESS é a soma dos resíduos entre o valor observado e o valor previsto para a amostra deixado de fora. O seu valor deve ser o menor possível para que a regressão seja considerada boa para predição. Outros índices que medem a capacidade de previsão da regressão linear, também são úteis para avaliar o modelo. O índice Q como a soma dos quadrados de cada linha (as amostras) da matriz de resíduos $\mathbf{E}$ (equação 45). Para a $i^{\text{ésima}}$

amostra de $\mathbf{X}$, $\mathbf{x}_i$:

$$Q_i = \mathbf{e}_i \mathbf{e}_i^T = \mathbf{x}_i (\mathbf{I} - \mathbf{P}_k \mathbf{P}_k^T) \mathbf{x}_i^T \quad (52)$$

Onde $\mathbf{e}_i$ é a $i^{\text{ésima}}$ linha de $\mathbf{E}$, $\mathbf{P}_k$ é a matriz dos k vetores 'loadings' retida na regressão (onde cada vetor é uma coluna de $\mathbf{P}_k$) e a matriz identidade $\mathbf{I}$, de tamanho apropriado. O valor estatístico de Q indica o quanto cada amostra está em conformidade com o modelo.

É uma medida da diferença, ou resíduo, entre uma amostra e sua projeção nas k variáveis latentes retidas no modelo.

A soma de quadrados normalizados dos ‘scores’, conhecida como T^2 estatístico de Hotelling, é uma medida da variação de cada amostra dentro do modelo. T^2 é definido da seguinte forma:

$$T^2_i = \mathbf{t}_i \boldsymbol{\lambda}^{-1} \mathbf{t}_i' = \mathbf{x}_i \mathbf{P} \boldsymbol{\lambda}^{-1} \mathbf{P}' \mathbf{x}_i' \quad (53)$$

Onde $\mathbf{t}_i$ agora refere-se à $i^{\text{ésima}}$ linha da matriz $\mathbf{T}_k$, dos vetores ‘scores’ do modelo, e $\boldsymbol{\lambda}^{-1}$ é a matriz diagonal que contém a inversa dos autovalores associados com os k autovetores das variáveis latentes guardadas pelo modelo.

Podem ser estabelecidos limites de confiança para os resíduos Q e T^2 . Para Q o limite de confiança pode ser calculado desde que todos os autovalores da matriz de covariância de $\mathbf{X}$, os λ_i tenham sido obtidos, da seguinte forma:

$$Q_\alpha = \Theta_1 \left[\frac{c_\alpha \sqrt{2\Theta_2 h_0^2} \Theta_1 + \Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{1/h_0} \quad (54)$$

$$\Theta_i = \sum_{j=k+1}^n \lambda_j^i \Leftrightarrow i = 1, 2, 3 \dots \quad (55)$$

E:

$$h_0 = 1 - \frac{2\Theta_1 \Theta_3}{2\Theta_2^2} \quad (56)$$

Na equação 54, c_α é o desvio padrão normal correspondente à parte superior $(1-\alpha)$. Na equação 55, ‘ k ’ é o número de variáveis latentes guardadas no modelo, e ‘ n ’ é o número total de variáveis latentes para acomodar a variância dos dados.

Limites de confiança estatísticos para os valores de T^2 podem ser calculados por meio da distribuição F , da seguinte maneira:

$$T^2_{k,m,\alpha} = \frac{k(m-1)}{m-k} F_{k,m-k,\alpha} \quad (57)$$

Aqui ‘ m ’ é o número de amostras usadas para desenvolver o modelo PLS e k é o número de variáveis latentes guardadas pelo modelo.

A interpretação geométrica de Q e T^2 é que Q é uma medida da variação dos dados fora das variáveis latentes incluídas no modelo PLS. Q é a distância do plano formado por

duas variáveis latentes. De fato a raiz quadrada de Q é a distância Euclidiana da amostra ao plano das VLs. O limite de Q define uma distância do plano que já pode ser considerada não usual, baseada nos dados utilizados para criar o modelo.

Por sua vez, T^2 é a medida da distância da projeção da amostra no espaço vetorial das VLs ao ponto médio multivariado, ou seja, da intersecção das duas VLs no plano formado.

O limite de T^2 define uma elipse no plano, dentro da qual os dados normalmente são projetados.

A análise hierárquica de agrupamentos (HCA) é realizada com os dados através de equações simples, que possibilitam o agrupamento das amostras no espaço vetorial.

Inicialmente as distâncias entre as amostras são calculadas:

$$d_{ab} = \sum_{j=1}^m [(x_{aj} - x_{bj})^m]^{1/m} \quad (58)$$

O agrupamento entre as amostras pode ser feito de várias maneiras. Um modo comumente utilizado é conhecido como conexão simples entre os objetos, feito segundo a equação que mostra a distância entre o bloco 'a' e 'b' que acaba de ser formado e o bloco 'c' já existente:

$$d_{ab \Rightarrow c} = 0,5d_{ac} + 0,5d_{bc} - 0,5|d_{ac} - d_{bc}| \quad (59)$$

Finalmente, a semelhança entre duas amostras, ou blocos de amostras é dada por sua similaridade, definida da seguinte forma:

$$s_{ab} = 1 - \frac{d_{ab}}{d_{\max}} \quad (60)$$

2.3.3. Redes neurais

A metodologia de treinamento de redes neurais é uma ferramenta baseada em princípios diferentes. Sua concepção foi sugerida pelo funcionamento dos neurônios no cérebro. A idéia básica de um neuro-computador vem de 1940, a mesma época da concepção das

idéias do computador digital de Neumann. Portanto “rede neural” é um nome que se origina de uma estrutura de ‘hardware’, como é a estrutura convencional de processamento serial.

Há vários tipos de redes de conexão como o tipo “perceptron”, o tipo de “Hopfield”, ou o tipo da “máquina de Boltzmann”. O método que será utilizado aqui é o “perceptron”. Foi inicialmente proposto por Rosenblatt em 1957 como um modelo de auto-organização. Com a recente adição de várias camadas de neurônios e com o método de treinamento de retropropagação, percebeu-se que o método tinha outras capacidades além das esperadas.

O modo básico de operação de uma rede neural é não-linear. Isto cria impressionante capacidade de ajuste e classificação, além de algumas dificuldades. Uma delas é que uma rede neural não pode fornecer coeficientes de correlação simples como os obtidos em QSAR com regressão multilinear.

Numa rede neural, a informação que é acumulada na etapa de aprendizagem é guardada na forma das intensidade das ligações entre os neurônios numa matriz de pesos. A matriz é reconstruída até que os padrões sejam reproduzidos na camada de saída. As redes construídas por este programa têm três camadas de neurônios: A primeira é onde ficam acumulados os dados de entrada, a segunda acumula valores obtidos durante a aprendizagem com as informações da matriz de pesos e dos padrões que serão reproduzidos, e a terceira tem os padrões, ou valores, obtidos na saída¹.

O valor de um neurônio na segunda ou terceira camada podem ser expressos pela equação (61).

$$O_j = 1/[1+\exp(-\alpha y_j)] \equiv f(y_j) \quad (61)$$

$$y_j = (\sum W_{ij} x_i) - \theta_j \quad (61^a)$$

Onde x_i é o valor de um neurônio da primeira ou da segunda camada; W_{ij} é um elemento da matriz de pesos, que expressa o valor atribuído à conexão entre os neurônios i e j , e

que pode ter valores negativos ou positivos; α é o parâmetro que expressa a não-linearidade da operação do neurônio; θ é o valor inicial do neurônio j , normalmente 1. A função sigmóide $f(y_j)$ não é única, e pode ser substituída por qualquer função para operação desejada, desde que seja diferenciável.

Dados N neurônios na primeira camada, um conjunto de entrada pode ser expresso como por um vetor com N elementos, chamados de “padrões de entrada”. Analogamente os dados de saída, agrupados num vetor serão chamados de “padrão de saída” (O_j). O vetor que é comparado com o padrão de saída para estabelecer a matriz de pesos (W_{ij}) é chamado “padrão de treinamento” (t_j).

O treinamento da rede é baseado nas seguintes equações:

$$\delta W_{ij} = - d_j x_i \varepsilon \quad (62)$$

$$d_j = (O_j - t_j) f'(y_j) \quad (62^a)$$

$$d_j = \left(\sum_i W_{ji}' d_i \right) f'(y_j) \quad (62b)$$

Aqui, ε é um parâmetro que determina a variação para correção durante a retropropagação. A equação 62^a é usada para corrigir a conexão entre a segunda e a terceira (saída) camadas, enquanto a equação 62b é para as outras conexões, onde W_{ji}' e d_j na segunda camada são W_{ij} e d_j da terceira camada. Se f é uma função sigmóide, sua derivada é

$$f'(y_j) = f(y_j) [1-f(y_j)] \alpha \quad (63)$$

Nas equações acima, tanto α como ε podem ser definidos independentemente para cada camada.

Como o valor de cada neurônio é definido entre 0 e 1, os dados de entrada devem ser apropriadamente escalados dentro desta região. Deve ser observado que se o valor de um neurônio é definido como zero na camada de entrada, as conexões a partir daquele

neurônio serão sempre nulas. Para contornar este problema, o menor valor assumido por um neurônio deve ser ligeiramente maior que zero, geralmente 0,1.

O programa usa a seguinte função de escalamento:

$$P = ((q_{\max} - q_{\min})p + q_{\min}p_{\max} - q_{\max}p_{\min}) / (p_{\max} - p_{\min}) \quad (64)$$

Onde $q_{\max}$ e $q_{\min}$ são os valores máximo e mínimo de escalamento, e $p_{\max}$ e $p_{\min}$ são os valores máximo e mínimo dos elementos do conjunto de entrada. Finalmente, o valor de cada elemento do conjunto de treinamento também deve estar entre 0 e 1.

O treinamento ocorre por retropropagação até que a função de erro se torne tão pequena quanto definido.

$$E = \sum (O_j - t_j)^2 \quad (65)$$

Capítulo 3

Estudos da conformação dos compostos

3.1. Estudo da conformação dos furanoditerpenos

O gráfico da figura 3.1 mostra a energia de todas as conformações criadas usando o Hamiltoniano do AM1, por rotação das ligações. Para o composto Plaunol E (figura 3.2) as ligações que tiveram o ângulo diedro variado, assim como os resultados para os demais métodos semi-empíricos partindo do confômero 25 (figura 3.2) são mostrados na

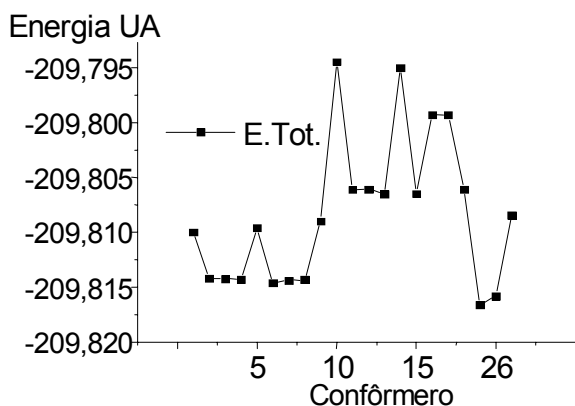


Figura 3.1: Busca sistemática de geometria com AM1 para o composto Plaunol E.

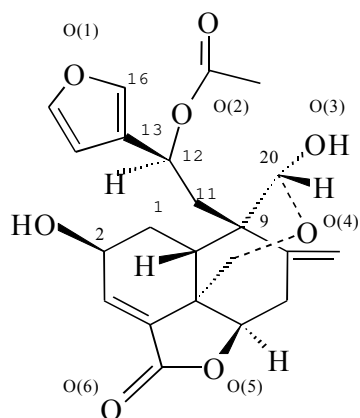
tabela 3.1. As conformações foram obtidas com a rotação das ligações indicadas na figura 3.2, em passos de 120 graus. Esta busca inicial tem como objetivo apenas encontrar as regiões onde se localizam os mínimos de energia, ficando a localização precisa para a minimização da energia do confômero de menor energia.

Tabela 3.1: Ângulos diedro para o mínimo de energia encontrado com os métodos semi-empíricos para o Plaunol E.

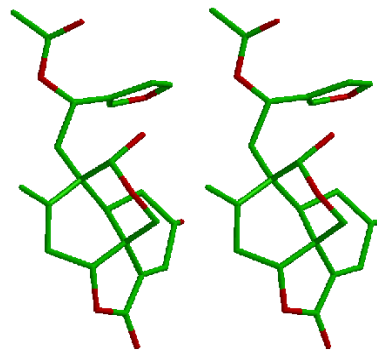
Plaunol E	AM1	PM3	MINDO3	MNDO
11_12_13_16	108	52	58	66
C1_C2_O_H	-63	-64	-64	-64
13_12_O_COCH ₃	-95	-90	-79	-75
C9_C20_O_H	-172	151	179	168

Para o composto Plaunol D (figura 3.3) o gráfico com as energias de todas as conformações criadas (figura 3.3) e os resultados para os métodos semi-empíricos a partir do confômero 23 (figura 3.3), mínimo local no gráfico, estão na tabela 3.2.

Todas as conformações com energia menor que $-187,618$ UA têm o mesmo ângulo diedro para a ligação C12-C13.



Estrutura bidimensional de Plaunol E e numeração dos átomos.

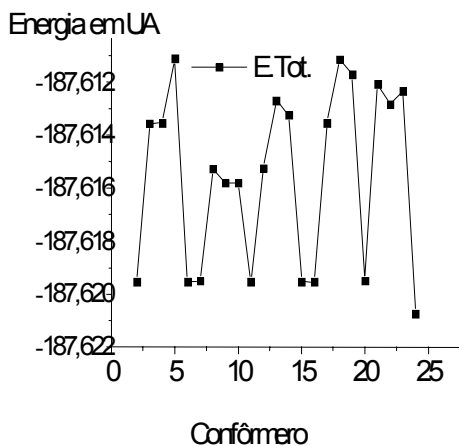


Estereograma do conformero 25 de Plaunol E.

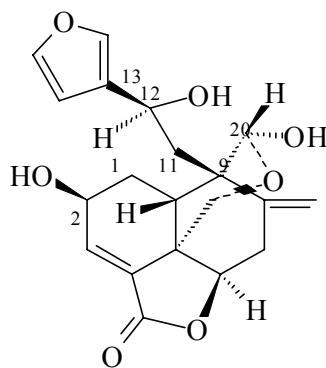
Figura 3.2: O composto Plaunol E.

Tabela 3.2: Ângulos diedro para as conformações de menor energia do composto Plaunol D.

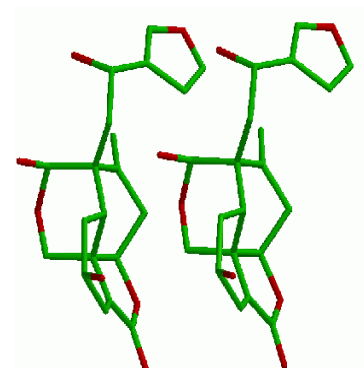
Plaunol D	AM1	PM3	MINDO3	MNDO
11_12_13_16	133	131	110	124
C1_C2_O_H	-64	-65	-63	-66
C13_C12_O_H	-54	-50	-71	-83
C9_C20_O_H	103	70	176	159



Busca sistemática de geometria com AM1. A conformação 23 é mostrada à direita.



Estrutura química do composto Plaunol D.

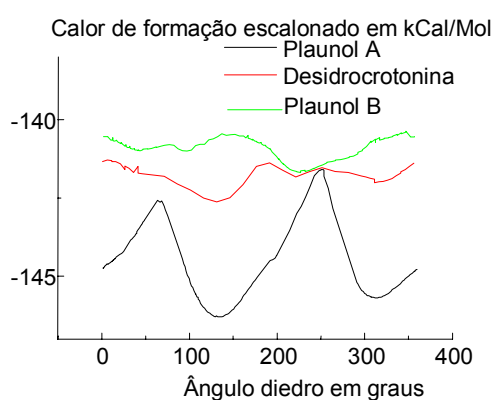


Estereograma do conformero 23 de Plaunol D.

Figura 3.3: O composto Plaunol D e o gráfico da energia total de algumas conformações.

3.1.1 Estudo da barreira rotacional de uma ligação em compostos diferentes

Para comparar os valores dos calores de formação de Plaunol A, Plaunol B, e desidrocrotonina (figura 3.5) foi necessário pré tratamento dos dados. Subtraiu-se 20 kcal/Mol de todos os valores em Plaunol B, e adicionou-se o mesmo valor para Desidrocrotonina, de modo que a faixa de valores fosse semelhante, e pudessem ser mostrados num mesmo gráfico sem grande perda da resolução. O gráfico da figura 3.4



mostra variação na posição de mínima energia do ângulo diedro, notadamente entre Plaunol A e Plaunol B.

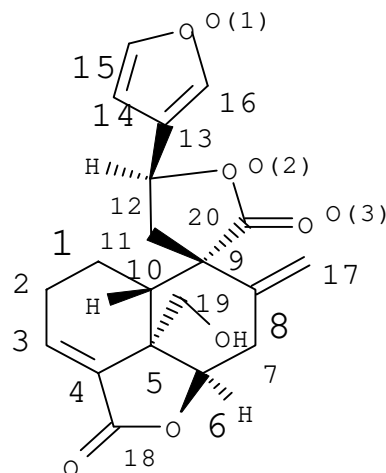
O fechamento do anel na posição C19-O(3) desloca e acentua a barreira rotacional para a ligação C12-C13 em Plaunol A. Observa-se que o carbono assimétrico C12 tem configuração **S** na desidrocrotonina, ao contrário dos demais compostos.

Figura 3.4: Energia rotacional da ligação C12-C13 para os compostos Plaunol A, Plaunol B e Desidrocrotonina, encontrada com

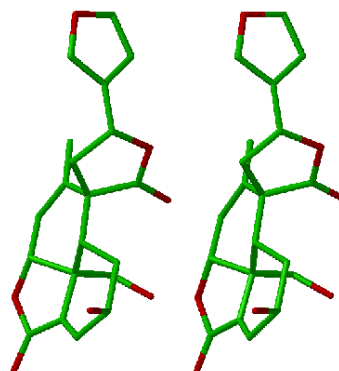
O composto denominado Plaunol C (figura 3.5), e teve as ligações C12-C13, C5-C19 e C2-OH estudadas. Os resultados com o Hamiltoniano AM1 mostram o ângulo diedro 11_12_13_16 é de $-58,7$ graus, não havendo variação maior que 1 kcal/Mol com a rotação das demais ligações. Para as otimizações com os demais métodos semi-empíricos, a partir do confômero mostrado, os resultados estão na tabela 3.3.

Tabela 3.3: Ângulos diedro calculados para Plaunol C por métodos semi-empíricos.

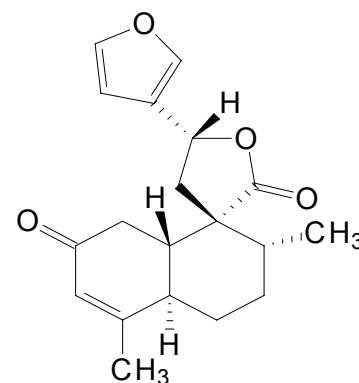
Plaunol C	AM1	PM3	MINDO3	MNDO
11_12_13_16	-59	-90	-161	-125
H_O_C2_C1	175	176	164	148
HO_C19_C5_C4	59	51	53	55



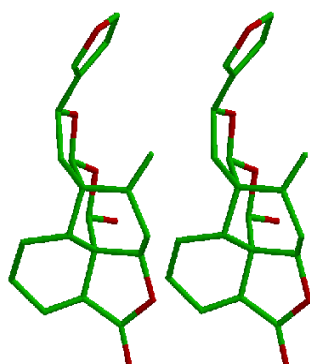
Estrutura química do composto Plaunol B.



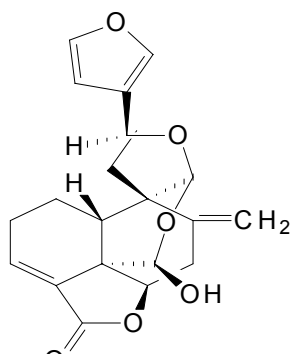
Estereograma para a conformação de mínimo local do composto Plaunol C.



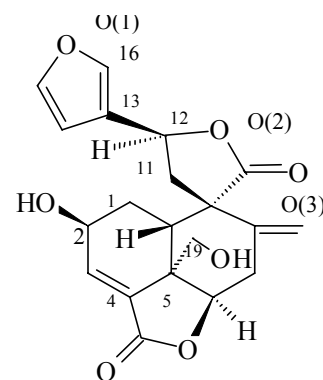
Estrutura química do composto Desidrocrotonina.



Estereograma para a conformação de mínimo local do composto Plaunol A.



Estrutura química do composto Plaunol A.



Estrutura química do composto Plaunol C.

Figura 3.5: Estruturas e estereogramas para as conformações de menor energia dos compostos Plaunol A, Plaunol B, Plaunol C, e Desidrocrotonina.

Figura 3.5: Estruturas e estereogramas para as conformações de menor energia dos compostos Plaunol A, Plaunol B, Plaunol C, e Desidrocrotonina.

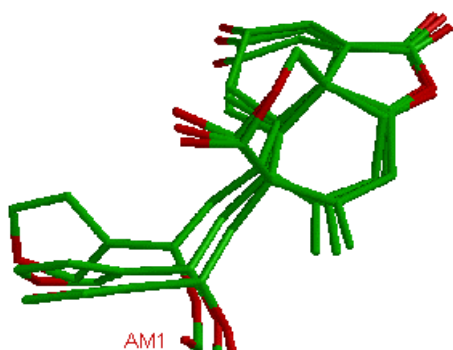


Figura 3.6: Superposição das conformações obtidas com métodos semi-empíricos para o composto Plaunol E.

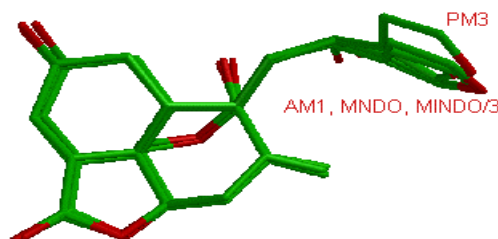
PM3, MINDO3 e MNDO têm grande superposição, a conformação original otimizada com AM1 destaca se das demais, como pode ser observado na figura 3.6.

Para montar esta figura foi feito o ajuste visual da

superposição dos átomos nas moléculas. Uma variação pequena na conformação produz uma grande variação nos ângulos diedro afetados, como pode ser notado comparando se a tabela 3.1 e a figura 3.6.

No composto Plaunol D, a única ligação em que há maior diferença no ângulo diedro é C9_C20_O_H (tabela 3.2). O resultado da superposição das estruturas pode ser visto na figura 3.7.

As
Plaunol B e
variação do
da ligação



conformações criadas para
Desidrocrotonina têm menor
calor de formação na rotação
(figura 3.4) por que não têm o

anel C5_C10_C9_C20_O4_C19 fechado como
em Plaunol A. No caso da Desidrocrotonina a
mudança na configuração do carbono assimétrico
desloca pouco o ângulo de mínimo calor de

formação. O composto Plaunol B é o que tem o mínimo calor de formação num ângulo
maior, próximo de 200 graus.

Figura 3.7: Superposição das conformações obtidas com os métodos semi-empíricos para o composto Plaunol D. As conformações obtidas com AM1, MNDO, e MINDO/3 estão quase totalmente superpostas, o que dificulta a identificação.

3.3. Compostos 4-fenil 2-guanidinothiazóis

Alguns dos 43 compostos com o esqueleto estrutural do composto 54 (figura 3.8), tiveram as ligações giradas para busca do mínimo de energia para cada caso. A escolha dos compostos levou em conta diferenças na estrutura química nas imediações da ligação estudada, de modo a abranger os possíveis substituintes da série.

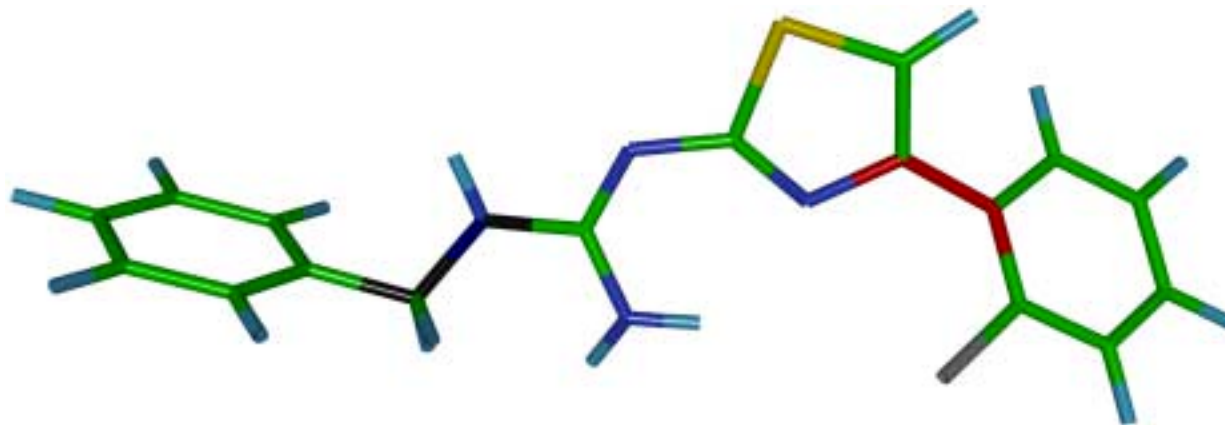
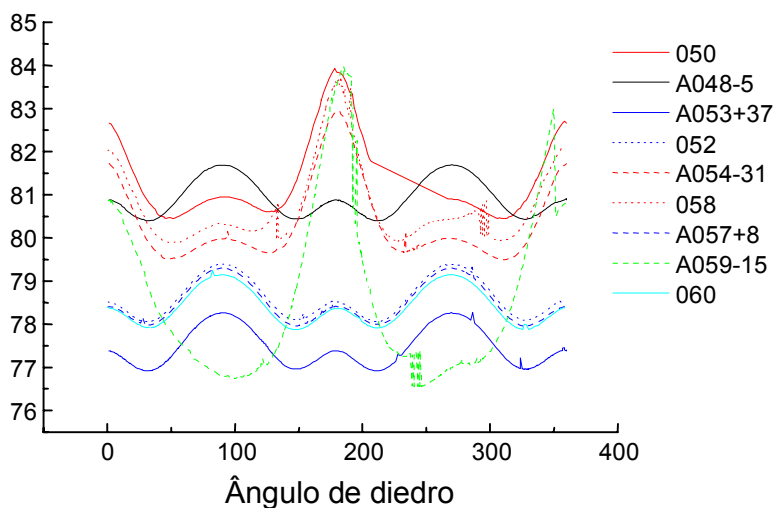


Figura 3.8: Estrutura química do composto 54 (esqueleto da figura 5.20). Ligação 1 em vermelho, e ligação 2 em preto. Átomos: C, verde; H, azul-claro; N, azul escuro; S, amarelo; Cl, cinza; N(4) e C(11), pretos; C(6) e C(7), vermelhos.

Calor de Formação em kcal/Mol



Para os compostos 48, 52, 53, 54, 57, 58, 59, 60, 62, 63 a rotação da ligação 1, entre o anel tiazol e o anel benzênico, que existe em todas as moléculas e está mostrada em vermelho na figura 3.8, produziu os resultados da figura 3.9.

Figura 3.9: Variação do calor de formação em função da rotação da ligação marcada em vermelho na figura 3.8, para alguns compostos do grupo.

A letra “A” antes do número do composto significa que o valor do calor de formação

mostrado no gráfico foi escalado, adicionando-se o valor escrito após o número do composto ao obtido nos cálculos. O composto 48, em preto, não tem substituintes no anel benzênico nem na posição R2 do anel tiazólico. Os compostos em vermelho têm as seguintes substituições: 50, *o*-cloro benzeno. 54, *o*-cloro em R3 e benzila em R2. 58, *o*-metil em R3. Os compostos em azul têm as substituições: 52, *p*-cloro em R3. 53, *p*-fluoro em R3. 57, benzila em R1 e *p*-fluoro em R3. O composto 60, em magenta, tem substituição *m*-metil em R3. O composto 59 tem substituição *o*-trimetil-amina em R3. Para as substituições *para* e *meta* não há orientação preferencial da posição do substituinte em relação ao nitrogênio sp^2 do anel tiazol, e a posição de equilíbrio acontece com os dois anéis formando ângulos em torno de 50 graus entre si. Os compostos *o*-substituídos têm o substituinte preferencialmente próximos ao nitrogênio do anel tiazol e os dois anéis em ângulos em torno de 50 graus, exceto o composto 59, que tem o substituinte volumoso, e fica preferencialmente próximo ao hidrogênio, com ângulo de quase 90 graus entre os anéis. Os gráficos dos compostos 58 e 59 têm variações abruptas no calor de formação nas regiões de 250 e 300 graus por causa da rotação dos grupos metila e dimetilamina

Calor de Formação em kcal/Mol

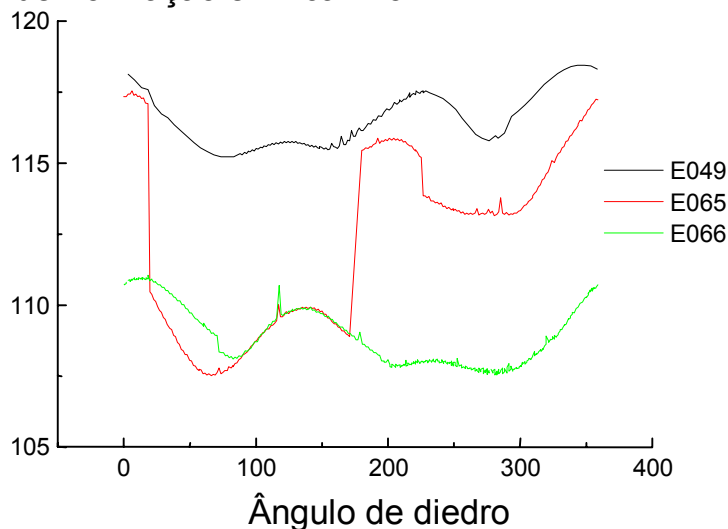


Figura 3.10: Variação do Calor de Formação em função da rotação da ligação mostrada em preto na figura 3.8 para alguns compostos do grupo.

Para a ligação entre o grupo benzila e o nitrogênio do grupo guanidina, mostrado em preto na figura 3.9, os resultados obtidos estão na figura 3.10. Os compostos 49, 65, e 66 foram os protótipos para esta ligação. A figura 3.10 mostra que o ângulo

de diedro para o menor Calor de Formação é aproximadamente 66 graus para os compostos 49 e 65. Este mesmo ângulo de diedro tem valor de aproximadamente -77 graus para o composto 66. A variação abrupta no Calor de Formação para o composto 65 ocorre por causa da inversão da amina secundária, que não aconteceu nos outros casos (figura 3.11).

3.4. Conclusões sobre as conformações dos 4-fenil 2-guanidinothiazóis

Para estes compostos, o estudo das conformações revelou a variação do calor de formação entre duas formas para a amina secundária, que vale 6,53 kcal/Mol (figura 3.10) entre 18,18 e 19,68 graus de rotação do ângulo diedro que liga o Nitrogênio ao

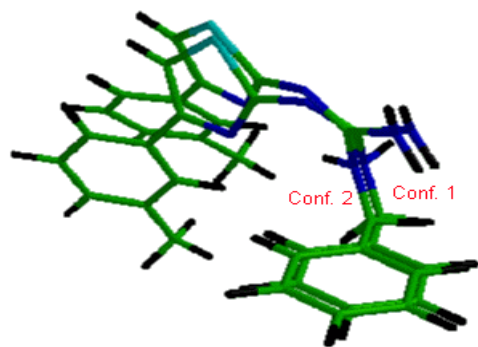


Figura 3.11: Superposição das conformações obtidas na busca sistemática para o mínimo local de energia no compostos 65, realizada com AM1 na ligação N-R1 (ver figura 1.3).

grupo R1 (marcado em verde na figura 3.8). A superposição das dos compostos pode ser observada na figura 3.11. A estrutura marcada como Conf. 1 é a conformação de maior energia, com calor de formação igual à 110,45kcal/Mol. Pode se notar que sua cadeia tem uma conformação mais distendida. A estrutura Conf. 2 tem calor de formação igual a 117,08 kcal/Mol, e há maior proximidade do anel benzênico de R1 em relação ao nitrogênio do.anel tiazol.

Capítulo 4

O estudo dos furanoditerpenos

4.1. A aplicação do Hamiltoniano MINDO3 nos furanoditerpenos

4.1.1. Regressão PCA e HCA

O modelo com parâmetros calculados pelo método MINDO3, autoescalados e com conexão simples para HCA é mostrado a seguir. O composto Plaunol A, inativo, pode ser isolado do grupo formado pelos demais, ativos, distando 1,806 UV da última conexão do grupo ativo. A similaridade entre os elementos do grupo ativo é 0,134 UV, entre Plaunol A

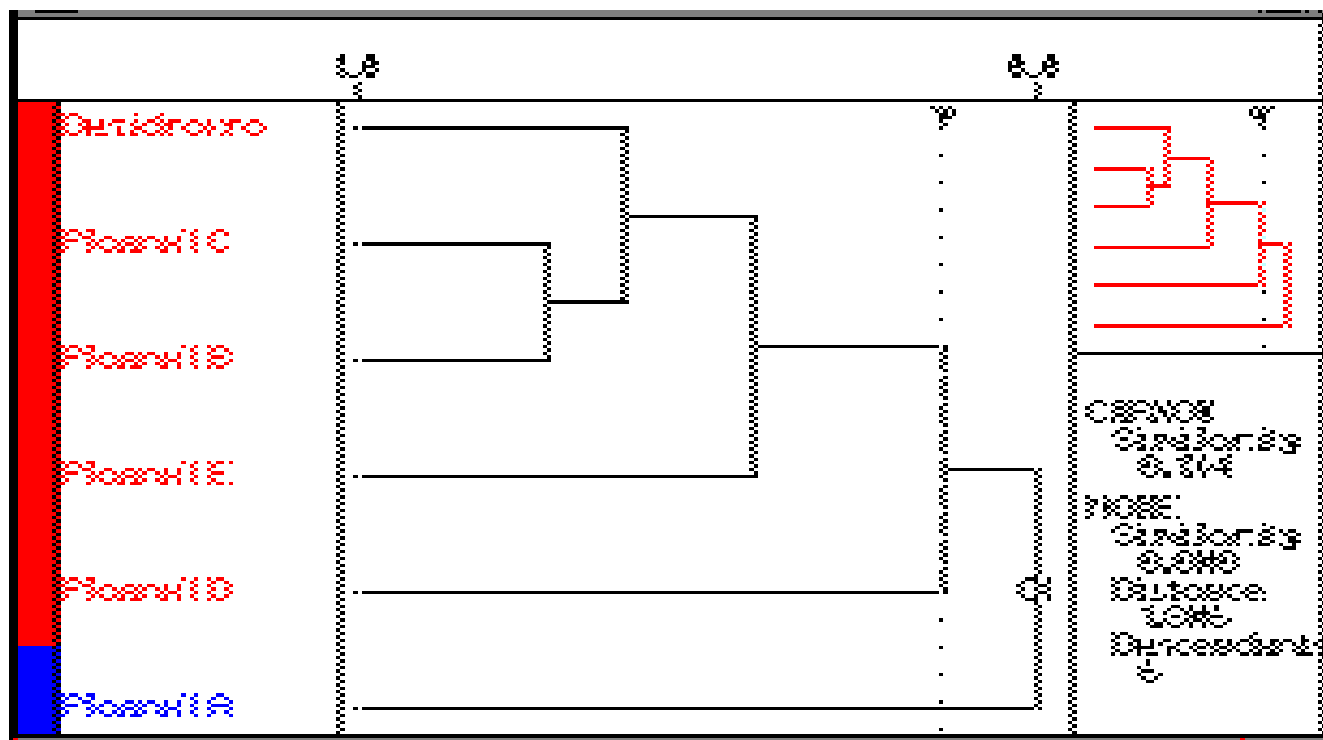


Figura 4.1: Dendrograma mostrando os compostos furanoditerpenos. O composto Plaunol A, em azul, é inativo, os demais, em vermelho, são ativos na inibição de úlcera em ratos.

e o grupo ativo é 0,0 UV como pode se observar no dendrograma da figura 4.1.

A seleção das variáveis foi feita graficando-se os dados calculados com métodos semi-empíricos dois a dois (carga em alguma posição e energia de LUMO, por exemplo), e observando-se quais eram as variáveis que isolavam o composto Plaunol A nas regiões

extremas do gráfico, ou seja, quais variáveis tinham valor muito diferente em Plaunol A. Estas variáveis foram utilizadas num modelo inicial, que foi melhorado por tentativa e erro até o melhor resultado, apresentado aqui. As variáveis que permaneceram após a seleção são os momentos de dipolo μ e μ^{car} . A energia de LUMO, e as cargas de Mulliken em dois átomos também foram selecionadas (ver tabela 4.3).

A primeira e a terceira componentes principais são capazes de isolar o composto Plaunol

A num ponto de valor extremo no gráfico dos 'scores' na figura 4.3. A primeira PC acumula em si uma variância de 57%, e a terceira acumula 15%, totalizando 73% da variância total dos dados originais. A variância de cada PC e seus coeficientes de combinação linear podem ser vistos na tabela 4.1. A primeira componente é o resultado da composição de todas as variáveis com coeficientes de módulo próximos. Não é possível fazer atribuições em relação à capacidade de separação para uma ou outra

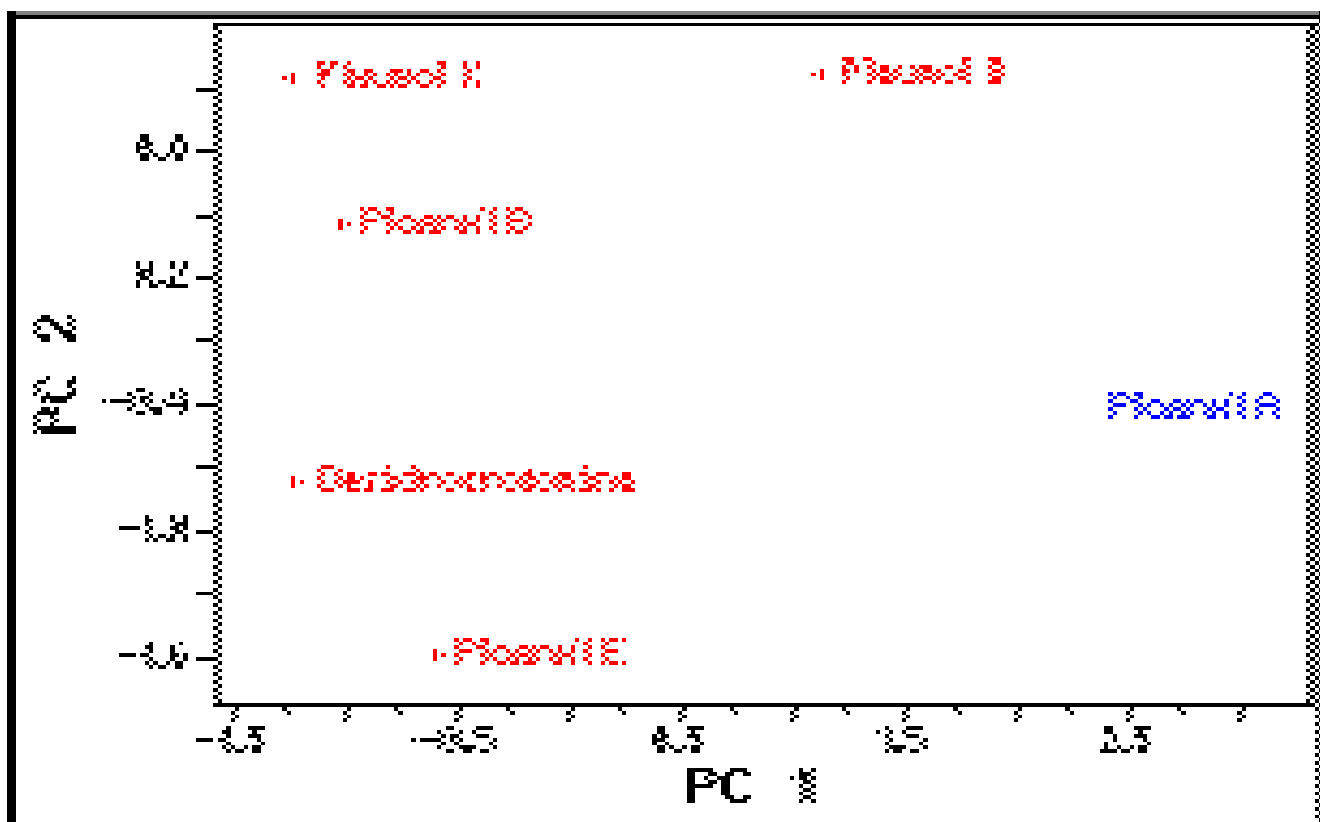


Figura 4.2: Gráficos das componentes principais mostrando os compostos furanoditerpenos. O composto Plaunol A, em azul, é inativo, os demais, em vermelho, são ativos.

variável relacionada, e sim às relações entre elas. Os coeficiente da variável $Q_{C(20)}^{Mult}$ têm influência discordante na primeira e na terceira componente. A ordem decrescente do valor da coordenada na primeira e na segunda componente corresponde a ordem das atividades biológicas dos compostos na figura 4.2.

Na tabela 4.2 estão todos os parâmetros calculados para os furanoditerpenos com o Hamiltoniano MINDO3.

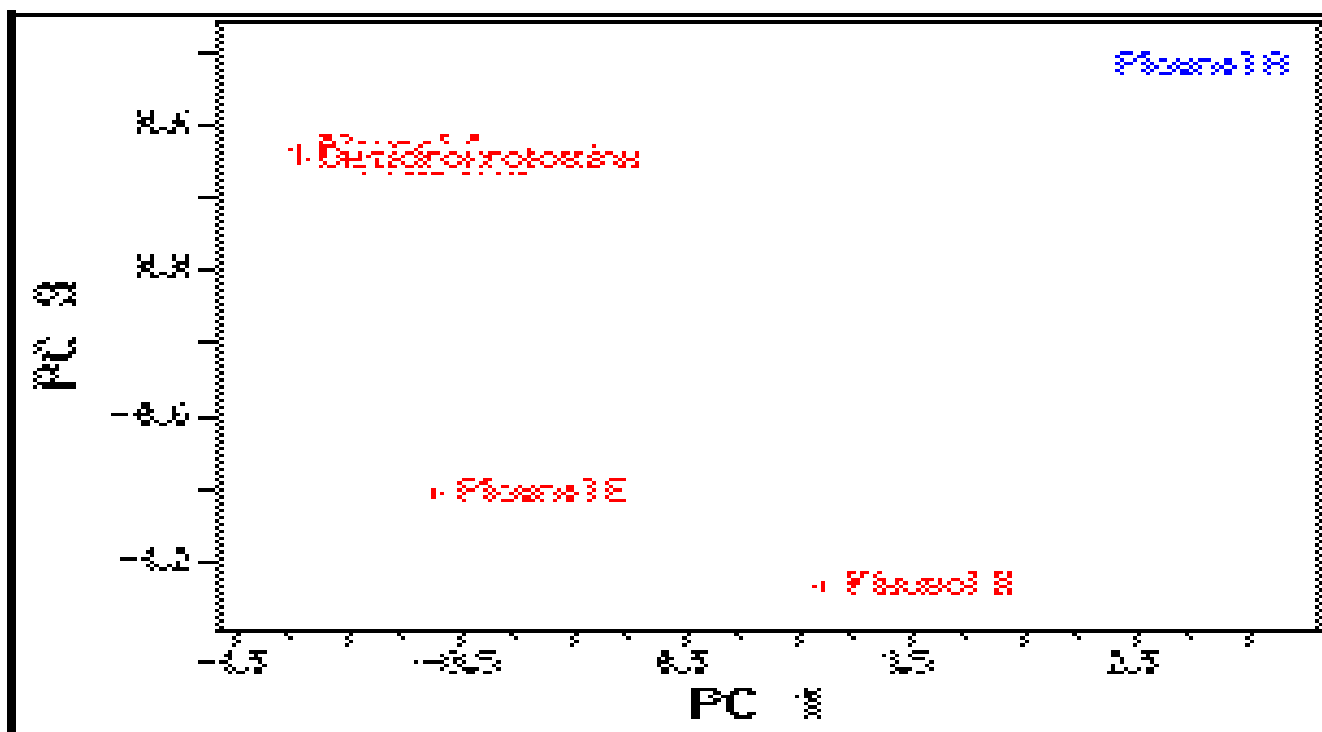


Figura 4.3: Gráfico das componentes principais dos compostos furanoditerpenos. O composto Plaunol A, em azul, é inativo, os demais, em vermelho, são ativos.

Tabela 4.1: Autovetores e variâncias acumuladas para as PCs no modelo construído com dados do MINDO3.

	PC 1	PC 2	PC 3
$\epsilon(\text{LUMO})$	-0,3082	0,75117	-0.0144
μ^{car}	0,50718	0,45800	-0,0069
μ	0,52924	0,35535	-0,0616
$Q^{\text{Mull}}_{\text{O}(1)}$	0,41441	-0,1745	0,76323
$Q^{\text{Mull}}_{\text{C}(20)}$	-0,4426	0,26308	0,64298
Variância Acumulada	57%	25%	15%

Tabela 4.2: Parâmetros calculados com MINDO3 para os furanoditerpenos.

(ΔH_f)	(μ^{hib})	(β^{E4})	$(Q^{Mull}_{o(2)})$	
-133,98399	1,080	40,302	-,3045	Plaunol A
-157,17092	,739	82,679	-,3054	Plaunol B
-189,58277	,217	119,662	-,3059	Plaunol C
-190,60403	1,139	-58,598	-,3108	Plaunol D
-241,97928	,408	19,950	-,3081	Plaunol E
-93,65634	,741	-81,155	-,3082	Desidroc.
$(\epsilon HOMO)$	(μ)	(β^{dip})	$(Q^{Mull}_{o(1)})$	
-8,71200	5,929	61,325	-,4678	Plaunol A
-8,76365	4,127	88,621	-,5053	Plaunol B
-8,78314	3,958	125,957	-,5043	Plaunol C
-8,62665	5,801	-46,458	-,5156	Plaunol D
-8,61304	3,417	28,497	-,5107	Plaunol E
-8,62821	3,770	-71,236	-,5052	Desidroc.
$(\epsilon LUMO)$	(α^{E4})	(γ^{E4})	$(Q^{Mull}_{o(3)})$	
,14041	28,125659	2,21226E+1	-,5331	Plaunol A
,29386	28,018571	2,04343E+1	-,5262	Plaunol B
,36658	28,525921	2,05957E+1	-,5342	Plaunol C
,28115	28,086598	1,91887E+1	-,5177	Plaunol D
,17177	31,655824	2,54635E+1	-,5275	Plaunol E
,20550	25,391543	1,73700E+1	-,5377	Desidroc.
(μ^{car})	(α^{dip})	(γ^{dip})	$(Q^{Mull}_{c(20)})$	
4,884	28,096795	2,67225E+1	,6691	Plaunol A
3,587	27,972435	2,50748E+1	,7901	Plaunol B
3,787	28,475562	2,76255E+1	,7928	Plaunol C
4,705	28,045434	3,08023E+1	,6715	Plaunol D
3,043	31,610688	3,47295E+1	,6703	Plaunol E
3,034	25,332490	2,65672E+1	,7894	Desidroc.

4.1.2. Regressão PLS

Para este caso a regressão PLS foi realizada com 5 variáveis independentes: energia do LUMO, momentos de dipolo μ e μ^{car} , a carga de Mulliken calculada sobre o oxigênio(1), e sobre o carbono(20). As propriedades que foram selecionadas com PCA/HCA são mostradas na tabela 4.3. As variáveis dependentes são as atividades biológicas medidas em dosagens de 3mg/Kg e 10mg/Kg. Os compostos que participam do conjunto de treinamento são Plaunol B, Plaunol C, Plaunol D e Plaunol E. O composto que teve o valor da atividade previsto foi a Desidrocrotonina. A seguir são apresentados os

procedimentos no programa MATLAB™, com o pacote complementar PLS_Toolbox instalado^{li}.

Tabela 4.3: Parâmetros calculados com MINDO3 utilizados na regressão PLS

	ϵ (LUMO)	μ	μ^{CAR}	$Q_{O(1)}^{Mult}$	$Q_{C(20)}^{Mult}$
Plaunol B	0,2939	3,5870	4,1270	0,5053	0,7901
Plaunol C	0,3666	3,7870	3,9580	0,5043	0,7928
Plaunol D	0,2812	4,7050	5,8010	0,5156	0,6715
PlaunolE	0,1718	3,0430	3,4170	0,5107	0,6703

Programa 4.1: Rotina de regressão PLS e predição de valores para MATLAB™.

```

1» load plmip1s2.txt -ascii
2» [min3_ax,min3_mx,min3_stdv]=auto(plmip1s2);
3» atividade2=[55 36 44 52 ; 85 88 61 82];
4» [atv2_ay,atv2_my,atv2_stdv]=auto(atividade2');
5» desi=[0,2055 3,0340 3,77 -0,5052 0,7894];
6» [desi_ax,desi_mx,desi_stdv]=auto(desi);
7» [p,q,w,t,u,b,ssqdif]=pls(min3_ax,atv2_ay,3);
8» ypred=plspred(desi_ax,b,p,q,w,3);
9» atividade=[85 88 61 82 52];
10» [atv_ay,atv_my,atv_stdv]=auto(atividade');
11» atv_ay
atv_ay =
0,7093
0,8960
0,7840
0,5227
1,3440

12» ypred
ypred =2,355      -1,295

13» tmp=ypred.*atv2_stdv
14» tmp+atv2_my
ans =64,131      63,133

```

O procedimento inicialmente carrega a matriz **plsmip1s2.txt** na memória (1). Esta matriz está mostrada na tabela 4.3. Em seguida a matriz é autoescalada, e o resultado é colocado na matriz **min3_ax** (2). As atividades biológicas medidas em dosagens de 10mg/Kg, e 3mg/Kg são carregadas via teclado na matriz **atividade2** (3), que também é autoescalada (a transposição da matriz é indicada pelo “ ’ ” após seu nome), e o resultado colocado na matriz **atv2_ay**, **atv2_my**, e **atv2_stdv**, que são os valores autoescalados, a média dos valores, e a estimativa do desvio padrão, respectivamente (4). Os parâmetros moleculares calculados com MINDO3 para o composto desidrocrotonina (última linha da

tabela 4.4) são colocados no vetor **desi** (5). O vetor é autoescalado, e os resultados colocados em **desi_ax**, **desi_mx**, e **desi_std** (6). A regressão PLS é obtida nas matrizes **b**, **p**, **q**, e **w**, usando os dados de **min_ax**, **atv2_ay**, com 3 variáveis latentes (7). Com estas matrizes e as propriedades autoescaladas da desidrocrotina, no vetor **desi_ax**, o programa calcula a atividade do composto. O valor obtido na predição para as atividades com 3mg/Kg e 10mg/Kg são colocados em **ypred** (8). Os valores das atividades para com dose de 10mg/Kg para todos os compostos foram escritos no vetor **atividade** (9), autoescalados com resultado em **atv_ay** (10). O valor grifado em **ypred** (12) pode ser diretamente comparados com o valor experimental correspondente, autoescalado, grifado na última linha do vetor **atv_ay** (11). Para retornar os valores à escala original utilizou-se a estimativa do desvio padrão, **atv2_std**, a média dos valores de **atividade2**, **atv2_my**, o valor da previsão **ypred** (13), e a variável **tmp** criada para armazenar os valores intermediários (14).

O valor medido experimentalmente para a atividade biológica da Desidrocrotina é 52, e o valor obtido com o modelo (63,13) está 21% deslocado em relação ao experimento. Um outro modelo foi construído com a rotina 'plsconvr', do pacote PLS_Toolbox, que cria diversos modelos eliminando sucessivamente cada uma das amostras, e ordena de diferentes maneiras as restantes para construir cada modelo, mostrando ao final qual o melhor caso. A validação cruzada para este caso usa a função SDEP para formar os conjuntos de teste. Neste procedimento o conjunto de dados é rearranjado repetidamente, e o PRESS é calculado para aquela ordem particular. Os dados utilizados como entrada na matriz, chamada **plminpls**, são mostrados na tabela 4.4.

Tabela 4.4: Parâmetros calculados com MINDO3 utilizados na construção do modelo “deixe um de fora” com regressão PLS.

	ϵ (LUMO)	μ	μ^{CAR}	$Q_{o(1)}^{Mul1}$	$Q_{c(20)}^{Mul1}$
Plaunol B	0,2939	3,5870	4,1270	-0,5053	0,7901
Plaunol C	0,3666	3,7870	3,9580	-0,5043	0,7928
Plaunol D	0,2812	4,7050	5,8010	-0,5156	0,6715

Plaunol E	0,1718	3,0430	3,4170	-0,5107	0,6703
Desidrocrotonina	0,2055	3,0340	3,7700	-0,5052	0,7894

Programa 4.2: Rotina para realizar teste de validação cruzada tipo “deixe um de fora” com MATLAB™.

```
1» load plminpls.txt -ascii
2» [min_ax,min_mx,min_std]=auto(plminpls);
3» [prs,cp,mlv,b1,w1,p1,q1,t21,tv]=plscvrd(min_ax,atv_ay,10,10,3,0,0);
```

Tabela 4.5: Percentual de variância capturado em cada uma das variáveis latentes (LVs) com o Modelo PLS para os dados calculados com MINDO3.

---Bloco X---			---Bloco Y---		
LV #	Esta	LV Total	Esta	LV Total	
1	51,04	51,04	24,31	24,31	
2	27,11	78,15	32,10	56,41	
3	21,85	100,00	22,74	79,15	

O programa 4.2 carrega os dados de **plminpls.txt** na memória (1), autoescala seus valores colocando o resultado no vetor **min_ax**, a média em **min_mx**, e o desvio padrão em **min_std** (2). A função *plscvrd* utiliza os valores **min_ax**, e **atv_ay** (a matriz **atv_ay** é a mesma utilizada no passo 10 do programa 4.1 para construir o modelo e testar as previsões (passo 3). Os valores em seguida; 10,10,3,0,0; são respectivamente o número de vezes que o modelo é reconstruído para cada arranjo diferente das amostras, o número de iterações realizadas pelo programa para cada reconstrução, o número máximo de variáveis latentes que podem ser usadas na regressão, e dois indicadores que orientam o programa a não parar para perguntar com quantas variáveis latentes será realizada a regressão, e não centrar na média os valores do subconjunto de calibração.

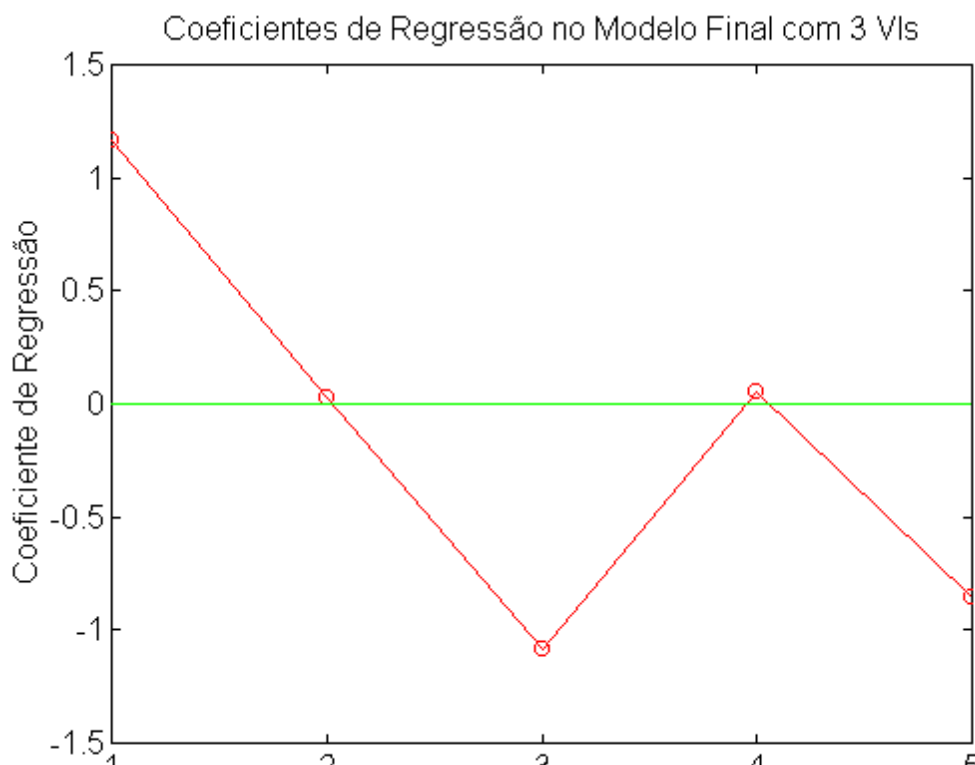


Figura 4.4: Coeficientes de regressão para as variáveis $\epsilon(\text{LUMO})$; μ ; μ^{CAR} ; $Q^{\text{Mull}}_{O(1)}$; $Q^{\text{Mull}}_{C(20)}$ utilizadas no modelo PLS de regressão para as atividades dos furanoditerpenos.

A saída de dados da função é obtida nas variáveis a seguir: **prs** é a soma dos quadrados do resíduo de predição (PRESS), **cp** é o PRESS cumulativo, **mlv** é o número de variáveis latentes com que se obtém o mínimo PRESS, os coeficientes de regressão são armazenados no vetor **b1** (3), e podem ser vistos na figura 4.4, **w1** é a matriz de pesos obtida na regressão, **p1** são os 'loadings' para cada variável do bloco X, **q1** é o limite aproximado dos resíduos do bloco X (figura 4.5), **t21** é o limite T^2 para o bloco X, e finalmente **tv** é a variância acumulada pelo modelo em cada variável latente (tabela 4.5).

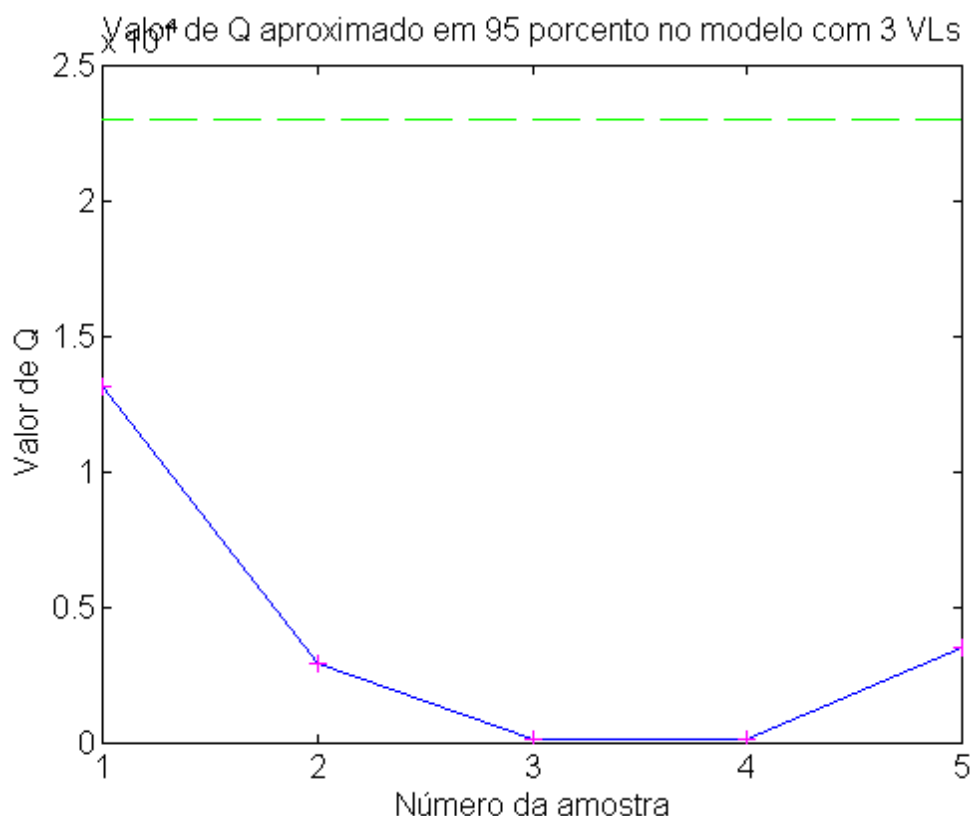


Figura 4.5: Gráfico do valor de Q para a exclusão de cada uma das amostras, e do limite para 95% de confiança, na linha tracejada superior.

Pode-se observar na figura 4.5 o gráfico dos valores de Q (equação 52 do capítulo 2) obtidos para os dados experimentais no modelo para cada amostra excluída. Os melhores resultados são obtidos com a exclusão de amostras com valores de atividade intermediários, que o modelo prevê por interpolação entre os pontos, o caso de Plaunol D e de Plaunol E (amostras 3 e 4 na figura 4.5). O composto usado para previsão nos modelos anteriores foi a Desidrocrotonina (amostra 5 na figura 4.5), e o motivo da escolha foi o fato deste composto ter a atividade biológica medida para apenas uma dosagem das drogas, e não a tentativa de se obter a melhor previsão possível (os compostos Plaunol B e Plaunol C são as amostras 1 e 2 na figura 4.5). Todos os valores são menores que o valor limite estatístico com 95% de confiança para o modelo, de $2,30 \times 10^{-4}$. A pior previsão

foi a da atividade do composto Plaunol B (amostra 1 na figura 4.5), que medida experimentalmente é 85% com dose 10mg/Kg e 55% com dose 3mg/Kg.

4.2. Aplicação do Hamiltoniano MNDO nos furanoditerpenos

4.2.1. Regressão PCA e HCA

Para os parâmetros calculados com o MNDO, autoescalados para PCA e com conexão simples para HCA, resultou no modelo a seguir. O composto Plaunol-A separa-se do centro de conexão com os demais no dendograma da figura 4.6, a uma distância de 1,656 unidades de variância (UV). O grupo formado pelos compostos ativos tem similaridade 0,191 UV, e a similaridade entre Plaunol A e o grupo é 0,0 UV.

Tanto a primeira como a segunda componente isolam o composto Plaunol A num extremo de valor no gráfico dos 'scores' na figura 4.7, acumulando 93% da variância total dos dados.

As variáveis μ^{hib} e μ , os momentos de dipolo calculados considerando os centros de carga com os orbitais híbridos e o momento dipolar total são importantes nas coordenadas em PC1 para Plaunol A e Plaunol D, principalmente.

As polarizabilidades de segunda ordem calculadas por expansão do momento dipolar e

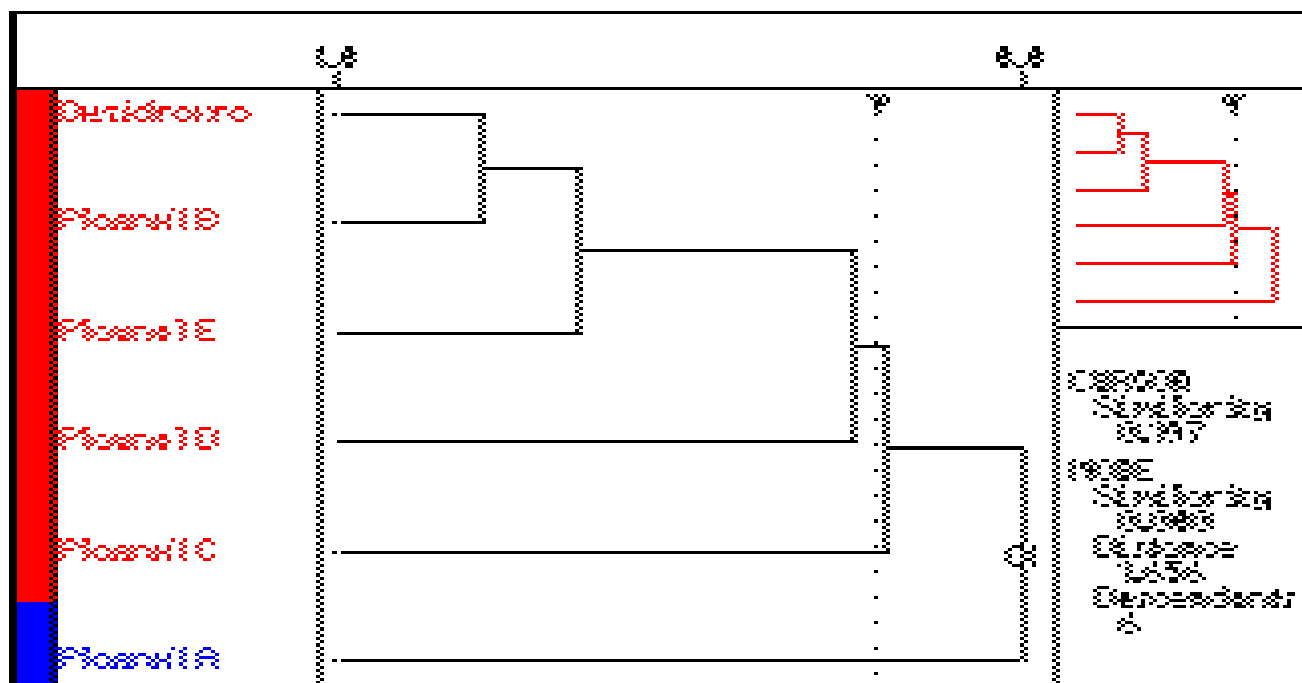


Figura 4.6: Dendrograma para análise HCA com parâmetros do MNDO para os furanoditerpenos. Plaunol A, inativo, aparece em azul.

da energia , β^{dip} e β^{E4} , são propriedades importantes para separar Plaunol B, Plaunol C, Desidrocrotonina, e Plaunol A, e podem ser vistas na tabela 4.7, onde são mostrados os valores de todas as propriedades moleculares calculadas para os compostos usando o método semi-empírico MNDO.

Tabela 4.6: Valores dos coeficientes de combinação linear de cada variável em cada componente principal, e a variância acumulada pela PC.

	PC#1	PC#2	PC#3
μ^{hib}	-0,3794	0,7732	-0,5064
μ	-0,5150	0,2795	0,8090
β^{E4}	0,5431	0,4107	0,1631
β^{dip}	0,5438	0,3940	0,2497
Variância acumulada	71%	22%	7%

A segunda componente tem todas as contribuições com o mesmo sinal, e separa o grupo ativo do composto inativo no gráfico da figura 4.7. Não é possível especificar apenas um parâmetro responsável pela separação entre os conjuntos ativo e o composto inativo. O modelo tem contribuição bem distribuída entre os parâmetros. Um maior número de parâmetros significa que o modelo será mais sensível à presença de anomalias no

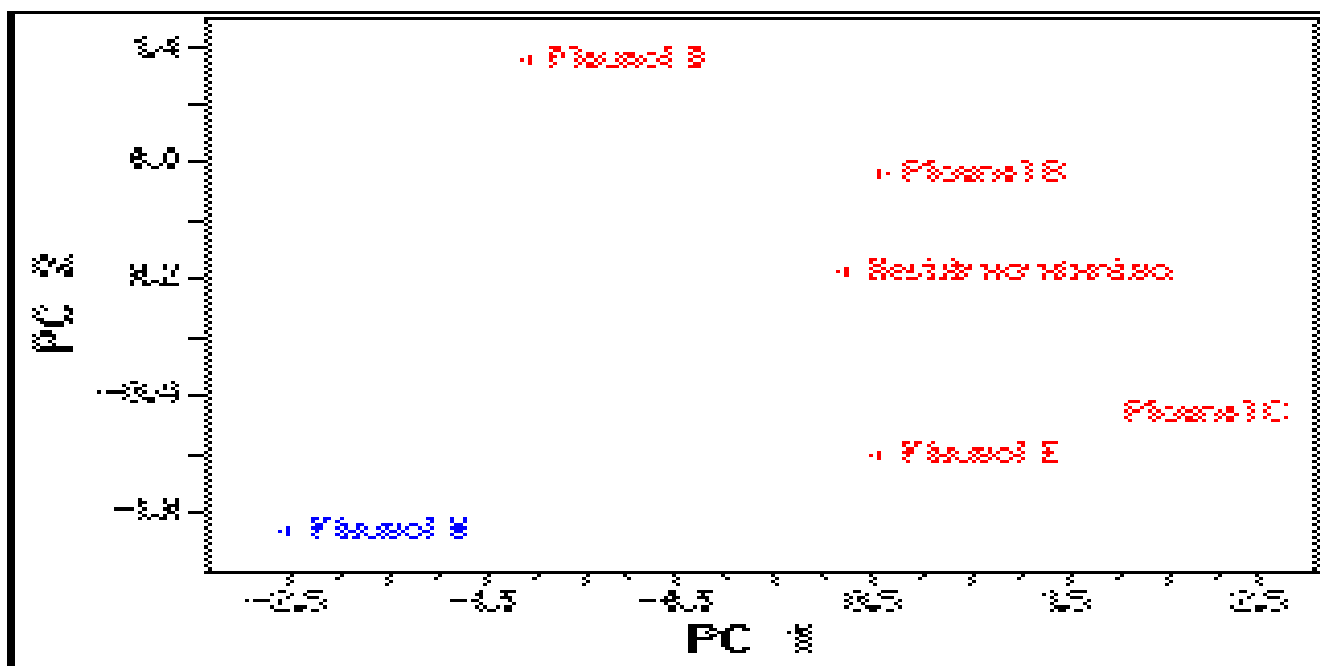


Figura 4.7: Primeira e segunda PCs do modelo PCA com parâmetros calculados com MNDO. O composto Plaunol A, inativo, está em azul.

conjunto de amostra.

Tabela 4.7: Parâmetros calculados com MNDO para os compostos furanoditerpenos.

(ΔH_f)	(ϵ_{HOMO})	(ϵ_{LUMO})	(μ^{car})	(μ^{hib})	
–	9,40663	,38575	4,704	,824	Plaunol A
136,28177					
–	9,50063	,42908	3,450	,992	Plaunol B
152,42970					
–	9,53643	,24817	3,817	,299	Plaunol C
181,27649					
–	9,44893	,42155	4,583	1,232	Plaunol D
193,40576					
–	9,39047	,51403	3,429	,659	Plaunol E
230,95145					
–61,51885	9,37477	,34624	3,056	,921	Desidrocl.
(μ)	(α^{E4})	(α^{dip})	(β^{E4})	(β^{dip})	
5,516	37,054487	25,790308	–91,333	–88,503	Plaunol A
4,146	36,902417	25,637633	12,282	11,313	Plaunol B
3,724	37,736414	26,239769	39,803	45,986	Plaunol C
5,743	37,883496	25,923386	–18,715	–10,498	Plaunol D
3,673	41,969707	28,794999	–20,752	–11,266	Plaunol E
3,966	34,499887	23,507641	–3,032	–4,406	Desidrocl.
(γ^{E4})	(γ^{dip})	$(Q^{Mull}_{O(2)})$	$(Q^{Mull}_{O(1)})$	$(Q^{Mull}_{O(3)})$	
1,20615E+4	1,30058E+4	–0,1186	–0,2828	–0,3250	Plaunol A
1,20901E+4	1,74405E+4	–0,1170	–0,2803	–0,3010	Plaunol B
1,18561E+4	7,38416E+3	–0,1181	–0,2766	–0,3008	Plaunol C
1,19143E+4	5,15569E+4	–0,1227	–0,3669	–0,3426	Plaunol D
1,41460E+4	3,42753E+4	–0,1211	–0,3678	–0,3455	Plaunol E
1,20688E+4	2,04912E+4	–0,1211	–0,2868	–0,2992	Desidrocl.
$(Q^{Mull}_{C(20)})$					
0,2888	Plaunol A				
0,3350	Plaunol B				
0,3295	Plaunol C				
0,3022	Plaunol D				
0,3002	Plaunol E				
0,3327	Desidrocl.				
.	.				

4.2.2. Regressão PLS com os dados selecionados com PCA/HCA

Na tabela 4.8 estão os dados selecionados por PCA para o modelo obtido com resultados do MNDO. O composto Plaunol A não foi usado na tentativa de se obter um modelo quantitativo da atividade, uma vez que o valor zero de atividade deste composto está muito distante do valor dos demais.

Tabela 4.8: Parâmetros calculados com MNDO e utilizados na obtenção do modelo regressão PLS tipo “deixe um de fora”.

	μ^{hib}	μ	β^{E4}	β^{DIP}
Plaunol B	0,9920	4,1460	12,282	11,313
Plaunol C	0,2990	3,7240	39,803	45,986
Plaunol D	1,2320	5,7430	-18,715	-10,498
Plaunol E	0,6590	3,6730	-20,752	-11,266
Desidrocrotonin	0,9210	3,9660	-3,0320	-4,4060
a				

Programa 4.3: Rotina para tratamento dos dados e previsão da atividade de um composto.

```

1» load mnd_pls.txt -ascii
2» atividade = [ 85 88 61 82 52 ];
3» atividade2 = [ 85 88 61 82 ; 55 36 44 52 ];
4» [x,mx,stdx]=auto(mnd_pls);
5» [y,my,stdy]=auto(atividade');
6» [p,q,w,t,u,b,ssqdif]=pls(x,y,3);
7» [y2,my2,stdy2]=auto(atividade2');
8» [x2,mx2,stdx2]=auto(mnd_pls2);
9» [p2,q2,w2,t2,u2,b2,ssqdif2]=pls(x2,y2,3);
10» cross=[0,9210 3,9660 -3,0320 -4,4060];
11» [across,mcross,stdcross]=auto(cross);
12» y2pred=plspred(across,b2,p2,q2,w2,3);
13» y2pred
y2pred =
1,3938 -0,7469

14» y
y =
0,7093
0,8960
0,7840
0,5227
1,3440

15» tmp=y2pred.*std2y;
16» tmp+m2y
ans = 61,9300 40,3725

```

Tabela 4.9: Percentual de variância capturada em cada variável latente (VL) pelo modelo PLS1 para os dados do MNDO.

Esta	LV	-----Bloco-X-----		-----Bloco-Y-----		LV #
		Total	Esta LV	Total		
1		75,12	75,12	41,66	41,66	
2		14,85	89,97	1,49	43,15	
3		9,94	99,91	1,70	44,86	

Tabela 4.10: Percentual de variância capturada em cada variável latente pelo modelo PLS2 para os dados do MNDO.

Esta	LV	-----Bloco-X-----		-----Bloco-Y-----		LV #
		Total	Esta LV	Total		
1		76,79	76,79	47,44	47,44	
2		18,50	95,29	30,59	78,03	
3		4,71	100,00	21,97	100,00	

Na tabela 4.8 são mostrados os dados de entrada utilizados na construção do modelo PLS, e que foram armazenados na matriz **mnd_pls**. A matriz **mnd_pls2**, que também é utilizada como conjunto de treinamento do modelo tem os dados para o composto Desidrocrotonina eliminados da última linha da tabela 4.8. Este composto foi eliminado porque tem sua atividade medida apenas na dose de 10mg/Kg, e para criar um modelo com duas respostas, que são as atividades biológicas medidas para as diferentes dosagens dos plaunóis, não poderia ser usado no conjunto de treinamento. O valor das atividades para os outros compostos é colocado em **atividade2**.

A rotina de cálculo realizada no MATLAB™ é mostrada no programa 4.3. Nos passos 1, 2, e 3 os dados são carregados na memória do computador a partir de um arquivo de texto (1), e do teclado (2 e 3). Em 4 e 5 as variáveis e as respostas são autoescaladas (note a transposição do vetor **atividade**), em 6 faz-se PLS com três variáveis latentes, para uma resposta (**y**, que contém os valores autoescalados de **atividade**). O termo resposta é usado como sinônimo de atividade biológica. A tabela 4.9 mostra que a variância capturada das variáveis independentes, do bloco X, na primeira variável latente é bem maior que a capturada nas variáveis dependentes, e que o aumento do número de variáveis latentes não soluciona o problema.

Nos passos 7 e 8 as variáveis e as respostas foram autoescaladas para construção do modelo com duas variáveis dependentes, em 9 foi feita a regressão PLS para o modelo com 3 variáveis latentes. A variância capturada nas duas primeiras variáveis latentes é bem mais aceitável para construção de um modelo de previsão, como pode ser visto na tabela 4.10. O aumento no tamanho do conjunto de variáveis dependentes, no bloco Y aumenta a variância capturada, mesmo que se considere apenas uma variável latente. Aqui fica clara a melhora na construção de um modelo com o aumento do número de variáveis, no caso as do bloco dependente, explorada pela metodologia de regressão PLS.

No passo 10 os valores dos dados da Desidrocrotonina (última linha na tabela 4.8) foram carregados via teclado, e os seus valores autoescalados (passo 11). O modelo criado (passo 12) para duas respostas foi usado para prever a atividade da Desidrocrotonina utilizando os dados calculados para este composto, autoescalados (**across**, passo 11), vetor de regressão **b2**, os loadings das variáveis latentes do bloco X **p2**, os loadings das variáveis do bloco Y **q2**, a matriz de pesos **w2** (passo 9), e o número de variáveis latentes utilizadas (três variáveis latentes).

O valor previsto (-1,3938), grifado, na primeira coluna de **y2pred** (passo 13) pode ser comparado com o valor autoescalado da atividade, medido experimentalmente (-1,3440), na última linha de **y**, e também grifado. A diferença entre os valores autoescalados é de 3,7%. O valor autoescalado deve passar pelo processo inverso para ser comparado com o valor experimental medido, obtido na literatura. O valor é obtido da função de autoescalamento (equação 36 do capítulo 2), isolando no primeiro membro x_{ij} , nos passos 15 e 16.

O valor medido experimentalmente para a atividade biológica da desidrocrotonina é 52% de inibição da úlcera induzida por reserpina para aplicação de 10mg/Kg da droga. Para dosagem de 3 mg/Kg não está disponível o resultado experimental para comparação. A

diferença entre o calculado (61,93; grifado na primeira coluna de da variável **ans**, no passo 16) e o medido é de 19% do valor experimental (52) para a dosagem de 10mg/Kg. Pode-se observar que aumenta a diferença percentual entre o valor experimental e o valor calculado ao ser revertido o processo de autoescalamiento. Isso ocorre porque a atividade da desidrocrotonina tem seu valor situado no limite inferior da escala de inibição. O reverso do autoescalamiento para um valor nesse extremo superestima esse valor, com os coeficientes de regressão de um conjunto que não contém este elemento como é o das atividades, ou Y, aumentando a proporcionalidade do valor autoescalado, em comparação com o caso em que se usa apenas quatro amostras. Para retornar o valor de **ypred** à escala normal foram usados os parâmetros de escalamento da matriz **atividade2**, a média **m2y** e a estimativa do desvio padrão **std2y**.

4.3. Regressão PCA e HCA para os dados do Hamiltoniano AM1 parametrizado para reproduzir solvatação implícita em água

Os dados obtidos com o Hamiltoniano parametrizado conforme descrito no capítulo 2, para as cargas derivadas do potencial eletrostáticoⁱⁱⁱ e cargas de Mulliken com o pacote Spartan4.1 foram selecionados, e submetidos à análise exploratória de dados, com autoescalamento para PCA e HCA, agrupados por conexão simples.

4.3.1. Modelo com cargas CHELP obtidas com Hamiltoniano AM1 parametrizado para reproduzir solvatação em água

Para as cargas derivadas do potencial eletrostático, os resultados mostram o composto Plaunol A separado do grupo ativo por 1,760 unidades de variância, e o grupo ativo com similaridade 0,107 UV no dendograma da figura 4.8. O Composto Plaunol A fica isolado num extremo de valor na primeira componente principal obtida por PCA (figura 4.9), que acumula 70% da variância total dos dados. A segunda acumula 17% e a terceira 12% (tabela 4.11), porém não separam o composto Plaunol A num extremo de valor.

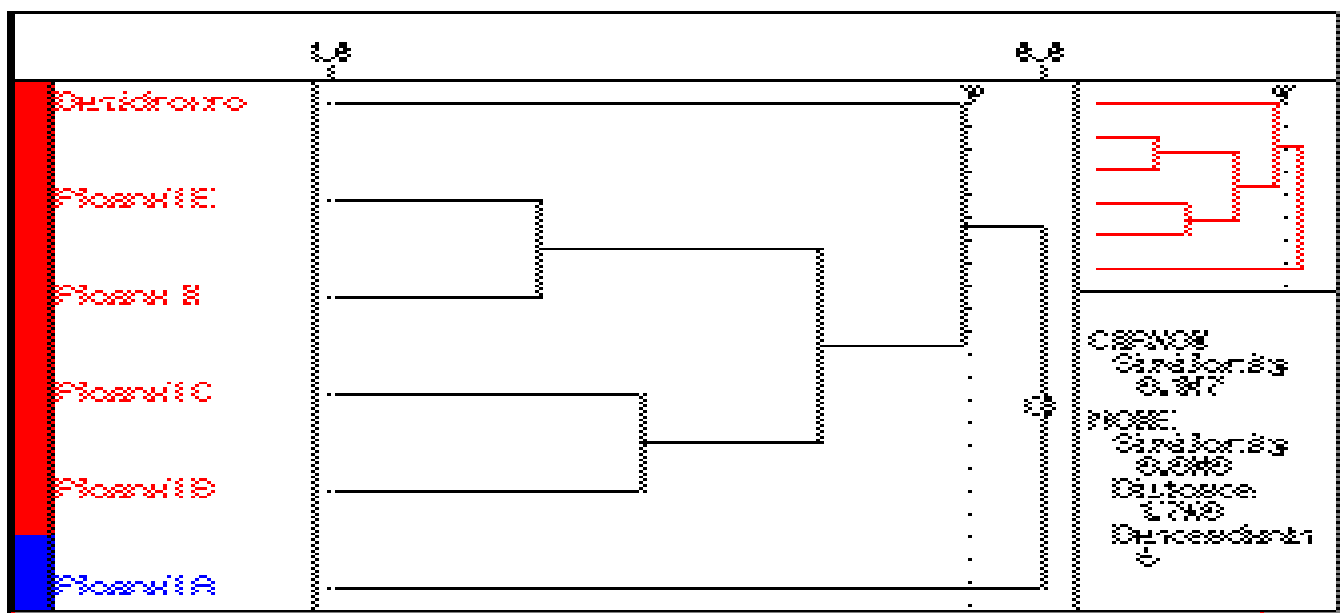


Figura 4.8: Dendrograma da HCA para os diterpenos com as cargas CHELP obtidas com Hamiltoniano AM1 parametrizado para reproduzir solvatação em água. O composto Plaunol A, inativo é mostrado em azul. Os demais em vermelho.

Os resultados mostram que variações da carga nas posições C1,C2,,C19,C20 e O3 relacionam-se com a atividade biológica. Valores muito negativos em C2 são relacionados com baixa atividade. A carga em C20 deve ser positiva, e quanto maior o seu valor, maior a atividade do composto na série. O carbono C19 não existe na Desidrocrotonina, e o valor mais positivo ocorre no Plaunol A, inativo. O carbono C1 deve ter carga negativa, porém o maior valor de atividade não ocorre com a carga mais negativa. Parece haver um ponto ótimo a partir do qual há diminuição da atividade. Na posição O3 a carga deve ser a mais negativa, considerando a presente série de compostos.

Na figura 4.10 são mostrados os ‘loadings’ de cada variável do modelo. O que se vê é que a posição das cargas nas posições C1, C19, e O3 é parecida, ou seja, a variação destes parâmetros tem quase a mesma influência na classificação de um dado composto. As cargas nas posições C2 e C20 estão em quadrantes opostos em PC2, e sua influência é inversa na classificação de um composto por esta PC.

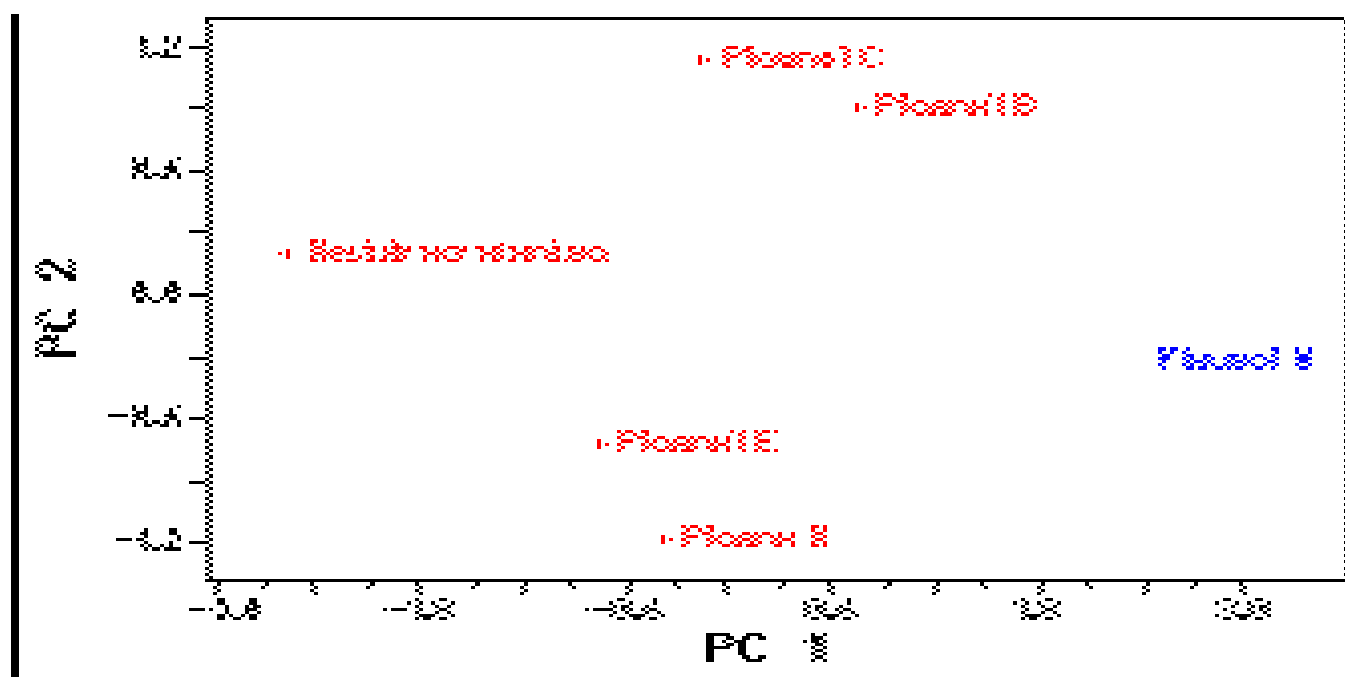


Figura 4.9: Gráfico dos ‘scores’ dos furanoditerpenos na primeira e segunda PCs no modelo com cargas CHELP obtidas com Hamiltoniano AM1 parametrizado para reproduzir solvatação em água.

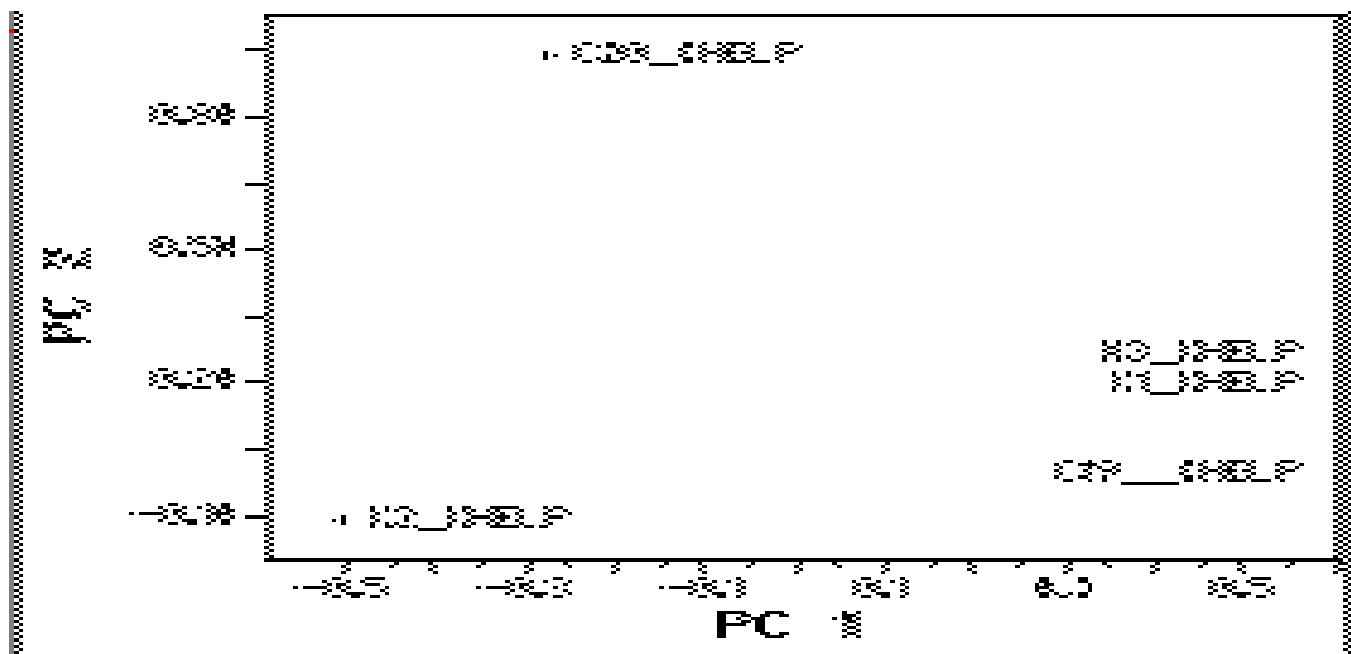


Figura 4.10: Gráfico mostrando os 'loadings' das variáveis na primeira e na segunda PCs no modelo com cargas CHELP calculadas com Hamiltoniano AM1 parametrizado para reproduzir solvatação em água.

4.3.2. Modelo com cargas de Mulliken calculadas com Hamiltoniano

AM1 parametrizado para reproduzir solvatação em água

Para as cargas de Mulliken, os resultados mostram o composto Plaunol A separado do grupo ativo por 1,649 unidades de variância dos demais compostos, e o grupo ativo com

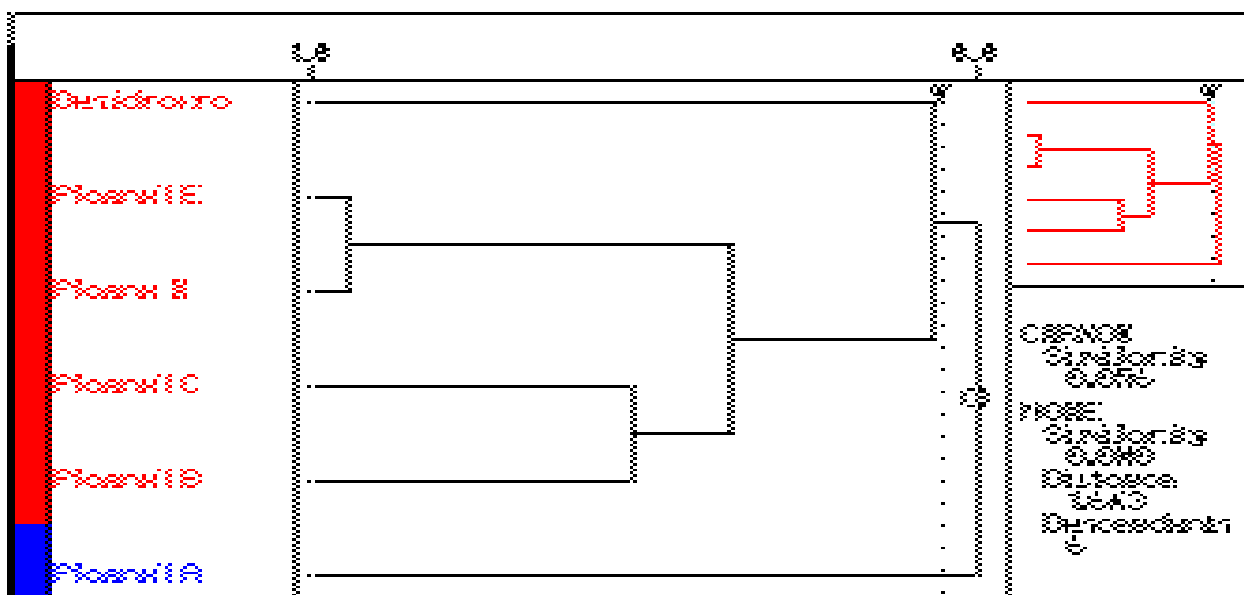


Figura 4.11: Dendrograma para HCA dos furanoditerpenos com cargas de Mulliken calculadas com Hamiltoniano AM1 parametrizado para reproduzir solvatação em água. Plaunol A, inativo, em azul.

similaridade 0,056 UV no dendograma da figura 4.11.

O Composto Plaunol A fica isolado num extremo de valor nas três primeiras componentes principais obtidas por PCA, que acumulam 99% da variância total dos dados (figuras 4.12 e 4.13). Os valores dos coeficientes de combinação e as variâncias acumuladas em cada PC podem ser vistos na tabela 4.11. Este modelo foi construído com as cargas de Mulliken nas posições C2, C3, C19, e C20.

Tabela 4.11: Coeficientes de combinação linear e variâncias acumuladas nas PCs para o modelo com cargas de Mulliken calculadas com Hamiltoniano parametrizado para reproduzir solvatação em água.

	PC#1	PC#2	PC#3
Q_{C2}^{Mull}	-0,5881	-0,3736	-0,1900
Q_{C3}^{Mull}	0,5644	0,4439	-0,2320
Q_{C19}^{Mull}	0,4227	-0,5506	0,6756
Q_{C20}^{Mull}	-0,3958	0,6001	0,6732
Variância acumulada	60%	24%	15%

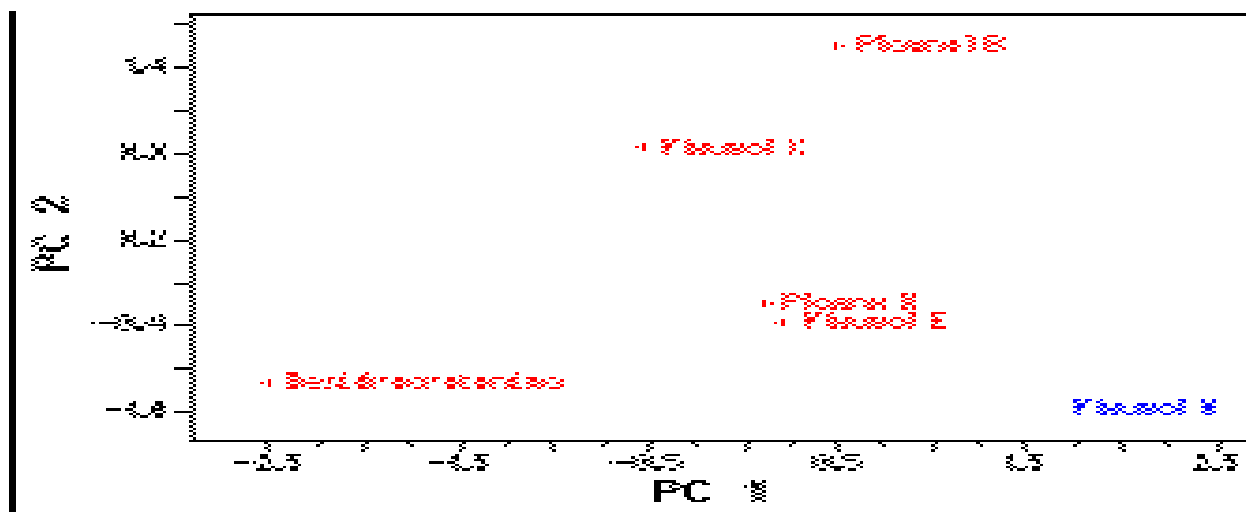


Figura 4.12: Gráfico das duas primeiras PCs para análise PCA dos furanoditerpenos com cargas de Mulliken calculadas com Hamiltoniano AM1 parametrizado para reproduzir solvatação em água.

A influência das cargas nas posições C2, C19 e C20 é idêntica ao que já foi falado sobre estas cargas, obtidas como derivadas do potencial eletrostático (figura 4.13).

Em C3 o maior valor de atividade não ocorre num extremo do valor de carga, e a maior atividade é obtida com cargas ligeiramente menores que zero. O valor mais positivo ocorre em Plaunol A, inativo. Todas as cargas calculadas para todos os átomos, pelos

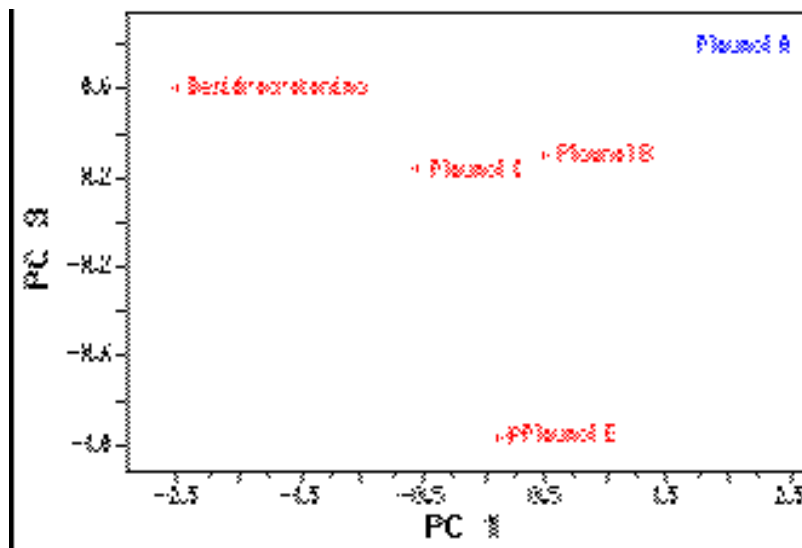


Figura 4.13: Gráfico dos 'scores' da primeira e da terceira PCs. O composto Plaunol A, inativo, aparece em azul, e os demais, ativos, em vermelho.

dois métodos, são mostradas na tabela 4.12.

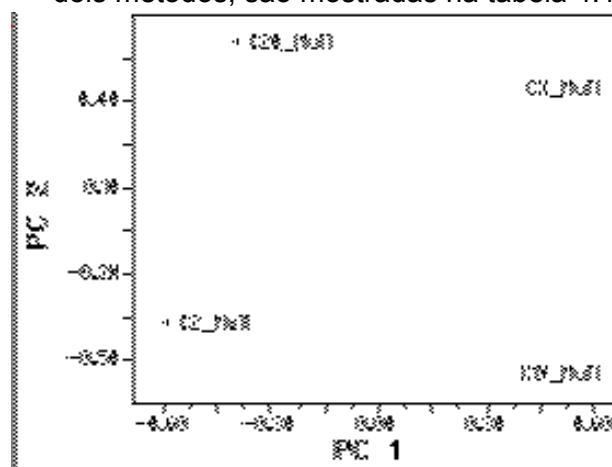


Figura 4.14: Gráfico dos 'loadings' da primeira e segunda PCs para cada variável

Tabela 4.12: Cargas calculadas para os átomos nos compostos furanoditerpenos com Hamiltoniano AM1 parametrizado para reproduzir solvatação em água.

(μ)	($Q^{\text{Mull}}_{\text{C}(2)}$)	($Q^{\text{Mull}}_{\text{C}(5)}$)	($Q^{\text{CHELP}}_{\text{C}(8)}$)	($Q^{\text{Mull}}_{\text{C}(10)}$)	($Q^{\text{CHELP}}_{\text{C}(13)}$)	($Q^{\text{Mull}}_{\text{C}(15)}$)	
6,24535	-0,17	-0,1	-0,03	-0,09	-0,06	-0,1	Plaunol A
7,637936	-0,16	-0,10	-0,10	-0,09	-0,04	-0,10	Plaunol B
8,186815	0,04	-0,10	-0,03	-0,09	-0,07	-0,09	Plaunol C
6,333518	0,05	-0,1	-0,04	-0,1	-0,15	-0,09	Plaunol D
6,372450	0,05	-0,09	-0,15	-0,09	-0,03	-0,1	Plaunol E
5,008195	0,29	-0,11	0,03	-0,08	-0,03	-0,09	Desidroc.
($Q^{\text{CHELP}}_{\text{C}(1)}$)	($Q^{\text{CHELP}}_{\text{C}(3)}$)	($Q^{\text{CHELP}}_{\text{C}(6)}$)	($Q^{\text{Mull}}_{\text{C}(8)}$)	($Q^{\text{CHELP}}_{\text{C}(11)}$)	($Q^{\text{Mull}}_{\text{C}(13)}$)	($Q^{\text{CHELP}}_{\text{C}(16)}$)	
-0,07	0,09	0,24	-0,12	-0,22	-0,16	0,01	Plaunol A
-0,11	0,05	0,10	-0,11	-0,27	-0,19	-0,02	Plaunol B
-0,21	-0,05	0,12	-0,11	-0,18	-0,20	0,07	Plaunol C
-0,29	-0,01	0,32	-0,12	-0,22	-0,23	0,03	Plaunol D
-0,24	-0,04	0,19	-0,13	-0,26	-0,17	0,02	Plaunol E
-0,46	-0,47	-0,12	-0,10	-0,33	-0,18	-0,01	Desidroc.
($Q^{\text{Mull}}_{\text{C}(1)}$)	($Q^{\text{Mull}}_{\text{C}(3)}$)	($Q^{\text{Mull}}_{\text{C}(6)}$)	($Q^{\text{CHELP}}_{\text{C}(9)}$)	($Q^{\text{Mull}}_{\text{C}(11)}$)	($Q^{\text{CHELP}}_{\text{C}(14)}$)	($Q^{\text{Mull}}_{\text{C}(16)}$)	
-0,18	-0,06	0,04	-0,08	-0,20	-0,24	-0,10	Plaunol A
-0,18	-0,05	0,02	-0,05	-0,22	-0,24	-0,09	Plaunol B
-0,22	-0,10	0,02	-0,31	-0,22	-0,32	-0,06	Plaunol C
-0,24	-0,1	0,03	-0,20	-0,20	-0,27	-0,09	Plaunol D
-0,24	-0,1	0,02	-0,11	-0,20	-0,27	-0,09	Plaunol E
-0,26	-0,27	-0,17	0,05	-0,22	-0,23	-0,09	Desidroc.
($Q^{\text{CHELP}}_{\text{C}(2)}$)	($Q^{\text{CHELP}}_{\text{C}(4)}$)	($Q^{\text{CHELP}}_{\text{C}(7)}$)	($Q^{\text{Mull}}_{\text{C}(9)}$)	($Q^{\text{CHELP}}_{\text{C}(12)}$)	($Q^{\text{Mull}}_{\text{C}(14)}$)	($Q^{\text{CHELP}}_{\text{C}(17)}$)	
-0,21	-0,41	-0,19	-0,06	0,14	-0,18	-0,24	Plaunol A
-0,13	-0,43	-0,10	-0,02	0,15	-0,19	-0,22	Plaunol B
0,20	-0,39	-0,09	-0,02	0,19	-0,21	-0,26	Plaunol C
0,24	-0,47	-0,23	-0,05	0,24	-0,21	-0,24	Plaunol D
0,25	-0,36	-0,17	0	0,1	-0,19	-0,12	Plaunol E
0,71	0,15	-0,17	-0,07	0,13	-0,19	-0,27	Desidroc.
($Q^{\text{CHELP}}_{\text{C}(5)}$)	($Q^{\text{Mull}}_{\text{C}(4)}$)	($Q^{\text{Mull}}_{\text{C}(7)}$)	($Q^{\text{CHELP}}_{\text{C}(10)}$)	($Q^{\text{Mull}}_{\text{C}(12)}$)	($Q^{\text{CHELP}}_{\text{C}(15)}$)	($Q^{\text{Mull}}_{\text{C}(17)}$)	
0	-0,18	-0,16	-0,06	0,09	0,04	-0,21	Plaunol A
0,18	-0,18	-0,14	-0,11	0,09	0,04	-0,23	Plaunol B
0,09	-0,16	-0,15	0,09	0,09	0,10	-0,23	Plaunol C
-0,04	-0,17	-0,16	0,25	0,1	0,08	-0,24	Plaunol D
-0,07	-0,17	-0,15	0,20	0,11	0,06	-0,23	Plaunol E
-0,15	-0,02	-0,19	0,04	0,10	0,03	-0,26	Desidroc.
($Q^{\text{Mull}}_{\text{C}(20)}$)	($Q^{\text{CHELP}}_{\text{O}(2)}$)	($Q^{\text{Mull}}_{\text{O}(3)}$)	($Q^{\text{CHELP}}_{\text{O}(5)}$)	($Q^{\text{Mull}}_{\text{O}(6)}$)	($Q^{\text{CHELP}}_{\text{O}(6)}$)	($Q^{\text{CHELP}}_{\text{O}(3)}$)	
0,16	-0,39	-0,30	-0,48	-0,31	-0,52	-0,45	Plaunol A
0,34	-0,44	-0,33	-0,48	-0,30	-0,51	-0,55	Plaunol B
0,34	-0,5	-0,33	-0,48	-0,29	-0,51	-0,55	Plaunol C
0,17	-0,55	-0,39	-0,52	-0,30	-0,52	-0,59	Plaunol D
0,16	-0,37	-0,39	-0,48	-0,30	-0,51	-0,62	Plaunol E
0,34	-0,47	-0,33	0	0	0	-0,56	Desidroc.

$(Q^{CHELP}_{O(1)})$	$(Q^{Mull}_{O(2)})$	$(Q^{CHELP}_{O(4)})$	$(Q^{Mull}_{O(5)})$	$(Q^{Mull}_{O(1)})$	$(Q^{Mull}_{O(4)})$	$(Q^{CHELP}_{C(20)})$	
-0,22	-0,27	-0,56	-0,26	-0,1	-0,33	0,44	Plaunol A
-0,2	-0,26	-0,57	-0,27	-0,1	-0,35	0,81	Plaunol B
-0,25	-0,26	-0,58	-0,27	-0,1	-0,35	0,91	Plaunol C
-0,22	-0,37	-0,47	-0,26	-0,1	-0,28	0,49	Plaunol D
-0,23	-0,28	-0,51	-0,27	-0,11	-0,32	0,60	Plaunol E
-0,20	-0,26	0	0	-0,1	0	0,83	Desidroc.
$(Q^{Mull}_{C(19)})$	$(Q^{CHELP}_{C(19)})$	$(Q^{Mull}_{C(18)})$	$(Q^{CHELP}_{C(18)})$				
0,19	0,34	0,38	0,84				
-0,02	0,15	0,38	0,85				Plaunol B
-0,02	0,17	0,38	0,85				Plaunol C
-0,02	0,13	0,37	0,90				Plaunol D
-0,01	0,09	0,37	0,83				Plaunol E
0	0	-0,23	-0,20				Desidroc.

4.4. Conclusões sobre o estudo dos furanoditerpenos

Os resultados obtidos com o Hamiltoniano MINDO3 mostram que a energia de LUMO,

as cargas de Mulliken nas posições C20 e O1 têm capacidade de modelagem da atividade antiúlcera. O aumento no momento de dipolo está relacionado com uma diminuição da atividade. As variáveis selecionadas têm interpretação conhecida na literatura, e são bastante usadas em estudos de estrutura e atividade. O momento de

dipolo informa sobre equilíbrio em fases líquidas,

relacionando-se com o parâmetro de Taft para

solventes. As energias dos orbitais de fronteira

relacionam-se com a capacidade de ligação química de

uma molécula, o seu caráter doador ou aceptor de

elétrons. Na figura 4.14 podem ser vistos os contornos

da densidade eletrônica total no composto Plaunol C

(pontilhado), e os contornos de LUMO (verde e azul).

O modelo obtido com os dados do Hamiltoniano MNDO

utilizou os momentos de dipolo total, e a sua

componente calculada com os orbitais híbridos, e as

polarizabilidades calculadas por expansão da energia

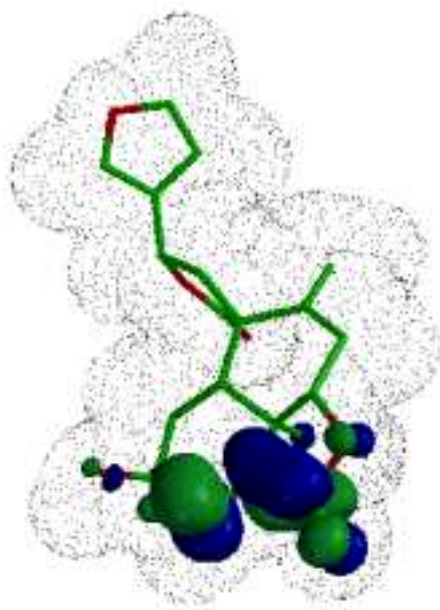


Figura 4.14: LUMO obtido com MINDO3, mostrado no composto Plaunol C (verde e azul), e densidade eletrônica total (pontos).

total e do momento de dipolo. A polarizabilidade também é um parâmetro relacionado com o equilíbrio entre fases líquidas.

A correlação entre estes parâmetros, portanto, faz sentido do ponto de vista químico, e pode ser útil para desvendar o mecanismo de ação destas drogas, ainda desconhecido por nós. A metodologia de análise de dados foi ferramenta fundamental na seleção das variáveis na abordagem do problema de correlação da atividade biológica dos compostos furanoditerpenos com mecanismo de ação desconhecido. Os modelos criados podem ser úteis na previsão da atividade de novos compostos, quantitativa ou qualitativamente, como mostram os resultados da regressão PLS e PCA para os furanoditerpenos, ainda que sejam estabelecidos com pequena base de validação estatística. O modelo com dados do MNDO foi capaz de prever aproximadamente a atividade biológica do composto Desidrocrotonina. O modelo obtido com dados do MINDO3, tipo de fora, fez previsões com alguma validade. Observando-se o gráfico da figura 4.5, pode-se ver que a única previsão pior que a da Desidrocrotonina é o primeiro composto, Plaunol B. É difícil a proposição de uma estrutura do farmacóforo ou sugerir alterações na estrutura para aumentar a potência da droga, que seriam os objetivos finais do trabalho.

As duas metodologias aplicadas fornecem parâmetros úteis na criação de modelos que distinguem os compostos ativos e inativos. Há uma ligeira preferência pelo Hamiltoniano MINDO3, que mostra um maior número de parâmetros capazes de se relacionar com os dados experimentais de atividade biológica. Para o desenvolvimento futuro do trabalho de SAR e QSAR com este grupo de compostos, deverei utilizar ainda os dois métodos, em busca de uma conclusão definitiva sobre qual será o melhor, ficando os resultados obtidos até agora colocados como um ponto de partida de onde deverão seguir pesquisas seguintes.

Foram calculadas cargas pelo método CHELP e de Mulliken, com o Hamiltoniano AM1 com efeito de solvatação implícita. O método CHELP para calcular as cargas permite que

se obtenha melhores resultados na HCA, permitindo a criação de um modelo que separe o composto Plaunol A a uma distância maior dos outros compostos que o modelos com cargas de Mulliken. No caso de PCA, o modelo com cargas de Mulliken permitiu que o composto Plaunol A ficasse isolado num extremo de valor nas três primeiras componentes principais, acumulando 99% da variância total dos dados. O modelo criado com as cargas CHELP separe o composto Plaunol A apenas na primeira PC, acumulando 70% da variância total dos dados. As posições atômicas indicadas por um ou outro modelo coincidem. Os resultados CHELP indicam que as cargas nas posições C1,C2, C19, C20, e O3 podem ser relacionadas com as atividades biológicas . Os resultados com cargas de Mulliken não indicam as posições C1 e O3.

Não há conclusão sobre qual a metodologia de estudo de carga é a mais adequada para este caso. Trata-se de um grupo muito pequeno, e as diferenças obtidas por um e outro método não são grandes, o que não permite suporte estatístico adequado às conclusões. Também neste caso o estudo deve continuar levando em conta as cargas obtidas com os dois métodos. Não foram calculados as polarizabilidades com o Hamiltoniano AM1 parametrizado para reproduzir solvatação em água porque este não faz parte do MOPAC ou do Gamess, e o programa Spartan não as calcula.

Os Hamiltonianos MINDO3 e MNDO parecem ser mais adequados ao estudo destes compostos que o AM1 parametrizado para reproduzir solvatação em água.

Capítulo 5

Os compostos 4-fenil 2-guanidinothiazóis

5.1 Modelos de classificação HCA e PCA

5.1.1. Modelo HCA/PCA com cargas de Mulliken para atividades *in vitro* dos 4-fenil 2-guanidinothiazóis

Os resultados dos testes farmacológicos laboratoriais foram obtidos com H^+, K^+ -ATPase isolada de mucosa gástrica canina. A determinação da atividade enzimática foi feita da maneira descrita por Beil^{liii} e colaboradores, com pequenas modificações. Foi determinado o percentual de inibição da atividade enzimática com uma concentração fixa da droga de 50 μM . Um composto foi considerado de baixa atividade quando o percentual de inibição na dosagem indicada era $\leq 39\%$.

O modelo obtido para o conjunto dos compostos 4-fenil-2-guanidinothiazóis usando todas as variáveis selecionadas na análise exploratória dos dados em cada modelo parcial, e com as cargas de Mulliken selecionadas (mostradas na tabela 5.5), mostrou os seguintes resultados na comparação com as respostas obtidas experimentalmente para inibição *in vitro* da atividade da bomba de prótons H^+, K^+ -ATPase:

O dendograma obtido com HCA dos dados autoescalados (figura 5.1), com conexão simples entre os compostos mostra os compostos com alta atividade agrupados em dois conjuntos. O maior tem 24 componentes com similaridade maior que 0,05UV, a uma distância de 1,51UV da conexão seguinte, com a outra parte do grupo de alta atividade. É mostrado em azul escuro, vermelho e preto na parte superior do dendograma. Neste grupo estão os compostos 63, 59, e 82, com inibição reportada de 25% , 27%, e 26% respectivamente, e que não foram corretamente classificados no conjunto de baixa atividade ao qual pertencem. O segundo conjunto contém o subgrupo de baixa atividade. Há 6 compostos de alta atividade agrupados antes da conexão com o subgrupo de baixa atividade, mostrados em vermelho claro na parte inferior do dendograma. As atividades reportadas para estes compostos variam bastante, havendo compostos próximos da

maior atividade observada, assim como do limite inferior (39%) para ser considerado de alta

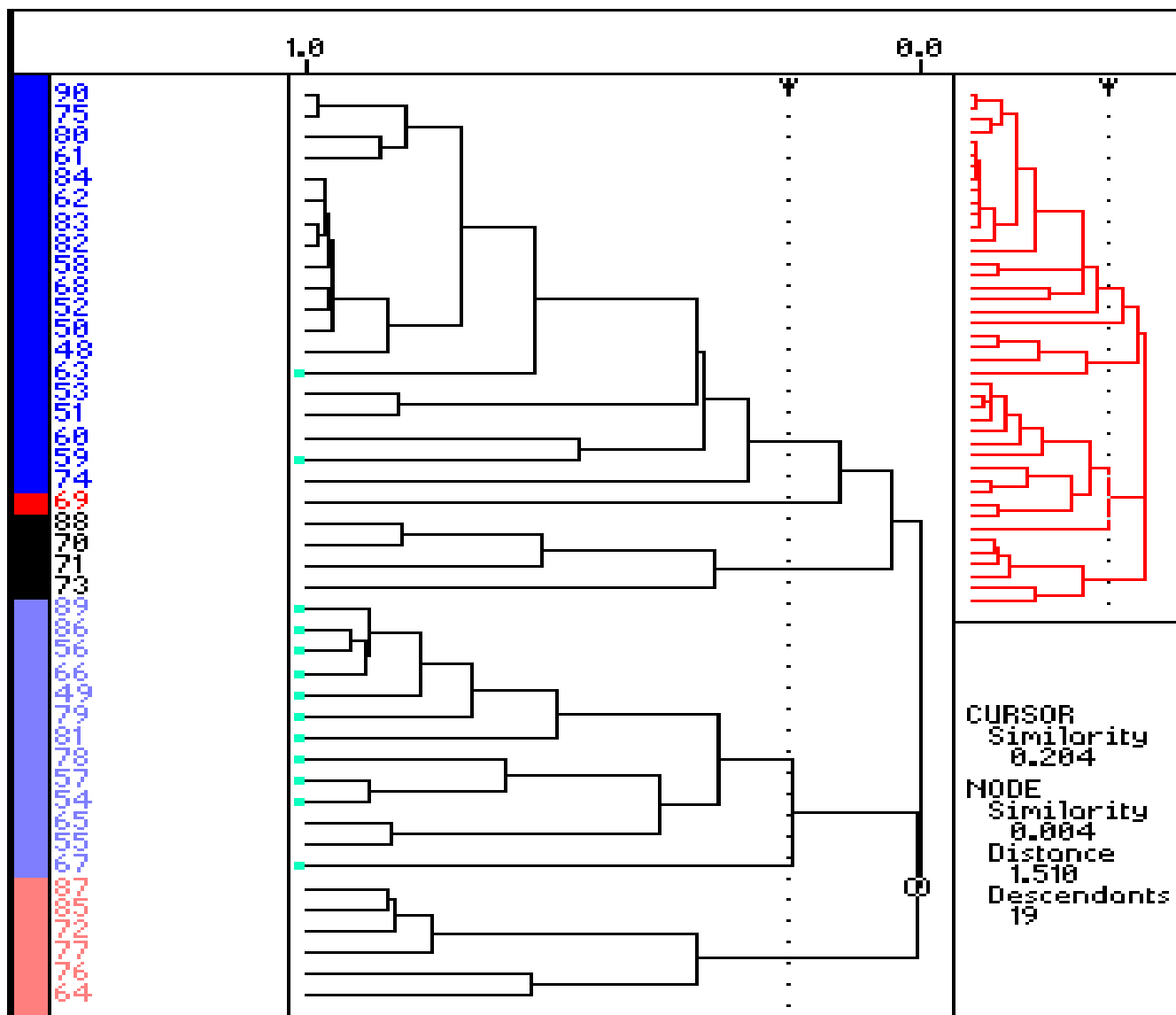


Figura 5.1: Dendrograma mostrando as conexões entre os compostos, com os parâmetros calculados com Hamiltoniano AM1 e cargas de Mulliken. O maior conjunto de alta atividade é mostrado em azul escuro, vermelho e preto na parte superior do dendrograma. Há 6 compostos de alta atividade agrupados antes da conexão com o subgrupo de baixa atividade, mostrados em vermelho claro. Os compostos de baixa atividade são mostrados em azul claro, na parte inferior do dendrograma. Os compostos marcados com um sinal verde são os classificados corretamente atividade.

Finalmente, estão mostrados os compostos de baixa atividade em azul claro, na parte inferior do dendrograma. Este conjunto tem 13 elementos com similaridade maior que 0,207UV a uma

distância de 1,510UV do centro de conexão com os compostos mais ativos. Os compostos 54 e 89 foram erradamente classificados neste conjunto no dendograma. Estes compostos tem atividade inibidora da H^+,K^+ -ATPase reportada de 99% e 111%, (as atividades são relativas ao composto 88, que tem 100% de atividade) medidas com dosagem de 50 μM da droga *in vitro*, e estão situados no extremo superior da atividade. Os compostos 55 e 65 não foram corretamente classificados na PCA, e aparecem corretamente classificados aqui no dendograma. Os compostos marcados com um sinal verde são os classificados no grupo de baixa atividade com PCA, usando 3 componentes principais para definir o espaço vetorial correspondente.

A PCA foi obtida com os dados autoescalados, em 5 PCs. Com 3PCs (figura 5.2) podem ser

separados 11 compostos de baixa atividade (vermelho forte), agrupados com 2 compostos de alta atividade classificados errado (55 e 65, em azul forte), dentro de um subespaço vetorial marcado graficamente no diagrama com as 3 PCs; e um classificado errado na região de alta atividade (82, que está sobreposto no gráfico da figura 5.2, e não

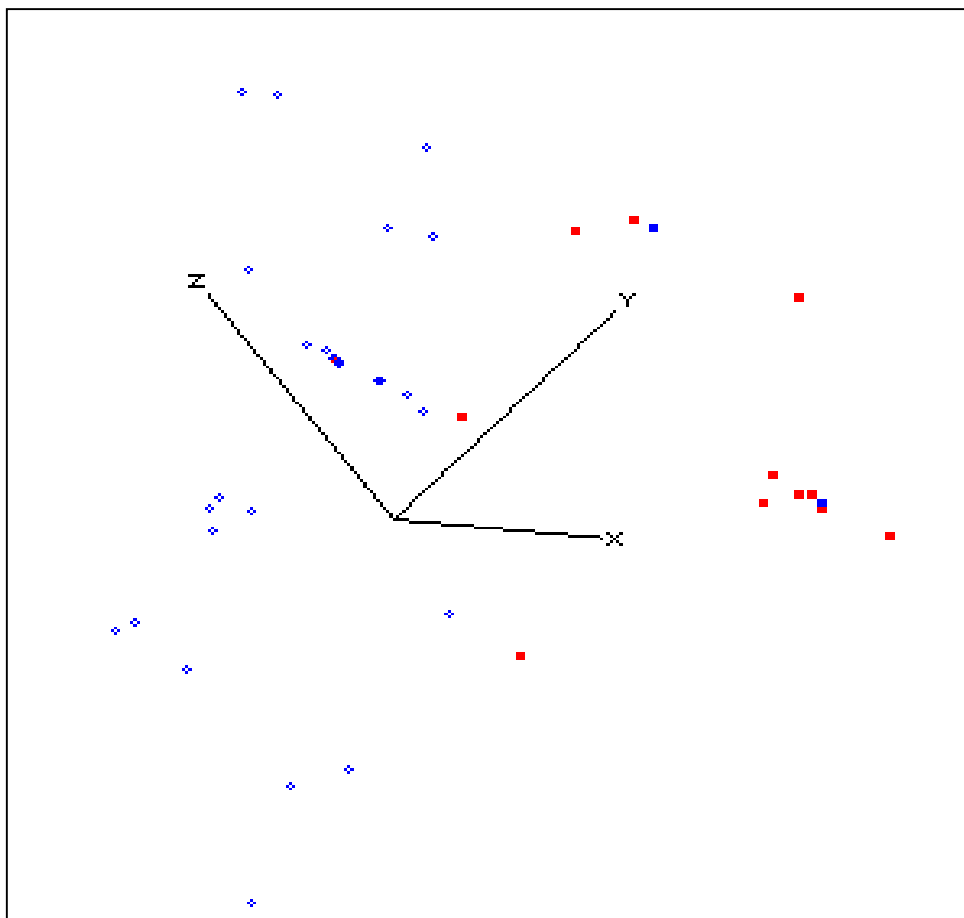


Figura 5.2: Gráfico das três primeiras PCs. Os compostos mais ativos são mostrados em azul, e em vermelho são os de baixa atividade. Os compostos 55 e 65 são de alta atividade e estão classificados errado no grupo menos ativo, em azul destacado. O composto 82 é de baixa atividade e está classificado errado no grupo de alta atividade. Não aparece no gráfico porque está sobreposto no

aparece), totalizando 7% de classificações erradas.

Fazendo a definição numérica do valor de corte para cada PC, o número de classificações erradas

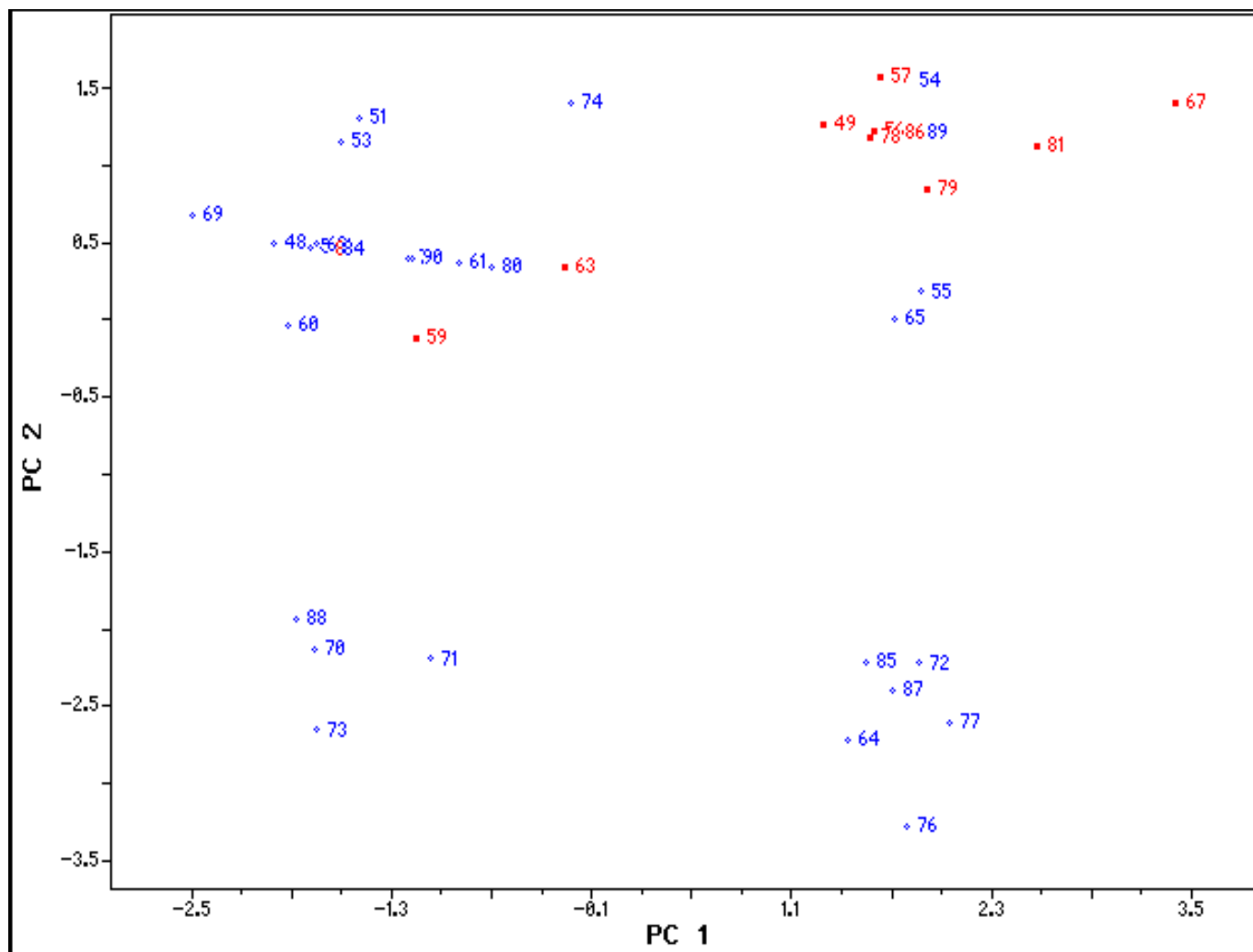


Figura 5.3: Primeira e segunda componentes principais para o modelo final com atividades *in vitro* e cargas de Mulliken. Os compostos de alta atividade são mostrados em azul, e os de baixa atividade em vermelho. O composto 82, que não aparece no gráfico da figura 5.2, aparece parcialmente sobreposto no canto superior esquerdo da figura.

sobe para 4 pouco ativos classificados erroneamente. A definição numérica foi feita da seguinte forma: (1) as coordenadas máxima e mínima dos 'scores' nas 3 primeiras PCs dos compostos classificados no grupo de baixa atividade foram encontradas, (2) os compostos com alta atividade que tinham coordenadas dentro da faixa de valores considerada, foram considerados classificações erradas. A grosso modo foi definido um cubo no espaço tridimensional onde só

deveriam haver compostos de baixa atividade. As coordenadas das arestas tais que $(-1,1567) \leq PC1 \leq (3,40735)$; $(-0,1234) \leq PC2 \leq (1,57318)$; e $(-1,4075) \leq PC3 \leq (0,67376)$. Graficamente também é definido um cubo, porém suas arestas não são necessariamente paralelas aos eixos das componentes principais, e uma delas estende-se de $-\infty$ a $+\infty$. As três PCs consideradas acumulam 95% da variância total dos dados autoescalados, e seus coeficientes e as variâncias acumuladas por cada PC são mostrados na tabela 5.1.

Tabela 5.1: Coeficientes de combinação linear das PCs para o modelo PCA com atividades in vitro e cargas de Mulliken.

↓Variável PC#→	PC1	PC2	PC3
Q_{N3}^{Mulliken}	0,18859	0,60791	-0,4244
Q_{N4}^{Mulliken}	0,18748	-0,6372	0,24207
Q_{C9}^{Mulliken}	0,19044	0,46028	0,85116
α	0,55201	-0,0124	-0,0400
Volume vdW	0,53413	-0,1048	-0,1874
Volume vdW R1	0,55047	-0,0362	-0,0095
Variância acumulada	50,47%	34,36%	10,37%

A primeira PC acumula 50% da variância total do sistema. Pode ser observado na tabela dos 'loadings' das variáveis que a composição desta componente principal é distribuída principalmente entre a polarizabilidade, os volumes de van der Waals da molécula e do grupo R1. O aumento da polarizabilidade é um fator que separa os compostos com menor atividade na parte positiva do gráfico cartesiano. Os compostos 54, 55, 65, e 89 são pouco ativos classificados errado na primeira componente principal (figura 5.3), e têm a polarizabilidade calculada maior que 32,00; enquanto a maioria dos inativos têm o valor desta propriedade calculada por volta de 24,00. Um grupo volumoso (benzila) na posição R1 também caracteriza estes compostos de baixa atividade classificados errado na primeira componente principal. Na Segunda componente principal as maiores contribuições vêm das cargas de Mulliken nas posições N3, N4, e C9. Os compostos menos ativos têm carga mais negativa nas posições N3 e C9, e carga menos negativa na posição N4 como contribuição para classificação correta nesta componente.

5.1.2. Modelo HCA/PCA com cargas CHELP para atividades *in vitro* dos 4-fenil-2-guanidino-1,3,4-oxadiazóis

O modelo obtido para o conjunto dos compostos 4-fenil-2-guanidino-1,3,4-oxadiazóis usando as variáveis selecionadas na análise exploratória dos dados (tabela 5.3) mostrou os seguintes resultados na comparação com as respostas obtidas experimentalmente para inibição *in vitro* da atividade da

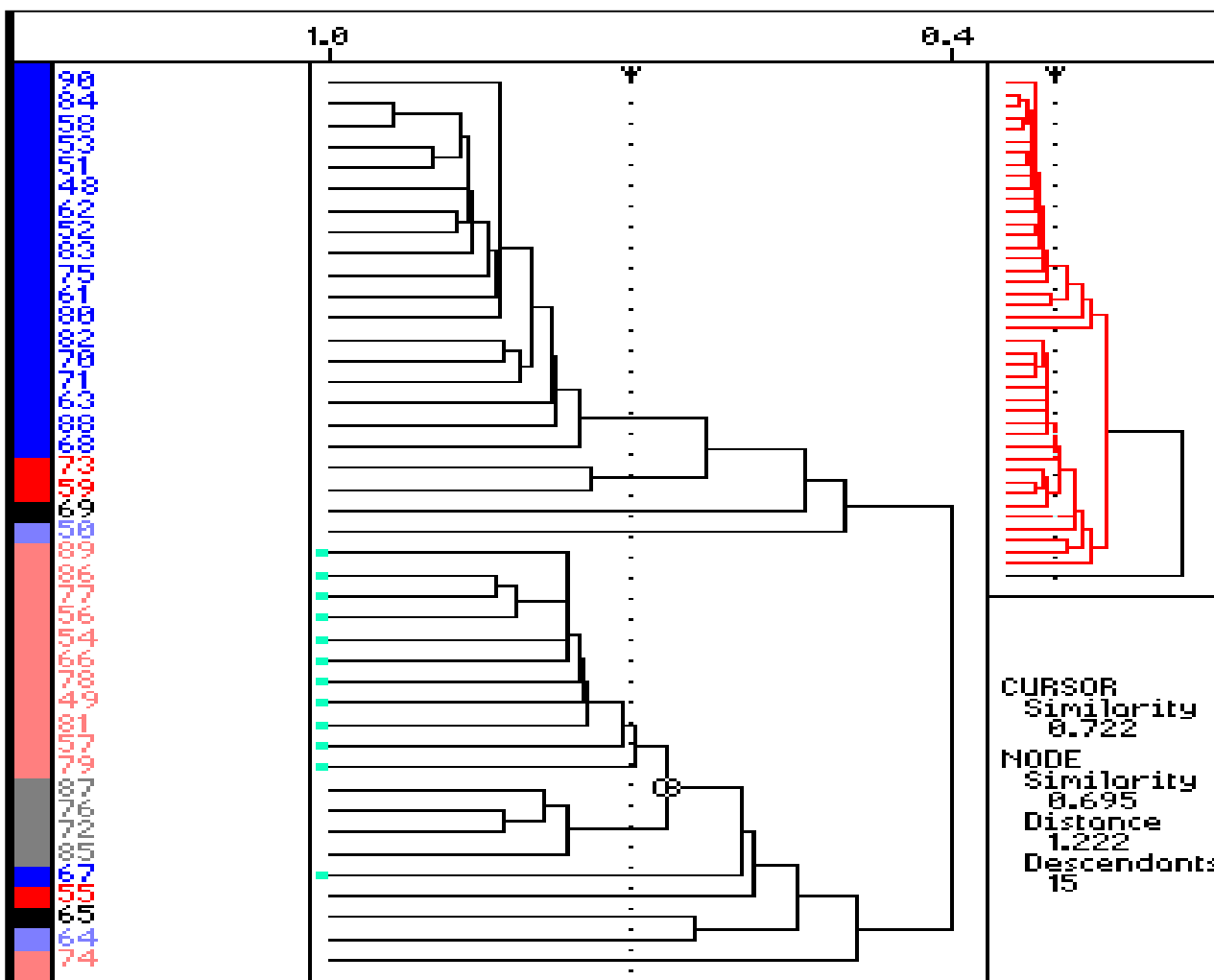


Figura 5.4: Dendrograma mostrando o agrupamento dos compostos no modelo com as variáveis selecionadas e as cargas CHELP. Em azul na parte superior, e em cinza na parte inferior são mostrados compostos de alta atividade. O grupo de baixa atividade é mostrado em vermelho claro, no centro do dendrograma. Os compostos na parte inferior, mostrados em diversas cores, não foram agrupados. No detalhe do canto superior direito é mostrado bomba de prótons H^+, K^+ -ATPase:

O dendograma obtido com HCA (figura 5.4) dos dados autoescalados, e conexão simples entre os compostos, mostra dois conjuntos principais. O maior tem 22 compostos de alta atividade com similaridade maior que 0,53UV, a uma distância de 1,51UV da conexão seguinte. Este é o grupo mostrado em azul na parte superior do dendograma. Neste grupo estão os compostos 63, 59, e 82, com inibição reportada de 25% , 27%, e 26% respectivamente, e que não foram corretamente classificados no conjunto de baixa atividade ao qual pertencem, na HCA nem na PCA.

O segundo conjunto contém o subgrupo de baixa atividade e outros compostos de alta atividade não agrupados no primeiro grupo. Os 11 compostos com baixa atividade agrupados antes da conexão com o subgrupo de alta atividade são mostrados em vermelho claro no centro do dendograma, e têm similaridade maior que 0,72UV, ficando a uma distância de 1,22UV do último

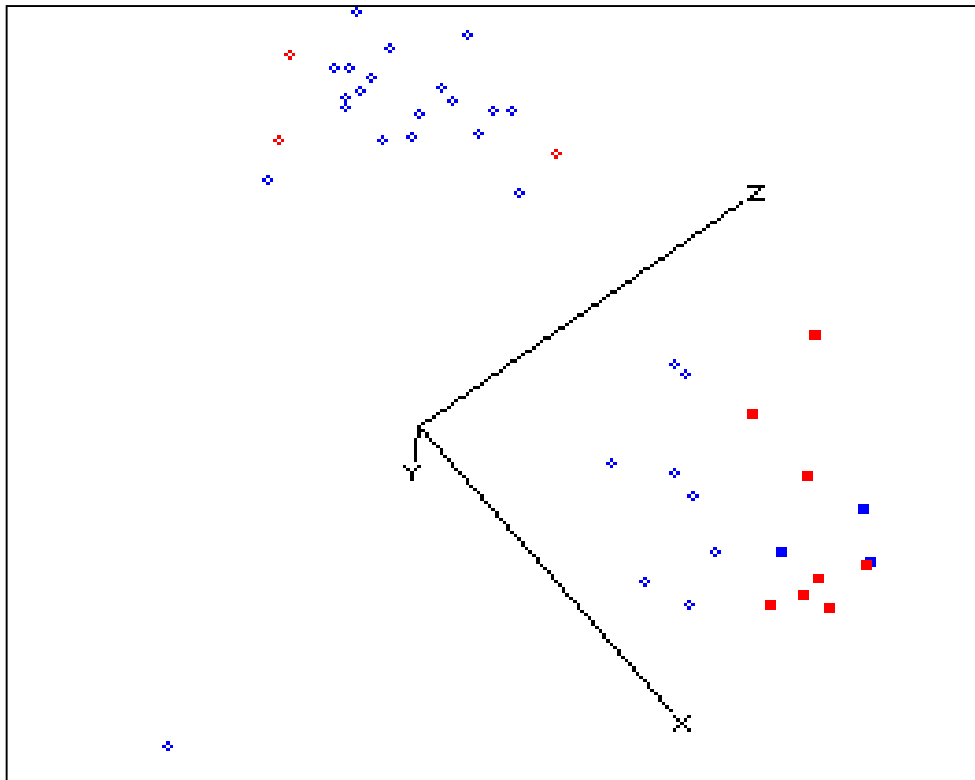


Figura 5.5: Gráfico das 3 primeiras PCs do modelo com cargas CHELP. Os compostos pouco ativos são mostrados em vermelho.

centro de conexão entre compostos de alta atividade. Os compostos 54, 77, e 89 são classificados errado neste grupo. A atividade inibidora da H^+, K^+ -ATPase reportada para estes compostos é respectivamente 99%, 85%, e 111% (o composto 88 tem 100% de inibição), havendo valores no extremo superior da atividade observada,

portanto.

Os demais compostos de alta atividade formam um grupo com 4 compostos de similaridade maior que 0,78UV, distantes 1,22UV do grupo de baixa atividade, restando mais 6 compostos não classificados nos grupos propostos, entre os quais um composto de baixa atividade (67), corretamente classificado na análise por componentes principais. O composto 60 foi o que mais distanciou-se dos demais, à 2,11UV, e não foi mostrado no gráfico principal para melhorar a visualização dos demais.

A PCA foi realizada com os dados autoescalados, em 5 PCs. Com 3 PCs podem ser separados 12 compostos de baixa atividade, agrupados com 3 compostos de alta atividade classificados errado, e 3 classificados errado na região de alta atividade (figura 5.5), dentro de um subespaço vetorial marcado graficamente no diagrama com as 3 PCs. Os compostos 57, 74, e 89 foram os mais ativos classificados errado, ainda que o composto 89 possa ser excluído do grupo de baixa atividade se o cubo que inscreve este grupo for cuidadosamente demarcado. Os classificados errado de baixa atividade, 59, 63, e 82, estão completamente misturados no grupo de alta atividade, não havendo qualquer variável que os diferencie dos mais ativos. Substituindo-se a terceira PC pela quarta os resultados continuam os mesmos, com as mesmas classificações erradas. A substituição pela quinta PC aumenta uma classificação errada, o composto 55. As 3 PCs consideradas acumulam quase 90% da variância total dos dados autoescalados, e criam um modelo que não classifica corretamente 14% dos compostos.

A primeira PC (figura 5.6) acumula quase 58% da variância total dos dados autoescalados. A análise dos 'loadings' desta primeira PC mostra contribuição significativa das variáveis Q_{N3}^{CHELP} , Q_{N4}^{CHELP} , Q_{C7}^{CHELP} , Q_{R2}^{CHELP} , α , vdWV, e R1vdWV principalmente (tabela 5.2). A contribuição da carga derivada do potencial eletrostático sobre o Nitrogênio 3 mostra que a classificação correta no grupo de baixa atividade acontece nos compostos em que a carga na posição é menos negativa. Os compostos de alta atividade pior classificados foram 59, 63, e 82, que têm a carga nesta posição mais negativa, com valor semelhante aos pouco ativos. O mesmo se aplica a carga na posição do Nitrogênio 4, inclusive quanto as classificações erradas. As cargas sobre os

carbonos C7 e R2 dos compostos pouco ativos bem classificados é menos positiva, e as más classificações correspondem às exceções a regra.

As polarizabilidades e os volumes de van der Waals da molécula e do substituinte R1 são os mesmos utilizados no modelo anterior e sua influência já foi discutida. A segunda PC acumula 21% da variância total dos dados autoescalados, e as principais contribuições são Q_{C6}^{CHELP} , e Q_{C7}^{CHELP} . A carga na posição C6 tem contribuição de tal modo que valores mais negativos correspondem a um 'score' mais alto para o composto na segunda PC. A carga na posição C7 contribui com aumento diretamente proporcional no 'score' do composto na segunda PC. Esta PC apenas não tem nenhuma capacidade de modelar a separação entre grupos de atividade diferente, mais é importante principalmente na classificação correta dos compostos 54, 55, 60, e 74, muito ativos e que ficam bem distantes se considerada esta PC.

A terceira PC tem papel secundário, com 10% da variância acumulada, e contribuição principal

das variáveis Q_{N4}^{CHELP} , Q_{R2}^{CHELP} , Q_{C6}^{CHELP} , α , e vdWV.

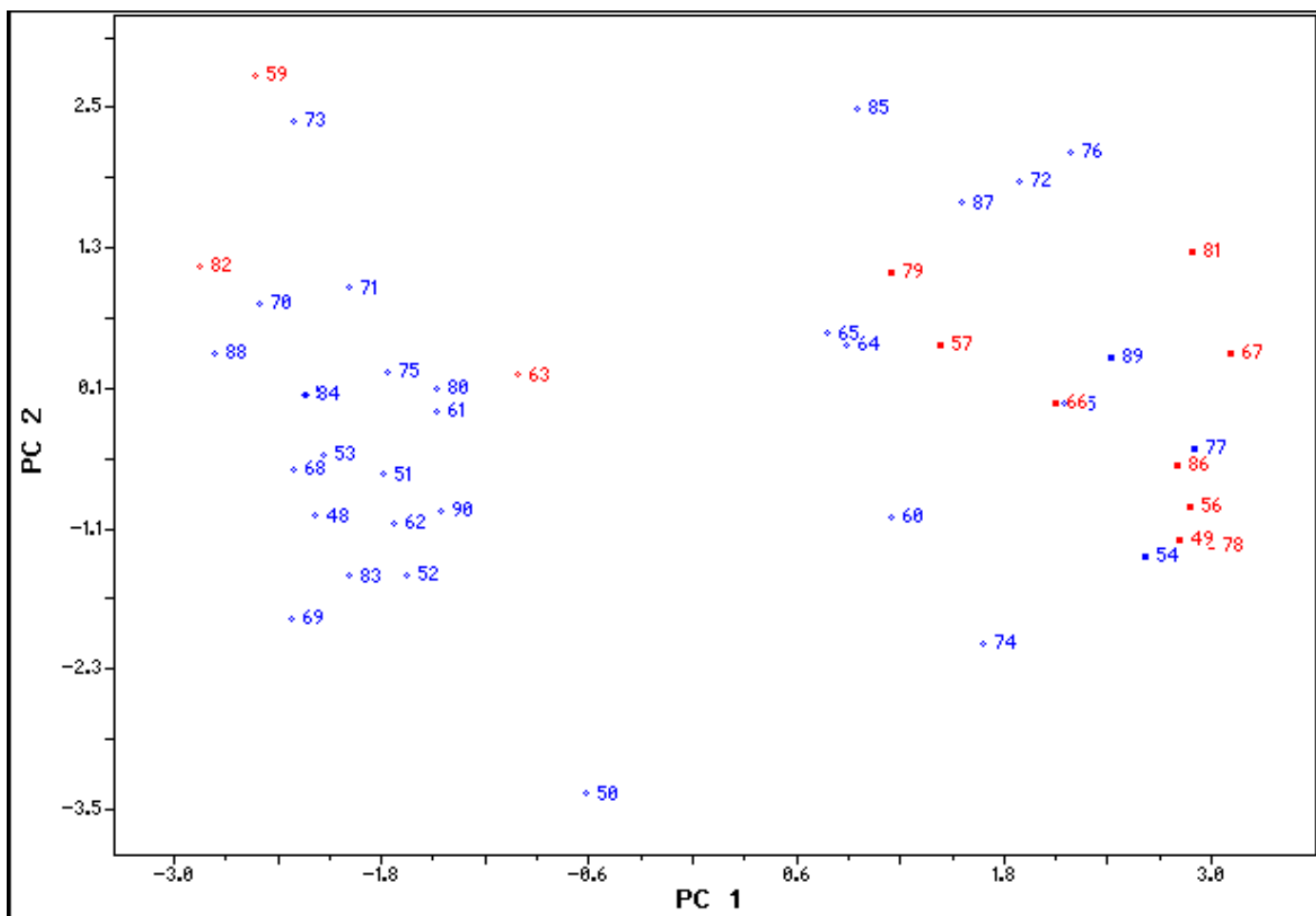


Figura 5.6: Primeira e segunda PCs para o modelo com as variáveis selecionadas e cargas CHELP. Os compostos mais ativos aparecem em azul no gráfico. Os compostos marcados com destaque são os que foram agrupados no grupo dos compostos menos ativos no gráfico com 3 PCs.

Tabela 5.2: Coeficientes de combinação linear das componentes principais, e as variâncias acumuladas, para o modelo com cargas CHELP.

↓Variável	PC#→	PC 1	PC 2	PC 3
Q_{N3}^{CHELP}		0,40323	-0,0148	0,07118
Q_{N4}^{CHELP}		0,40288	0,23128	0,31220
Q_{C6}^{CHELP}		0,37352	0,33421	0,29705
Q_{C7}^{CHELP}		0,41576	0,19675	0,19745
Q_{R2}^{CHELP}		0,37220	-0,0506	-0,5569
α		0,13649	-0,6661	0,32050
VdWV		-0,2573	0,58426	-0,1311
R1vdWV		-0,3733	0,09030	0,58325
Variância acumulada		57,65%	21,51%	10,68%

Tabela 5.3: Variáveis calculadas com Hamiltoniano AM1, com cargas CHELP, selecionadas com HCA/PCA.

No.	Q_{N3}^{CHELP} Q_{C7}^{CHELP}	α	vdWv	R1vdWv	Q_{N4}^{CHELP}	Q_{C6}^{CHELP}	Q_{R2}^{CHELP}
48	-0,9600	21,8195	183,9000	7,2300	-0,9400	-0,0700	
0,1500							1,0700
49	-0,8100	31,1784	273,4000	99,2600	-0,7800	-0,0100	
0,0700							0,9500
50	-0,9500	22,6049	197,7000	7,2300	-0,8600	0,1200	
0,0400							1,0400
51	-0,9500	23,1656	199,0000	7,2300	-0,9000	-0,0900	
0,1600							1,0700
52	-0,9500	23,4521	199,0000	7,2300	-0,9300	0	
0,1300							1,0700
53	-0,9600	22,4181	190,8000	7,2300	-0,9200	-0,1100	
0,1700							1,0700
54	-0,8200	32,8028	290,8000	99,2600	-0,8200	0,0700	
0,0800							0,9900
55	-0,9200	33,2900	295,1000	99,2600	-0,7200	-0,0900	
0,1000							1,0200
56	-0,8100	32,2269	293,6000	99,2600	-0,8300	0,0200	
0,1000							0,9200
57	-0,8200	32,2844	279,0000	99,2600	-0,8800	-0,0800	
0,1600							1,0200
58	-0,9600	22,8810	202,2000	7,2300	-0,9400	-0,1400	
0,1900							1,0700
59	-0,9700	26,3259	244,6000	7,2300	-0,9300	-0,3000	
0,3100							1,0800
60	-0,9400	23,2545	199,9000	7,2300	-0,6300	-0,1200	
0,1600							0,8200
61	-0,9600	26,6694	244,7000	7,2300	-0,9300	-0,0700	
0,1800							1,0700
62	-0,9600	23,5759	204,0000	7,2300	-0,9200	-0,0500	
0,1400							1,0700
63	-0,9600	31,6972	255,5000	7,2300	-0,9300	-0,0700	
0,1800							1,0700
64	-0,9900	32,0625	304,3000	99,2600	-0,8500	-0,1200	
0,1200							1,0300
65	-0,9800	32,8602	288,7000	99,2600	-0,8000	-0,0400	
0,1800							1,0600
66	-0,8500	33,0913	310,2000	99,2600	-0,8300	-0,0400	
0,1200							1,0100
67	-0,8300	42,1141	353,5000	99,2600	-0,8600	0	
0,1200							1,0100
68	-0,9600	23,2381	195,0000	7,2300	-0,9500	-0,0100	
0,2100							1,0800
69	-0,9600	26,2517	86,5000	7,2300	-0,9200	0	
0,1500							1,0700
70	-0,9700	24,0035	220,8000	7,2300	-0,9600	-0,1800	
0,2200							1,0900

71	-0,9600	26,6246	252,9000	7,2300	-0,9600	-0,1500	1,0900
	0,2200						
72	-0,8600	34,2347	301,8000	99,2600	-0,7900	-0,2300	0,9800
	0,1900						
73	-0,9500	24,3254	241,4000	7,2300	-0,9500	-0,3300	1,0400
	0,2700						
74	-0,7500	30,8774	228,8200	7,2300	-0,8400	0,0600	0,9500
	0,0900						
75	-0,9500	24,9693	231,9000	7,2300	-0,9300	-0,1300	1,0700
	0,1900						
76	-0,8200	34,1975	323,6000	99,2600	-0,8100	0,2500	0,9700
	0,1900						
77	-0,8100	34,6882	316,3000	99,2600	-0,8200	0,0100	0,9700
	0,1100						
78	-0,7500	30,8774	306,1000	93,3000	-0,8400	0,0600	0,9500
	0,0900						
79	-0,8600	34,4674	311,9000	99,2600	-0,9200	-0,1300	1,0600
	0,1500						
80	-0,9500	27,3243	257,5000	7,2300	-0,9400	-0,0300	1,0800
	0,2100						
81	-0,8000	37,5912	335,2000	99,2600	-0,8100	-0,100	0,9700
	0,1800						
82	-0,9700	22,6912	206,2000	7,2300	-0,9700	-0,2500	1,0900
	0,2200						
83	-0,9700	22,8992	207,2000	7,2300	-0,9500	0,0300	1,0900
	0,1400						
84	-0,9600	23,2203	205,2000	7,2300	-0,9400	-0,1500	1,0800
	0,1800						
85	-0,8700	32,8822	295,2000	99,2600	-0,8400	-0,2600	1,0200
	0,2300						
86	-0,8000	32,9490	298,1000	99,2600	-0,8100	0,0300	0,9500
	0,1300						
87	-0,8400	33,3815	300,9000	99,2600	-0,8300	-0,1800	1,0100
	0,2000						
88	-0,9800	23,4393	210,6000	7,2300	-0,9700	-0,0900	1,1000
	0,2400						
89	-0,8100	34,1889	298,1000	99,2600	-0,8200	-0,0500	0,9700
	0,1600						
90	-0,9600	25,1892	232,2000	7,2300	-0,9300	-0,0100	1,0700
	0,1500						

5.1.3. Modelo da atividade *in vivo* dos 4-fenil 2-guanidinothiazóis

O modelo obtido para relacionar as atividades *in vivo* com os parâmetros calculados é mostrado a seguir. Os parâmetros selecionados, os que tinham boa capacidade de modelagem, são mostrados na tabela 5.7. Os dados de atividade foram medidos como a concentração μM necessária à inibição de 50% da atividade enzimática. Um composto foi considerado com alta atividade quando sua $\text{IC}_{50} \leq 15 \mu\text{M}$. Esta determinação é disponível para 23 compostos desta série.

A HCA foi realizada com os dados autoescalados, utilizando conexão simples entre as amostras.

O dendograma na figura 5.7 separa o grupo menos ativo com 14 compostos marcado em vermelho no centro, e quatro classificações erradas. Os compostos 71, 84, 88, e 90 são compostos com alta atividade classificados errado neste conjunto, com IC_{50} de 14; 15; 1,5; e 1,6 μM respectivamente. A similaridade entre os elementos é maior que 0,197UV, distando 1,725UV da conexão com os demais compostos. Os compostos marcados com um ponto verde são os classificados no grupo de alta atividade em PCA (71, e 90 corretamente classificados com PCA). Os compostos de alta atividade estão agrupados no canto superior esquerdo do dendograma, marcados em azul. Foram classificados 7 compostos com alta atividade, agrupados com mais de 0,284UV de similaridade, a uma distância de 1,629UV da conexão com o grupo de baixa atividade. Aqui não houve classificação errada.

Fica isolado dos grupos de alta ou baixa atividade o composto 74, de baixa atividade, que aparece em preto no canto inferior esquerdo do dendograma. Está a 1,845UV de distância da conexão entre os grupos definidos de alta e baixa atividade. O composto 54 é muito ativo, mostrado em vermelho na parte inferior do gráfico, dista 1,925UV da conexão entre 74 e o centro de conexão com o grupo menos ativo.

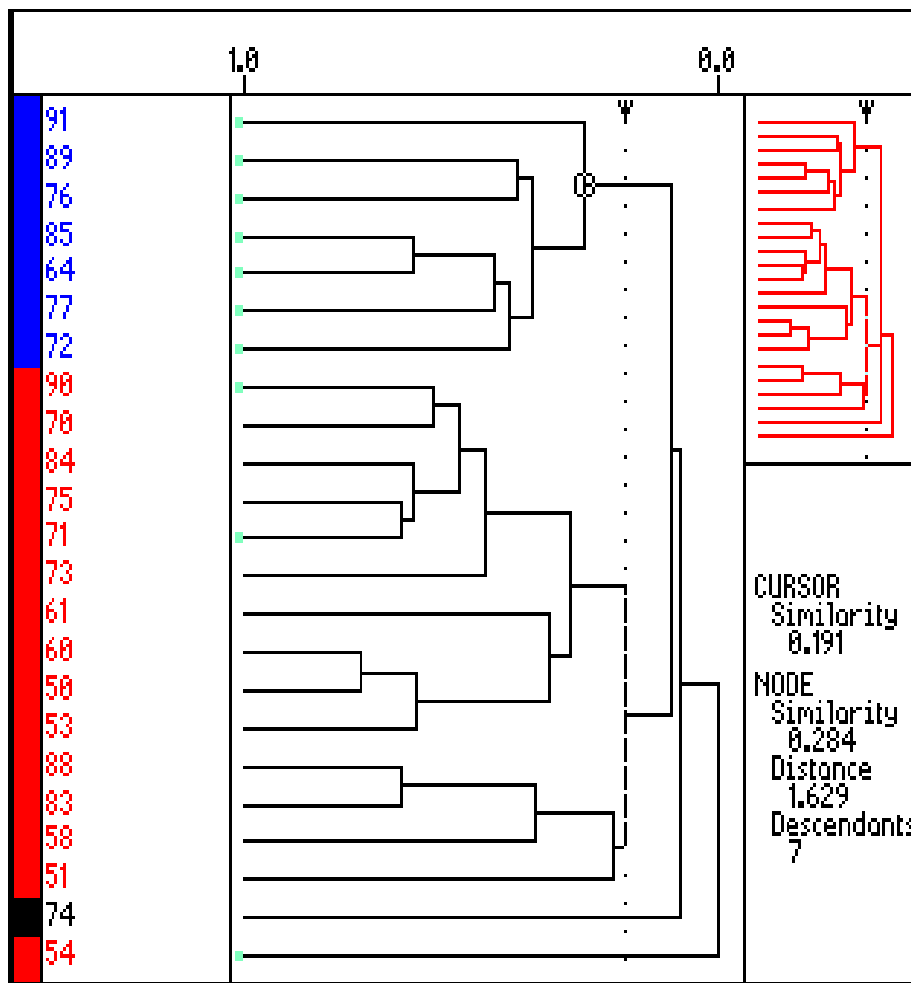


Figura 5.7:.Dendograma obtido da HCA final para as atividades *in vivo* dos compostos guanidinothiazóis. Os compostos menos ativos estão marcados em vermelho no centro, os mais ativos em azul na parte superior. Os compostos 74, pouco ativo, e 54, muito ativo, em preto e vermelho respectivamente, não se agrupam bem.

A análise PCA foi feita com autoescalamento dos dados, em 5 PCs. Para este conjunto de 23 compostos, foram formados dois grupos: 12 mais ativos e 11 menos ativos, e foram usadas as 3 primeiras PCs (PC1⇒X, PC2⇒Y, PC3⇒Z), totalizando cerca de 85% da variância total dos dados autoescalados (figura 5.8). As únicas classificações erradas são os compostos 84 e 88. A substituição da 3ª PC pela 4ª não muda o resultado, e pela 5ª aumenta uma classificação errada. O grupo mais ativo com 10 compostos, todos classificados certo com 3 PCs (figura 5.8), tem os compostos 71 e 90 classificados errado na primeira PC. A primeira PC acumula 37% da variância total dos dados autoescalados (tabela 5.4). É uma combinação dos volumes de van der Waals das

moléculas, do substituinte R1, da área disponível para o solvente na superfície de van der Waals, e da polarizabilidade calculada, principalmente. O valor da polarizabilidade dos compostos bem classificados nesta 1ª PC é alto, maior que 32,00; enquanto os mau classificados têm valores até 26,60 aproximadamente (composto 71). Os volumes dos compostos bem classificados são

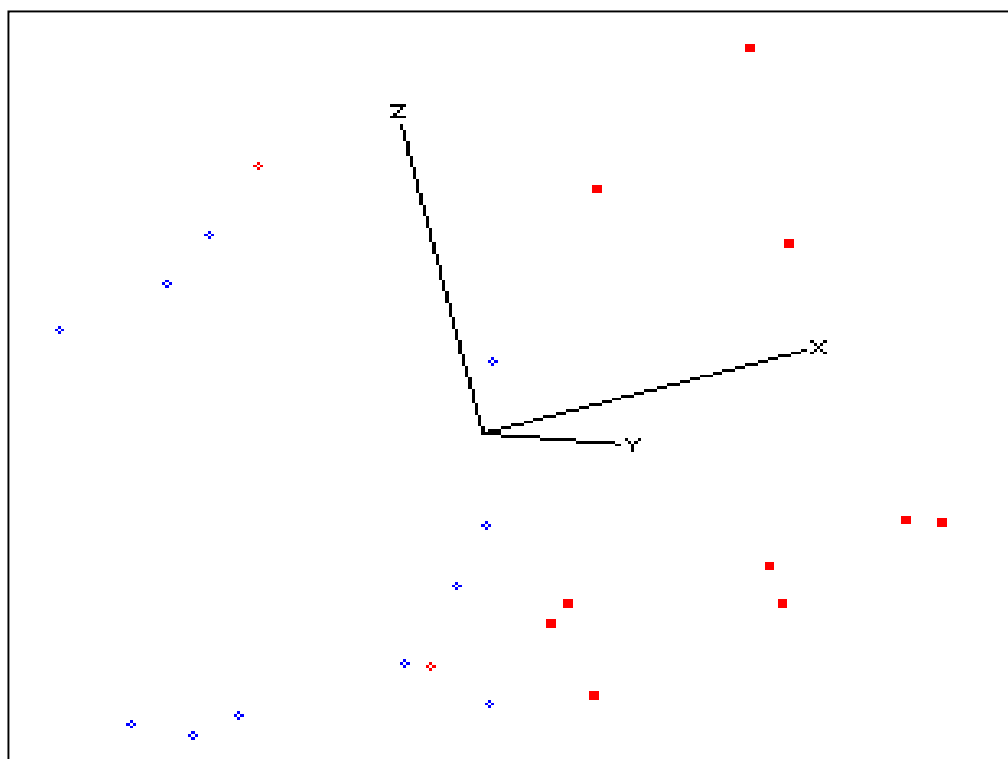


Figura 5.8: Gráfico das 3 primeiras PCs para a PCA com dados obtidos com AM1 e cargas CHELP, e as atividades *in vivo* dos guanidinothiazóis. Os mais ativos em vermelho. Os compostos 71 e 90, em vermelho não destacado são muito ativos e não foram bem classificados com este

modelo
maiores que $295 \text{ \AA}^3$, e a área acessível para o solvente maior que $324 \text{ \AA}^2$ (composto 85). O volume do substituinte R1 dos bem classificados corresponde ao grupo benzila nesta posição. A 2ª PC acumula 28% da variância dos dados, e não cria 2 classes para os compostos mais ou menos ativos. Sua maior contribuição é auxiliar na separação dos compostos 71 e 90 que têm alta atividade e não foram bem classificados somente com a primeira PC, e dos compostos 61 e 74 que têm baixa atividade e também não foram bem classificados somente com a primeira PC. Sua composição tem grande influência dos índices de densidade de orbitais de fronteira, normalizados

pela energia do orbital correspondente (HOMO ou LUMO), nas posições C5, C9, e C10; O momento de dipolo calculado com as orbitais híbridas é outro fator importante, seguido pelas cargas calculadas na posição N3, e R2. A densidade orbital em LUMO sobre o carbono 5, em HOMO sobre o carbono 9 e em HOMO sobre o carbono 10 é menos negativa nos compostos com alta atividade bem classificados na segunda PC. Um valor baixo na componente nos orbitais híbridos do momento de dipolo aproxima os compostos do grupo de alta atividade. Uma carga mais positiva na posição R2, e mais negativa na posição N3 também ajudam na classificação correta dos compostos mais ativos pela 2ª PC. Na 3ª PC as principais contribuições são as cargas nas posições R2 e N3, da mesma maneira que na 2ª PC. As densidades de elétrons em HOMO e LUMO atuam de maneira inversa na 2ª e na 3ª PCs, ainda que sua atuação na 3ª PC seja pequena.

Tabela 5.4: Coeficientes de combinação linear e variâncias acumuladas por cada PC no modelo para atividades *in vivo*.

↓Variável	PC#→	PC 1	PC 2	PC 3
F_{C5}^{LUMO}		0,17492	0,46177	0,17318
F_{C9}^{HOMO}		0,07369	0,48690	0,18637
F_{C10}^{HOMO}		-0,0229	0,48564	0,17117
Q_{N3}^{CHELP}		0,04165	-0,1689	0,67662
Q_{R2}^{CHELP}		-0,0087	0,20221	-0,6617
$\mu^{híbrido}$		-0,0034	-0,4737	0,02820
α		0,49379	-0,0109	-0,0846
Volume vdW		0,50654	-0,0085	-0,0060
Área acessível ao Solvente de vdW		0,48803	-0,1303	-0,0451
Volume vdW de R1		0,47227	-0,0529	-0,0150
Variância acumulada		37%	28%	20%

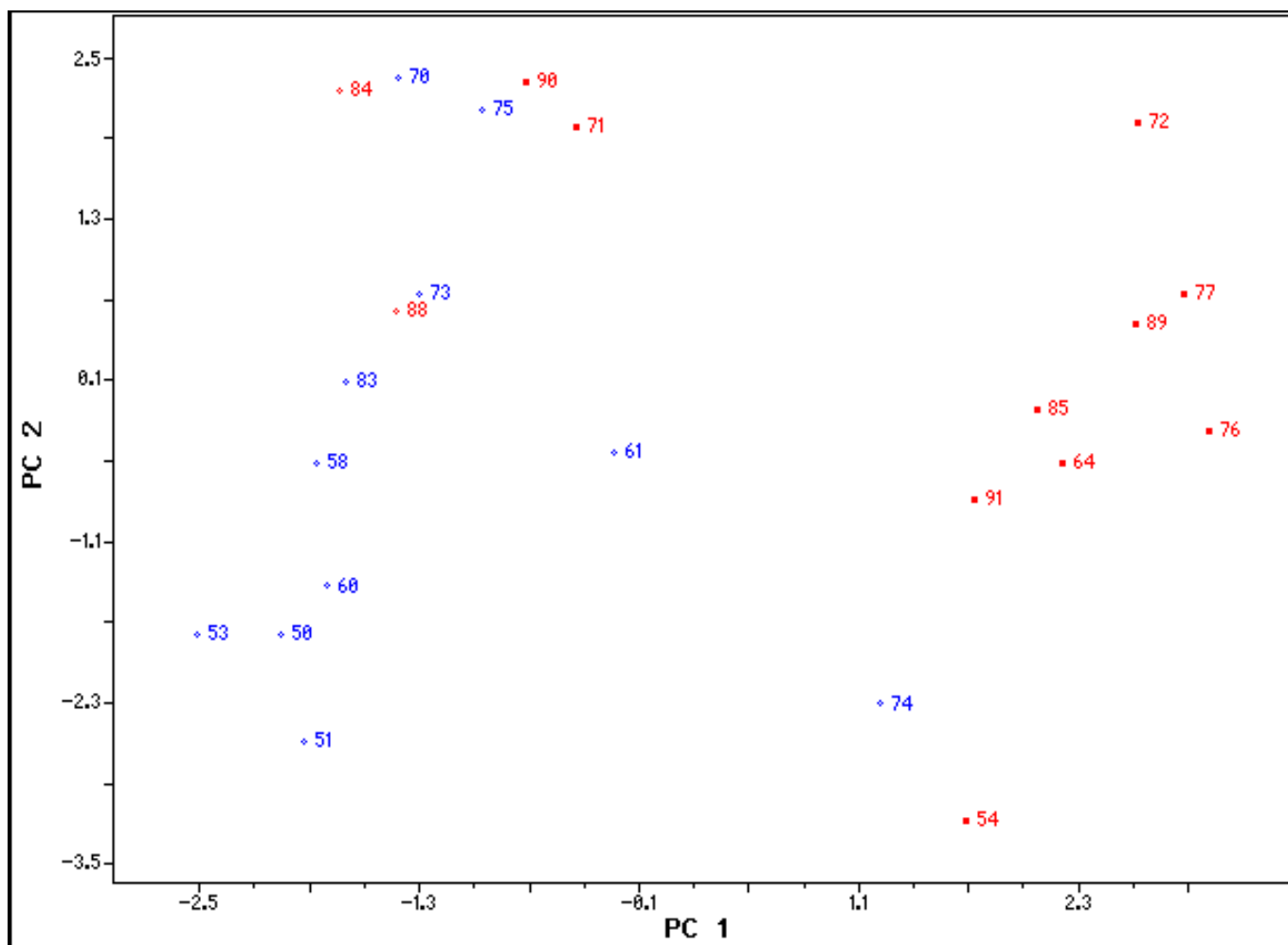


Figura 5.9: Gráfico das duas primeiras PCs do modelo PCA com atividades *in vivo*. Os compostos de alta atividade estão marcados em vermelho, e os de baixa em azul. Os compostos que aparecem em destaque são os que foram marcados no grupo de alta atividade no gráfico com 3 PCs. Os compostos 84 e 88 não foram corretamente agrupados, e aparecem sem destaque.

5.2. Modelos de regressão PLS com as variáveis selecionadas com HCA e PCA

5.2.1. Modelo de regressão PLS para atividades *in vitro*

Para os compostos 4-fenil 2-guanidinothiazóis foi construído um modelo de correlação quantitativa

com as atividades *in vitro*. As variáveis usadas neste modelo foram Q_{N3}^{Mulliken} , Q_{N4}^{Mulliken} , Q_{C9}^{Mulliken} ,

α , vdWV, e R1vdWV, respectivamente, são mostradas na tabela 5.5. No programa 5.1 são

mostrados os comandos executados no MATLAB para análise dos dados:

Programa 5.1: Comandos para executar regressão PLS com teste de validação cruzada tipo 'deixe um de fora' para o MATLAB™.

```
1» load am1_fim.mat
2» [fim0_ax,fim0_mx,fim0_std]=auto(am1_fim0);
3» [vitro_y,vitro_my,vitro_std]=auto(vitro');
4» [prs,cp,mlv,b,w,p,q,t21,tv]=plscvrand(fim0_ax,vitro_y,50,30,6,0,1);
```

No passo 1 do programa 5.1 a matriz com os dados da tabela

5.5 são carregados na

memória. No passo 2 os dados

são autoescalados. No passo 3

são autoescaladas as

atividades *in vitro* dos

guanidinothiazóis, mostradas na

tabela 1, do capítulo 1. Estes

dados já tinham sido

carregados previamente na

memória via teclado. No passo

4 é realizada a regressão com 6 variáveis latentes (VLs), e o teste de validação cruzada.

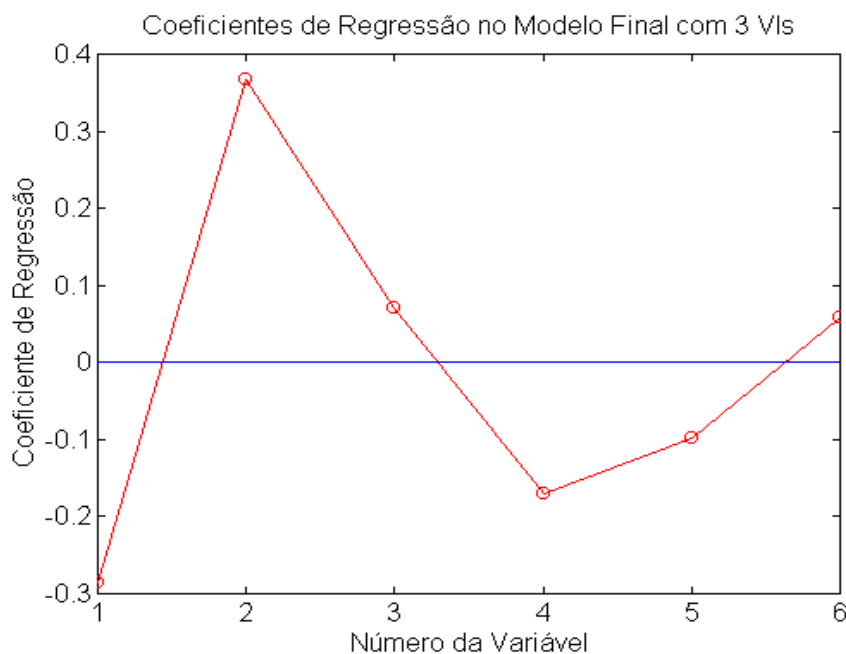


Figura 5.10: Valor do coeficiente de cada variável na matriz de regressão do modelo PLS que relaciona quantitativamente as atividades *in vitro* com os dados selecionados, utilizando cargas de Mulliken.

Tabela 5.5: Parâmetros calculados com Hamiltoniano AM1, e cargas de Mulliken, selecionados com PCA/HCA e utilizados para regressão PLS das atividades *in vitro* dos guanidinothiazóis

No.	Q_{N3}^{Mulliken}	Q_{N4}^{Mulliken}	Q_{C9}^{Mulliken}	α	vdWV
48	-0,3300	-0,3800	-0,1400	21,8195	183,9000
18,3200					
49	-0,2700	-0,3800	-0,1400	31,1784	273,4000
131,8800					
50	-0,3300	-0,3800	-0,1400	22,6049	197,7000
18,3200					
51	-0,3300	-0,3900	-0,1300	23,1656	199,0000
18,3200					
52	-0,3300	-0,3800	-0,1400	23,4521	199,0000
18,3200					
53	-0,3300	-0,3800	-0,1300	22,4181	190,8000
18,3200					
54	-0,2900	-0,3800	-0,1300	32,8028	290,8000
131,8800					
55	-0,3300	-0,3400	-0,1300	33,2900	295,1000
131,8800					
56	-0,2700	-0,3800	-0,1400	32,2269	293,6000
131,8800					
57	-0,2900	-0,3800	-0,1300	32,2844	279,0000
131,8800					
58	-0,3300	-0,3800	-0,1400	22,8810	202,2000
18,3200					
59	-0,3200	-0,3800	-0,1500	26,3259	244,6000
18,3200					
60	-0,3200	-0,3800	-0,1500	23,2545	199,9000
18,3200					
61	-0,3300	-0,3800	-0,1400	26,6694	244,7000
18,3200					
62	-0,3300	-0,3800	-0,1400	23,5759	204,0000
18,3200					
63	-0,3300	-0,3800	-0,1400	31,6972	255,5000
18,3200					
64	-0,3600	-0,2800	-0,1500	32,0625	304,3000
131,8800					
65	-0,3400	-0,3400	-0,1300	32,8602	288,7000
131,8800					
66	-0,2700	-0,3800	-0,1400	33,0913	310,2000
131,8800					
67	-0,2900	-0,3800	-0,1300	42,1141	353,5000
131,8800					
68	-0,3300	-0,3800	-0,1400	23,2381	195,0000
18,3200					
69	-0,3300	-0,3800	-0,1400	26,2517	228,3200
18,3200					
70	-0,3900	-0,3300	-0,1500	24,0035	220,8000
18,3200					

71	-0,3800	-0,3200	-0,1500	26,6246	252,9000
18,3200					
72	-0,3700	-0,2800	-0,1400	34,2347	301,8000
131,8800					
73	-0,3800	-0,3300	-0,1600	24,3254	241,4000
18,3200					
74	-0,3100	-0,3800	-0,1300	30,8774	228,8200
18,3200					
75	-0,3300	-0,3800	-0,1400	24,9693	231,9000
18,3200					
76	-0,3800	-0,2700	-0,1500	34,1975	323,6000
131,8800					
77	-0,3800	-0,2700	-0,1400	34,6882	316,3000
131,8800					
78	-0,3100	-0,3800	-0,1300	30,8774	306,1000
141,6200					
79	-0,2900	-0,3800	-0,1400	34,4674	311,9000
131,8800					
80	-0,3300	-0,3800	-0,1400	27,3243	257,5000
18,3200					
81	-0,2700	-0,3800	-0,1400	37,5912	335,2000
131,8800					
82	-0,3300	-0,3800	-0,1400	22,6912	206,2000
18,3200					
83	-0,3300	-0,3800	-0,1400	22,8992	207,2000
18,3200					
84	-0,3300	-0,3800	-0,1400	23,2203	205,2000
18,3200					
85	-0,3800	-0,2900	-0,1400	32,8822	295,2000
131,8800					
86	-0,2700	-0,3800	-0,1400	32,9490	298,1000
131,8800					
87	-0,3800	-0,2800	-0,1400	33,3815	300,9000
131,8800					
88	-0,3800	-0,3300	-0,1500	23,4393	210,6000
18,3200					
89	-0,2700	-0,3800	-0,1400	34,1889	298,1000
131,8800					
90	-0,3300	-0,3800	-0,1400	25,1892	232,2000
18,3200					

Uma explicação mais completa da sintaxe e dos detalhes de armazenamento dos dados nas variáveis, assim como dos sinalizadores da função PLSCVRND foi feita no item 4.1.2 (capítulo4).

O PRESS mínimo cumulativo é obtido com 3 VLs, seu valor é 33,62. O modelo com 6 variáveis obtidas na seleção com PCA/HCA teve desempenho apenas razoável na predição dos valores de atividade biológica dos compostos. Os coeficientes de regressão mostram (figura 5.10) que há contribuição significativa de todas as variáveis selecionadas para obtenção do modelo. O percentual da variância capturada com o modelo PLS é mostrado na tabela 5.6. A variância

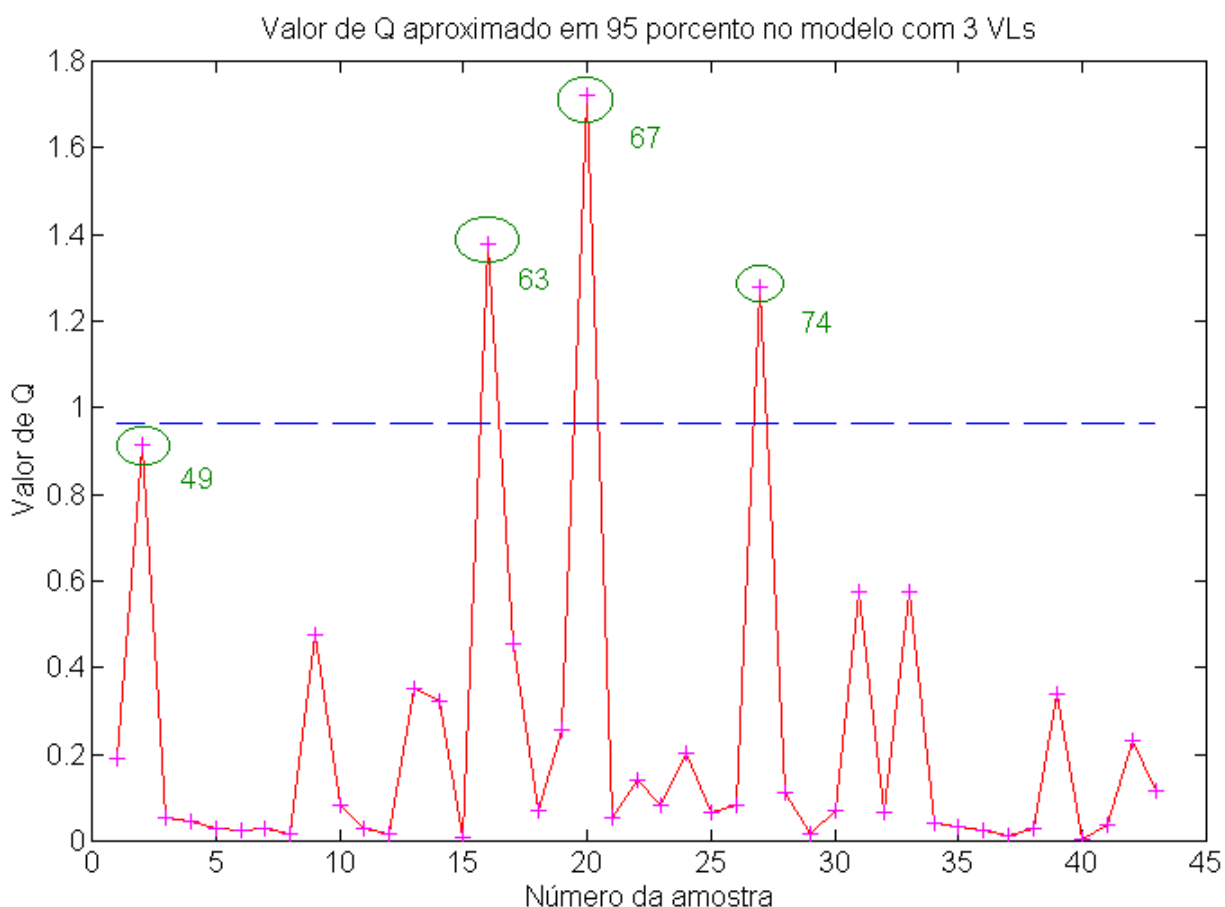


Figura 5.11: Valor de Q para cada um dos compostos no modelo PLS de relação quantitativa entre atividades *in vitro* e os dados selecionados, utilizando cargas de Mulliken. O valor de Q é uma medida da distância de cada amostra dos novos eixos de coordenadas criados, chamados variáveis latentes. O valor para limite de confiança no resultado é a distância em que deveriam estar situados 95% das amostras.

acumulada para as variáveis dependentes, o bloco Y, é pequena nestas três variáveis latentes.

Tabela 5.6: Variâncias acumuladas nas VLs para o modelo com variáveis calculadas com AM1 e cargas CHELP, e as atividades *in vitro* para os guanidinoiazóis.

VL #	-----Bloco-X-----		-----Bloco-Y-----	
	Esta VL	Total	Esta VL	Total
1	37,00	37,00	31,68	31,68
2	43,29	80,28	1,97	33,65
3	15,49	95,77	2,17	35,82

O limite de Q para 95% de confiança é 0,96. Os compostos 63, 67, e 74 têm o valor de Q obtido maior ou igual ao valor limite, e o composto 49 se aproxima bastante do limite (figura 5.11). Os compostos 63 e 67 têm em comum o fato de serem os únicos da série que têm o substituinte p-fenil na posição R3, além de pertencerem ao grupo de baixa atividade. A diferença entre eles é o substituinte R1, que é benzila no composto 67 e hidrogênio no composto 63. O composto 67 é tem o segundo menor valor de atividade *in vitro* medido para esta série de compostos, e o 49 o menor.

O composto 74 pertence ao grupo de alta atividade, situada na faixa intermediária dos valores para o seu grupo. Tem substituição *m*-metoxi na posição R3, e hidrogênio nas demais.

O limite de T^2 para 95 por cento de confiança é 8,94. Os valores podem ser vistos na figura 5.12.

Os valores de Q e T^2 mostram que não há grande dispersão das variáveis das amostras no conjunto de treinamento.

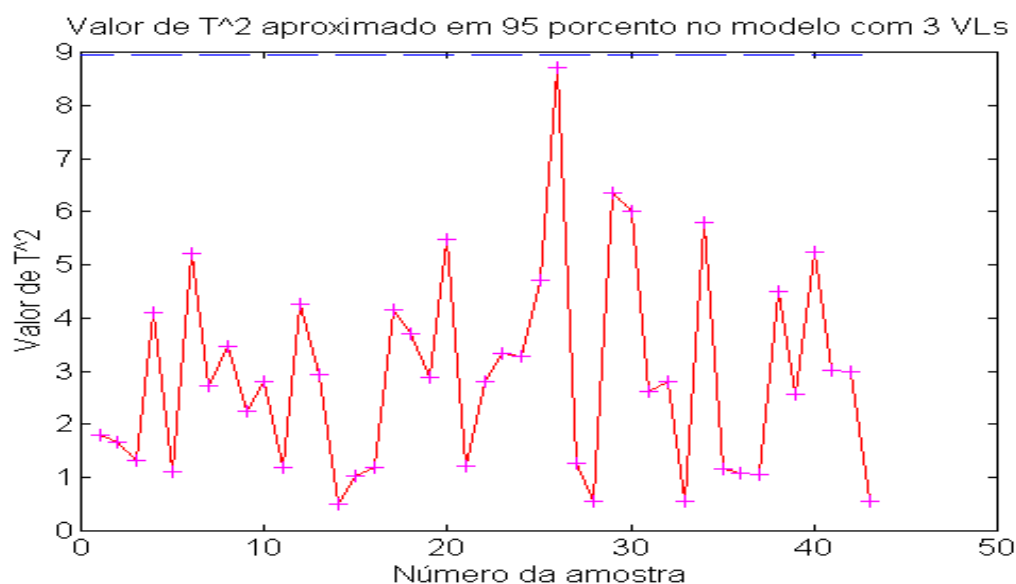


Figura 5.12: Valor de T^2 para cada um dos compostos no modelo PLS para relação quantitativa. O valor de T^2 é uma medida da distância de cada amostra ao ponto onde se cruzam os novos eixos (variáveis latentes). O valor para limite de confiança no resultado mostra a distância onde deveriam estar

5.3.2. Regressão PLS para atividades *in vivo*

O modelo PLS foi obtido com as variáveis selecionadas com PCA usadas como variáveis independentes no programa 5.1. As variáveis usadas neste caso são F_{C5}^{LUMO} , F_{C9}^{HOMO} , F_{C10}^{HOMO} , Q_{N3}^{CHELP} , Q_{R2}^{CHELP} , $\mu^{híbrido}$, α , vdWV, vdWSAV, e vdWVR1, mostradas na tabela 5.7. O gráfico da figura 5.13 mostra o coeficiente linear de cada uma das variáveis, na ordem em que aparecem na tabela 5.7. O valor autoescalado de $\log IC_{50}$, foi usado como variável dependente.

O coeficiente de regressão mostra que as variáveis que têm menor importância no ajuste do

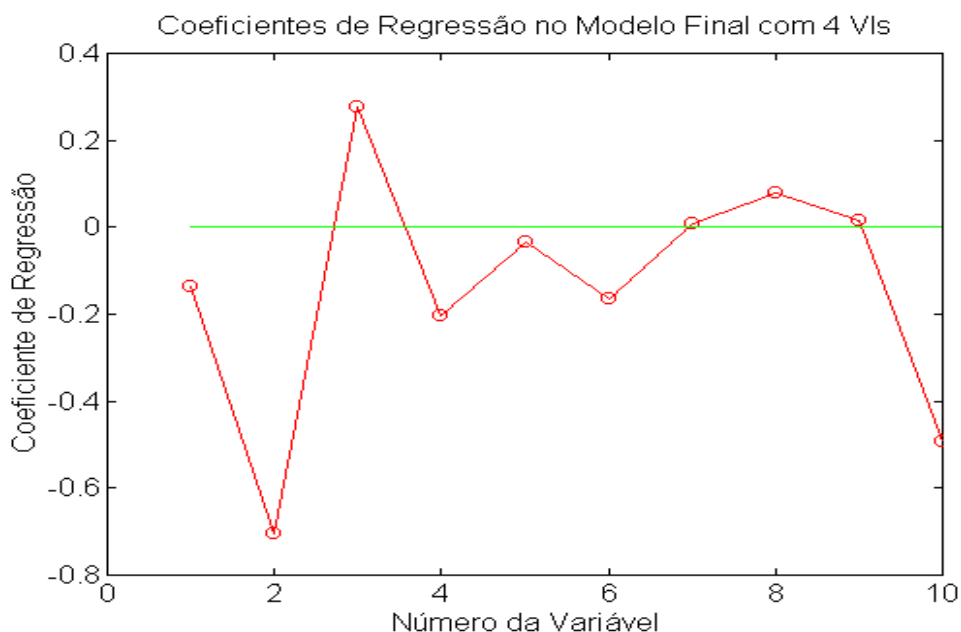


Figura 5.13: Valor do coeficiente de cada variável na matriz de regressão do modelo PLS com dados obtidos do Hamiltoniano AM1, com cargas CHELP, e o $\log IC_{50}$, a concentração que inibe 50% da atividade *in vivo* da H^+, K^+ -ATPase.

modelo aos dados experimentais são a carga no substituinte R2, a polarizabilidade e o volume de van der Waals acessível ao solvente, variáveis 5, 7, e 9, e foram mantidas com o intuito de comparar o desempenho dos modelos (PCA, PLS, Neural) com o mesmo conjunto de variáveis independentes. O PRESS mínimo cumulativo com 4 variáveis latentes vale 14,91, e o percentual da variância capturada com o modelo PLS é mostrado na tabela 5.8.

Tabela 5.7: Variáveis calculadas com Hamiltoniano AM1, com cargas CHELP, selecionadas com HCA/PCA e utilizadas para regressão PLS com as atividades *in vivo* dos guanidinothiazóis

No.	F _{C5} ^{LUMO}	F _{C9} ^{HOMO}	F _{C10} ^{HOMO}	Q _{N3} ^{CHELP}	Q _{R2} ^{CHELP}	μ ^{híbrido}	α _r	vdWV	vdWSAV	vdWVR1
50	-0,0549	-0,0207	-0,0179	-0,9600	1,0700	1,8610		197,7000		246,2000
22,6049								7,2300		
51	-0,0617	-0,0220	-0,0137	-0,8100	0,9500	1,8870		199,0000		252,3000
23,1656								7,2300		
53	-0,0665	-0,0213	-0,0126	-0,9500	1,0700	1,9920		190,8000		223,0000
22,4181								7,2300		
54	-0,0832	-0,0212	-0,0194	-0,9500	1,0700	1,8830		290,8000		342,9000
32,8028								99,2600		
58	-0,0593	-0,0165	-0,0015	-0,8100	0,9800	1,9090		202,2000		250,3000
22,8810								7,2300		
60	-0,0459	-0,0218	-0,0145	-0,9600	1,0700	1,8530		199,9000		257,3000
23,2545								7,2300		
61	-0,0335	-0,0215	-0,0141	-0,9700	1,0800	1,2410		244,7000		317,3000
26,6694								7,2300		
64	-0,0250	-0,0208	-0,0081	-0,9600	1,0700	1,8500		304,3000		337,7000
32,0625								99,2600		
70	-0,0287	-0,0124	-0,0024	-0,9600	1,0800	1,1850		220,8000		233,3000
24,0035								7,2300		
71	-0,0230	-0,0172	-0,0064	-0,9600	1,0700	0,5800		252,9000		270,7000
26,6246								7,2300		
72	-0,0218	-0,0120	-0,0024	-0,9700	1,0900	1,3600		301,8000		339,3000
34,2347								99,2600		
73	-0,0389	-0,0171	-0,0068	-0,9600	1,0900	1,4250		241,4000		238,0000
24,3254								7,2300		
74	-0,0452	-0,0216	-0,0150	-0,8600	0,9800	1,8850		306,1000		356,9000
30,8774								7,2300		
75	-0,0216	-0,0155	-0,0044	-0,9500	1,0400	0,8100		231,9000		254,2000
24,9693								7,2300		
76	-0,0211	-0,0168	-0,0099	-0,7500	0,9500	1,2570		323,6000		347,2000
34,1975								99,2600		
77	-0,0204	-0,0175	-0,0080	-0,9500	1,0700	1,2320		316,3000		350,7000
34,6882								99,2600		
83	-0,0228	-0,0175	-0,0078	-0,8000	0,9700	1,5680		207,2000		223,9000
22,8992								7,2300		
84	-0,0234	-0,0163	-0,0052	-0,9700	1,0900	0,7540		205,2000		229,0000
23,2203								7,2300		
85	-0,0325	-0,0169	-0,0087	-0,9700	1,0900	1,8750		295,2000		324,1000
32,8822								99,2600		
88	-0,0268	-0,0143	-0,0048	-0,8000	0,9500	1,5530		210,6000		240,3000
23,4409								7,2300		
89	-0,0212	-0,0147	-0,0056	-0,8400	1,0100	1,5690		298,1000		338,3000
34,1923								99,2600		
90	-0,0175	-0,0106	-0,0067	-0,9800	1,1000	1,3610		232,2000		263,6000
25,1913								7,2300		
91	-0,0204	-0,0155	-0,0165	-0,8100	0,9700	1,5860		303,7600		331,5200
25,3428								99,2600		

Tabela 5.8: Variâncias acumuladas nas VLs para o modelo com dados obtidos do Hamiltoniano AM1, com cargas CHELP, e o log IC₅₀, a concentração que inibe 50% da atividade *in vivo* da H⁺,K⁺-ATPase.

VL #	-----Bloco-X-----		-----Bloco-Y-----	
	Esta VL	Total	Esta VL	Total
1	34,70	34,70	46,11	46,11
2	20,21	54,90	10,58	56,70
3	12,38	67,28	9,85	66,55
4	21,12	88,40	2,35	68,90

Com quatro variáveis latentes o modelo acumula 88% da variância total das variáveis

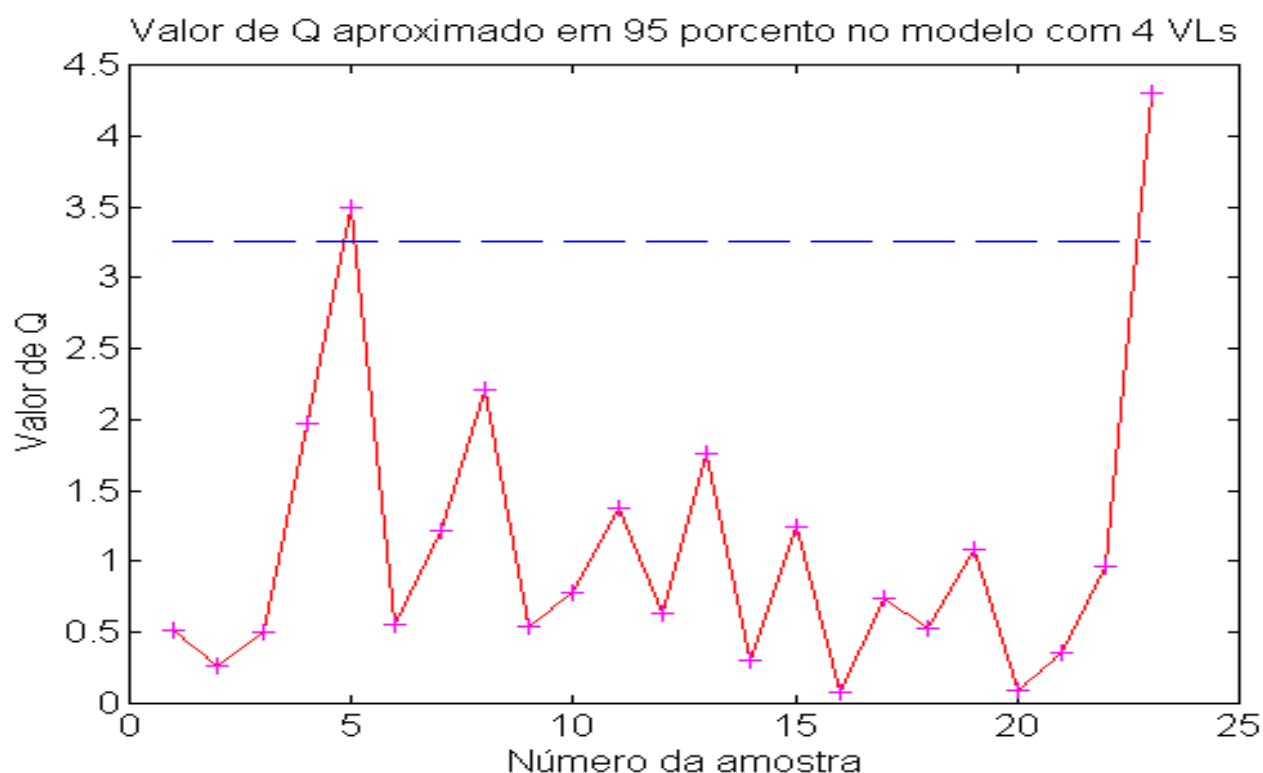


Figura 5.14: Valor de Q para cada um dos compostos no modelo PLS com dados obtidos com cargas CHELP do Hamiltoniano AM1, e o log IC₅₀, (a concentração que inibe 50% da atividade *in vivo* da H⁺,K⁺-ATPase). A linha tracejada é o valor para limite de confiança de 95% independentes autoescaladas, e 69% da variância total das variáveis dependentes.

O limite de Q para 95% de confiança é 3,26. Apenas dois compostos ultrapassam este limite no conjunto de dados utilizado na predição de seus valores de atividade, 54, e 91, a quinta e a vigésima terceira amostras respectivamente (figura 5.14). Têm em comum a presença de um grupo benzila na posição R1. Estes compostos estão situados no extremo superior da atividade

biológica, e ainda que a predição dos seus valores não seja obtida por extrapolação, a retirada destes compostos do conjunto afeta bastante a média e dificulta a previsão.

O limite de T^2 para 95 por cento de confiança é 8,94. Os valores obtidos para cada um dos compostos podem ser vistos na figura 5.15.

Neste modelo os valores de Q e de T^2 mostram uma maior dispersão no valor das variáveis relacionadas com as atividades dos compostos no conjunto de treinamento, que no caso do modelo com cargas de Mulliken e atividade *in vitro* estudado.

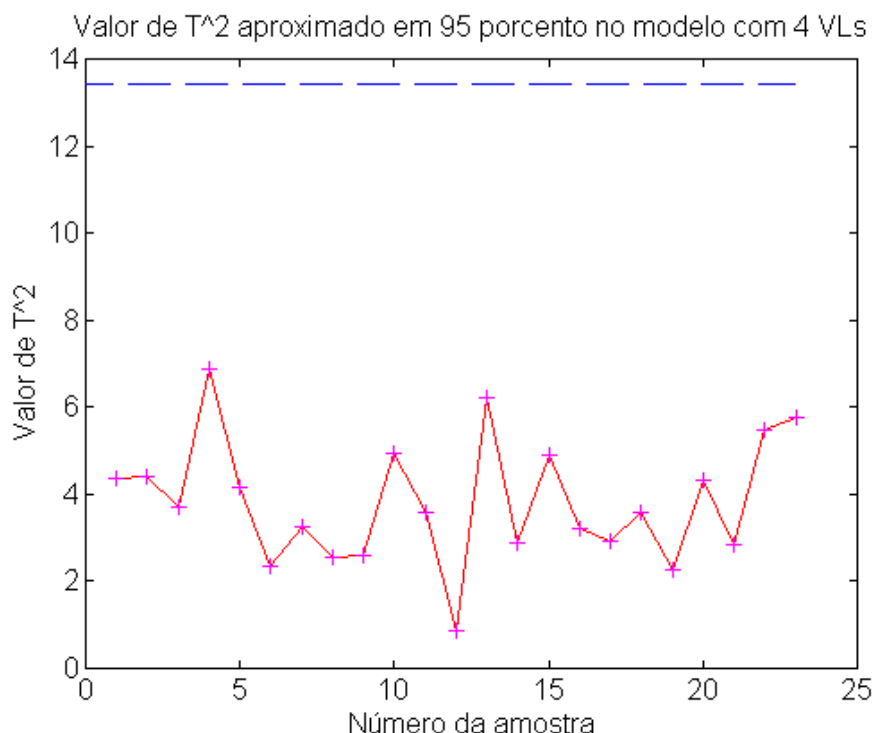


Figura 5.15: Valor de T^2 para cada um dos compostos no modelo PLS com dados obtidos com cargas CHELP do Hamiltoniano AM1, e o log IC_{50} , (a concentração que inibe 50% da atividade *in vivo* da H^+, K^+ -ATPase). A linha tracejada é o valor para limite de confiança de 95%.

5.3. Redes neurais usando os parâmetros selecionados com PCA/HCA

Para comparação entre as metodologias foram realizados testes de treinamento de redes neurais para os dados obtidos. A partir da seleção de variáveis feito por PCA e HCA para cada tipo de variável, foram treinadas redes neurais para reproduzir o padrão encontrado nesta série de compostos. Os resultados obtidos para os modelos com rede neural para as variáveis consideradas e a atividade *in vivo* e *in vitro* serão mostrados a seguir.

5.3.1. Modelos de classificação das atividades *in vitro*, usando cargas de Mulliken

O modelo inicialmente criado com o programa PSDD obtido do QCPE, foi obtido com uma rede de 6 neurônios na primeira camada, 10 neurônios na segunda, e dois na terceira, treinada por retropropagação, sem utilização de bias. Os parâmetros definidos para a criação da rede podem ser vistos na tabela 5.9, são iguais aos padrões internos do programa, e utilizados para todos os modelos de correlação com atividades *in vitro*, salvo menção contrária. As variáveis utilizadas na primeira camada foram Q_{N3}^{Mulliken} , Q_{N4}^{Mulliken} , Q_{C9}^{Mulliken} , α , vdWV, e R1vdWV. Os valores das variáveis podem ser vistos na tabela 5.5 .

Tabela 5.9:Valores dos parâmetros utilizados para criação da rede neural definidos internamente no programa PSDD, e aplicados nas equações 56 até 60.

Parâmetro	Valor
Valor de Alpha	1,00
Valor de Halpha	1,00
Valor deBeta (parâmetro de não linearidade)	1,00
Valor Inicial de eprison	0,15
Valor inicial de Hepsiron	0,15
Valor máximo de eprison	0,15
Valor máximo de Hepsiron	0,15
Controlador da fixação da matriz de pesos (XOPT)	0,50
Faixa de escalamento: (min. e max.)	0,10<r<0,90

O composto 82 foi o único classificado errado no treinamento da rede neural, como muito ativo

(tabela 5.10, classificação resultante).

Para previsão sucessiva de cada composto foi treinada uma rede com 20 neurônios na segunda camada, e limite na função de erro (equação 59) de 10^{-6} para cada ciclo de treinamento. Este modelo neural não foi capaz de prever acuradamente 10 compostos entre os 43 do conjunto todo, ou 23%, usando estas variáveis para treinamento da rede. Os compostos marcados em vermelho, 48, 54, 60, 69, e 89, são os que pertencem ao grupo de alta atividade e foram previstos no grupo de baixa atividade. Os compostos marcados em azul, 57, 59, 63, 78, e 82, pertencem ao grupo de baixa atividade e foram previstos no grupo de alta atividade (tabela 5.10, classificação prevista 1).

Foi criado um outro modelo com 20 neurônios na segunda camada, e limite da função de erro fixado em 0,01 para as mesmas variáveis do caso anterior, e os demais parâmetros de criação da rede também iguais.

Neste caso o aumento do limite de erro aplicado a cada ciclo melhorou a capacidade de predição do modelo. Foram classificados errado nove compostos do conjunto de 43, ou seja 21% dos compostos foram previstos erradamente (tabela 5.10, classificação prevista 2).

O valor de Beta igual a 1 implica nenhuma adição de função não linear à equação 56.

Os parâmetros não lineares da equação 56, ALPHA para a conexão entre a primeira e a segunda camadas, e HALPHA para a conexão entre a segunda e a terceira, o foram variados entre 2 e 7. O

melhor resultado obtido, com $\text{ALPHA}=3$ e $\text{HALPHA}=5$; 20 neurônios na segunda camada e função de erro igual a 0,001 está mostrado como classificação prevista 3, na tabela 5.10. Este modelo é melhor que os anteriores, com apenas 8 previsões erradas, ou seja, 18,6% de previsões erradas

Tabela 5.10: Resultados obtidos com redes neurais para as atividades in vitro dos 4-fenil 2-guanidinothiazóis com os dados selecionados com PCA/HCA, com cargas de Mulliken. Os resultados mostrados em vermelho são para os compostos muito ativos (classe 1) erradamente classificados. Os mostrados em azul são os pouco ativos (classe 2) classificados errado.

Padrões de Treinamento			Classificação		Classificação		Classificação		Classificação	
No.	Classe		Resultado		Prevista 1		Prevista 2		Prevista 3	
Padrão			1	2	1	2	1	2	1	2
48	1	1 0	,996		,252		,927	,074	,998	,002
			,004		,748					
49	2	0 1	,000		,000		,000	1,000	,012	,990
			1,000		1,000					
50	1	1 0	1,000		,691		1,000		,097	,905
			,000		,309		,000			
51	1	1 0	1,000		,899		1,000		1,000	,000
			,000		,101		,000			
52	1	1 0	1,000		,953		1,000		1,000	,000
			,000		,047		,000			
53	1	1 0	1,000		,985		1,000		,997	,003
			,000		,015		,000			
54	1	1 0	,883		,001		,000	1,000	,001	,999
			,116		,999					
55	1	1 0	1,000		,980		1,000		1,000	,000
			,000		,020		,000			
56	2	0 1	,004		,009		,008	,992	,000	1,000
			,996		,991					
57	2	0 1	,114		,950		,989		,999	,001
			,886		,050		,011			
58	1	1 0	1,000		,668		1,000		,733	,269
			,000		,332		,000			
59	2	0 1	,034		1,000		,951		,995	,005
			,966		,000		,048			
60	1	1 0	,983		,050		,021		,643	,340
			,017		,950		,980			
61	1	1 0	1,000		1,000		1,000		1,000	,000
			,000		,000		,000			
62	1	1 0	1,000		,945		1,000		1,000	,000
			,000		,055		,000			
63	2	0 1	,040		,994		1,000		1,000	,000
			,960		,006		,000			
64	1	1 0	,985		,898		,962		1,000	,000
			,015		,102		,043			
65	1	1 0	1,000		,995		1,000		1,000	,000
			,000		,005		,000			
66	2	0 1	,056		,139		,122		,373	,630
			,944		,861		,878			
67	2	0 1	,002		,011		,027		,000	1,000
			,998		,989		,974			
68	1	1 0	1,000		,917		1,000		1,000	,000
			,000		,083		,000			

69	1	1 0	1,000 ,000	,000 1,000	1,000 ,000	1,000 ,000	,000
70	1	1 0	1,000 ,000	,997 ,003	1,000 ,000	,999 ,000	,001
71	1	1 0	,997 ,004	,948 ,052	1,000 ,000	1,000 ,000	,000
72	1	1 0	1,000 ,000	1,000 ,000	1,000 ,000	,996 ,000	,004
73	1	1 0	,989 ,012	,991 ,008	,889 ,110	1,000 ,000	,000
74	1	1 0	1,000 ,000	,756 ,245	1,000 ,000	1,000 ,000	,000
75	1	1 0	1,000 ,000	,999 ,001	1,000 ,000	1,000 ,000	,000
76	1	1 0	,990 ,010	,714 ,285	,980 ,022	,977 ,000	,023
77	1	1 0	,998 ,002	1,000 ,000	1,000 ,000	,995 ,000	,005
78	2	0 1	,003 ,997	,787 ,212	,011 ,990	,000 ,000	1,000
79	2	0 1	,064 ,936	,311 ,689	,641 ,365	,151 ,000	,849
80	1	1 0	,969 ,031	1,000 ,000	,182 ,814	,780 ,000	,216
81	2	0 1	,003 ,997	,010 ,990	,010 ,990	,000 ,000	1,000
82	2	0 1	1,000 ,000	,999 ,001	1,000 ,000	1,000 ,000	,000
83	1	1 0	1,000 ,000	,586 ,414	1,000 ,000	,979 ,000	,020
84	1	1 0	1,000 ,000	,805 ,195	1,000 ,000	,997 ,000	,003
85	1	1 0	1,000 ,000	1,000 ,000	1,000 ,000	,997 ,000	,003
86	2	0 1	,086 ,914	,095 ,905	,240 ,758	,971 ,000	,029
87	1	1 0	1,000 ,000	1,000 ,000	1,000 ,000	1,000 ,000	,000
88	1	1 0	1,000 ,000	1,000 ,000	1,000 ,000	1,000 ,000	,000
89	1	1 0	,891 ,109	,005 ,995	,012 ,988	,002 ,000	,999
90	1	1 0	1,000 ,000	1,000 ,000	1,000 ,000	1,000 ,000	,000
Percentual de classificações erradas			2%	23%	21%	18,6%	

5.3.2. Modelo quantitativo para as atividades *in vitro*, com cargas de Mulliken

Esta rede foi treinada com os mesmas variáveis utilizadas para classificação PCA/HCA,

mostradas na tabela 5.5. Os resultados obtidos podem ser vistos na tabela 5.11 .

Foram usados 20 neurônios na segunda camada, e limite de erro de 10^{-5} para cada ciclo de treinamento. Testes com outros limites de erro foram feitos, sem nenhuma melhora significativa no modelo. Os parâmetros Alpha e Halpha são modificados pelo programa para 4,0 em modelos quantitativos, com um único neurônio na última camada, pelo uso da palavra chave QSAR incluída no arquivo de entrada de dados.

Foi feita a regressão linear entre os valores da tabela 5.11, para quantificar o ajuste obtido dos dados experimentais pelo modelo neural. Os resultados são mostrados na tabela 5.12, e mostrados graficamente na figura 5.16.

Tabela 5.11: Padrões de treinamento e valores ajustados pelo modelo neural para as atividades *in vitro* dos guanidinothiazóis utilizando Hamiltoniano Am1 e cargas de Mulliken.

• 48	49	50	51	52	53	54	55	56
Valor:52,000	-1,000	71,000	48,000	48,000	61,000	99,000	80,000	38,000
Calc: 51,583	4,729	54,932	51,045	56,529	63,108	106,912	78,871	33,288
• 57	58	59	60	61	62	63	64	65
Valor:35,000	78,000	27,000	60,000	56,000	61,000	25,000	111,000	53,000
Calc: 41,021	58,703	27,738	58,753	60,984	62,513	26,600	109,697	51,496
• 66	67	68	69	70	71	72	73	74
Valor:16,000	5,000	45,000	52,000	63,000	75,000	77,000	70,000	48,000
Calc: 13,2383	3,900	53,391	49,391	64,324	75,066	71,594	70,053	52,277
• 75	76	77	78	79	80	81	82	83
Valor:54,000	99,000	85,000	25,000	24,000	41,000	25,000	26,000	69,000
Calc: 81,942	101,540	77,138	27,157	32,194	48,097	29,672	62,668	64,598
• 84	85	86	87	88	89	90		
Valor:75,000	85,000	8,000	49,000	100,000	111,000	97,000		
Calc: 62,989	72,497	39,225	49,280	105,673	133,534	81,649		

Tabela 5.12: Resultados da regressão linear entre os valores calculados e os valores medidos experimentalmente da atividade *in vitro* dos guanidínotiazóis.

$Y = A + B * X$		
Parâmetro	Valor	Erro
A	8,99457	
	3,64264	
B	0,87998	
	0,05766	

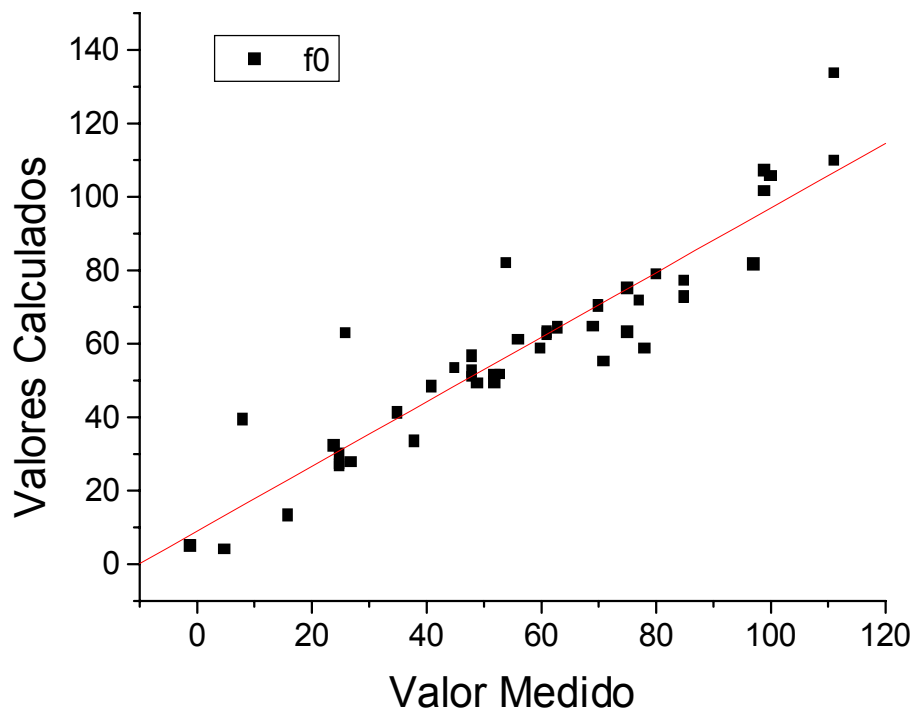


Figura 5.16: Gráfico da regressão entre os valores experimentais da atividade *in vitro* e os valores obtidos nos cálculos com a rede neural treinada com parâmetros calculados com AM1 e cargas de Mulliken.

R	SD	N	P
0,92214	10,75039	43	
<0,0001			

5.3.3. Modelo com as atividades *in vitro* com cargas CHELP.

Foram treinadas redes neurais que relacionam as atividades medidas *in vitro* com as variáveis selecionadas no modelo com PCA/HCA, utilizando-se o Hamiltoniano AM1 e as cargas derivadas do potencial eletrostático. As variáveis utilizadas; Q_{N3}^{CHELP} , Q_{N4}^{CHELP} , Q_{C6}^{CHELP} , Q_{C7}^{CHELP} , Q_{R2}^{CHELP} , α , vdWV, vdWVR1 são mostradas na tabela 5.3; e os resultados, são mostrados na tabela 5.13.

Os padrões de treinamento foram acuradamente reproduzidos pelo modelo neural, e estão mostrados como “classificação resultante” na tabela 5.13. A predição um a um com uma rede treinada com limite de erro de 10^{-5} e o mesmo número de neurônios na primeira e segunda camada (8) é mostrada como “previsão 1” na tabela 5.13. O resultado para uma rede treinada com o dobro de neurônios na segunda camada (16) é mostrado como “previsão 2” na tabela 5.13. Os compostos pouco ativos classificados errado são marcados em azul, e os muito ativos classificados errado são marcados em vermelho.

O modelo com mais neurônios na segunda camada foi mais eficiente na previsão da classe dos compostos retirados no treinamento do modelo, diminuindo de 16 para 14 as previsões erradas. Significa que o modelo errou 32% das previsões com este conjunto de variáveis e padrões de treinamento da rede neural.

A variação dos valores de Alpha e Halpha para 3 e 5, respectivamente, a partir do modelo “previsão 2” criou o modelo “previsão 3”, que faz 12 classificações erradas, ou 28% das amostras.

Tabela 5.13: Resultados dos modelos com redes neurais para as atividades *in vitro* dos 4-fenil 2-guanidinothiazóis, usando as cargas CHELP.

Padrão de Treinamento.		Classificação o Resultante		Classificação o Prevista 1		Classificação o Prevista 2		Classificação o Prevista 3	
Num. Classe	Padrão	1	2	1	2	1	2	1	2
48	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000		1,000 ,000	
49	2	0 1	,009 ,991	,998 ,002		,996 ,004		,151 ,842	
50	1	1 0	1,000 ,000	,000 1,000		,001 ,999		1,000 ,000	
51	1	1 0	,999 ,001	1,000 ,000		1,000 ,000		1,000 ,000	
52	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000		1,000 ,000	
53	1	1 0	,999 ,001	1,000 ,000		,999 ,001		1,000 ,000	
54	1	1 0	,983 ,017	,000 1,000		,000 1,000		,000 1,000	
55	1	1 0	,991 ,008	,936 ,065		1,000 ,000		,999 ,001	
56	2	0 1	,007 ,993	,387 ,614		,168 ,833		,001 ,999	
57	2	0 1	,017 ,984	,000 1,000		,000 1,000		,071 ,936	
58	1	1 0	,999 ,001	1,000 ,000		,995 ,005		1,000 ,000	
59	2	0 1	,020 ,980	1,000 ,000		1,000 ,000		,999 ,001	
60	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000		1,000 ,000	
61	1	1 0	1,000 ,000	,829 ,171		,976 ,024		,997 ,004	
62	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000		1,000 ,000	
63	2	0 1	,020 ,979	1,000 ,000		,058 ,947		1,000 ,000	
64	1	1 0	,995 ,005	1,000 ,000		1,000 ,000		,983 ,024	
65	1	1 0	,992 ,008	1,000 ,000		1,000 ,000		,998 ,002	
66	2	0 1	,021 ,979	1,000 ,000		,999 ,001		1,000 ,000	
67	2	0 1	,018 ,982	,000 1,000		,000 1,000		,000 1,000	
68	1	1 0	1,000 ,000	1,000 ,000		,905 ,094		1,000 ,000	
69	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000		,984 ,012	
70	1	1 0	,971 ,029	,270 ,730		,592 ,409		,900 ,122	
71	1	1 0	,997 ,003	,998 ,002		,999 ,001		,991 ,011	
72	1	1 0	,996 ,004	1,000 ,000		1,000 ,000		1,000 ,000	
73	1	1 0	,979 ,021	,003 ,997		,009 ,991		,000 1,000	
74	1	1 0	1,000 ,000	,000 1,000		,000 1,000		,068 ,943	
75	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000		1,000 ,000	
76	1	1 0	,993 ,007	,836 ,162		,986 ,014		1,000 ,000	
77	1	1 0	,972 ,028	,005 ,995		,015 ,985		,000 1,000	
78	2	0 1	,004 ,996	,999 ,001		,997 ,002		,029 ,972	
79	2	0 1	,000 1,000	,002 ,998		,004 ,996		,000 1,000	
80	1	1 0	1,000 ,000	,913 ,087		,936 ,065		1,000 ,000	
81	2	0 1	,020 ,980	1,000 ,000		1,000 ,000		1,000 ,000	
82	2	0 1	,026 ,974	1,000 ,000		1,000 ,000		1,000 ,000	
83	1	1 0	1,000 ,000	,713 ,281		1,000 ,000		1,000 ,000	
84	1	1 0	,991 ,009	1,000 ,000		,998 ,002		1,000 ,000	
85	1	1 0	,995 ,005	1,000 ,000		,992 ,008		,999 ,001	
86	2	0 1	,028 ,972	,026 ,974		,266 ,738		,867 ,135	
87	1	1 0	,980 ,020	,000 1,000		,000 1,000		,973 ,031	
88	1	1 0	1,000 ,000	,000 1,000		,000 1,000		1,000 ,000	
89	1	1 0	,974 ,025	,431 ,560		,430 ,565		,000 1,000	
90	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000		1,000 ,000	
Percentual de		0%		37%		32%		28%	

classificações
erradas

5.3.4. Modelo quantitativo para as atividades *in vitro*, com cargas CHELP

O modelo foi ajustado para reproduzir os dados quantitativos de atividade biológica para os dados selecionados. Trata-se do treinamento de uma rede neural, e não de previsão com exclusão sucessiva de cada amostra. Os resultados reproduzem fielmente as atividades medidas experimentalmente, como pode ser visto no gráfico da figura 5.17 e na tabela 5.14, a regressão linear entre os dados medidos experimentalmente e os ajustados pelo modelo mostra os resultados da tabela 5.15.

Tabela 5.14: Padrões de treinamento e valores ajustados pelo modelo neural para as atividades *in vitro* dos guanidinothiazóis utilizando Hamiltoniano Am1 e cargas CHELP.

•	48	49	50	51	52	53	54	55	56
Valor:	52,000	-1,000	71,000	48,000	48,000	61,000	99,000	80,000	38,000
Calc.:	52,370	-,965	71,057	48,703	47,615	60,309	98,801	80,322	38,008
•	57	58	59	60	61	62	63	64	65
Valor:	35,000	78,000	27,000	60,000	56,000	61,000	25,000	111,00	53,000
Calc.:	35,081	77,134	26,890	59,618	55,320	60,296	25,058	110,970	53,066
•	66	67	68	69	70	71	72	73	74
Valor:	16,000	5,000	45,000	52,000	63,000	75,000	77,000	70,000	48,000
Calc.:	16,034	5,053	44,723	51,207	62,983	74,825	76,936	69,901	47,688
•	75	76	77	78	79	80	81	82	83
Valor:	54,000	99,000	85,000	25,000	24,000	41,000	25,000	26,000	69,000
Calc.:	53,575	98,854	84,942	24,661	24,026	41,228	24,973	25,720	68,768
•	84	85	86	87	88				
Valor:	75,000	85,000	8,000	49,000	100,000				
Calc.:	75,274	84,851	7,755	48,953	99,977				
•	89	90							
Valor:	111,000	97,000							
Calc.:	110,950	97,010							

Tabela 5.15: Resultados da regressão linear entre os valores calculados e os valores medidos experimentalmente da atividade *in vitro* dos guanidinothiazóis.

Y = A + B * X			
Parâmetro	Valor	Erro	
A	-0,09443		
	0,10712		
B	0,99942	0,0017	
R	SD	N	P
0,99994	0,31613	43	
	<0,0001		

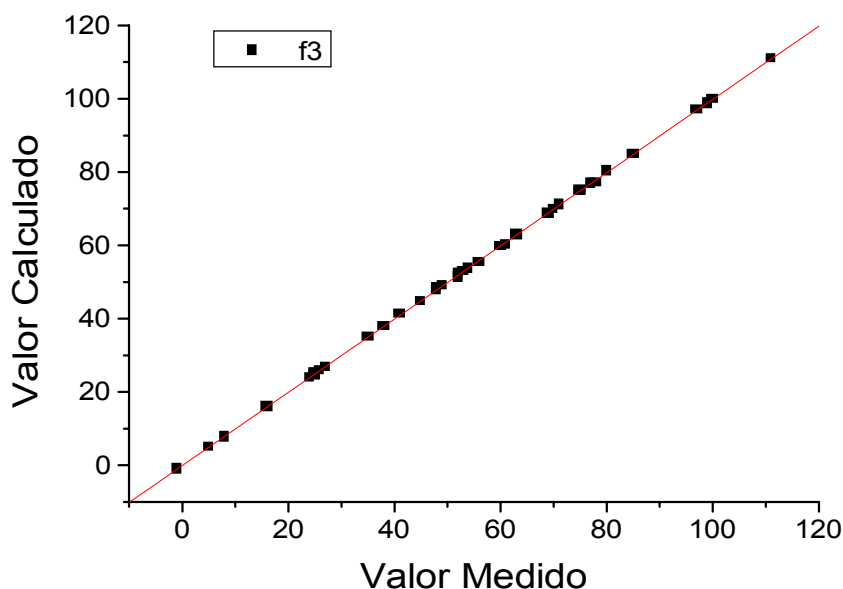


Figura 5.17: Gráfico da regressão entre os valores experimentais da atividade *in vitro* e os valores obtidos nos cálculos com a rede neural treinada com parâmetros calculados com AM1 e cargas CHELP.

5.3.5. Modelos de classificação *in vivo*

Para relacionar as atividades *in vivo* reportadas para os 23 compostos da série, foram treinadas redes neurais. Foi também estudada a influência da variação do limite de erro aplicado em cada ciclo de treinamento. Os resultados obtidos são mostrados na tabela 5.16. As 10 variáveis utilizadas na primeira camada de neurônios foram selecionadas por PCA/HCA. São: F_{C5}^{LUMO} , F_{C9}^{HOMO} , F_{C10}^{HOMO} , Q_{N3}^{CHELP} , Q_{R2}^{CHELP} , $\mu^{hibrido}$, α , VvdW, vdWSA, e R1VvdW, podem ser vistas na tabela 5.7.

Tabela 5.16: Resultados dos modelos com redes neurais para as atividades *in vivo* dos 4-fenil 2-guanidino-1,2,3,4-tetrazóis, usando as variáveis selecionadas com PCA/HCA.

Padrões de treinamento		Padrão Calculado		Padrão Previsto 1		Padrão Previsto 2	
Num.	Classe	1	2	1	2	1	2
50	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000	
51	1	1 0	,997 ,003	,996 ,004		,991 ,008	
53	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000	
54	2	0 1	,000 1,000	,000 1,000		,000 1,000	
58	1	1 0	,997 ,004	,996 ,004		,988 ,012	
60	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000	
61	1	1 0	,991 ,010	,906 ,094		,390 ,610	
64	2	0 1	,007 ,992	1,000 ,000		1,000 ,001	
70	1	1 0	,982 ,017	,004 ,996		,026 ,975	
71	2	0 1	,018 ,983	1,000 ,000		,996 ,004	
72	2	0 1	,000 1,000	,000 1,000		,000 1,000	
73	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000	
74	1	1 0	1,000 ,000	1,000 ,000		1,000 ,000	
75	1	1 0	,975 ,025	,000 1,000		,000 1,000	
76	2	0 1	,000 1,000	,000 1,000		,000 1,000	
77	2	0 1	,000 1,000	,000 1,000		,000 1,000	
83	1	1 0	,982 ,019	,000 1,000		,002 ,998	
84	2	0 1	,020 ,980	1,000 ,000		,999 ,001	
85	2	0 1	,000 1,000	,000 1,000		,000 1,000	
88	2	0 1	,022 ,978	1,000 ,000		1,000 ,000	
89	2	0 1	,000 1,000	,000 1,000		,000 1,000	
90	2	0 1	,013 ,987	,823		,915 ,088	

91	2	0	1	,000	1,000	,173 ,000 1,000	,000	1,000
Percentual de classificações erradas		0%		35%		39%		

Os padrões de atividade do conjunto de treinamento foram fielmente reproduzidos pelo modelo neural, não havendo nenhuma classificação errada nos padrões calculados pelo modelo.

O modelo neural inicial foi treinado para reproduzir as classificações com 22 neurônios na segunda camada, sem bias, por retropropagação, com limite de erro fixado em 10^{-5} . Os parâmetros para estabelecer os limites das variáveis internas na formação da rede neural utilizados foram os padrões do programa.

Para as mesmas variáveis foi testado o modelo de previsão sucessiva, que retira seqüencialmente cada um dos compostos do conjunto e prevê a sua posição com o modelo criado a partir dos restantes. Os parâmetros para a criação da rede foram os mesmos que no caso anterior.

Para o caso da previsão dos compostos por exclusão sucessiva de cada um deles o modelo errou 8 classificações, no conjunto de 23 compostos, totalizando 35% de erro. As previsões erradas foram os compostos 64, 71, 84, 88, e 90 previstos como pouco ativos no modelo, e os compostos 70, 75, e 83 previstos como muito ativos no modelo. Não há relação clara entre a classificação e alguma estrutura química presente nos compostos.

Outro modelo foi construído para as mesmas variáveis, e os mesmos parâmetros para a criação da rede, exceto o limite de erro, fixado em 0,01. O resultado é mostrado nos padrões previstos 2. Houveram 9 classificações erradas, uma a mais que no caso anterior (o composto 61 que é pouco ativo foi erradamente classificado no grupo muito ativo) totalizando 39% de classificações erradas.

5.3.6. Ajuste de rede quantitativa para as atividades *in vivo*

O modelo quantitativo foi obtido com as mesmas variáveis que o modelo classificatório para as atividades *in vivo*, mostrados na tabela 1.2 (capítulo 1). O programa foi instruído a realizar regressão não linear múltipla, com a palavra chave QSAR. O limite de erro fixado em 10^{-4} , com 22 neurônios na segunda camada. O ajuste da rede aos dados experimentais (tabela 5.17) é

muito bom, como pode ser observado no gráfico (figura 5.18 e tabela 5.18) dos valores calculados versus valores experimentais.

Tabela 5.17: Padrões de treinamento e valores ajustados pelo modelo neural para as atividades *in vivo* dos guanidinothiazóis utilizando Hamiltoniano AM1 e cargas CHELP.

• 50	51	53	54	58	60	61	64	70
Valor: 2,890	3,970	3,457	1,792	2,996	3,466	3,611	2,303	3,466
Calc: 2,886	3,966	3,446	1,784	2,989	3,454	3,602	2,321	3,466
• 71	72	73	74	75	76	77	83	84
Valor: 2,639	1,649	3,296	4,094	3,555	,833	2,303	3,135	2,708
Calc: 2,654	1,686	3,327	4,109	3,558	,838	2,307	3,142	2,709
• 85	88	89	90	91				

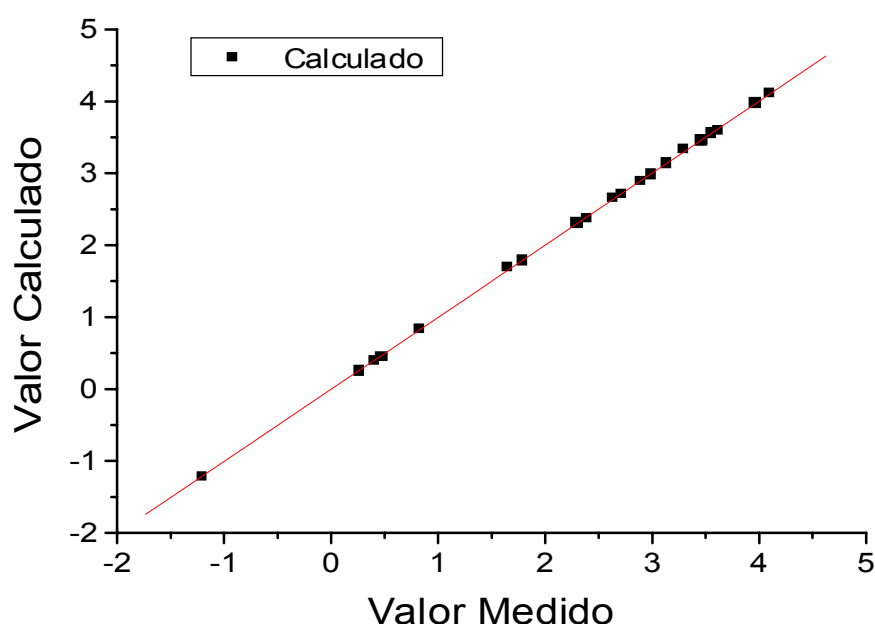


Figura 5.18: Gráfico da regressão entre os valores experimentais da atividade *in vivo* e os valores ajustados nos cálculos com a rede neural treinada com parâmetros calculados com AM1 e cargas CHELP.

Valor: 2,398	,406	-1,204	,470	,262
Calc: 2,375	,397	-1,216	,451	,254

Tabela 5.18: Resultados da regressão linear entre os valores calculados e os valores medidos experimentalmente da atividade *in vivo* dos guanidinothiazóis.

$Y = A + B * X$			
Parâmetro	Valor	Erro	
A	-0,00524	0,00625	
B	1,00237	0,00229	
R	SD	N	P

0,99995	0,01489	23	<0,0001
---------	---------	----	---------

5.3.7. Modelo de previsão quantitativa das atividades *in vivo*

O modelo de previsão quantitativa foi obtido com os mesmos dados do ajuste da rede, porém com limite de erro fixado em 0,01. O maior valor do limite de convergência produziu melhor resultado na predição dos valores de atividade (tabela 5.19). Este modelo tem desempenho razoável na predição quantitativa dos valores. O valor de R, o coeficiente de correlação entre os conjuntos de dados comparados, é obtido na previsão da atividade dos compostos pelo modelo (tabela 5.20), traçando-se o gráfico do valor previsto versus o valor medido experimentalmente (figura 5.19), pode ser comparado ao valor denominado q^2 , utilizado em modelos de CoMFA para testar o ajuste de uma regressão PLS ^{liv}. Para predição de compostos, com a mesma metodologia de “deixar um de fora”, é considerado bom se $q^2 > 0,3$ (no caso $R^2 = 0,305$). Para este tipo de regressão não deve-se esperar resultados como os obtidos na ajuste simples de uma função aos dados experimentais, quer seja função linear ou polinomial.

Tabela 5.19: Padrões de treinamento e classificações resultantes da rede treinada, com método “deixe um de fora”.

• 50	51	53	54	58	60	61	64	70
Valor: 2,890	3,970	3,457	1,792	2,996	3,466	3,611	2,303	3,466
Previ: 2,874	3,722	4,354	1,982	2,475	3,304	2,661	3,832	2,577
• 71	72	73	74	75	76	77	83	84
Valor: 2,639	1,649	3,296	4,094	3,555	,833	2,303	3,135	2,708
Previ: 4,298	1,699	2,767	4,569	2,149	-2,388	1,075	,461	4,676
• 85	88	89	90	91				
Valor: 2,398	,406	-1,204	,470	,262				
Previ: 1,034	2,570	2,540	2,706	-1,127				

Tabela 5.20: Regressão linear entre os valores previstos para cada um dos compostos, e o log dos resultados experimentais de atividade *in vivo*.

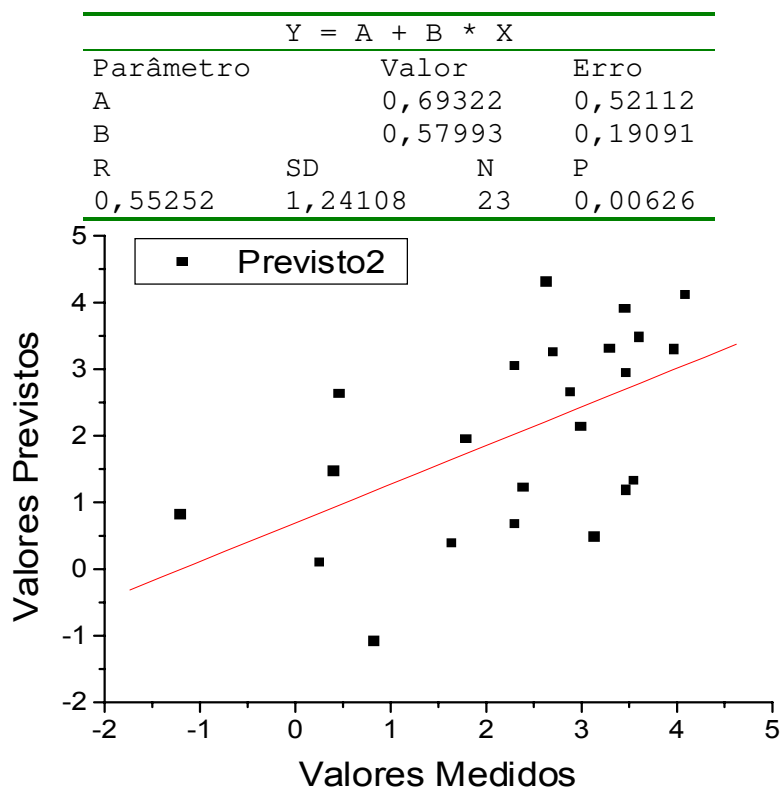


Figura 5.19: Gráfico da regressão entre os valores experimentais da atividade *in vivo* e os valores previstos com o método “deixe um de fora” com a rede neural treinada com

5.4. Seleção preliminar dos parâmetros

5.4.1. Análise exploratória das cargas obtidas para os 4-fenil 2-guanidinotiazóis

Os dados das cargas nas diferentes posições atômicas (figura 5.20) foram calculados para o conjunto inicial de 43 compostos e são mostrados na tabela 5.21. Para este conjunto de moléculas foram calculadas as cargas derivadas do potencial eletrostático (CHELP) e de Mulliken. A seleção de variáveis foi feita diretamente no conjunto de dados,

observando-se quais variáveis agrupavam corretamente os compostos mais ou menos ativos, em gráficos das variáveis duas a duas. Uma variável que isolava boa parte dos compostos de um dos grupos de atividade (alto ou baixo) foi considerada com grande poder de modelagem.

A grosso modo, verificou-se quais variáveis tinham uma faixa de valor característica num dos grupos, que poderia ser utilizada para determinar que compostos com aqueles valores pertenceriam ao tal grupo.

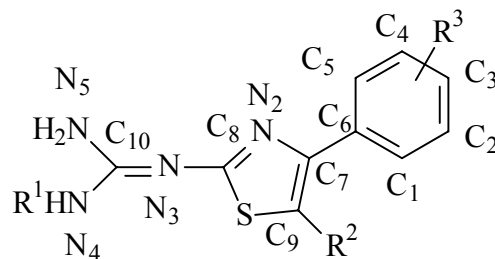


Figura 5.20: A numeração dos átomos adotada nas análises de cargas e densidades eletrônicas, diferente da numeração sistemática

Tabela 5.21 Cargas CHELP (Ax) e de Mulliken (AxM), em negrito, calculadas para os 43 guanidinothiazóis, nas linhas.

S1	S1M	N1	N1M	N2	N2M	N3	N3M	N4	N4M	R2	No.
0,06	0,52	-0,52	-0,2	-0,87	-0,3	-0,96	-0,33	-0,94	-0,38	1,07	48
0,06	0,52	-0,51	-0,2	-0,81	-0,29	-0,81	-0,27	-0,78	-0,38	0,95	49
0,06	0,51	-0,52	-0,19	-0,88	-0,3	-0,95	-0,33	-0,86	-0,38	1,04	50
0,07	0,52	-0,57	-0,21	-0,89	-0,3	-0,95	-0,33	-0,9	-0,39	1,07	51
0,07	0,52	-0,55	-0,2	-0,88	-0,3	-0,95	-0,33	-0,93	-0,38	1,07	52
0,07	0,52	-0,55	-0,21	-0,87	-0,3	-0,96	-0,33	-0,92	-0,38	1,07	53
0,06	0,51	-0,51	-0,19	-0,83	-0,30	-0,82	-0,29	-0,82	-0,38	0,99	54
0,07	0,51	-0,5	-0,21	-0,85	-0,31	-0,92	-0,33	-0,72	-0,34	1,02	55
0,07	0,52	-0,51	-0,2	-0,82	-0,29	-0,81	-0,27	-0,83	-0,38	0,98	56
0,08	0,52	-0,53	-0,21	-0,85	-0,3	-0,82	-0,29	-0,88	-0,38	1,02	57
0,07	0,51	-0,53	-0,2	-0,86	-0,3	-0,96	-0,33	-0,94	-0,38	1,07	58
0,08	0,52	-0,57	-0,2	-0,86	-0,3	-0,97	-0,33	-0,93	-0,38	1,08	59
0,06	0,52	-0,49	-0,2	-0,73	-0,3	-0,94	-0,32	-0,63	-0,38	0,82	60
0,07	0,52	-0,54	-0,20	-0,87	-0,30	-0,96	-0,33	-0,93	-0,38	1,07	61
0,06	0,51	-0,53	-0,21	-0,87	-0,30	-0,96	-0,33	-0,92	-0,38	1,07	62
0,06	0,52	-0,56	-0,2	-0,88	-0,3	-0,96	-0,33	-0,93	-0,38	1,07	63
0,06	0,51	-0,39	-0,14	-0,76	-0,24	-0,99	-0,36	-0,85	-0,28	1,03	64
0,06	0,5	-0,57	-0,21	-0,86	-0,31	-0,98	-0,34	-0,8	-0,34	1,06	65
0,06	0,52	-0,51	-0,2	-0,83	-0,29	-0,85	-0,27	-0,83	-0,38	1,01	66
0,05	0,51	-0,52	-0,21	-0,84	-0,3	-0,83	-0,29	-0,86	-0,38	1,01	67
0,07	0,51	-0,56	-0,2	-0,87	-0,3	-0,96	-0,33	-0,95	-0,38	1,08	68
0,06	0,51	-0,53	-0,2	-0,86	-0,3	-0,96	-0,33	-0,92	-0,38	1,07	69
0,05	0,5	-0,59	-0,18	-0,89	-0,29	-0,97	-0,39	-0,96	-0,33	1,09	70
0,05	0,5	-0,6	-0,19	-0,88	-0,29	-0,96	-0,38	-0,96	-0,32	1,09	71
0,06	0,51	-0,52	-0,2	-0,82	-0,29	-0,86	-0,37	-0,79	-0,28	0,98	72
0,07	0,5	-0,55	-0,18	-0,81	-0,28	-0,95	-0,38	-0,95	-0,33	1,04	73
0,05	0,51	-0,5	-0,21	-0,84	-0,31	-0,75	-0,31	-0,84	-0,38	0,95	74
0,07	0,52	-0,65	-0,21	-0,87	-0,3	-0,95	-0,33	-0,93	-0,38	1,07	75
0,06	0,51	-0,51	-0,2	-0,81	-0,29	-0,82	-0,38	-0,81	-0,27	0,97	76
0,06	0,52	-0,51	-0,20	-0,82	-0,29	-0,81	-0,38	-0,82	-0,27	0,97	77
0,05	0,51	-0,51	-0,21	-0,84	-0,31	-0,75	-0,31	-0,84	-0,38	0,95	78
0,07	0,51	-0,52	-0,21	-0,85	-0,3	-0,86	-0,29	-0,92	-0,38	1,06	79
0,07	0,52	-0,57	-0,2	-0,88	-0,3	-0,95	-0,33	-0,94	-0,38	1,08	80
0,07	0,52	-0,52	-0,2	-0,8	-0,29	-0,8	-0,27	-0,81	-0,38	0,97	81
0,07	0,51	-0,56	-0,2	-0,87	-0,29	-0,97	-0,33	-0,97	-0,38	1,09	82
0,07	0,52	-0,55	-0,2	-0,89	-0,3	-0,97	-0,33	-0,95	-0,38	1,09	83
0,07	0,52	-0,54	-0,21	-0,87	-0,30	-0,96	-0,33	-0,94	-0,38	1,08	84
0,07	0,51	-0,54	-0,20	-0,85	-0,29	-0,87	-0,38	-0,84	-0,29	1,02	85
0,06	0,52	-0,52	-0,20	-0,8	-0,29	-0,8	-0,27	-0,81	-0,38	0,95	86
0,07	0,51	-0,53	-0,2	-0,82	-0,29	-0,84	-0,38	-0,83	-0,28	1,01	87
0,07	0,51	-0,6	-0,18	-0,89	-0,28	-0,98	-0,38	-0,97	-0,33	1,1	88
0,07	0,52	-0,52	-0,2	-0,81	-0,29	-0,81	-0,27	-0,82	-0,38	0,97	89
0,04	0,5	-0,59	-0,2	-0,86	-0,3	-0,96	-0,33	-0,93	-0,38	1,07	90

Continua.

Continuação da tabela 5.21:

C1	C1M	C2	C2M	C3	C3M	C4	C4M	C5	C5M	R2M	No.
-0,03	-0,1	-0,11	-0,14	-0,1	-0,13	-0,1	-0,13	-0,08	-0,12	0,25	48
-0,07	-0,1	-0,09	-0,14	-0,1	-0,13	-0,1	-0,13	-0,1	-0,13	0,25	49
-0,14	-0,04	0,03	-0,13	-0,15	-0,12	-0,09	-0,13	-0,11	-0,11	0,25	50
-0,02	-0,1	-0,15	-0,13	0,02	-0,12	-0,14	-0,06	0,03	-0,11	0,26	51
-0,09	-0,09	0,0	-0,13	-0,14	-0,06	0,02	-0,13	-0,14	-0,11	0,25	52
0	-0,08	-0,24	-0,17	0,31	0,09	-0,24	-0,17	-0,03	-0,09	0,25	53
-0,13	-0,04	0,03	-0,13	-0,14	-0,12	-0,08	-0,13	-0,09	-0,13	0,26	54
-0,01	-0,1	-0,15	-0,13	0,01	-0,12	-0,14	-0,06	0,04	-0,11	0,26	55
-0,1	-0,09	0	-0,13	-0,15	-0,06	0,05	-0,13	-0,16	-0,11	0,25	56
-0,02	-0,08	-0,24	-0,17	0,31	0,09	-0,23	-0,17	-0,04	-0,09	0,26	57
0,09	-0,04	-0,16	-0,14	-0,06	-0,12	-0,16	-0,14	-0,01	-0,11	0,25	58
0,28	-0,09	-0,2	-0,13	-0,09	-0,13	-0,1	-0,13	-0,04	-0,11	0,25	59
-0,05	-0,1	0,04	-0,08	-0,15	-0,13	-0,09	-0,13	-0,05	-0,12	0,24	60
-0,03	-0,1	-0,12	-0,09	0,01	-0,1	-0,14	-0,14	-0,08	-0,12	0,25	61
-0,06	-0,1	-0,14	-0,14	0,04	-0,07	-0,14	-0,13	-0,08	-0,11	0,25	62
-0,03	-0,10	-0,09	-0,13	-0,04	-0,03	-0,07	-0,12	-0,1	-0,12	0,25	63
0,06	-0,03	-0,13	-0,14	-0,07	-0,12	-0,15	-0,14	-0,02	-0,11	0,21	64
-0,1	-0,1	0,05	-0,08	-0,16	-0,13	-0,07	-0,13	-0,11	-0,12	0,27	65
-0,03	-0,1	-0,16	-0,14	0,03	-0,07	-0,13	-0,14	-0,09	-0,11	0,24	66
-0,1	-0,1	-0,08	-0,13	-0,05	-0,04	-0,05	-0,12	-0,15	-0,12	0,26	67
-0,15	-0,14	-0,01	-0,1	-0,35	-0,18	0,57	0,05	-0,33	-0,18	0,25	68
-0,26	-0,12	0,36	0,04	-0,14	-0,17	-0,15	-0,11	-0,11	-0,14	0,25	69
-0,02	-0,07	-0,31	-0,2	0,52	0,05	-0,35	-0,19	0,04	-0,08	0,25	70
-0,05	-0,07	-0,23	-0,19	0,42	0,06	-0,38	-0,17	0,05	-0,09	0,24	71
0,03	-0,06	-0,35	-0,2	0,56	0,06	-0,39	-0,19	0,08	-0,08	0,25	72
0,06	-0,07	-0,21	-0,17	-0,02	-0,09	-0,29	-0,21	0,44	0,09	0,24	73
-0,12	-0,14	-0,08	-0,1	-0,25	-0,20	0,37	0,07	-0,29	-0,15	0,27	74
-0,01	-0,07	-0,25	-0,21	0,37	0,08	-0,28	-0,16	-0,01	-0,09	0,25	75
0,05	-0,08	-0,19	-0,17	-0,05	-0,09	-0,25	-0,21	0,37	0,09	0,24	76
-0,07	-0,14	-0,12	-0,1	-0,2	-0,2	0,33	0,07	-0,25	-0,15	0,25	77
-0,12	-0,14	-0,08	-0,1	-0,25	-0,2	0,37	0,07	-0,29	-0,15	0,27	78
-0,01	-0,07	-0,26	-0,21	0,37	0,08	-0,3	-0,16	0	-0,09	0,25	79
-0,07	-0,12	-0,21	-0,13	0,14	-0,01	0,3	0,07	-0,26	-0,14	0,25	80
-0,05	-0,09	-0,19	-0,19	0,14	0,04	0,18	0,04	-0,14	-0,11	0,24	81
0,05	-0,07	-0,18	-0,17	-0,05	-0,09	-0,27	-0,22	0,38	0,09	0,24	82
-0,12	-0,14	-0,06	-0,1	-0,28	-0,21	0,37	0,07	-0,27	-0,15	0,25	83
0,02	-0,06	-0,29	-0,22	0,36	0,08	-0,27	-0,16	-0,01	-0,08	0,25	84
0,04	-0,07	-0,2	-0,17	-0,02	-0,09	-0,3	-0,22	0,41	0,09	0,25	85
-0,12	-0,14	-0,05	-0,1	-0,29	-0,21	0,39	0,07	-0,28	-0,15	0,25	86
0,01	-0,06	-0,28	-0,22	0,36	0,08	-0,27	-0,16	0,03	-0,08	0,25	87
-0,1	-0,1	-0,23	-0,18	0,2	0	0,2	0,06	-0,16	-0,12	0,24	88
-0,07	-0,1	-0,23	-0,18	0,18	0	0,24	0,06	-0,21	-0,12	0,25	89
-0,14	-0,1	-0,21	-0,18	0,18	0	0,21	0,06	-0,17	-0,12	0,25	90

Continua.

Continuação da tabela 5.21:

C6	C6M	C7	C7M	C8	C8M	C9	C9M	R1	R1M	No.
-0,07	-0,01	0,15	-0,01	-0,25	-0,49	0,60	-0,14	0,16	0,18	48
-0,01	-0,01	0,07	-0,02	-0,2	-0,49	0,59	-0,14	0,14	0,18	49
0,12	0,0	0,04	-0,02	-0,22	-0,48	0,64	-0,14	0,14	0,18	50
-0,09	0,01	0,16	-0,02	-0,27	-0,49	0,65	-0,13	0,16	0,18	51
0	0	0,13	-0,02	-0,25	-0,49	0,62	-0,14	0,15	0,18	52
-0,11	-0,02	0,17	-0,02	-0,26	-0,49	0,61	-0,13	0,16	0,18	53
0,07	0	0,08	-0,02	-0,22	-0,48	0,59	-0,13	0,14	0,18	54
-0,09	0	0,1	-0,02	-0,24	-0,49	0,57	-0,13	0,15	0,18	55
0,02	0	0,1	-0,02	-0,22	-0,49	0,58	-0,14	0,15	0,18	56
-0,08	-0,02	0,16	-0,02	-0,28	-0,49	0,6	-0,13	0,16	0,18	57
-0,14	-0,01	0,19	-0,01	-0,26	-0,49	0,58	-0,14	0,15	0,18	58
-0,3	0	0,31	-0,02	-0,28	-0,49	0,57	-0,14	0,14	0,18	59
-0,12	0	0,16	-0,02	-0,22	-0,49	0,51	-0,15	0,14	0,18	60
-0,07	-0,1	0,18	-0,01	-0,28	-0,49	0,61	-0,14	0,16	0,18	61
-0,05	-0,01	0,14	-0,01	-0,26	-0,49	0,63	-0,14	0,16	0,18	62
-0,07	0	0,18	-0,02	-0,28	-0,49	0,63	-0,14	0,17	0,18	63
-0,12	-0,01	0,12	-0,03	-0,2	-0,48	0,5	-0,15	0,14	0,18	64
-0,04	0	0,18	-0,02	-0,27	-0,49	0,61	-0,13	0,16	0,18	65
-0,04	-0,01	0,12	-0,01	-0,23	-0,49	0,59	-0,14	0,15	0,18	66
0	0	0,12	-0,02	-0,25	-0,49	0,59	-0,13	0,15	0,18	67
-0,01	0,03	0,21	-0,02	-0,28	-0,49	0,6	-0,14	0,16	0,18	68
0	0,01	0,15	-0,02	-0,25	-0,49	0,6	-0,14	0,15	0,18	69
-0,18	-0,05	0,22	-0,02	-0,28	-0,48	0,65	-0,15	0,16	0,18	70
-0,15	-0,04	0,22	-0,01	-0,27	-0,49	0,65	-0,15	0,15	0,18	71
-0,23	-0,05	0,19	-0,01	-0,25	-0,49	0,57	-0,14	0,15	0,18	72
-0,33	-0,03	0,27	-0,02	-0,26	-0,46	0,55	-0,16	0,16	0,19	73
0,06	0,03	0,09	-0,02	-0,22	-0,48	0,59	-0,13	0,15	0,18	74
-0,13	-0,04	0,19	-0,01	-0,28	-0,49	0,62	-0,14	0,16	0,18	75
-0,25	-0,03	0,19	-0,02	-0,24	-0,46	0,56	-0,15	0,17	0,19	76
0,01	0,02	0,11	-0,02	-0,23	-0,48	0,59	-0,14	0,15	0,18	77
0,06	0,03	0,09	-0,02	-0,23	-0,48	0,6	-0,13	0,15	0,18	78
-0,13	-0,4	0,15	-0,1	-0,27	-0,49	0,59	-0,14	0,16	0,18	79
-0,03	0,01	0,21	-0,02	-0,29	-0,49	0,62	-0,14	0,17	0,18	80
-0,1	-0,02	0,18	-0,01	-0,26	-0,49	0,56	-0,14	0,16	0,18	81
-0,25	-0,03	0,22	-0,02	-0,25	-0,46	0,6	-0,14	0,17	0,19	82
0,03	0,03	0,14	-0,02	-0,26	-0,48	0,64	-0,14	0,16	0,18	83
-0,15	-0,04	0,18	-0,01	-0,28	-0,49	0,6	-0,14	0,16	0,18	84
-0,26	-0,03	0,23	-0,02	-0,26	-0,46	0,59	-0,14	0,17	0,19	85
0,03	0,03	0,13	-0,02	-0,24	-0,48	0,58	-0,14	0,16	0,18	86
-0,18	-0,04	0,2	-0,01	-0,27	-0,49	0,56	-0,14	0,15	0,18	87
-0,09	-0,01	0,24	-0,02	-0,28	-0,48	0,64	-0,15	0,16	0,18	88
-0,05	0	0,16	-0,02	-0,26	-0,49	0,58	-0,14	0,16	0,18	89
-0,01	0	0,15	-0,02	-0,14	-0,4	0,62	-0,14	-0,14	-0,15	90

diretamente relacionadas a atividade biológica no sentido que maiores volumes levam à menor atividade dos compostos. O volume dos substituintes não tem relação com as atividades *in vitro*.

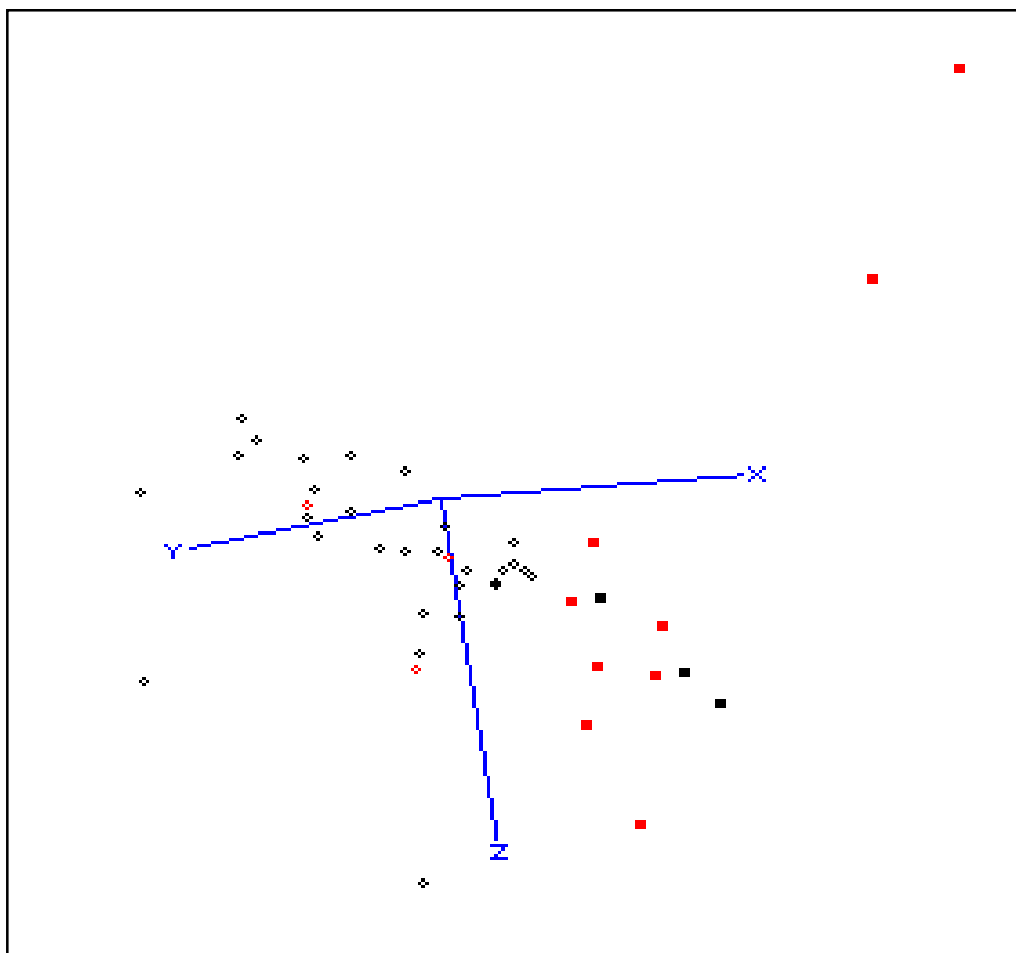


Figura 5.22: Gráfico das três primeiras componentes principais do modelo obtido com os volumes dos compostos. Os compostos em preto e cinza são os mais ativos.

Os compostos 54, 74 e 89 foram classificados errado no grupo dos pouco ativos no modelo com volumes. Os compostos 89 e 54 têm o substituinte benzila na posição R1.

Tabela 5.22: Volumes ($\text{\AA}^3$) e superfícies ($\text{\AA}^2$) para os compostos e para os substituintes, calculados com o programa SURF.

vdWV	vdWSA	SAV	SAS	R1vdWV	R1vdWS	R1SAV	R5vdWS	No.
					A		A	
183,90	235,00	682,20	440,60	7,23	18,32	81,32	18,32	48
273,40	337,50	967,10	596,80	99,26	131,88	405,33	18,32	49
197,70	246,20	721,70	457,60	7,23	18,32	81,32	18,32	50
199,00	252,30	729,20	466,50	7,23	18,32	81,32	18,32	51
199,00	252,50	729,30	466,10	7,23	18,32	81,32	43,05	52
190,80	223,00	693,70	447,50	7,23	18,32	81,32	22,75	53
290,80	342,90	1006,0	615,50	99,26	131,88	405,33	18,32	54
		0						
295,10	324,70	1004,0	602,20	99,26	131,88	405,33	18,32	55
		0						
293,60	373,00	1085,0	705,00	99,26	131,88	405,33	43,05	56
		0						
279,00	330,80	978,10	604,90	99,26	131,88	405,33	22,75	57
202,20	250,30	727,30	459,90	7,23	18,32	81,32	18,32	58
244,60	314,20	862,00	523,50	7,23	18,32	81,32	18,32	59
199,90	257,30	739,80	473,80	7,23	18,32	81,32	18,32	60
244,70	317,30	886,20	551,60	7,23	18,32	81,32	18,32	61
204,00	241,60	739,50	473,70	7,23	18,32	81,32	43,11	62
255,50	314,80	900,30	554,70	7,23	18,32	81,32	102,12	63
304,30	337,70	983,00	570,40	99,26	131,88	405,33	18,32	64
288,70	359,40	1011,0	618,20	99,26	131,88	405,33	18,32	65
		0						
310,20	336,70	1004,0	606,80	99,26	131,88	405,33	43,11	66
		0						
353,50	430,40	1291,0	835,50	99,26	131,88	405,33	102,12	67
		0						
195,00	250,10	722,30	463,90	7,23	18,32	81,32	18,32	68
86,50	114,60	357,00	230,60	7,23	18,32	81,32	18,32	69
220,80	233,30	747,00	473,70	7,23	18,32	81,32	48,06	70
252,90	270,70	848,00	528,60	7,23	18,32	81,32	88,22	71
301,80	339,30	1021,0	625,10	99,26	131,88	405,33	48,06	72
		0						
241,40	238,00	775,00	484,00	7,23	18,32	81,32	18,32	73
306,10	356,90	1025,0	620,20	7,23	18,32	81,32	18,32	74
		0						
231,90	254,20	780,00	492,50	7,23	18,32	81,32	87,29	75
323,60	347,20	1045,0	634,60	99,26	131,88	405,33	18,32	76
		0						
316,30	350,70	1054,0	642,60	99,26	131,88	405,33	18,32	77
		0						
306,10	356,90	1025,0	620,20	93,30	141,62	412,25	18,32	78
		0						
311,90	356,60	1055,0	641,70	99,26	131,88	405,33	87,29	79
		0						
257,50	274,30	853,00	530,70	7,23	18,32	81,32	87,29	80
335,20	388,80	1128,0	679,70	99,26	131,88	405,33	87,29	81

		0							
206,20	231,30	716,00	456,50	7,23	18,32	81,32	18,32	82	
207,20	223,90	720,00	458,30	7,23	18,32	81,32	18,32	83	
205,20	229,00	720,00	460,90	7,23	18,32	81,32	34,30	84	
295,20	324,10	991,00	606,70	99,26	131,88	405,33	18,32	85	
298,10	327,90	994,00	608,40	99,26	131,88	405,33	18,32	86	
300,90	324,80	995,00	606,90	99,26	131,88	405,33	34,30	87	
210,60	240,30	742,00	468,40	7,23	18,32	81,32	34,30	88	
298,10	338,30	1016,0	617,40	99,26	131,88	405,33	34,30	89	
		0							
232,20	263,60	798,00	501,10	7,23	18,32	81,32	34,30	90	

Continua.

Continuação da tabela 5.22:

R1SAS	R2vdWV	R2vdWS	R2SAV	R2SAS	R3vdWV	R3vdWS	R5SAV	No.
		A				A		
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	48
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	49
85,36	7,23	18,32	81,32	85,36	26,50	43,05	81,32	50
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	51
85,36	7,23	18,32	81,32	85,36	7,23	18,32	155,84	52
85,36	7,23	18,32	81,32	85,36	7,23	18,32	96,89	53
285,44	7,23	18,32	81,32	85,36	26,50	43,05	81,32	54
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	55
285,44	7,23	18,32	81,32	85,36	7,23	18,32	155,84	56
285,44	7,23	18,32	81,32	85,36	7,23	18,32	96,89	57
85,36	7,23	18,32	81,32	85,36	27,88	43,11	81,32	58
85,36	7,23	18,32	81,32	85,36	112,82	160,25	81,32	59
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	60
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	61
85,36	7,23	18,32	81,32	85,36	7,23	18,32	170,26	62
85,36	7,23	18,32	81,32	85,36	7,23	18,32	300,00	63
285,44	7,23	18,32	81,32	85,36	27,88	43,11	81,32	64
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	65
285,44	7,23	18,32	81,32	85,36	7,23	18,32	170,26	66
285,44	7,23	18,32	81,32	85,36	7,23	18,32	300,00	67
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	68
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	69
85,36	7,23	18,32	81,32	85,36	7,23	18,32	172,20	70
85,36	7,23	18,32	81,32	85,36	7,23	18,32	281,83	71
285,44	7,23	18,32	81,32	85,36	7,23	18,32	172,20	72
85,36	7,23	18,32	81,32	85,36	59,67	87,29	81,32	73
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	74
85,36	7,23	18,32	81,32	85,36	7,23	18,32	279,14	75
285,44	7,23	18,32	81,32	85,36	59,67	87,29	81,32	76
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	77
292,99	7,23	18,32	81,32	85,36	7,23	18,32	81,32	78
285,44	7,23	18,32	81,32	85,36	7,23	18,32	279,14	79
85,36	7,23	18,32	81,32	85,36	7,23	18,32	279,14	80
285,44	7,23	18,32	81,32	85,36	7,23	18,32	279,14	81
85,36	7,23	18,32	81,32	85,36	17,35	34,30	81,32	82
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	83
85,36	7,23	18,32	81,32	85,36	7,23	18,32	130,35	84
285,44	7,23	18,32	81,32	85,36	17,35	34,30	81,32	85
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	86
285,44	7,23	18,32	81,32	85,36	7,23	18,32	130,35	87
85,36	7,23	18,32	81,32	85,36	7,23	18,32	130,35	88
285,44	7,23	18,32	81,32	85,36	7,23	18,32	130,35	89
85,36	27,88	43,11	170,26	152,27	7,23	18,32	130,35	90

Continua.

Continuação da tabela 5.22:

R3SAV	R3SAS	R4vdWV	R4vdWS	R4SAV	R4SAS	R5vdWV	R5SAS	No.
A								
81,32	85,36	7,23	18,32	81,32	85,36	7,23	85,36	48
81,32	85,36	7,23	18,32	81,32	85,36	7,23	85,36	49
155,84	133,29	7,23	18,32	81,32	85,36	7,23	85,36	50
81,32	85,36	26,50	43,05	155,84	133,29	7,23	85,36	51
81,32	85,36	7,23	18,32	81,32	85,36	26,50	133,29	52
81,32	85,36	7,23	18,32	81,32	85,36	10,30	101,39	53
155,84	133,29	7,23	18,32	81,32	85,36	7,23	85,36	54
81,32	85,36	26,50	43,05	155,84	133,29	7,23	85,36	55
81,32	85,36	7,23	18,32	81,32	85,36	26,50	133,29	56
81,32	85,36	7,23	18,32	81,32	85,36	10,30	101,39	57
170,26	152,27	7,23	18,32	81,32	85,36	7,23	85,36	58
510,72	364,43	7,23	18,32	81,32	85,36	7,23	85,36	59
81,32	85,36	27,88	43,11	170,26	152,27	7,23	85,36	60
81,32	85,36	112,82	160,25	510,72	364,43	7,23	85,36	61
81,32	85,36	7,23	18,32	81,32	85,36	27,88	152,27	62
81,32	85,36	7,23	18,32	81,32	85,36	86,03	175,05	63
170,26	152,27	7,23	18,32	81,32	85,36	7,23	85,36	64
81,32	85,36	27,88	43,11	170,26	152,27	7,23	85,36	65
81,32	85,36	7,23	18,32	81,32	85,36	27,88	152,27	66
81,32	85,36	7,23	18,32	81,32	85,36	86,03	175,05	67
81,32	85,36	30,56	48,06	172,20	150,38	7,23	85,36	68
81,32	85,36	60,32	88,22	281,83	217,14	7,23	85,36	69
81,32	85,36	7,23	18,32	81,32	85,36	30,56	150,38	70
81,32	85,36	7,23	18,32	81,32	85,36	60,32	217,14	71
81,32	85,36	7,23	18,32	81,32	85,36	30,56	150,38	72
279,14	215,34	7,23	18,32	81,32	85,36	7,23	85,36	73
81,32	85,36	59,67	87,29	279,14	215,34	7,23	85,36	74
81,32	85,36	7,23	18,32	81,32	85,36	59,67	215,34	75
279,14	215,34	7,23	18,32	81,32	85,36	7,23	85,36	76
81,32	85,36	59,67	87,29	279,14	215,34	7,23	85,36	77
81,32	85,36	59,67	87,29	279,14	215,34	7,23	85,36	78
81,32	85,36	7,23	18,32	81,32	85,36	59,67	215,34	79
81,32	85,36	59,67	87,29	279,14	215,34	59,67	215,34	80
81,32	85,36	59,67	87,29	279,14	215,34	59,67	215,34	81
130,35	125,91	7,23	18,32	81,32	85,36	7,23	85,36	82
81,32	85,36	17,35	34,30	130,35	125,91	7,23	85,36	83
81,32	85,36	7,23	18,32	81,32	85,36	17,35	125,91	84
130,35	125,91	7,23	18,32	81,32	85,36	7,23	85,36	85
81,32	85,36	17,35	34,30	130,35	125,91	7,23	85,36	86
81,32	85,36	7,23	18,32	81,32	85,36	17,35	125,91	87
81,32	85,36	17,35	34,30	130,35	125,91	17,35	125,91	88
81,32	85,36	17,35	34,30	130,35	125,91	17,35	125,91	89
81,32	85,36	17,35	34,30	130,35	125,91	17,35	125,91	90

5.4.3. Análise das densidades de elétron nos orbitais de fronteira

Foi criado um modelo com todos os índices de densidade eletrônica nos orbitais de fronteira para auxiliar na seleção de variáveis. Inicialmente foram calculadas 64 variáveis, que passaram por esta triagem inicial, que revelou quais tinham o maior poder de

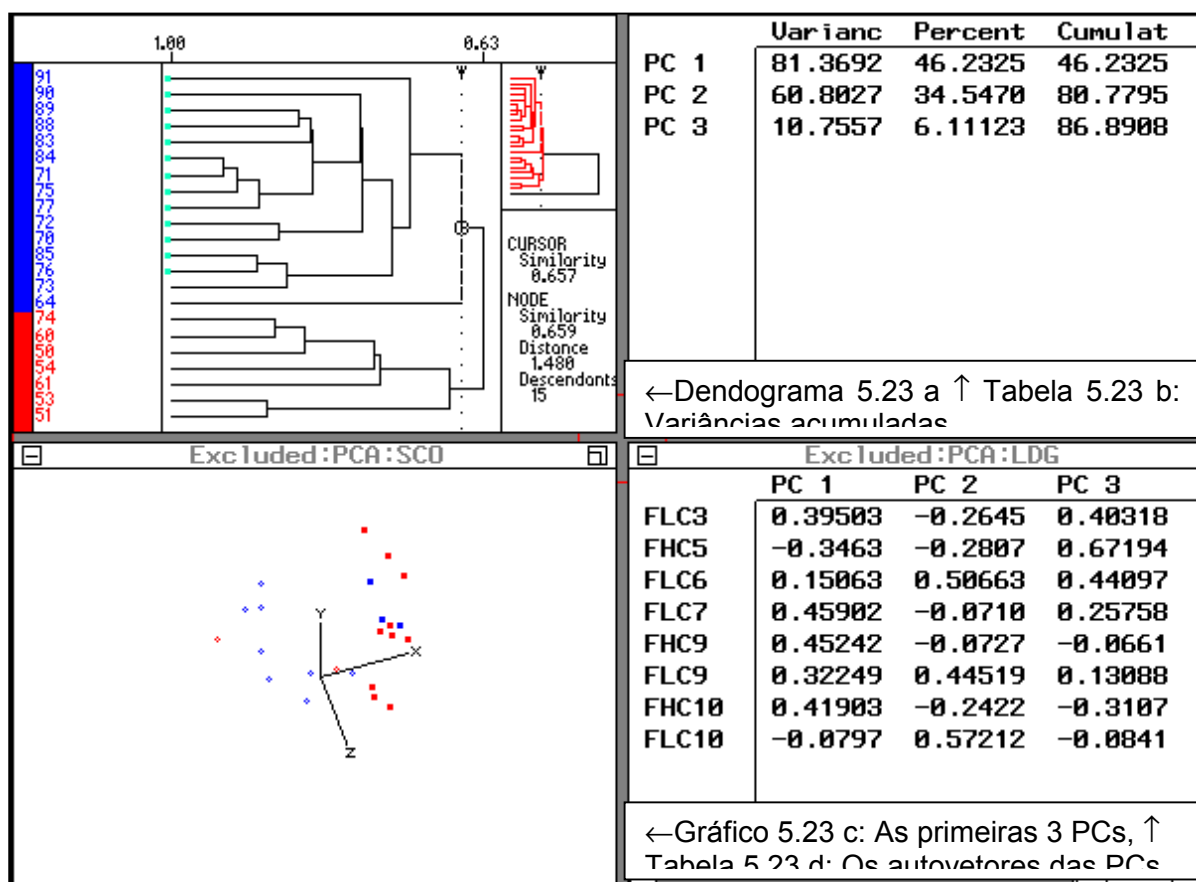


Figura 5.23: Resultados da análise exploratória dos dados das densidades eletrônicas nos orbitais de fronteira para os 4-fenil 2-guanidinothiazóis. Esta análise foi realizada somente para auxiliar a seleção de variáveis previsão para as atividades medidas *in vivo*.

Na figura 5.23 a são mostrados os gráficos correspondentes HCA com dados autoescalados para os 23 compostos. O bloco marcado em azul tem 15 elementos, 11 são compostos mais ativos, com similaridade superior à 0,648UV distante 1,534UV do outro bloco. Os quatro menos ativos no bloco azul são os compostos 70, 73, 75, e 83. O bloco vermelho tem sete compostos, seis menos ativos, com similaridade maior que

0,674. O composto 54 foi erradamente classificado neste bloco. O composto 58 não aparece no dendograma principal (aparece em preto, no detalhe vermelho do canto superior esquerdo da figura 5.23 a), está situado à distância de 2,534UV dos demais blocos, e é pouco ativo. Os compostos que aparecem com um ponto verde na figura 5.23 a são os que estão em destacados no gráfico da figura 5.23 c

O gráfico das três componentes utilizadas em PCA é mostrado na figura 5.23 c. Elas acumulam 87% da variância total dos dados, e classificam errado cinco compostos. No grupo mais ativo, mostrado em vermelho, com ponto maiores estão classificados errado os compostos 70, 75, e 83, que aparecem em azul. No grupo menos ativo, mostrado em azul, com pontos menores, estão erradamente classificados os compostos 54, e 64. Todas as variáveis selecionadas têm contribuição expressiva na primeira componente principal, exceto a densidade em lumo da camada de valência sobre R5, denominada FLC10. A tabela 5.23 mostra todas as densidades eletrônicas nos orbitais de fronteira calculadas para os guanidinothiazóis.

Para as atividades *in vitro* o modelo realizado para seleção de variáveis não resultou nenhuma variável com boa capacidade de previsão.

Tabela 5.23: Densidades nos orbitais de fronteira para os átomos, seguindo a numeração para os átomos apresentada na figura 5.20.

HOMO	L-H/2	FHC2	FLC3	FHC5	FLN5	No.
-8,25245	3,97291	- 0,000359301	-0,079246	-0,00229525	-0,103249	50
-8,23943	3,88334	- 0,000332243	-0,124786	-0,00370626	-0,0475783	51
-8,20147	3,85682	- 0,000799971	-0,149524	-0,00433153	-0,0419736	52
-8,22589	3,86031	-0,00106959	-0,141631	-0,00428557	-0,0424858	53
-8,15928	3,93954	- 0,000473245	-0,113449	-0,00248707	-0,0569864	54
-8,20259	3,9026	- 0,000104903	-	-	-0,100392	55
-8,21754	3,89291	- 0,000844918	0,00316125	0,000627178	-0,0687726	56
-8,25341	3,86875	-0,00104039	-0,0960254	-0,00389388	-0,0662593	57
-8,24392	3,99361	-6,19027e- 05	-0,0145536	-	0	58
-8,28062	3,98493	- 0,000231148	-0,0199773	0,000393583	-0,0733141	59
-8,14074	3,91727	- 0,000647691	-0,0747453	-0,00355452	-0,0630898	60
-8,22775	3,91811	- 0,000756878	-0,0551191	-0,0035766	-0,0811808	61
-8,10104	3,89855	- 0,000923636	-0,0876768	-0,00372797	-0,103592	62
-8,12078	3,86176	-0,00156019	-0,16329	-0,00405856	-0,0314514	63
-8,38106	4,02375	-0,00195132	-0,0293352	-0,00563509	-	64
-8,01407	3,90092	- 0,000521382	-0,132346	-0,00314345	0,00950046	65
-8,11745	3,88594	-0,00106823	-0,0782108	-0,00384967	-0,135831	66
-8,05438	3,84342	-0,00139822	-0,184497	-0,00393596	-0,0486055	67
-8,10947	3,91204	-8,35326e- 05	-0,0736434	-	-0,029914	68
-8,12818	3,92518	-0,0020253	-0,070906	0,000687531	-0,110269	69
-8,00353	3,77185	-0,00845113	-0,070906	-0,00696879	-0,116587	70
-8,28582	3,85446	-0,00237612	-0,0460418	-0,00491673	-0,0328982	71
-7,98894	3,76648	-0,00825389	-0,0416145	-0,0061609	-	72
-8,17152	3,89851	-9,56891e- 06	-0,0327339	-0,00496198	0,00137184	73
-8,16352	3,92341	- 0,000938322	-0,0557498	-0,00307131	-0,027876	74
-8,17308	3,80998	-0,00496742	-0,0902292	-0,00499187	-0,0417172	75
-8,27457	3,92813	- 0,000356301	-0,0315756	-0,00673826	-0,0567092	76
-8,35760	3,88277	-0,00189711	-0,0323853	-0,00332126	-0,0501627	77
-8,38917	3,90379	-0,00322282	-0,0280483	-0,00705756	-0,045478	78

					0,00138266	
-8,27175	3,89798	-0,00445855	-0,0287984	-0,00607709	-0,0638038	79
-8,19440	3,822	-0,00843867	-0,105896	-0,0115803	-0,0097706	80
-8,15780	3,82502	-0,00732104	-0,0359196	-0,0103964	-0,0308466	81
-8,22002	3,91008	-2,41705e- 05	-0,0497929	-0,00292704	-0,0480181	82
-8,35779	3,89495	-0,00268819	-0,0397475	-0,00943751	-0,0263603	83
-8,19792	3,81717	-0,00445141	-0,0381655	-0,00629194	-0,0351716	84
-8,19258	3,89831	-2,93541e- 05	-0,0466685	-0,00276155	-0,0453888	85
-8,37862	3,88275	-0,00213989	-0,032767	-0,00828844	-0,0292333	86
-8,18703	3,80587	-0,00439014	-0,031833	-0,00626462	-0,0498451	87
-8,32062	3,91671	-0,00637545	-0,0520048	-0,0109412	-0,0558743	88
-8,20692	3,80873	-0,00594718	-0,0407307	-0,0104763	-0,0304377	89
-8,27604	3,86746	-0,00591914	-0,0281775	-0,00938687	-0,0521721	90
-8,01003	3,80898	-0,00179671	-0,0374137	-0,00446576	-0,0353327	91

Continua.

Continuação da tabela 5.23:

LUMO	FHC1	FLC2	FHC4	FLC5	FHN5	No.
-,30663	-0,00184468	-0,0196647	-	-0,0549353	-	50
			0,00018920		0,00629811	
-,47276	-0,00485667	-0,0191873	-	-0,0617224	-	51
			0,00086689		0,00473391	
-,48784	-0,00450929	-0,0276308	-	-0,0715253	-	52
			0,00104163		0,00441058	
-,50526	-0,00441439	-0,0219358	-	-0,0665453	-	53
			0,00130934		0,00429152	
-,28020	-0,00177664	-0,0312261	-	-0,0832273	-	54
			0,00011774		0,00453485	
-,39739	-	-0,00106709	-6,6459e-	-0,00451426	-	55
	0,000600982		05		0,00799046	
-,43171	-0,00337098	-0,0199815	-	-0,0524693	-0,0063597	56
			0,00062648			
-,51592	-0,00400959	-0,0141938	-	-0,0453692	-	57
			0,00128335		0,00686643	
-,25669	-	-0,0513218	-	-0,0592731	-	58
	0,000470543		0,00010544		0,00453613	
-,31076	-0,00173458	-0,00255306	-	-0,0136401	-0,00393	59
			0,00026775			
-,30620	-0,00387702	-0,00909719	-	-0,0458566	-0,0031258	60
			0,00063982			
-,39153	-0,00388095	-0,00594128	-	-0,0334692	-	61
			0,00073088		0,00496955	
-,30394	-0,00405786	-0,0113507	-	-0,0475123	-0,0049531	62
			0,00121698			
-,39726	-0,00383233	-0,0433029	-	-0,0654488	-	63
			0,00125339		0,00266924	
-,33355	-0,00298585	-0,00569699	-0,000251	-0,0249829	-	64
					0,00067482	
-,21222	-0,00343935	-0,0161865	-	-0,073069	-	65
			0,00047659		0,00663549	
-,34557	-0,00362084	-0,010484	-0,0008276	-0,0438238	-	66
					0,00336571	
-,36753	-0,00367891	-0,0495704	-	-0,0726251	-	67
			0,00112035		0,00321228	
-,28539	-0,00900102	-0,0129104	-	-0,0542163	-	68
			0,00248403		0,00438243	
-,27782	-0,00108806	-0,0118744	-8,8592e-	-0,04404	-	69
			05		0,00477718	
-,45983	-0,0049307	-0,00187451	-	-0,0286886	-	70
			0,00721991		0,00064267	
-,57690	-0,0051934	-0,00341591	-	-0,0230236	-	71
			0,00158291		0,00067955	
-,45598	-0,00480576	-0,00135845	-	-0,0217771	-0,0006523	72
			0,00746001			
-,37450	-0,00806209	-0,00941984	-	-0,0389423	-0,0019723	73
			0,00360768			

-,31670	-0,0024103	-0,0140035	-	-0,0452005	-	74
			0,00015534		0,00317976	
-,55312	-0,00450883	-0,00131596	-	-0,0216427	-	75
			0,00361337		0,00135792	
-,41831	-0,00508036	-0,0045592	-	-0,0211412	-	76
			0,00121369		0,00402948	
-,59206	-0,00348769	-0,00336405	-	-0,0204439	-	77
			0,00038904		0,00382512	
-,58159	-0,00212603	-0,00283729	-	-0,0165412	-	78
			0,00010166		0,00026386	
-,47579	-0,00397421	-0,00141218	-	-0,0205448	-	79
			0,00318638		0,00166826	
-,55040	-0,00156912	-0,00867371	-	-0,0538408	-	80
			0,00207176		0,00024736	
-,50777	-0,00124878	-0,00256031	-	-0,0218174	-	81
			0,00165868		0,00159596	
-,39986	-0,00733369	-0,00909221	-	-0,0345361	-	82
			0,00279988		0,00278608	
-,56788	-0,00196337	-0,00485051	-8,9137e-	-0,0227772	-	83
			05		0,00217067	
-,56358	-0,00475671	-0,00152591	-	-0,0233578	-0,0017305	84
			0,00322644			
-,39595	-0,00680024	-0,00862738	-	-0,0325296	-	85
			0,00253106		0,00339319	
-,61311	-0,00221363	-0,00371651	-9,852e-05	-0,0183404	-	86
					0,00374229	
-,57529	-0,00475111	-0,00125092	-	-0,0197262	-	87
			0,00320634		0,00156686	
-,48719	-0,00130121	-0,00366754	-	-0,0267924	-	88
			0,00122069		0,00157306	
-,58945	-0,00134293	-0,00293487	-0,0011571	-0,0211526	-	89
					0,00234217	
-,54113	-	-0,00339403	-	-0,0174797	-	90
	0,000851379		0,00083371		0,00230926	
-,39207	-0,00125431	-0,00550152	-	-0,0203657	-	91
			0,00046091		0,00708484	

Continua.

Continuação da tabela 5.23:

L-H	FLC1	FHC3	FLC4	FHC6	No.
7,94582	-0,033862	-0,00239003	-	-0,0021334	50
			0,00667423		
7,76667	-0,0662154	-0,00623451	-0,01796	-0,00440648	51
7,71363	-0,0656543	-0,00697148	-0,0227074	-0,00545211	52
7,72063	-0,0615564	-0,00696011	-0,017972	-0,0059939	53
7,87908	-0,0428244	-0,00246187	-	-0,00205242	54
			0,00784868		
7,8052	-0,00321455	-0,00017502	-	-0,00025819	55
			0,00087556		
7,78583	-0,0521072	-0,00532227	-0,0203712	-0,0041794	56
7,73749	-0,043506	-0,00642488	-0,0124964	-0,00567356	57
7,98723	-0,00295576	-0,00054018	-0,03909	-0,00046603	58
7,96986	-0,0155114	-0,00185441	-	-0,0017893	59
			0,00349746		
7,83454	-0,0453736	-0,00566816	-0,0101239	-0,00449513	60
7,83622	-0,0355579	-0,00576561	-	-0,00476311	61
			0,00638906		
7,7971	-0,0505192	-0,00660977	-0,0118515	-0,00570258	62
7,72352	-0,0653914	-0,00727409	-0,0451362	-0,00622774	63
8,04751	-0,0169434	-0,00648524	-	-0,00704122	64
			0,00190291		
7,80185	-0,078649	-0,00477334	-0,0203821	-0,00358549	65
7,77188	-0,0424166	-0,006149	-0,0104542	-0,00515938	66
7,68685	-0,0722983	-0,0068572	-0,0521508	-0,0057361	67
7,82408	-0,0418921	-0,00965284	-	-0,00477104	68
			0,00972921		
7,85036	-0,0447502	-0,00769218	-0,0093544	-0,00436537	69
7,5437	-0,0308033	-0,0132869	-	-0,0194454	70
			0,00283185		
7,70892	-0,0273392	-0,0113416	-	-0,0102835	71
			0,00560123		
7,53296	-0,0216764	-0,0132593	-	-0,0193528	72
			0,00177537		
7,79702	-0,0282441	-0,00808906	-	-0,0109031	73
			0,00450299		
7,84682	-0,0625535	-0,00605884	-0,0172926	-0,00375184	74
7,61996	-0,0213408	-0,0117861	-	-0,014716	75
			0,00220115		
7,85626	-0,0198281	-0,00593595	-	-0,00623834	76
			0,00350013		
7,76554	-0,0217042	-0,00930935	-	-0,00739424	77
			0,00453645		
7,80758	-0,0230898	-0,0122005	-	-0,00825784	78
			0,00478769		
7,79596	-0,0191807	-0,0104752	-	-0,0129368	79
			0,00196636		
7,644	-0,0648343	-0,0164692	-0,0139964	-0,0150515	80
7,65003	-0,0247104	-0,0143975	-	-0,0126707	81

			0,00373086		
7,82016	-0,0243192	-0,00711396	-	-0,00921882	82
			0,00365178		
7,78991	-0,029319	-0,0107448	-0,0073511	-0,00715217	83
7,63434	-0,0266484	-0,0112655	-	-0,0137073	84
			0,00304805		
7,79663	-0,0229917	-0,00654648	-0,003289	-0,00842893	85
7,76551	-0,0249001	-0,00966002	-	-0,00638572	86
			0,00626287		
7,61174	-0,022401	-0,0111944	-	-0,0135759	87
			0,00249615		
7,83343	-0,0369205	-0,0147058	-	-0,0118154	88
			0,00753265		
7,61747	-0,0288338	-0,0140017	-0,0057813	-0,0112005	89
7,73491	-0,0190435	-0,0121297	-	-0,00931748	90
			0,00303805		
7,61796	-0,022444	-0,00537834	-	-0,00381421	91
			0,00460987		

Continua.

Continuação da tabela 5.23:

FLC6	FLC7	FLN2	FLS1	FLC9	No.
-0,0765026	-0,170136	-0,183574	-0,636533	-0,450171	50
-0,102652	-0,15095	-0,0568018	-0,324621	-0,284662	51
-0,121262	-0,148052	-0,043049	-0,288914	-0,26811	52
-0,106215	-0,139846	-0,0462619	-0,289077	-0,260096	53
-0,107196	-0,208461	-0,152689	-0,629475	-0,459929	54
-	-0,107587	-0,186683	-0,545716	-0,338753	55
0,00303397					
-0,0935105	-0,143751	-0,0811818	-0,382009	-0,311783	56
-0,0702968	-0,118825	-0,0682697	-0,319053	-0,268482	57
-0,584826	-0,10818	-0,407963	-0,669352	-0,726024	58
-0,0167599	-0,155927	-0,2315	-0,711269	-0,445154	59
-0,0560565	-0,187956	-0,174936	-0,640771	-0,456189	60
-0,0399778	-0,144547	-0,140222	-0,505509	-0,357791	61
-0,0643821	-0,192551	-0,169214	-0,632852	-0,457699	62
-0,146248	-0,150732	-0,0609535	-0,338944	-0,295313	63
-0,0212143	-0,148333	-0,204106	-0,5676	-0,424359	64
-0,103424	-0,280661	-0,210687	-0,864179	-0,616615	65
-0,0581115	-0,156119	-0,136781	-0,518237	-0,37551	66
-0,16622	-0,157718	-0,0587732	-0,343871	-0,304362	67
-0,0627478	-0,201818	-0,185905	-0,69137	-0,487726	68
-0,0562648	-0,198262	-0,203374	-0,719074	-0,4996	69
-0,021922	-0,128837	-0,151734	-0,417704	-0,268786	70
-0,0256429	-0,0961933	-0,111341	-0,310882	-0,219222	71
-0,0148175	-0,105923	-0,163428	-0,40918	-0,271725	72
-0,0336267	-0,127259	-0,200363	-0,488166	-0,327567	73
-0,0801244	-0,193153	-0,148972	-0,590559	-0,435033	74
-0,0159721	-0,0904169	-0,132225	-0,333396	-0,226432	75
-0,0218434	-0,0902178	-0,1938	-0,423253	-0,281222	76
-0,021812	-0,0836537	-0,112722	-0,294364	-0,211754	77
-0,0203083	-0,0858804	-0,122258	-0,313076	-0,230375	78
-0,0146596	-0,097894	-0,157405	-0,386105	-0,251189	79
-0,0748003	-0,174017	-0,0536219	-0,305798	-0,245016	80
-0,0214859	-0,101776	-0,135531	-0,360346	-0,242988	81
-0,0297212	-0,108225	-0,193754	-0,448504	-0,302961	82
-0,0309254	-0,103088	-0,109503	-0,320453	-0,226173	83
-0,0198431	-0,0975977	-0,120399	-0,325556	-0,214967	84
-0,0275813	-0,102927	-0,192905	-0,438578	-0,29262	85
-0,0250518	-0,085981	-0,10299	-0,286663	-0,201776	86
-0,0163006	-0,0850269	-0,122862	-0,309607	-0,212347	87
-0,0341967	-0,116876	-0,126974	-0,369088	-0,262163	88
-0,0265595	-0,0910742	-0,10543	-0,296339	-0,206435	89
-0,0178269	-0,104879	-0,129968	-0,352802	-0,227457	90
-0,0272972	-0,10187	-0,122176	-0,396682	-0,268062	91

Continua.

Continuação da tabela 5.23:

FHC7	FHN2	FHS1	FHC9	FHN3	No.
-0,0187098	-0,00298187	-0,0125372	-0,0207382	-0,0216932	50
-0,0180517	-0,00121621	-0,0108734	-0,0219989	-0,0220395	51
-0,0172457	-	-0,0102916	-0,0214451	-0,0218706	52
	0,000984976				
-0,0169701	-	-0,0102111	-0,0212869	-0,0220041	53
	0,000923212				
-0,0172906	-0,00341609	-0,0116139	-0,0212432	-0,0219158	54
-0,0184712	-0,00444374	-0,0137965	-0,0206701	-0,02214	55
-0,0172791	-0,00166995	-0,0110157	-0,0211303	-0,0216691	56
-0,0168529	-0,0010143	-0,0105758	-0,0205796	-0,0214983	57
-	-0,00912867	-0,0270888	-0,0165431	-0,0235571	58
0,00574255					
-0,0181822	-0,00319863	-0,0125162	-0,0209461	-0,022315	59
-0,0179604	-0,00151114	-0,0114354	-0,0218075	-0,0226872	60
-0,0176757	-0,00141169	-0,0112626	-0,0214543	-0,0224729	61
-0,0175429	-0,00120964	-0,0110641	-0,0214098	-0,0223895	62
-0,0163782	-0,00105166	-0,010125	-0,0202695	-0,0211421	63
-0,0194708	-	-0,0131971	-0,0207772	-0,0208704	64
	0,000399214				
-0,017332	-0,00231586	-0,0107373	-0,0220164	-0,0228336	65
-0,0170513	-0,00153416	-0,0106995	-0,0212929	-0,0224433	66
-0,0162487	-0,0013536	-0,010067	-0,0203993	-0,0215396	67
-0,0161872	-0,00132853	-0,0102972	-0,0197377	-0,0208694	68
-0,0168373	-0,00157557	-0,0108785	-0,0201407	-0,0213461	69
-	-	-0,00527142	-0,0123947	-0,0167038	70
0,00966802	0,000161194				
-0,0156851	-	-0,00980061	-0,017246	-0,0232431	71
	0,000224195				
-0,0096768	-	-0,00565872	-0,0119677	-0,016616	72
	0,000176824				
-0,0173394	-	-0,010667	-0,017135	-0,0232607	73
	0,000318116				
-0,0179145	-0,00175942	-0,0112024	-0,0215868	-0,02218	74
-0,0133811	-3,02147e-05	-0,00802629	-0,0154836	-0,0214354	75
-0,0181462	-0,00139975	-0,0126004	-0,016784	-0,0245728	76
-0,0166147	-	-0,0109504	-0,0174862	-0,0236762	77
	0,000619928				
-0,0163166	-0,00023628	-0,0106554	-0,0164607	-0,0220243	78
-0,0135015	-	-0,00848091	-0,0154583	-0,0216484	79
	0,000272073				
-0,0124184	-	-0,00670531	-0,0143964	-0,0159632	80
	0,000223344				
-0,0122173	-	-0,00762544	-0,0137942	-0,0189987	81
	0,000100402				
-0,0177247	-	-0,0114469	-0,0170953	-0,0242001	82
	0,000621389				
-0,0167749	-	-0,0109897	-0,0174556	-0,0227242	83

	0,000574772				
-0,0138584	-7,22706e-05	-0,00822088	-0,0162811	-0,022424	84
-0,0174656	-	-0,0114891	-0,0169428	-0,0246157	85
	0,000899907				
-0,0161469	-	-0,0107747	-0,0169062	-0,0234221	86
	0,000934718				
-0,0139985	-5,33622e-05	-0,00843454	-0,0162599	-0,0220247	87
-0,0127275	-7,07495e-05	-0,00766388	-0,0143319	-0,0192118	88
-0,0129795	-	-0,00790471	-0,0147209	-0,0203646	89
	0,000191989				
-0,0144481	-	-0,00975877	-0,0105997	-0,0209725	90
	0,000895298				
-0,0166014	-0,00294868	-0,010379	-0,01545	-0,022453	91

Continua.

Continuação da tabela 5.23:

FLN3	FHC10	FLC10	FHN4	FLN4	No.
-0,70498	-0,0178685	-0,0335241	-	-0,598917	50
			0,00581387		
-0,520112	-0,013669	-0,01286	-	-0,274765	51
			0,00471525		
-0,501806	-0,0128787	-0,00950719	-0,0044213	-0,238962	52
-0,477097	-0,0125793	-0,00986298	-	-0,242847	53
			0,00436623		
-0,754636	-0,0193989	-0,0302209	-	-0,578123	54
			0,00535773		
-0,458567	-0,021072	-0,0483852	-	-0,541949	55
			0,00640832		
-0,530106	-0,0147087	-0,0179985	-	-0,339123	56
			0,00505556		
-0,446976	-0,0130785	-0,0166497	-	-0,00772257	57
			0,00159534		
-0,603117	-0,00154508	-0,191896	-0,0263366	-0,210392	58
-0,637463	-0,0184598	-0,0383161	-	-0,683499	59
			0,00586199		
-0,721973	-0,014539	-0,0330265	-	-0,598847	60
			0,00504917		
-0,560981	-0,0141173	-0,0266496	-	-0,475031	61
			0,00495325		
-0,73147	-0,013458	-0,0328233	-0,0047803	-0,587787	62
-0,535355	-0,0123597	-0,0128124	-	-0,29084	63
			0,00435136		
-0,621478	-0,00808849	-0,115921	-	-0,589462	64
			0,00522167		
-1,0115	-0,0175219	-0,0329002	-	-0,788323	65
			0,00452465		
-0,598258	-0,0145846	-0,0302092	-	-0,500475	66
			0,00456376		
-0,556072	-0,0136544	-0,0138332	-	-0,300248	67
			0,00421954		
-0,773191	-0,0129207	-0,0354378	-	-0,643744	68
			0,00441964		
-0,780642	-0,0138466	-0,0387478	-	-0,680765	69
			0,00475505		
-0,482335	-0,00240621	-0,051544	-	-0,406998	70
			0,00175741		
-0,396154	-0,00641124	-0,0378629	-	-0,033298	71
			0,00241755		
-0,452974	-0,00241758	-0,0621418	-	-0,439206	72
			0,00197844		
-0,596067	-0,00676694	-0,0753087	-	-0,523223	73
			0,00416793		
-0,721039	-0,0149883	-0,0274262	-0,0050661	-0,539593	74
-0,390816	-0,00436593	-0,0477531	-	-0,374398	75
			0,00329752		

-0,490837	-0,00994971	-0,0731709	-	-0,515584	76
			0,00552592		
-0,370631	-0,00802435	-0,0454488	-	-0,333586	77
			0,00482683		
-0,394316	-0,00595472	-0,049109	-	-0,337996	78
			0,00472504		
-0,428914	-0,00590698	-0,0563053	-	-0,0407989	79
			0,00255748		
-0,501486	-0,00270022	-0,0259787	-	-0,133921	80
			0,00124877		
-0,43	-0,00460098	-0,0469086	-	-0,380097	81
			0,00291063		
-0,546874	-0,00766349	-0,0698637	-	-0,517844	82
			0,00487164		
-0,412769	-0,00782307	-0,0384817	-	-0,317389	83
			0,00448067		
-0,391208	-0,00519639	-0,0392848	-	-0,35451	84
			0,00335259		
-0,525609	-0,00871232	-0,07264	-	-0,526589	85
			0,00491597		
-0,36462	-0,00904747	-0,036016	-	-0,312304	86
			0,00470945		
-0,368407	-0,00500934	-0,0481564	-	-0,0276408	87
			0,00205448		
-0,477867	-0,00477656	-0,0435027	-	-0,0349001	88
			0,00165161		
-0,377713	-0,00562151	-0,0363452	-	-0,321249	89
			0,00325595		
-0,416296	-0,00666791	-0,0415832	-	-0,0311931	90
			0,00218334		
-0,418237	-0,0165335	-0,0319234	-	-0,453845	91
			0,00387504		

5.5. Conclusões sobre a análise dos compostos 4-fenil 2-guanidinothiazóis

5.5.1. Análise exploratória dos dados com HCA e PCA e regressão

PLS

No caso das atividades *in vitro* dos 4-fenil-2-guanidinothiazóis os modelos criados com cargas de Mulliken tiveram resultado melhor que os modelos criados com cargas CHELP.

O modelo PCA obtido em 3 PCs, com as cargas de Mulliken tem menos classificações erradas (55,65, e 82, na figura 5.24), porém forma conjuntos mais esparsos. O modelo análogo obtido com cargas CHELP faz mais classificações erradas tanto do grupo mais ativo (57, 74, e 89, na figura 5.25) como do grupo menos ativo (59, 63, e 82, na figura 5.26), porém cria um maior agrupamento. A conclusão em favor das cargas de Mulliken pode ser obtida levando se em consideração regressões PLS, que foi mal sucedida utilizando cargas CHELP.

Há coincidência nas posições apontadas pelos modelos onde a carga tem importância na atividade biológica dos compostos, exatamente no caso dos nitrogênios N3, e N4. A carga

CHELP na posição R2 tem importância, e a de Mulliken na posição C9, que são adjacentes como pode ser observado na estrutura numerada da figura 5.20. As posições C6 e C7 não foram apontadas pelas cargas de Mulliken.

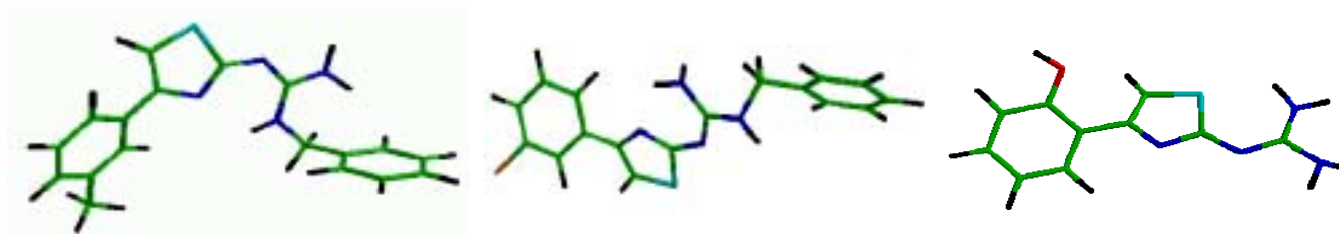


Figura 5.24: Conformações criadas com Hamiltoniano AM1 para os compostos de alta atividade 55(dir.) e 65 (centro), e para o composto de baixa atividade 82(esq.). Estes compostos foram mal classificados na PCA final com cargas de Mulliken

Para a correlação entre as atividades *in vivo* dos guanidinothiazóis as cargas CHELP foram mais úteis, sem dúvida. Os parâmetros que mostraram correlação também têm bastante coincidência com os que produziam melhores resultados para as atividades *in vitro*, ainda

que não seja possível traçar paralelismo direto entre os mecanismos que levam a uma ou outra ação. Os ensaios *in vivo* incluem fatores que não têm influência determinada nos resultados dos ensaios *in vitro*, como a decomposição da droga, sua excreção, ou os mecanismos ligados ao transporte da droga do aparelho digestivo ao sítio receptor intracelular. Os parâmetros acrescentados ao modelo das atividades *in vivo*, μ^{hib} , vdWSA, $F_{\text{C5}}^{\text{LUMO}}$, $F_{\text{C9}}^{\text{HOMO}}$, e $F_{\text{C10}}^{\text{HOMO}}$ devem estar relacionados de alguma forma a estes fatores que não estavam presentes na medição das atividades *in vitro*.

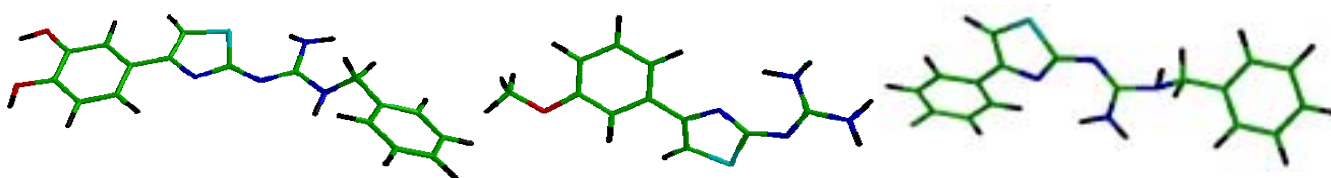


Figura 5.25: Compostos 89(esq.), 74(centro), e 57(dir.), que têm baixa atividade e foram mal classificados na PCA final com cargas CHELP.

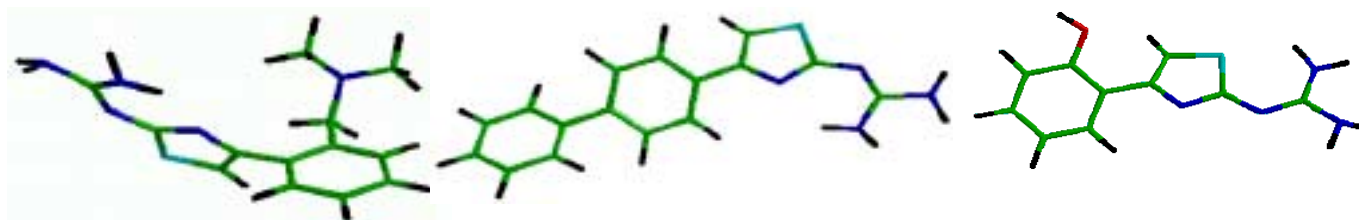


Figura 5.26: Os compostos 59(esq.), 63(centro), e 82(dir.), que têm alta atividade e foram erradamente classificados na PCA final com cargas CHELP.

Os modelos foram capazes de classificar corretamente 21 dos 23 compostos. As classificações erradas são os compostos 84 e 88(fig.59), que são muito ativos e foram classificados como pouco ativos.

Os modelos criados para esta classe de compostos, para os dois tipos de atividade biológica descritos, também podem ser úteis na previsão classificatória das atividades dos compostos, com uma base estatística bem mais sólida que no caso dos furanoditerpenos

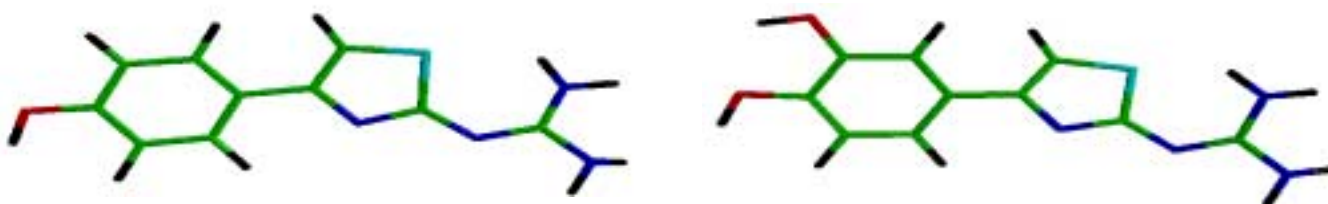


Figura 5.27: Compostos 84(esq.) e 88(dir.) que são muito ativos *in vivo* e foram classificados incorretamente na PCA.

(capítulo 5, parte 2).

5.5.2. Modelos de regressão com redes neurais

Os modelos baseados no treinamento de redes neurais são muito poderosos na classificação da atividade dos compostos 4-fenil-2-guanidinothiazóis. As redes neurais treinadas são capazes de ajustes perfeitos às condições impostas. Pode-se ver que o único composto classificado errado é o número 82, na tabela 5.10.

Os resultados de previsão são bons, considerando-se que são experiências preliminares com as técnicas de ajuste deste tipo de modelo, porém com excessiva sensibilidade à retirada de alguma amostra do modelo. Isto afeta grandemente sua capacidade de previsão para amostras não incluídas no conjunto de treinamento, prejudicando a qualidade dos resultados das previsões tipo “deixe um de fora”. Pode-se concluir até o momento que o aumento da acuidade numérica da função de erro do ajuste da rede aumenta a qualidade dos resultados na classificação.

Para classificação de grupos mais ou menos ativos, o melhor resultado mostra capacidade de previsão das grupo com índice de acertos de 82% com modelos tipo “deixe um de fora” (tabela 5.10).

Os resultados de ajuste de modelos para previsão de valores dão resultados excelentes, com índice de correlação R entre os valores experimentais e os ajustados de 0,99. Os modelos quantitativos tipo “deixe um de fora” indicam que há um limite em torno de 0,01 onde se obtém os melhores resultados. Este resultado pode ser devido ao supertreinamento da rede, que faz com que ela perca sua flexibilidade para ajustar amostras desconhecidas. A rede treinada para prever quantitativamente os resultados de atividade *in vitro* dos guanidinothiazóis teve um desempenho animador, e foi capaz de prever o valor das atividades dos compostos *in vivo* com uma correlação entre os valores previstos, e os valores experimentais de 0,55

A interpretação dos resultados no sentido de sugerir modificações químicas nas moléculas para melhorar certas características da atividade de um medicamento deve levar em conta a estrutura de farmacóforo ^{lv} para a classe de compostos que agem pelo mesmo mecanismo, além da estrutura do receptor, de maneira que possam ser identificados sítios da molécula que liguem-se a pontos específicos do receptor. Isso não foi possível obter com os modelos de regressão com redes neurais.

5.5.3. Relações entre os resultados obtidos e os trabalhos anteriormente publicados

O trabalho publicado por LaMattina e colaboradores mostra as atividades biológicas medidas para a série de compostos, e algumas considerações sobre a influência da estrutura química na atividade, baseadas nos resultados obtidos na substituição dos grupos guanidina, tiazol, R1, R2, e R3. (capítulo 1, parte 4).

As cargas nos nitrogênios N3 e N4 (ver figura 5.20), do grupo guanidina, são apontadas em todos os modelos com atividades *in vitro* como importante para atividade. O resultado concorda com os estudos anteriores, onde foi demonstrado que este grupo é essencial para a atividade biológica destes compostos. O modelo com atividades *in vivo* não aponta a carga na posição N4.

Na região do substituinte R2, as cargas sobre o substituinte e sobre o carbono C9 são apontadas. No carbono C9 o modelo com cargas de Mulliken e atividades *in vitro* mostra que a carga nesta posição deve ser mais negativa.

Estas observações são concordantes com as realizadas por LaMattina e seus colaboradores, responsáveis pela síntese dos compostos e medição das suas atividades biológicas. As demais variáveis selecionadas não têm paralelo aos resultados obtidos anteriormente, e poderão ser consideradas melhoramentos na compreensão da interação dos compostos com o receptor, se melhor entendidas e interpretadas em conjunto com o conhecimento da estrutura do receptor ligado às drogas, que ainda não é disponível.

Capítulo 5

Os compostos 4-fenil-2-guanidinothiazóis.

5.1. Seleção preliminar dos parâmetros.

5.1.1.1. Análise exploratória das cargas obtidas para os 4-fenil 2-guanidinothiazóis.

Os dados das cargas nas diferentes posições dos compostos (fig.30) foi calculado para o conjunto inicial de 43 compostos (tab.xx). Os dados foram autoescalados para HCA, feita com ligação simples entre os grupos. Os

compostos 89, 54 e 74 foram classificados errado no grupo dos compostos inativos. Este grupo é formado por

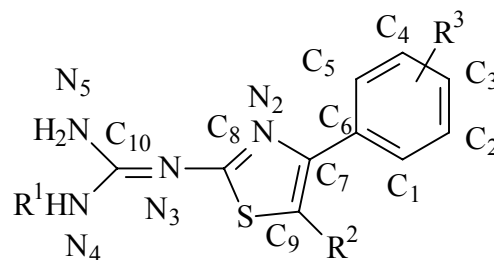


Figura 30: A numeração dos átomos adotada nas análises de cargas e

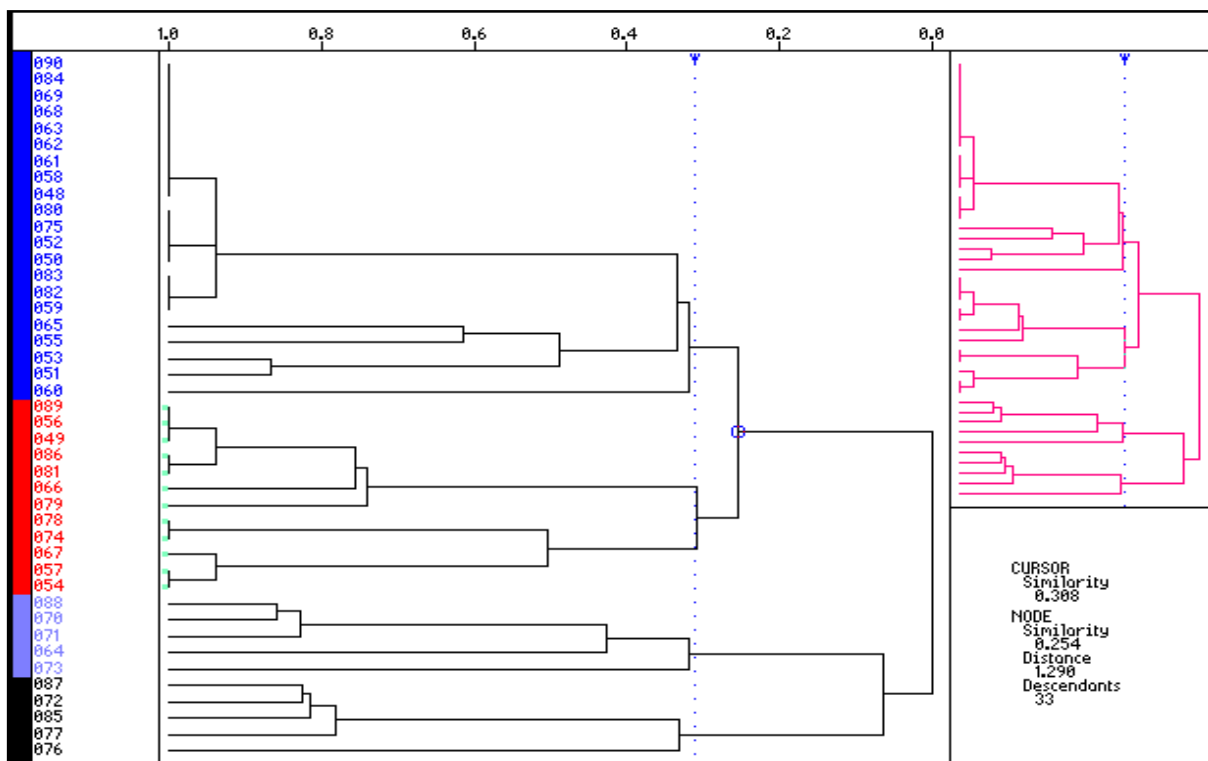


Figura 31: Dendrograma mostrando a conexão entre os compostos obtida no agrupamento hierárquico de blocos, com as cargas CHELP e de Mulliken. 12 elementos com similaridade maior que 0,309, e dista 1,290 UV do grupo ativo mais

próximo. Este grupo ativo tem 21 elementos com similaridade maior que 0,319, e nele estão classificados errado como ativos os compostos 59, 63 e 82. Distando 1,494 do centro de conexão entre os primeiros grupos há outro agrupamento ativo de 10 elementos com similaridade maior que 0,065 que não tem nenhum composto classificado errado. O dendograma é mostrado na figura 31.

A PCA foi feita em três componentes principais com os dados autoescalados. O gráfico dos 'scores' dos compostos na primeira e segunda componentes, na figura 32, mostra que a primeira componente principal isola num extremo de valor os elementos agrupados como inativos na HCA, inclusive os classificados errado. Os compostos mais próximos são os elementos do primeiro grupo ativo criado na HCA, incluindo também os classificados errado. A segunda e terceira componentes principais ajudam a separar o

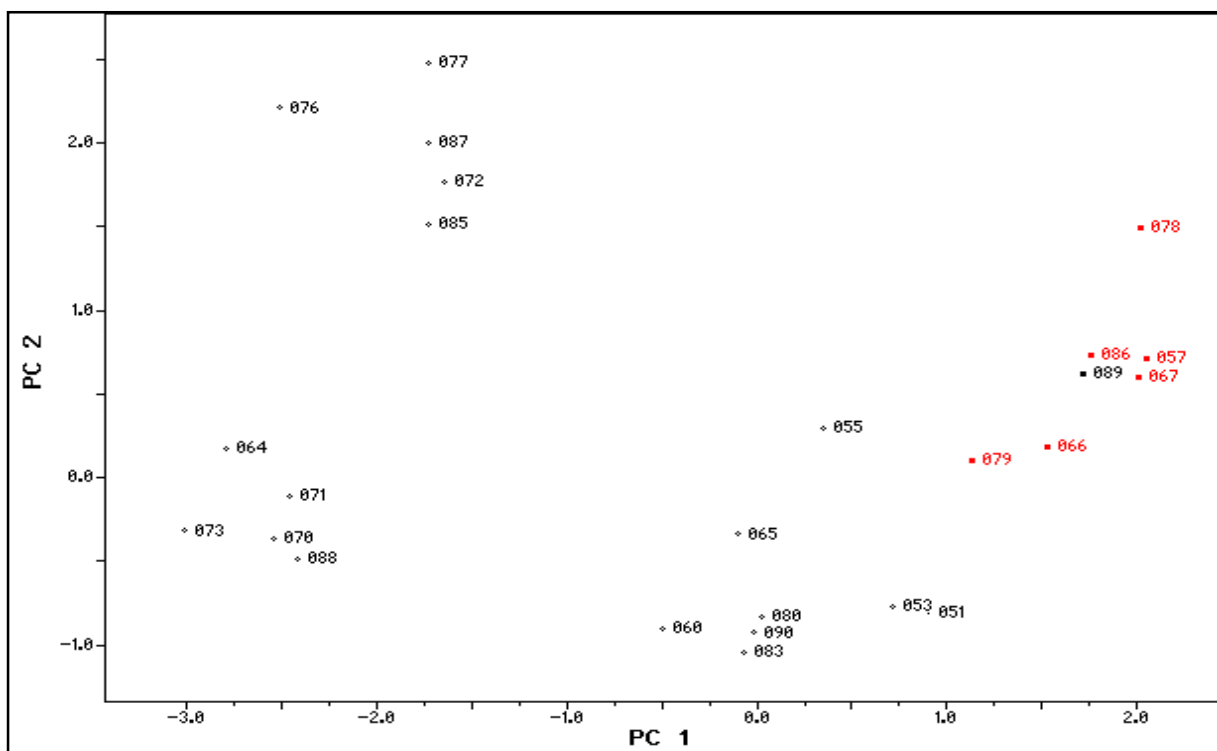


Figura 32: Gráfico dos 'scores' das duas primeiras componentes principais da análise PCA com atividades *in vitro*. Os compostos mais ativos são mostrados em preto, e os menos ativos em vermelho.

grupo ativo (PCA), de modo que no gráfico das três componentes este grupo aparece

mais destacado dos demais, ainda que com as mesmas classificações erradas já citadas.

É justificável a utilização de três componentes principais na separação de um conjunto com mais de 15 elementos, que é o caso.

As variáveis com capacidade de modelagem foram as cargas de Mulliken obtidas nas posições N4, C9 e N3 e a carga obtida como derivada do potencial eletrostático na posição N3. Os resultados indicam que cargas mais negativas nas posições N3 e C9 são relacionadas à maior atividade biológica. Na posição N4 cargas menos negativas favorecem a atividade. Os resultados obtidos por comparação com os valores do percentual de inibição competitiva da enzima, *in vitro*, são relacionados apenas à variação constante de equilíbrio entre a enzima e o substrato natural, e o complexo enzima-substrato. Não se pode associar estes resultados a mecanismos de transporte, do sistema digestivo até o meio intracelular. Deste modo é razoável atribuir deslocamento do equilíbrio químico à variação da carga nas posições. A carga elétrica num determinado átomo pode afetar a reatividade de um sítio de reação num composto, além de influir na aproximação, já que o potencial eletrostático é uma força que age a longa distância entre

Tabela 10: 'Loadings' das amostras, e da variância acumulada por cada uma das componentes principais no modelo com cargas para o grupo dos inibidores reversíveis da os átomos.

		PC1	PC2	PC3
Loading da variável	N3	0,33318	0,80180	-0,2689
	N3n	0,62011	-0,0430	-0,4019
	N4n	-0,5093	0,58469	0,16925
	C9n	0,49503	0,11577	0,85873
Variância acumulada		55,8265%	26,9284%	14,8473

De fato a presença dos nitrogênios nas posições N3 e N4 parece ser muito importante para a atividade biológica dos compostos. Os compostos 2, 3, 5, e 6 que não têm esta funcionalização e apresentam muito pouca inibição da enzima (3, 5, 6), ou nenhuma (2), o que é uma indicação de sua participação na interação com o receptor. Estes compostos

não foram incluídos no estudo presente. A carga na posição C9 parece não ter papel decisivo no processo, atuando paralelamente.

Na tabela 11 estão listados todos os parâmetros de carga calculados para este conjunto de moléculas.

Tabela 11: Cargas CHEPL(Ax) e de Mulliken(Axn), em negrito, calculadas para os 43 guanidinothiazóis, nas linhas.

S1	S1n	N1	N1n	N2	N2n	N3	N3n	N4	N4n	C11	No.
0,06	0,52	-0,52	-0,2	-0,87	-0,3	-0,96	-0,33	-0,94	-0,38	1,07	48
0,06	0,52	-0,51	-0,2	-0,81	-0,29	-0,81	-0,27	-0,78	-0,38	0,95	49
0,06	0,51	-0,52	-0,19	-0,88	-0,3	-0,95	-0,33	-0,86	-0,38	1,04	50
0,07	0,52	-0,57	-0,21	-0,89	-0,3	-0,95	-0,33	-0,9	-0,39	1,07	51
0,07	0,52	-0,55	-0,2	-0,88	-0,3	-0,95	-0,33	-0,93	-0,38	1,07	52
0,07	0,52	-0,55	-0,21	-0,87	-0,3	-0,96	-0,33	-0,92	-0,38	1,07	53
0,06	0,51	-0,51	-0,19	-0,83	-0,30	-0,82	-0,29	-0,82	-0,38	0,99	54
0,07	0,51	-0,5	-0,21	-0,85	-0,31	-0,92	-0,33	-0,72	-0,34	1,02	55
0,07	0,52	-0,51	-0,2	-0,82	-0,29	-0,81	-0,27	-0,83	-0,38	0,98	56
0,08	0,52	-0,53	-0,21	-0,85	-0,3	-0,82	-0,29	-0,88	-0,38	1,02	57
0,07	0,51	-0,53	-0,2	-0,86	-0,3	-0,96	-0,33	-0,94	-0,38	1,07	58
0,08	0,52	-0,57	-0,2	-0,86	-0,3	-0,97	-0,33	-0,93	-0,38	1,08	59
0,06	0,52	-0,49	-0,2	-0,73	-0,3	-0,94	-0,32	-0,63	-0,38	0,82	60
0,07	0,52	-0,54	-0,20	-0,87	-0,30	-0,96	-0,33	-0,93	-0,38	1,07	61
0,06	0,51	-0,53	-0,21	-0,87	-0,30	-0,96	-0,33	-0,92	-0,38	1,07	62
0,06	0,52	-0,56	-0,2	-0,88	-0,3	-0,96	-0,33	-0,93	-0,38	1,07	63
0,06	0,51	-0,39	-0,14	-0,76	-0,24	-0,99	-0,36	-0,85	-0,28	1,03	64
0,06	0,5	-0,57	-0,21	-0,86	-0,31	-0,98	-0,34	-0,8	-0,34	1,06	65
0,06	0,52	-0,51	-0,2	-0,83	-0,29	-0,85	-0,27	-0,83	-0,38	1,01	66
0,05	0,51	-0,52	-0,21	-0,84	-0,3	-0,83	-0,29	-0,86	-0,38	1,01	67
0,07	0,51	-0,56	-0,2	-0,87	-0,3	-0,96	-0,33	-0,95	-0,38	1,08	68
0,06	0,51	-0,53	-0,2	-0,86	-0,3	-0,96	-0,33	-0,92	-0,38	1,07	69
0,05	0,5	-0,59	-0,18	-0,89	-0,29	-0,97	-0,39	-0,96	-0,33	1,09	70
0,05	0,5	-0,6	-0,19	-0,88	-0,29	-0,96	-0,38	-0,96	-0,32	1,09	71
0,06	0,51	-0,52	-0,2	-0,82	-0,29	-0,86	-0,37	-0,79	-0,28	0,98	72
0,07	0,5	-0,55	-0,18	-0,81	-0,28	-0,95	-0,38	-0,95	-0,33	1,04	73
0,05	0,51	-0,5	-0,21	-0,84	-0,31	-0,75	-0,31	-0,84	-0,38	0,95	74
0,07	0,52	-0,65	-0,21	-0,87	-0,3	-0,95	-0,33	-0,93	-0,38	1,07	75
0,06	0,51	-0,51	-0,2	-0,81	-0,29	-0,82	-0,38	-0,81	-0,27	0,97	76
0,06	0,52	-0,51	-0,20	-0,82	-0,29	-0,81	-0,38	-0,82	-0,27	0,97	77
0,05	0,51	-0,51	-0,21	-0,84	-0,31	-0,75	-0,31	-0,84	-0,38	0,95	78
0,07	0,51	-0,52	-0,21	-0,85	-0,3	-0,86	-0,29	-0,92	-0,38	1,06	79
0,07	0,52	-0,57	-0,2	-0,88	-0,3	-0,95	-0,33	-0,94	-0,38	1,08	80
0,07	0,52	-0,52	-0,2	-0,8	-0,29	-0,8	-0,27	-0,81	-0,38	0,97	81
0,07	0,51	-0,56	-0,2	-0,87	-0,29	-0,97	-0,33	-0,97	-0,38	1,09	82
0,07	0,52	-0,55	-0,2	-0,89	-0,3	-0,97	-0,33	-0,95	-0,38	1,09	83
0,07	0,52	-0,54	-0,21	-0,87	-0,30	-0,96	-0,33	-0,94	-0,38	1,08	84
0,07	0,51	-0,54	-0,20	-0,85	-0,29	-0,87	-0,38	-0,84	-0,29	1,02	85

0,06	0,52	-0,52	-0,20	-0,8	-0,29	-0,8	-0,27	-0,81	-0,38	0,95	86
0,07	0,51	-0,53	-0,2	-0,82	-0,29	-0,84	-0,38	-0,83	-0,28	1,01	87
0,07	0,51	-0,6	-0,18	-0,89	-0,28	-0,98	-0,38	-0,97	-0,33	1,1	88
0,07	0,52	-0,52	-0,2	-0,81	-0,29	-0,81	-0,27	-0,82	-0,38	0,97	89
0,04	0,5	-0,59	-0,2	-0,86	-0,3	-0,96	-0,33	-0,93	-0,38	1,07	90
C1	C1n	C2	C2n	C3	C3n	C4	C4n	C5	C5n	C11n	No.
-0,03	-0,1	-0,11	-0,14	-0,1	-0,13	-0,1	-0,13	-0,08	-0,12	0,25	48
-0,07	-0,1	-0,09	-0,14	-0,1	-0,13	-0,1	-0,13	-0,1	-0,13	0,25	49
-0,14	-0,04	0,03	-0,13	-0,15	-0,12	-0,09	-0,13	-0,11	-0,11	0,25	50
-0,02	-0,1	-0,15	-0,13	0,02	-0,12	-0,14	-0,06	0,03	-0,11	0,26	51
-0,09	-0,09	0,0	-0,13	-0,14	-0,06	0,02	-0,13	-0,14	-0,11	0,25	52
0	-0,08	-0,24	-0,17	0,31	0,09	-0,24	-0,17	-0,03	-0,09	0,25	53
-0,13	-0,04	0,03	-0,13	-0,14	-0,12	-0,08	-0,13	-0,09	-0,13	0,26	54
-0,01	-0,1	-0,15	-0,13	0,01	-0,12	-0,14	-0,06	0,04	-0,11	0,26	55
-0,1	-0,09	0	-0,13	-0,15	-0,06	0,05	-0,13	-0,16	-0,11	0,25	56
-0,02	-0,08	-0,24	-0,17	0,31	0,09	-0,23	-0,17	-0,04	-0,09	0,26	57
0,09	-0,04	-0,16	-0,14	-0,06	-0,12	-0,16	-0,14	-0,01	-0,11	0,25	58
0,28	-0,09	-0,2	-0,13	-0,09	-0,13	-0,1	-0,13	-0,04	-0,11	0,25	59
-0,05	-0,1	0,04	-0,08	-0,15	-0,13	-0,09	-0,13	-0,05	-0,12	0,24	60
-0,03	-0,1	-0,12	-0,09	0,01	-0,1	-0,14	-0,14	-0,08	-0,12	0,25	61
-0,06	-0,1	-0,14	-0,14	0,04	-0,07	-0,14	-0,13	-0,08	-0,11	0,25	62
-0,03	-0,10	-0,09	-0,13	-0,04	-0,03	-0,07	-0,12	-0,1	-0,12	0,25	63
0,06	-0,03	-0,13	-0,14	-0,07	-0,12	-0,15	-0,14	-0,02	-0,11	0,21	64
-0,1	-0,1	0,05	-0,08	-0,16	-0,13	-0,07	-0,13	-0,11	-0,12	0,27	65
-0,03	-0,1	-0,16	-0,14	0,03	-0,07	-0,13	-0,14	-0,09	-0,11	0,24	66
-0,1	-0,1	-0,08	-0,13	-0,05	-0,04	-0,05	-0,12	-0,15	-0,12	0,26	67
-0,15	-0,14	-0,01	-0,1	-0,35	-0,18	0,57	0,05	-0,33	-0,18	0,25	68
-0,26	-0,12	0,36	0,04	-0,14	-0,17	-0,15	-0,11	-0,11	-0,14	0,25	69
-0,02	-0,07	-0,31	-0,2	0,52	0,05	-0,35	-0,19	0,04	-0,08	0,25	70
-0,05	-0,07	-0,23	-0,19	0,42	0,06	-0,38	-0,17	0,05	-0,09	0,24	71
0,03	-0,06	-0,35	-0,2	0,56	0,06	-0,39	-0,19	0,08	-0,08	0,25	72
0,06	-0,07	-0,21	-0,17	-0,02	-0,09	-0,29	-0,21	0,44	0,09	0,24	73
-0,12	-0,14	-0,08	-0,1	-0,25	-0,20	0,37	0,07	-0,29	-0,15	0,27	74
-0,01	-0,07	-0,25	-0,21	0,37	0,08	-0,28	-0,16	-0,01	-0,09	0,25	75
0,05	-0,08	-0,19	-0,17	-0,05	-0,09	-0,25	-0,21	0,37	0,09	0,24	76
-0,07	-0,14	-0,12	-0,1	-0,2	-0,2	0,33	0,07	-0,25	-0,15	0,25	77
-0,12	-0,14	-0,08	-0,1	-0,25	-0,2	0,37	0,07	-0,29	-0,15	0,27	78
-0,01	-0,07	-0,26	-0,21	0,37	0,08	-0,3	-0,16	0	-0,09	0,25	79
-0,07	-0,12	-0,21	-0,13	0,14	-0,01	0,3	0,07	-0,26	-0,14	0,25	80
-0,05	-0,09	-0,19	-0,19	0,14	0,04	0,18	0,04	-0,14	-0,11	0,24	81
0,05	-0,07	-0,18	-0,17	-0,05	-0,09	-0,27	-0,22	0,38	0,09	0,24	82
-0,12	-0,14	-0,06	-0,1	-0,28	-0,21	0,37	0,07	-0,27	-0,15	0,25	83
0,02	-0,06	-0,29	-0,22	0,36	0,08	-0,27	-0,16	-0,01	-0,08	0,25	84
0,04	-0,07	-0,2	-0,17	-0,02	-0,09	-0,3	-0,22	0,41	0,09	0,25	85
-0,12	-0,14	-0,05	-0,1	-0,29	-0,21	0,39	0,07	-0,28	-0,15	0,25	86
0,01	-0,06	-0,28	-0,22	0,36	0,08	-0,27	-0,16	0,03	-0,08	0,25	87
-0,1	-0,1	-0,23	-0,18	0,2	0	0,2	0,06	-0,16	-0,12	0,24	88
-0,07	-0,1	-0,23	-0,18	0,18	0	0,24	0,06	-0,21	-0,12	0,25	89
-0,14	-0,1	-0,21	-0,18	0,18	0	0,21	0,06	-0,17	-0,12	0,25	90
C6	C6n	C7	C7n	C8	C8n	C9	C9n	R10	R10n	No.	
-0,07	-0,01	0,15	-0,01	-0,25	-0,49	0,60	-0,14	0,16	0,18	48	

-0,01	-0,01	0,07	-0,02	-0,2	-0,49	0,59	-0,14	0,14	0,18	49
0,12	0,0	0,04	-0,02	-0,22	-0,48	0,64	-0,14	0,14	0,18	50
-0,09	0,01	0,16	-0,02	-0,27	-0,49	0,65	-0,13	0,16	0,18	51
0	0	0,13	-0,02	-0,25	-0,49	0,62	-0,14	0,15	0,18	52
-0,11	-0,02	0,17	-0,02	-0,26	-0,49	0,61	-0,13	0,16	0,18	53
0,07	0	0,08	-0,02	-0,22	-0,48	0,59	-0,13	0,14	0,18	54
-0,09	0	0,1	-0,02	-0,24	-0,49	0,57	-0,13	0,15	0,18	55
0,02	0	0,1	-0,02	-0,22	-0,49	0,58	-0,14	0,15	0,18	56
-0,08	-0,02	0,16	-0,02	-0,28	-0,49	0,6	-0,13	0,16	0,18	57
-0,14	-0,01	0,19	-0,01	-0,26	-0,49	0,58	-0,14	0,15	0,18	58
-0,3	0	0,31	-0,02	-0,28	-0,49	0,57	-0,14	0,14	0,18	59
-0,12	0	0,16	-0,02	-0,22	-0,49	0,51	-0,15	0,14	0,18	60
-0,07	-0,1	0,18	-0,01	-0,28	-0,49	0,61	-0,14	0,16	0,18	61
-0,05	-0,01	0,14	-0,01	-0,26	-0,49	0,63	-0,14	0,16	0,18	62
-0,07	0	0,18	-0,02	-0,28	-0,49	0,63	-0,14	0,17	0,18	63
-0,12	-0,01	0,12	-0,03	-0,2	-0,48	0,5	-0,15	0,14	0,18	64
-0,04	0	0,18	-0,02	-0,27	-0,49	0,61	-0,13	0,16	0,18	65
-0,04	-0,01	0,12	-0,01	-0,23	-0,49	0,59	-0,14	0,15	0,18	66
0	0	0,12	-0,02	-0,25	-0,49	0,59	-0,13	0,15	0,18	67
-0,01	0,03	0,21	-0,02	-0,28	-0,49	0,6	-0,14	0,16	0,18	68
0	0,01	0,15	-0,02	-0,25	-0,49	0,6	-0,14	0,15	0,18	69
-0,18	-0,05	0,22	-0,02	-0,28	-0,48	0,65	-0,15	0,16	0,18	70
-0,15	-0,04	0,22	-0,01	-0,27	-0,49	0,65	-0,15	0,15	0,18	71
-0,23	-0,05	0,19	-0,01	-0,25	-0,49	0,57	-0,14	0,15	0,18	72
-0,33	-0,03	0,27	-0,02	-0,26	-0,46	0,55	-0,16	0,16	0,19	73
0,06	0,03	0,09	-0,02	-0,22	-0,48	0,59	-0,13	0,15	0,18	74
-0,13	-0,04	0,19	-0,01	-0,28	-0,49	0,62	-0,14	0,16	0,18	75
-0,25	-0,03	0,19	-0,02	-0,24	-0,46	0,56	-0,15	0,17	0,19	76
0,01	0,02	0,11	-0,02	-0,23	-0,48	0,59	-0,14	0,15	0,18	77
0,06	0,03	0,09	-0,02	-0,23	-0,48	0,6	-0,13	0,15	0,18	78
-0,13	-0,4	0,15	-0,1	-0,27	-0,49	0,59	-0,14	0,16	0,18	79
-0,03	0,01	0,21	-0,02	-0,29	-0,49	0,62	-0,14	0,17	0,18	80
-0,1	-0,02	0,18	-0,01	-0,26	-0,49	0,56	-0,14	0,16	0,18	81
-0,25	-0,03	0,22	-0,02	-0,25	-0,46	0,6	-0,14	0,17	0,19	82
0,03	0,03	0,14	-0,02	-0,26	-0,48	0,64	-0,14	0,16	0,18	83
-0,15	-0,04	0,18	-0,01	-0,28	-0,49	0,6	-0,14	0,16	0,18	84
-0,26	-0,03	0,23	-0,02	-0,26	-0,46	0,59	-0,14	0,17	0,19	85
0,03	0,03	0,13	-0,02	-0,24	-0,48	0,58	-0,14	0,16	0,18	86
-0,18	-0,04	0,2	-0,01	-0,27	-0,49	0,56	-0,14	0,15	0,18	87
-0,09	-0,01	0,24	-0,02	-0,28	-0,48	0,64	-0,15	0,16	0,18	88
-0,05	0	0,16	-0,02	-0,26	-0,49	0,58	-0,14	0,16	0,18	89
-0,01	0	0,15	-0,02	-0,14	-0,4	0,62	-0,14	-0,14	-0,15	90

5.1.2. Análise dos volumes moleculares

Outro modelo foi construído com parâmetros de volume para os compostos e para os substituintes. Os dados foram autoescalados para HCA com conexão simples, e os resultados não foram bons. A PCA em três componentes com os dados autoescalados foi capaz de mostrar o grupo inativo com algum destaque dos demais compostos, embora

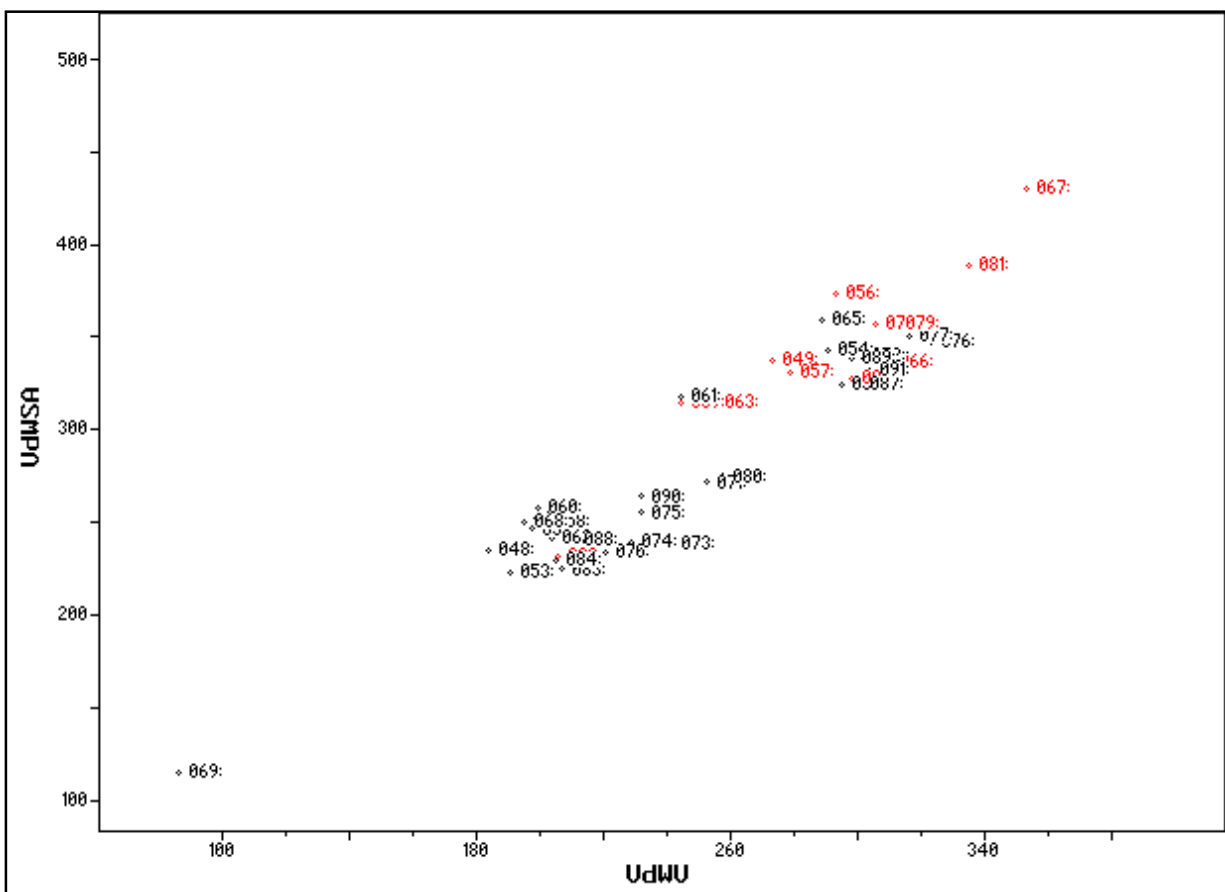


Figura 33: Gráfico das variáveis Volume de van der Waals e Área Acessível ao Solvente de van der Waals, onde se observa a correlação entre estas variáveis. Há tendência dos compostos ativos, em preto, a ter menor volume molecular.

não tenha ficado nem extremo de valor em nenhuma das componentes. O gráfico das três componentes principais, na figura 34 é capaz de isolar os compostos muito ativos, classificando errado os compostos 66 82 e 86 no grupo muito ativo, e os compostos 65, 61, 54 e 74 no grupo pouco ativo. A análise dos gráficos das variáveis duas a duas, na figura 33, permite observar que o aumento do volume molecular e da superfície molecular acessível ao solvente, são grandezas correlacionadas, e diretamente relacionadas a

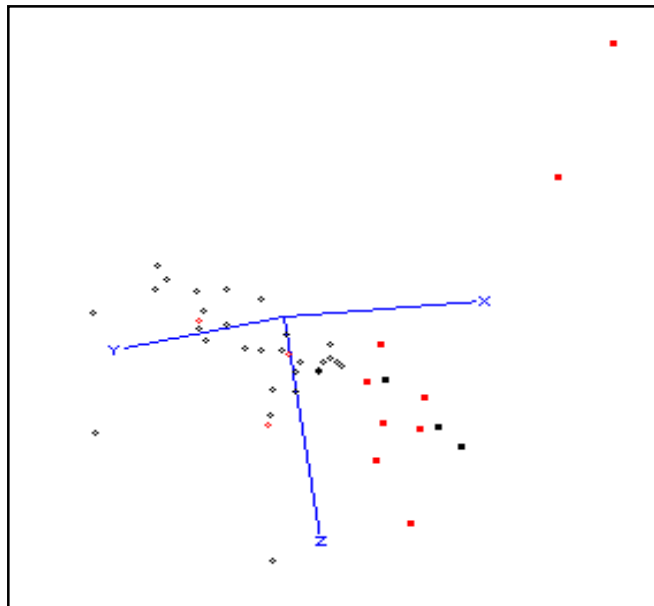


Figura 34: Gráfico das três primeiras componentes principais do modelo obtido com os volumes dos compostos. Os

atividade biológica no sentido que maiores volumes levam à menor atividade dos compostos. O volume dos substituintes não tem relação com as atividades *in vitro*. Foram classificados errado como compostos inativos, no modelo com volumes e no modelo com cargas, os

compostos 54 e 74, que não têm nenhuma semelhança estrutural marcante. Os

compostos 89 e 54, que também foram classificados errado nos dois modelos, têm o substituinte benzila na posição R1. Nenhum composto foi classificado errado como ativo nos dois modelos consecutivos.

Na tabela 11 estão listados todos os volumes calculados para os compostos e seus substituintes.

Tabela 11: Volumes e superfícies para os compostos, calculados com o programa SURF.

vdWV	vdWSA	SAV	SAS	R1vdWV	R1vdWS	R1SAV	R5vdWS	No.
				A				
183,90	235,00	682,20	440,60	7,23	18,32	81,32	18,32	48
273,40	337,50	967,10	596,80	99,26	131,88	405,33	18,32	49
197,70	246,20	721,70	457,60	7,23	18,32	81,32	18,32	50
199,00	252,30	729,20	466,50	7,23	18,32	81,32	18,32	51
199,00	252,50	729,30	466,10	7,23	18,32	81,32	43,05	52
190,80	223,00	693,70	447,50	7,23	18,32	81,32	22,75	53
290,80	342,90	1006,0	615,50	99,26	131,88	405,33	18,32	54
		0						
295,10	324,70	1004,0	602,20	99,26	131,88	405,33	18,32	55
		0						
293,60	373,00	1085,0	705,00	99,26	131,88	405,33	43,05	56
		0						
279,00	330,80	978,10	604,90	99,26	131,88	405,33	22,75	57
202,20	250,30	727,30	459,90	7,23	18,32	81,32	18,32	58
244,60	314,20	862,00	523,50	7,23	18,32	81,32	18,32	59
199,90	257,30	739,80	473,80	7,23	18,32	81,32	18,32	60
244,70	317,30	886,20	551,60	7,23	18,32	81,32	18,32	61
204,00	241,60	739,50	473,70	7,23	18,32	81,32	43,11	62
255,50	314,80	900,30	554,70	7,23	18,32	81,32	102,12	63

304,30	337,70	983,00	570,40	99,26	131,88	405,33	18,32	64
288,70	359,40	1011,0	618,20	99,26	131,88	405,33	18,32	65
		0						
310,20	336,70	1004,0	606,80	99,26	131,88	405,33	43,11	66
		0						
353,50	430,40	1291,0	835,50	99,26	131,88	405,33	102,12	67
		0						
195,00	250,10	722,30	463,90	7,23	18,32	81,32	18,32	68
86,50	114,60	357,00	230,60	7,23	18,32	81,32	18,32	69
220,80	233,30	747,00	473,70	7,23	18,32	81,32	48,06	70
252,90	270,70	848,00	528,60	7,23	18,32	81,32	88,22	71
301,80	339,30	1021,0	625,10	99,26	131,88	405,33	48,06	72
		0						
241,40	238,00	775,00	484,00	7,23	18,32	81,32	18,32	73
306,10	356,90	1025,0	620,20	7,23	18,32	81,32	18,32	74
		0						
231,90	254,20	780,00	492,50	7,23	18,32	81,32	87,29	75
323,60	347,20	1045,0	634,60	99,26	131,88	405,33	18,32	76
		0						
316,30	350,70	1054,0	642,60	99,26	131,88	405,33	18,32	77
		0						
306,10	356,90	1025,0	620,20	93,30	141,62	412,25	18,32	78
		0						
311,90	356,60	1055,0	641,70	99,26	131,88	405,33	87,29	79
		0						
257,50	274,30	853,00	530,70	7,23	18,32	81,32	87,29	80
335,20	388,80	1128,0	679,70	99,26	131,88	405,33	87,29	81
		0						
206,20	231,30	716,00	456,50	7,23	18,32	81,32	18,32	82
207,20	223,90	720,00	458,30	7,23	18,32	81,32	18,32	83
205,20	229,00	720,00	460,90	7,23	18,32	81,32	34,30	84
295,20	324,10	991,00	606,70	99,26	131,88	405,33	18,32	85
298,10	327,90	994,00	608,40	99,26	131,88	405,33	18,32	86
300,90	324,80	995,00	606,90	99,26	131,88	405,33	34,30	87
210,60	240,30	742,00	468,40	7,23	18,32	81,32	34,30	88
298,10	338,30	1016,0	617,40	99,26	131,88	405,33	34,30	89
		0						
232,20	263,60	798,00	501,10	7,23	18,32	81,32	34,30	90
R1SAS	R2vdWV	R2vdWS	R2SAV	R2SAS	R3vdWV	R3vdWS	R5SAV	No.
		A				A		
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	48
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	49
85,36	7,23	18,32	81,32	85,36	26,50	43,05	81,32	50
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	51
85,36	7,23	18,32	81,32	85,36	7,23	18,32	155,84	52
85,36	7,23	18,32	81,32	85,36	7,23	18,32	96,89	53
285,44	7,23	18,32	81,32	85,36	26,50	43,05	81,32	54
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	55
285,44	7,23	18,32	81,32	85,36	7,23	18,32	155,84	56
285,44	7,23	18,32	81,32	85,36	7,23	18,32	96,89	57
85,36	7,23	18,32	81,32	85,36	27,88	43,11	81,32	58

85,36	7,23	18,32	81,32	85,36	112,82	160,25	81,32	59
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	60
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	61
85,36	7,23	18,32	81,32	85,36	7,23	18,32	170,26	62
85,36	7,23	18,32	81,32	85,36	7,23	18,32	300,00	63
285,44	7,23	18,32	81,32	85,36	27,88	43,11	81,32	64
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	65
285,44	7,23	18,32	81,32	85,36	7,23	18,32	170,26	66
285,44	7,23	18,32	81,32	85,36	7,23	18,32	300,00	67
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	68
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	69
85,36	7,23	18,32	81,32	85,36	7,23	18,32	172,20	70
85,36	7,23	18,32	81,32	85,36	7,23	18,32	281,83	71
285,44	7,23	18,32	81,32	85,36	7,23	18,32	172,20	72
85,36	7,23	18,32	81,32	85,36	59,67	87,29	81,32	73
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	74
85,36	7,23	18,32	81,32	85,36	7,23	18,32	279,14	75
285,44	7,23	18,32	81,32	85,36	59,67	87,29	81,32	76
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	77
292,99	7,23	18,32	81,32	85,36	7,23	18,32	81,32	78
285,44	7,23	18,32	81,32	85,36	7,23	18,32	279,14	79
85,36	7,23	18,32	81,32	85,36	7,23	18,32	279,14	80
285,44	7,23	18,32	81,32	85,36	7,23	18,32	279,14	81
85,36	7,23	18,32	81,32	85,36	17,35	34,30	81,32	82
85,36	7,23	18,32	81,32	85,36	7,23	18,32	81,32	83
85,36	7,23	18,32	81,32	85,36	7,23	18,32	130,35	84
285,44	7,23	18,32	81,32	85,36	17,35	34,30	81,32	85
285,44	7,23	18,32	81,32	85,36	7,23	18,32	81,32	86
285,44	7,23	18,32	81,32	85,36	7,23	18,32	130,35	87
85,36	7,23	18,32	81,32	85,36	7,23	18,32	130,35	88
285,44	7,23	18,32	81,32	85,36	7,23	18,32	130,35	89
85,36	27,88	43,11	170,26	152,27	7,23	18,32	130,35	90
R3SAV	R3SAS	R4vdWV	R4vdWS	R4SAV	R4SAS	R5vdWV	R5SAS	No.
A								
81,32	85,36	7,23	18,32	81,32	85,36	7,23	85,36	48
81,32	85,36	7,23	18,32	81,32	85,36	7,23	85,36	49
155,84	133,29	7,23	18,32	81,32	85,36	7,23	85,36	50
81,32	85,36	26,50	43,05	155,84	133,29	7,23	85,36	51
81,32	85,36	7,23	18,32	81,32	85,36	26,50	133,29	52
81,32	85,36	7,23	18,32	81,32	85,36	10,30	101,39	53
155,84	133,29	7,23	18,32	81,32	85,36	7,23	85,36	54
81,32	85,36	26,50	43,05	155,84	133,29	7,23	85,36	55
81,32	85,36	7,23	18,32	81,32	85,36	26,50	133,29	56
81,32	85,36	7,23	18,32	81,32	85,36	10,30	101,39	57
170,26	152,27	7,23	18,32	81,32	85,36	7,23	85,36	58
510,72	364,43	7,23	18,32	81,32	85,36	7,23	85,36	59
81,32	85,36	27,88	43,11	170,26	152,27	7,23	85,36	60
81,32	85,36	112,82	160,25	510,72	364,43	7,23	85,36	61
81,32	85,36	7,23	18,32	81,32	85,36	27,88	152,27	62
81,32	85,36	7,23	18,32	81,32	85,36	86,03	175,05	63
170,26	152,27	7,23	18,32	81,32	85,36	7,23	85,36	64

81,32	85,36	27,88	43,11	170,26	152,27	7,23	85,36	65
81,32	85,36	7,23	18,32	81,32	85,36	27,88	152,27	66
81,32	85,36	7,23	18,32	81,32	85,36	86,03	175,05	67
81,32	85,36	30,56	48,06	172,20	150,38	7,23	85,36	68
81,32	85,36	60,32	88,22	281,83	217,14	7,23	85,36	69
81,32	85,36	7,23	18,32	81,32	85,36	30,56	150,38	70
81,32	85,36	7,23	18,32	81,32	85,36	60,32	217,14	71
81,32	85,36	7,23	18,32	81,32	85,36	30,56	150,38	72
279,14	215,34	7,23	18,32	81,32	85,36	7,23	85,36	73
81,32	85,36	59,67	87,29	279,14	215,34	7,23	85,36	74
81,32	85,36	7,23	18,32	81,32	85,36	59,67	215,34	75
279,14	215,34	7,23	18,32	81,32	85,36	7,23	85,36	76
81,32	85,36	59,67	87,29	279,14	215,34	7,23	85,36	77
81,32	85,36	59,67	87,29	279,14	215,34	7,23	85,36	78
81,32	85,36	7,23	18,32	81,32	85,36	59,67	215,34	79
81,32	85,36	59,67	87,29	279,14	215,34	59,67	215,34	80
81,32	85,36	59,67	87,29	279,14	215,34	59,67	215,34	81
130,35	125,91	7,23	18,32	81,32	85,36	7,23	85,36	82
81,32	85,36	17,35	34,30	130,35	125,91	7,23	85,36	83
81,32	85,36	7,23	18,32	81,32	85,36	17,35	125,91	84
130,35	125,91	7,23	18,32	81,32	85,36	7,23	85,36	85
81,32	85,36	17,35	34,30	130,35	125,91	7,23	85,36	86
81,32	85,36	7,23	18,32	81,32	85,36	17,35	125,91	87
81,32	85,36	17,35	34,30	130,35	125,91	17,35	125,91	88
81,32	85,36	17,35	34,30	130,35	125,91	17,35	125,91	89
81,32	85,36	17,35	34,30	130,35	125,91	17,35	125,91	90

5.1.3. Análise das densidades de elétron nos orbitais de
fronteira.

Foi realizado um modelo com todos os índices de densidade eletrônica nos orbitais de fronteira. Ao todo foram obtidas 64 variáveis, que passaram por esta triagem inicial que revelou quais tinham o maior poder de previsão para as atividades medidas *in vivo*.

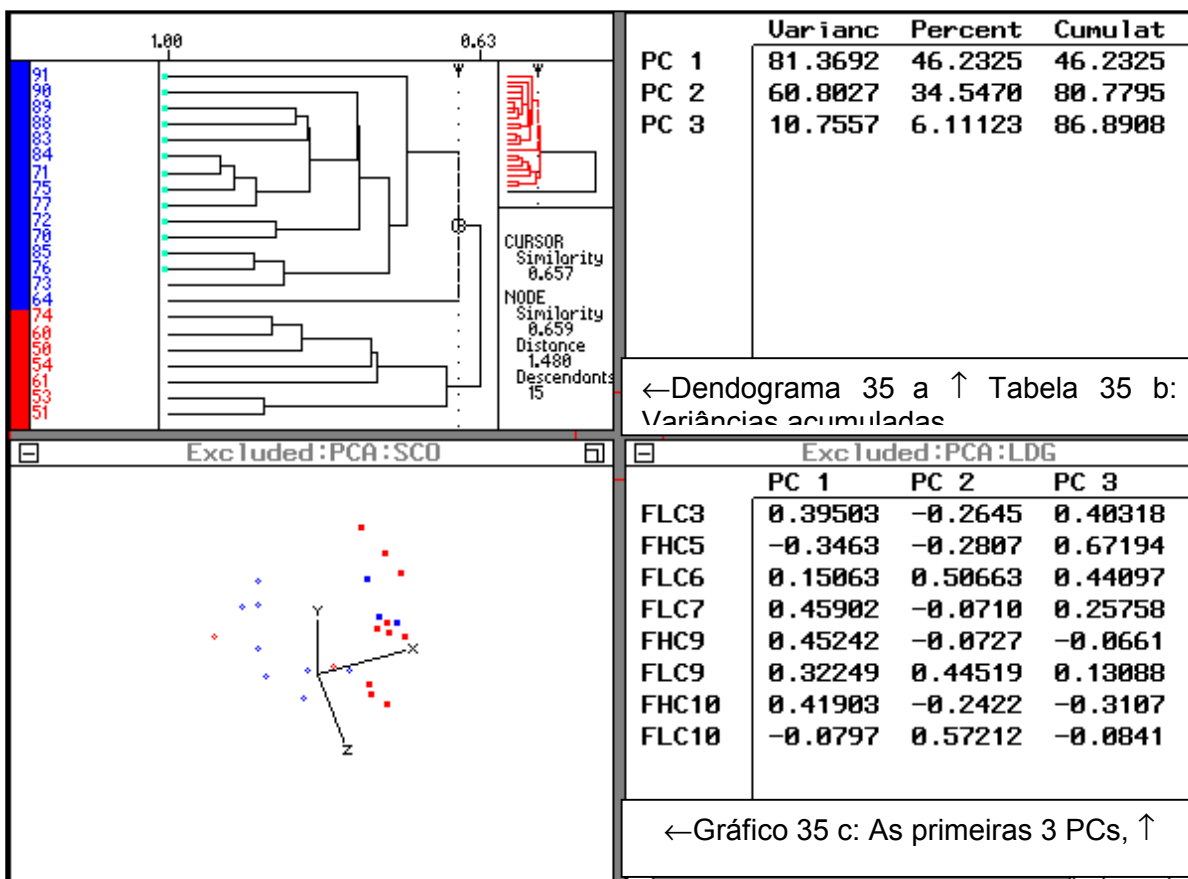


Figura 35: Resultados da análise exploratória dos dados das densidades eletrônicas nos orbitais de fronteira para os 4-fenil 2-guanidinothiazóis.

Na figura 35^a são mostrados os gráficos correspondentes HCA com dados autoescalados para os 23 compostos. O bloco marcado em azul tem 15 elementos, 11 são compostos mais ativos, com similaridade superior à 0,648UV distante 1,534UV do outro bloco. Os quatro menos ativos no bloco azul são os compostos 70, 73, 75, e 83. O bloco vermelho tem sete compostos, seis menos ativos, com similaridade maior que 0,674. O composto 54 foi erradamente classificado neste bloco. O composto 58 não aparece no diagrama, está situado à distância de 2,534UV dos demais blocos, e é pouco ativo. Os compostos

que aparecem com um ponto verde são os que estão em em negrito no gráfico da figura

35c

O gráfico das três componentes utilizadas em PCA é mostrado na figura 35c. Elas acumulam 87% da variância total dos dados, e classificam errado cinco compostos. No grupo mais ativo, mostrado em vermelho, com ponto maiores estão classificados errado os compostos 70, 75, e 83, que aparecem em azul. No grupo menos ativo, mostrado em azul, com pontos menores, estão erradamente classificados os compostos 54, e 64. Todas as variáveis selecionadas têm contribuição expressiva na primeira componente principal, exceto a densidade em lume da camada de valência sobre R5, denominada FLC10. A tabela 12 mostra todas as densidades eletrônicas nos orbitais de fronteira calculadas para os guanidínotiazóis.

Para as atividades *in vitro* o modelo realizado não resultou nenhuma variável com boa capacidade de previsão.

Tabela 13: Densidades nos orbitais de fronteira para os átomos, seguindo a numeração da figura 30.

HOMO	L-H/2	FHC2	FLC3	FHC5	FLN5	No.
-8,25245	3,97291	-	-0,079246	-0,00229525	-0,103249	50
		0,000359301				
-8,23943	3,88334	-	-0,124786	-0,00370626	-0,0475783	51
		0,000332243				
-8,20147	3,85682	-	-0,149524	-0,00433153	-0,0419736	52
		0,000799971				
-8,22589	3,86031	-0,00106959	-0,141631	-0,00428557	-0,0424858	53
-8,15928	3,93954	-	-0,113449	-0,00248707	-0,0569864	54
		0,000473245				
-8,20259	3,9026	-	-	-	-0,100392	55
		0,000104903	0,00316125	0,000627178		
-8,21754	3,89291	-	-0,115627	-0,00354757	-0,0687726	56
		0,000844918				
-8,25341	3,86875	-0,00104039	-0,0960254	-0,00389388	-0,0662593	57
-8,24392	3,99361	-6,19027e-	-0,0145536	-	0	58
		05		0,000393583		
-8,28062	3,98493	-	-0,0199773	-0,00165699	-0,0733141	59
		0,000231148				
-8,14074	3,91727	-	-0,0747453	-0,00355452	-0,0630898	60
		0,000647691				
-8,22775	3,91811	-	-0,0551191	-0,0035766	-0,0811808	61

		0,000756878				
-8,10104	3,89855	-	-0,0876768	-0,00372797	-0,103592	62
		0,000923636				
-8,12078	3,86176	-0,00156019	-0,16329	-0,00405856	-0,0314514	63
-8,38106	4,02375	-0,00195132	-0,0293352	-0,00563509	-	64
					0,00950046	
-8,01407	3,90092	-	-0,132346	-0,00314345	-0,135831	65
		0,000521382				
-8,11745	3,88594	-0,00106823	-0,0782108	-0,00384967	-0,0486055	66
-8,05438	3,84342	-0,00139822	-0,184497	-0,00393596	-0,029914	67
-8,10947	3,91204	-8,35326e-	-0,0736434	-	-0,110269	68
		05		0,000687531		
-8,12818	3,92518	-0,0020253	-0,070906	-0,00696879	-0,116587	69
-8,00353	3,77185	-0,00845113	-0,0460418	-0,00491673	-0,0328982	70
-8,28582	3,85446	-0,00237612	-0,0416145	-0,0061609	-	71
					0,00137184	
-7,98894	3,76648	-0,00825389	-0,0327339	-0,00496198	-0,027876	72
-8,17152	3,89851	-9,56891e-	-0,0557498	-0,00307131	-0,0417172	73
		06				
-8,16352	3,92341	-	-0,0902292	-0,00499187	-0,0567092	74
		0,000938322				
-8,17308	3,80998	-0,00496742	-0,0315756	-0,00673826	-0,0501627	75
-8,27457	3,92813	-	-0,0323853	-0,00332126	-0,045478	76
		0,000356301				
-8,35760	3,88277	-0,00189711	-0,0326516	-0,00705756	-0,0307273	77
-8,38917	3,90379	-0,00322282	-0,0280483	-0,0103782	-	78
					0,00138266	
-8,27175	3,89798	-0,00445855	-0,0287984	-0,00607709	-0,0638038	79
-8,19440	3,822	-0,00843867	-0,105896	-0,0115803	-0,0097706	80
-8,15780	3,82502	-0,00732104	-0,0359196	-0,0103964	-0,0308466	81
-8,22002	3,91008	-2,41705e-	-0,0497929	-0,00292704	-0,0480181	82
		05				
-8,35779	3,89495	-0,00268819	-0,0397475	-0,00943751	-0,0263603	83
-8,19792	3,81717	-0,00445141	-0,0381655	-0,00629194	-0,0351716	84
-8,19258	3,89831	-2,93541e-	-0,0466685	-0,00276155	-0,0453888	85
		05				
-8,37862	3,88275	-0,00213989	-0,032767	-0,00828844	-0,0292333	86
-8,18703	3,80587	-0,00439014	-0,031833	-0,00626462	-0,0498451	87
-8,32062	3,91671	-0,00637545	-0,0520048	-0,0109412	-0,0558743	88
-8,20692	3,80873	-0,00594718	-0,0407307	-0,0104763	-0,0304377	89
-8,27604	3,86746	-0,00591914	-0,0281775	-0,00938687	-0,0521721	90
-8,01003	3,80898	-0,00179671	-0,0374137	-0,00446576	-0,0353327	91
LUMO	FHC1	FLC2	FHC4	FLC5	FHN5	No.
- ,30663	-0,00184468	-0,0196647	-	-0,0549353	-	50
			0,00018920		0,00629811	
- ,47276	-0,00485667	-0,0191873	-	-0,0617224	-	51
			0,00086689		0,00473391	
- ,48784	-0,00450929	-0,0276308	-	-0,0715253	-	52
			0,00104163		0,00441058	
- ,50526	-0,00441439	-0,0219358	-	-0,0665453	-	53
			0,00130934		0,00429152	

- ,28020	-0,00177664	-0,0312261	-	-0,0832273	-	54
			0,00011774		0,00453485	
- ,39739	-0,000600982	-0,00106709	-6,6459e-	-0,00451426	-	55
			05		0,00799046	
- ,43171	-0,00337098	-0,0199815	-	-0,0524693	-0,0063597	56
			0,00062648			
- ,51592	-0,00400959	-0,0141938	-	-0,0453692	-	57
			0,00128335		0,00686643	
- ,25669	-0,000470543	-0,0513218	-	-0,0592731	-	58
			0,00010544		0,00453613	
- ,31076	-0,00173458	-0,00255306	-	-0,0136401	-0,00393	59
			0,00026775			
- ,30620	-0,00387702	-0,00909719	-	-0,0458566	-0,0031258	60
			0,00063982			
- ,39153	-0,00388095	-0,00594128	-	-0,0334692	-	61
			0,00073088		0,00496955	
- ,30394	-0,00405786	-0,0113507	-	-0,0475123	-0,0049531	62
			0,00121698			
- ,39726	-0,00383233	-0,0433029	-	-0,0654488	-	63
			0,00125339		0,00266924	
- ,33355	-0,00298585	-0,00569699	-0,000251	-0,0249829	-	64
					0,00067482	
- ,21222	-0,00343935	-0,0161865	-	-0,073069	-	65
			0,00047659		0,00663549	
- ,34557	-0,00362084	-0,010484	-0,0008276	-0,0438238	-	66
					0,00336571	
- ,36753	-0,00367891	-0,0495704	-	-0,0726251	-	67
			0,00112035		0,00321228	
- ,28539	-0,00900102	-0,0129104	-	-0,0542163	-	68
			0,00248403		0,00438243	
- ,27782	-0,00108806	-0,0118744	-8,8592e-	-0,04404	-	69
			05		0,00477718	
- ,45983	-0,0049307	-0,00187451	-	-0,0286886	-	70
			0,00721991		0,00064267	
- ,57690	-0,0051934	-0,00341591	-	-0,0230236	-	71
			0,00158291		0,00067955	
- ,45598	-0,00480576	-0,00135845	-	-0,0217771	-0,0006523	72
			0,00746001			
- ,37450	-0,00806209	-0,00941984	-	-0,0389423	-0,0019723	73
			0,00360768			
- ,31670	-0,0024103	-0,0140035	-	-0,0452005	-	74
			0,00015534		0,00317976	
- ,55312	-0,00450883	-0,00131596	-	-0,0216427	-	75
			0,00361337		0,00135792	
- ,41831	-0,00508036	-0,0045592	-	-0,0211412	-	76
			0,00121369		0,00402948	
- ,59206	-0,00348769	-0,00336405	-	-0,0204439	-	77
			0,00038904		0,00382512	
- ,58159	-0,00212603	-0,00283729	-	-0,0165412	-	78
			0,00010166		0,00026386	
- ,47579	-0,00397421	-0,00141218	-	-0,0205448	-	79

			0,00318638		0,00166826	
- ,55040	-0,00156912	-0,00867371	-	-0,0538408	-	80
			0,00207176		0,00024736	
- ,50777	-0,00124878	-0,00256031	-	-0,0218174	-	81
			0,00165868		0,00159596	
- ,39986	-0,00733369	-0,00909221	-	-0,0345361	-	82
			0,00279988		0,00278608	
- ,56788	-0,00196337	-0,00485051	-8,9137e-	-0,0227772	-	83
			05		0,00217067	
- ,56358	-0,00475671	-0,00152591	-	-0,0233578	-0,0017305	84
			0,00322644			
- ,39595	-0,00680024	-0,00862738	-	-0,0325296	-	85
			0,00253106		0,00339319	
- ,61311	-0,00221363	-0,00371651	-9,852e-05	-0,0183404	-	86
					0,00374229	
- ,57529	-0,00475111	-0,00125092	-	-0,0197262	-	87
			0,00320634		0,00156686	
- ,48719	-0,00130121	-0,00366754	-	-0,0267924	-	88
			0,00122069		0,00157306	
- ,58945	-0,00134293	-0,00293487	-0,0011571	-0,0211526	-	89
					0,00234217	
- ,54113	-0,000851379	-0,00339403	-	-0,0174797	-	90
			0,00083371		0,00230926	
- ,39207	-0,00125431	-0,00550152	-	-0,0203657	-	91
			0,00046091		0,00708484	
L-H	FLC1	FHC3	FLC4	FHC6		No.
7,94582	-0,033862	-0,00239003	-	-0,0021334		50
			0,00667423			
7,76667	-0,0662154	-0,00623451	-0,01796	-0,00440648		51
7,71363	-0,0656543	-0,00697148	-0,0227074	-0,00545211		52
7,72063	-0,0615564	-0,00696011	-0,017972	-0,0059939		53
7,87908	-0,0428244	-0,00246187	-	-0,00205242		54
			0,00784868			
7,8052	-0,00321455	-0,00017502	-	-0,00025819		55
			0,00087556			
7,78583	-0,0521072	-0,00532227	-0,0203712	-0,0041794		56
7,73749	-0,043506	-0,00642488	-0,0124964	-0,00567356		57
7,98723	-0,00295576	-0,00054018	-0,03909	-0,00046603		58
7,96986	-0,0155114	-0,00185441	-	-0,0017893		59
			0,00349746			
7,83454	-0,0453736	-0,00566816	-0,0101239	-0,00449513		60
7,83622	-0,0355579	-0,00576561	-	-0,00476311		61
			0,00638906			
7,7971	-0,0505192	-0,00660977	-0,0118515	-0,00570258		62
7,72352	-0,0653914	-0,00727409	-0,0451362	-0,00622774		63
8,04751	-0,0169434	-0,00648524	-	-0,00704122		64
			0,00190291			
7,80185	-0,078649	-0,00477334	-0,0203821	-0,00358549		65
7,77188	-0,0424166	-0,006149	-0,0104542	-0,00515938		66
7,68685	-0,0722983	-0,0068572	-0,0521508	-0,0057361		67
7,82408	-0,0418921	-0,00965284	-	-0,00477104		68

			0,00972921		
7,85036	-0,0447502	-0,00769218	-0,0093544	-0,00436537	69
7,5437	-0,0308033	-0,0132869	-	-0,0194454	70
			0,00283185		
7,70892	-0,0273392	-0,0113416	-	-0,0102835	71
			0,00560123		
7,53296	-0,0216764	-0,0132593	-	-0,0193528	72
			0,00177537		
7,79702	-0,0282441	-0,00808906	-	-0,0109031	73
			0,00450299		
7,84682	-0,0625535	-0,00605884	-0,0172926	-0,00375184	74
7,61996	-0,0213408	-0,0117861	-	-0,014716	75
			0,00220115		
7,85626	-0,0198281	-0,00593595	-	-0,00623834	76
			0,00350013		
7,76554	-0,0217042	-0,00930935	-	-0,00739424	77
			0,00453645		
7,80758	-0,0230898	-0,0122005	-	-0,00825784	78
			0,00478769		
7,79596	-0,0191807	-0,0104752	-	-0,0129368	79
			0,00196636		
7,644	-0,0648343	-0,0164692	-0,0139964	-0,0150515	80
7,65003	-0,0247104	-0,0143975	-	-0,0126707	81
			0,00373086		
7,82016	-0,0243192	-0,00711396	-	-0,00921882	82
			0,00365178		
7,78991	-0,029319	-0,0107448	-0,0073511	-0,00715217	83
7,63434	-0,0266484	-0,0112655	-	-0,0137073	84
			0,00304805		
7,79663	-0,0229917	-0,00654648	-0,003289	-0,00842893	85
7,76551	-0,0249001	-0,00966002	-	-0,00638572	86
			0,00626287		
7,61174	-0,022401	-0,0111944	-	-0,0135759	87
			0,00249615		
7,83343	-0,0369205	-0,0147058	-	-0,0118154	88
			0,00753265		
7,61747	-0,0288338	-0,0140017	-0,0057813	-0,0112005	89
7,73491	-0,0190435	-0,0121297	-	-0,00931748	90
			0,00303805		
7,61796	-0,022444	-0,00537834	-	-0,00381421	91
			0,00460987		
FLC6	FLC7	FLN2	FLS1	FLC9	No.
-0,0765026	-0,170136	-0,183574	-0,636533	-0,450171	50
-0,102652	-0,15095	-0,0568018	-0,324621	-0,284662	51
-0,121262	-0,148052	-0,043049	-0,288914	-0,26811	52
-0,106215	-0,139846	-0,0462619	-0,289077	-0,260096	53
-0,107196	-0,208461	-0,152689	-0,629475	-0,459929	54
-	-0,107587	-0,186683	-0,545716	-0,338753	55
0,00303397					
-0,0935105	-0,143751	-0,0811818	-0,382009	-0,311783	56
-0,0702968	-0,118825	-0,0682697	-0,319053	-0,268482	57

-0,584826	-0,10818	-0,407963	-0,669352	-0,726024	58
-0,0167599	-0,155927	-0,2315	-0,711269	-0,445154	59
-0,0560565	-0,187956	-0,174936	-0,640771	-0,456189	60
-0,0399778	-0,144547	-0,140222	-0,505509	-0,357791	61
-0,0643821	-0,192551	-0,169214	-0,632852	-0,457699	62
-0,146248	-0,150732	-0,0609535	-0,338944	-0,295313	63
-0,0212143	-0,148333	-0,204106	-0,5676	-0,424359	64
-0,103424	-0,280661	-0,210687	-0,864179	-0,616615	65
-0,0581115	-0,156119	-0,136781	-0,518237	-0,37551	66
-0,16622	-0,157718	-0,0587732	-0,343871	-0,304362	67
-0,0627478	-0,201818	-0,185905	-0,69137	-0,487726	68
-0,0562648	-0,198262	-0,203374	-0,719074	-0,4996	69
-0,021922	-0,128837	-0,151734	-0,417704	-0,268786	70
-0,0256429	-0,0961933	-0,111341	-0,310882	-0,219222	71
-0,0148175	-0,105923	-0,163428	-0,40918	-0,271725	72
-0,0336267	-0,127259	-0,200363	-0,488166	-0,327567	73
-0,0801244	-0,193153	-0,148972	-0,590559	-0,435033	74
-0,0159721	-0,0904169	-0,132225	-0,333396	-0,226432	75
-0,0218434	-0,0902178	-0,1938	-0,423253	-0,281222	76
-0,021812	-0,0836537	-0,112722	-0,294364	-0,211754	77
-0,0203083	-0,0858804	-0,122258	-0,313076	-0,230375	78
-0,0146596	-0,097894	-0,157405	-0,386105	-0,251189	79
-0,0748003	-0,174017	-0,0536219	-0,305798	-0,245016	80
-0,0214859	-0,101776	-0,135531	-0,360346	-0,242988	81
-0,0297212	-0,108225	-0,193754	-0,448504	-0,302961	82
-0,0309254	-0,103088	-0,109503	-0,320453	-0,226173	83
-0,0198431	-0,0975977	-0,120399	-0,325556	-0,214967	84
-0,0275813	-0,102927	-0,192905	-0,438578	-0,29262	85
-0,0250518	-0,085981	-0,10299	-0,286663	-0,201776	86
-0,0163006	-0,0850269	-0,122862	-0,309607	-0,212347	87
-0,0341967	-0,116876	-0,126974	-0,369088	-0,262163	88
-0,0265595	-0,0910742	-0,10543	-0,296339	-0,206435	89
-0,0178269	-0,104879	-0,129968	-0,352802	-0,227457	90
-0,0272972	-0,10187	-0,122176	-0,396682	-0,268062	91
FHC7	FHN2	FHS1	FHC9	FHN3	No.
-0,0187098	-0,00298187	-0,0125372	-0,0207382	-0,0216932	50
-0,0180517	-0,00121621	-0,0108734	-0,0219989	-0,0220395	51
-0,0172457	-0,000984976	-0,0102916	-0,0214451	-0,0218706	52
-0,0169701	-0,000923212	-0,0102111	-0,0212869	-0,0220041	53
-0,0172906	-0,00341609	-0,0116139	-0,0212432	-0,0219158	54
-0,0184712	-0,00444374	-0,0137965	-0,0206701	-0,02214	55
-0,0172791	-0,00166995	-0,0110157	-0,0211303	-0,0216691	56
-0,0168529	-0,0010143	-0,0105758	-0,0205796	-0,0214983	57
-	-0,00912867	-0,0270888	-0,0165431	-0,0235571	58
0,00574255					
-0,0181822	-0,00319863	-0,0125162	-0,0209461	-0,022315	59
-0,0179604	-0,00151114	-0,0114354	-0,0218075	-0,0226872	60
-0,0176757	-0,00141169	-0,0112626	-0,0214543	-0,0224729	61
-0,0175429	-0,00120964	-0,0110641	-0,0214098	-0,0223895	62
-0,0163782	-0,00105166	-0,010125	-0,0202695	-0,0211421	63
-0,0194708	-0,000399214	-0,0131971	-0,0207772	-0,0208704	64

-0,017332	-0,00231586	-0,0107373	-0,0220164	-0,0228336	65
-0,0170513	-0,00153416	-0,0106995	-0,0212929	-0,0224433	66
-0,0162487	-0,0013536	-0,010067	-0,0203993	-0,0215396	67
-0,0161872	-0,00132853	-0,0102972	-0,0197377	-0,0208694	68
-0,0168373	-0,00157557	-0,0108785	-0,0201407	-0,0213461	69
-	-0,000161194	-0,00527142	-0,0123947	-0,0167038	70
0,00966802					
-0,0156851	-0,000224195	-0,00980061	-0,017246	-0,0232431	71
-0,0096768	-0,000176824	-0,00565872	-0,0119677	-0,016616	72
-0,0173394	-0,000318116	-0,010667	-0,017135	-0,0232607	73
-0,0179145	-0,00175942	-0,0112024	-0,0215868	-0,02218	74
-0,0133811	-3,02147e-05	-0,00802629	-0,0154836	-0,0214354	75
-0,0181462	-0,00139975	-0,0126004	-0,016784	-0,0245728	76
-0,0166147	-0,000619928	-0,0109504	-0,0174862	-0,0236762	77
-0,0163166	-0,00023628	-0,0106554	-0,0164607	-0,0220243	78
-0,0135015	-0,000272073	-0,00848091	-0,0154583	-0,0216484	79
-0,0124184	-0,000223344	-0,00670531	-0,0143964	-0,0159632	80
-0,0122173	-0,000100402	-0,00762544	-0,0137942	-0,0189987	81
-0,0177247	-0,000621389	-0,0114469	-0,0170953	-0,0242001	82
-0,0167749	-0,000574772	-0,0109897	-0,0174556	-0,0227242	83
-0,0138584	-7,22706e-05	-0,00822088	-0,0162811	-0,022424	84
-0,0174656	-0,000899907	-0,0114891	-0,0169428	-0,0246157	85
-0,0161469	-0,000934718	-0,0107747	-0,0169062	-0,0234221	86
-0,0139985	-5,33622e-05	-0,00843454	-0,0162599	-0,0220247	87
-0,0127275	-7,07495e-05	-0,00766388	-0,0143319	-0,0192118	88
-0,0129795	-0,000191989	-0,00790471	-0,0147209	-0,0203646	89
-0,0144481	-0,000895298	-0,00975877	-0,0105997	-0,0209725	90
-0,0166014	-0,00294868	-0,010379	-0,01545	-0,022453	91
FLN3	FHC10	FLC10	FHN4	FLN4	No.
-0,70498	-0,0178685	-0,0335241	-	-0,598917	50
			0,00581387		
-0,520112	-0,013669	-0,01286	-	-0,274765	51
			0,00471525		
-0,501806	-0,0128787	-0,00950719	-0,0044213	-0,238962	52
-0,477097	-0,0125793	-0,00986298	-	-0,242847	53
			0,00436623		
-0,754636	-0,0193989	-0,0302209	-	-0,578123	54
			0,00535773		
-0,458567	-0,021072	-0,0483852	-	-0,541949	55
			0,00640832		
-0,530106	-0,0147087	-0,0179985	-	-0,339123	56
			0,00505556		
-0,446976	-0,0130785	-0,0166497	-	-0,00772257	57
			0,00159534		
-0,603117	-0,00154508	-0,191896	-0,0263366	-0,210392	58
-0,637463	-0,0184598	-0,0383161	-	-0,683499	59
			0,00586199		
-0,721973	-0,014539	-0,0330265	-	-0,598847	60
			0,00504917		
-0,560981	-0,0141173	-0,0266496	-	-0,475031	61
			0,00495325		

-0,73147	-0,013458	-0,0328233	-0,0047803	-0,587787	62
-0,535355	-0,0123597	-0,0128124	-	-0,29084	63
			0,00435136		
-0,621478	-0,00808849	-0,115921	-	-0,589462	64
			0,00522167		
-1,0115	-0,0175219	-0,0329002	-	-0,788323	65
			0,00452465		
-0,598258	-0,0145846	-0,0302092	-	-0,500475	66
			0,00456376		
-0,556072	-0,0136544	-0,0138332	-	-0,300248	67
			0,00421954		
-0,773191	-0,0129207	-0,0354378	-	-0,643744	68
			0,00441964		
-0,780642	-0,0138466	-0,0387478	-	-0,680765	69
			0,00475505		
-0,482335	-0,00240621	-0,051544	-	-0,406998	70
			0,00175741		
-0,396154	-0,00641124	-0,0378629	-	-0,033298	71
			0,00241755		
-0,452974	-0,00241758	-0,0621418	-	-0,439206	72
			0,00197844		
-0,596067	-0,00676694	-0,0753087	-	-0,523223	73
			0,00416793		
-0,721039	-0,0149883	-0,0274262	-0,0050661	-0,539593	74
-0,390816	-0,00436593	-0,0477531	-	-0,374398	75
			0,00329752		
-0,490837	-0,00994971	-0,0731709	-	-0,515584	76
			0,00552592		
-0,370631	-0,00802435	-0,0454488	-	-0,333586	77
			0,00482683		
-0,394316	-0,00595472	-0,049109	-	-0,337996	78
			0,00472504		
-0,428914	-0,00590698	-0,0563053	-	-0,0407989	79
			0,00255748		
-0,501486	-0,00270022	-0,0259787	-	-0,133921	80
			0,00124877		
-0,43	-0,00460098	-0,0469086	-	-0,380097	81
			0,00291063		
-0,546874	-0,00766349	-0,0698637	-	-0,517844	82
			0,00487164		
-0,412769	-0,00782307	-0,0384817	-	-0,317389	83
			0,00448067		
-0,391208	-0,00519639	-0,0392848	-	-0,35451	84
			0,00335259		
-0,525609	-0,00871232	-0,07264	-	-0,526589	85
			0,00491597		
-0,36462	-0,00904747	-0,036016	-	-0,312304	86
			0,00470945		
-0,368407	-0,00500934	-0,0481564	-	-0,0276408	87
			0,00205448		
-0,477867	-0,00477656	-0,0435027	-	-0,0349001	88

			0,00165161		
-0,377713	-0,00562151	-0,0363452	-	-0,321249	89
			0,00325595		
-0,416296	-0,00666791	-0,0415832	-	-0,0311931	90
			0,00218334		
-0,418237	-0,0165335	-0,0319234	-	-0,453845	91
			0,00387504		

5.2 Modelos finais com todos os parâmetros selecionados com PCA.

5.2.1. Modelo HCA/PCA final para atividades *in vitro* dos 4-fenil-2-guanidino-tiazóis.

Os resultados dos testes farmacológicos laboratoriais foram obtidos com H⁺,K⁺-ATPase isolada de mucosa gástrica canina. A determinação da atividade enzimática foi feita da maneira descrita por Beil^{lv} e colaboradores, com pequenas modificações. Foi determinado o percentual de inibição da atividade enzimática com uma concentração fixa da droga de 50 µM. Um composto foi considerado de baixa atividade quando o percentual de inibição na dosagem indicada era β 39%.

O modelo final obtido para o conjunto dos compostos 4-fenil-2-guanidino-tiazóis usando todas as variáveis selecionadas na análise exploratória dos dados em cada modelo

parcial, e com as cargas de Mulliken

selecionadas, mostrou os seguintes resultados na comparação com as respostas obtidas experimentalmente para inibição *in vitro* da atividade da bomba de prótons H⁺,K⁺-ATPase:

O dendograma obtido com HCA dos dados autoescalados (fig.36), com conexão simples entre os compostos mostra os compostos com alta atividade agrupados em dois conjuntos. O maior tem 24 componentes com similaridade maior que 0,050UV, a uma distância de

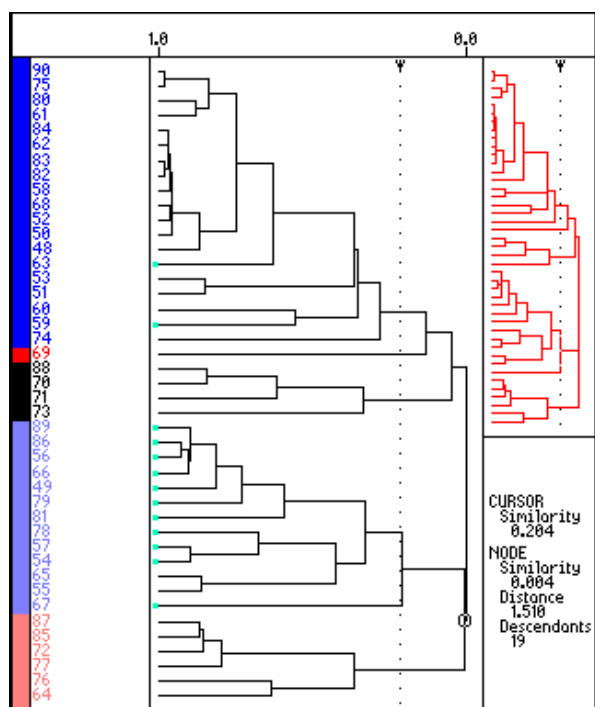


Figura 36: Dendograma mostrando as conexões entre os compostos com os parâmetros selecionados .

1,513UV da conexão seguinte, com a outra parte do grupo de alta atividade. É mostrado em azul escuro, vermelho e preto na parte superior do dendograma. Neste grupo estão os compostos 63, 59, e 82, com inibição reportada de 25% , 27%, e 26% respectivamente, e que não foram corretamente classificados no conjunto de baixa atividade ao qual pertencem. O segundo conjunto contém o subgrupo de baixa atividade. Há 6 compostos de alta atividade agrupados antes da conexão com o subgrupo de baixa atividade, mostrados em vermelho claro na parte inferior do dendograma. As atividades reportadas para estes compostos variam bastante, havendo compostos próximos da maior atividade observada, assim como do limite inferior (39%) para ser considerado de alta atividade. Finalmente, é mostrada a separação do grupo de compostos de baixa atividade. Este conjunto tem 13 elementos com similaridade maior que 0,207UV a uma distância de 1,510UV do centro de conexão com os compostos mais ativos. Os compostos 54 e 89 foram erradamente classificados neste conjunto no dendograma. Estes compostos tem atividade inibidora da H^+, K^+ -ATPase reportada de 99% e 111%, medidas com dosagem de 50 μM da droga *in vitro*, e estão situados no extremo superior da atividade. A má classificação destes compostos é um grande equívoco do modelo, e pode ser causada justamente por se situarem no extremo da escala. Os compostos 55 e 65 não foram corretamente classificados na PCA, e aparecem corretamente classificados aqui no dendograma. Os compostos marcados com um sinal verde são os classificados no grupo de baixa atividade com PCA, usando 3 componentes principais para definir o espaço vetorial correspondente.

A PCA foi obtida com os dados autoescalados, em 5 PCs. Com 3PCs (fig.37) podem ser separados 11 compostos de baixa atividade (vermelho forte), agrupados com 2 compostos de alta atividade classificados errado (azul forte), dentro de um subespaço vetorial marcado graficamente no diagrama com as 3 PCs; e um classificado errado na região de alta atividade. Fazendo a definição numérica do valor de corte para cada PC, o

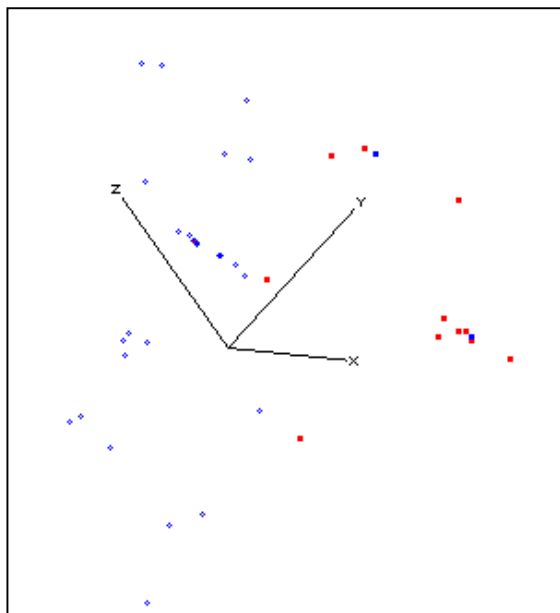


Figura 37: Gráfico das três primeiras PCs. Os compostos mais ativos são mostrados em vermelho são os de

número de classificações erradas sobe para 4 pouco ativos classificados errado. A definição numérica foi feita da seguinte forma: (1) as coordenadas máxima e mínima nos 'scores' nas 3 primeiras PCs dos compostos classificados graficamente no grupo de baixa atividade foram consideradas, (2) os compostos com alta atividade que tinham coordenadas dentro da faixa de valores considerada, foram considerados classificações erradas. A grosso modo foi definido um cubo no espaço tridimensional onde só deveriam haver compostos de baixa atividade. As coordenadas das

arestas tais que $(-1,1567) \leq PC1 \leq (3,40735)$; $(-0,1234) \leq PC2 \leq (1,57318)$; e $(-1,4075) \leq PC3 \leq (0,67376)$. Graficamente também é definido um cubo, porém suas arestas não são necessariamente paralelas aos eixos das componentes principais, e uma delas estende se de $-\infty$ a $+\infty$. As três PCs consideradas acumulam 95% da variância total dos dados autoescalados, e seus autovetores são mostrados na tabela 14.

Tabela 14: Autovetores das PCs para o modelo PCA final com atividades *in vitro* e cargas de Mulliken.

↓Variável PC#→	PC1	PC2	PC3
Q_{N3}^{Mulliken}	0,18859	0,60791	-0,4244
Q_{N4}^{Mulliken}	0,18748	-0,6372	0,24207
Q_{C9}^{Mulliken}	0,19044	0,46028	0,85116
α	0,55201	-0,0124	-0,0400
Volume vdW	0,53413	-0,1048	-0,1874
Volume vdW R1	0,55047	-0,0362	-0,0095
Variância acumulada	50,47%	34,36%	10,37%

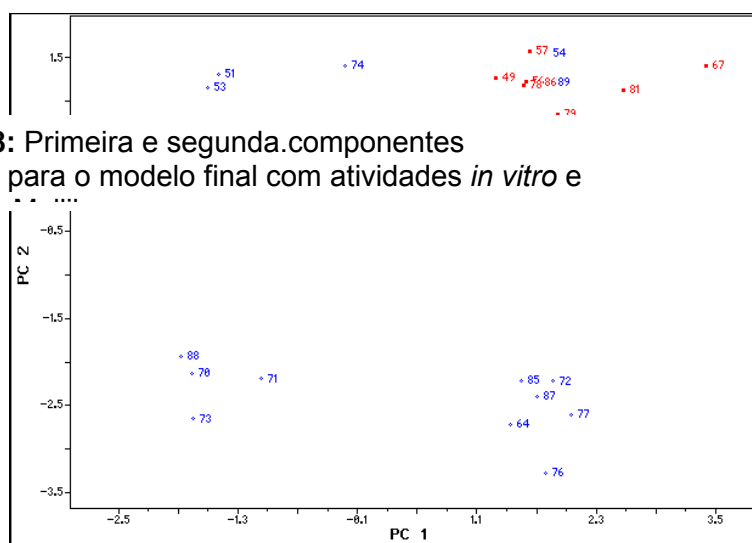
A primeira PC acumula 52% da variância total do sistema. Pode ser observado na tabela dos 'loadings' das variáveis que a composição desta componente principal é distribuída

principalmente entre a polarizabilidade, os volumes de van der Waals da molécula e do grupo R1, e a carga obtida como derivada do potencial eletrostático no nitrogênio 3. O aumento da polarizabilidade é um fator que separa os compostos com menor atividade na parte positiva do gráfico cartesiano. Os compostos 54, 55, 65, e 89 são pouco ativos classificados errado na primeira componente principal (fig.38), e têm a polarizabilidade calculada maior que 32.000, enquanto a maioria dos inativos têm o valor desta propriedade calculada por volta de 24.000. Um grupo volumoso (benzila) na posição R1

também caracteriza estes

compostos de baixa atividade classificados errado na primeira componente principal. Na Segunda componente principal as maiores contribuições vêm das cargas de Mulliken nas posições N3, N4, e C9. Os compostos menos ativos têm carga mais negativa nas posições

Figura 38: Primeira e segunda componentes principais para o modelo final com atividades *in vitro* e



N3 e C9, e carga menos negativa na posição N4 como contribuição para classificação correta nesta componente.

5.2.2. Modelo HCA/PCA final para atividades *in vitro* dos 4-fenil-2-guanidino-tiazóis.

O modelo final obtido para o conjunto dos compostos 4-fenil-2-guanidino-tiazóis usando todas as variáveis selecionadas na análise exploratória dos dados em cada modelo parcial para os tipos de variável, e as cargas derivadas do potencial eletrostático, mostrou os seguintes resultados na comparação com as respostas obtidas experimentalmente para inibição *in vitro* da atividade da bomba de prótons H^+, K^+ -ATPase:

O dendograma obtido com HCA (fig.38) dos dados autoescalados, e conexão simples entre os compostos, mostra dois conjuntos principais. O maior tem 22 componentes com similaridade maior que 0,532UV, a uma distância de 1,512UV da conexão seguinte, com a outra parte do grupo de alta atividade. Este é o grupo mostrado em azul na parte superior do dendograma. Neste grupo estão os compostos 63, 59, e 82, com inibição reportada de 25% , 27%, e 26% respectivamente, e que não foram corretamente classificados no conjunto de baixa atividade ao qual pertencem, nem na HCA nem na PCA.

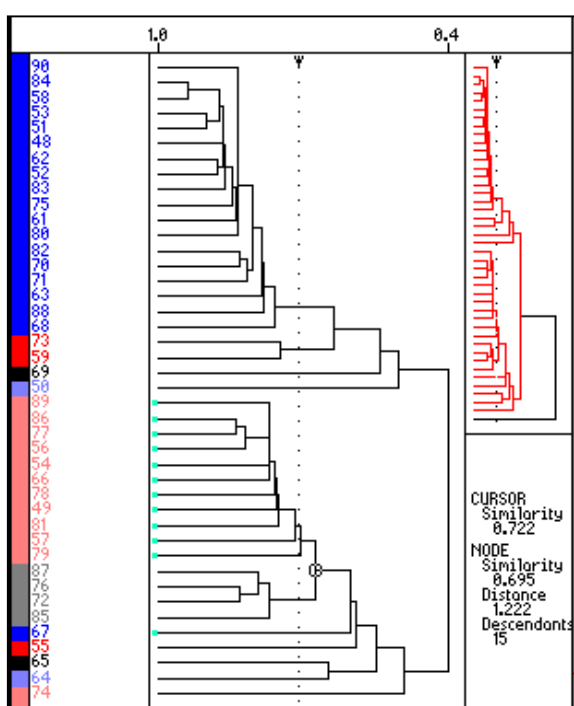


Figura 38: Dendograma mostrando o agrupamento dos compostos no modelo com as variáveis selecionadas

O segundo conjunto contém o subgrupo de baixa atividade e outros compostos de alta atividade não agrupados no primeiro grupo. Os 11 compostos com baixa atividade agrupados antes da conexão com o subgrupo de alta atividade são mostrados em vermelho claro no centro do dendograma, e têm similaridade maior que 0,722UV, ficando a uma distância de 1,222UV do último centro de conexão entre compostos de alta atividade. Os compostos 54, 77, e 89 são classificados errado neste grupo. A atividade inibidora da H^+, K^+ -ATPase reportada para estes compostos é respectivamente 99%,

85%, e 111% (o composto 88 tem 100% de inibição), havendo valores no extremo superior da atividade observada, portanto.

Os demais compostos de alta atividade formam um grupo com 4 compostos de similaridade maior que

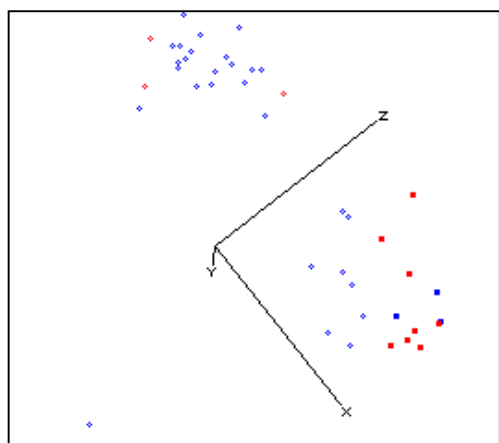


Figura 39: Gráfico das 3 primeiras PCs do modelo. Os compostos pouco ativos são

0,783UV, distantes 1,222UV do grupo de baixa atividade, restando mais 6 compostos não classificados nos grupos propostos, entre os quais um compostos de baixa atividade (67), corretamente classificado na análise por componentes principais.

A PCA foi realizada com os dados autoescalados, em 5 PCs. Com 3 PCs podem ser separados 12 compostos de baixa atividade, agrupados com 3 compostos de alta atividade classificados errado, e 3 classificado errado na região de alta atividade (fig.39), dentro de um subespaço vetorial marcado graficamente no diagrama com as 3 PCs. Os compostos 57, 74, e 89 foram os mais ativos classificados errado, ainda que o composto 89 possa ser excluído do grupo de baixa atividade se o cubo que inscreve este grupo for cuidadosamente demarcado. Os classificados errado de baixa atividade, 59, 63, e 82, estão completamente misturados no grupo de alta atividade, não havendo qualquer variável que os diferencie dos mais ativos. Substituindo se a terceira PC pela quarta os resultados continuam os mesmos, com as mesmas classificações erradas. A substituição pela quinta PC aumenta uma classificação errada, o composto 55. As 3 PCs consideradas acumulam quase 90% da variância total dos dados autoescalados.

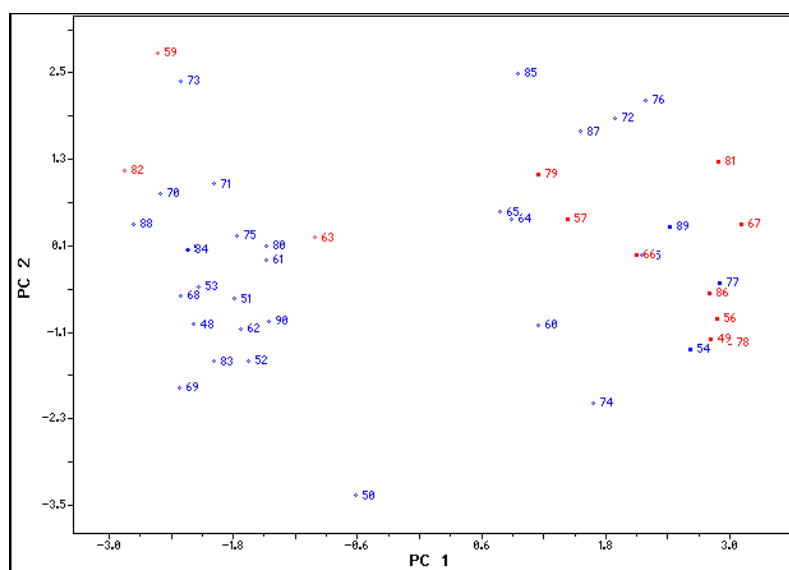


Figura 40: Primeira e segunda PCs para o modelo com as variáveis selecionadas e cargas CHELP

A primeira PC (fig.40) acumula quase 58% da variância total dos dados autoescalados. A análise dos 'loadings' desta primeira PC mostra contribuição significativa das variáveis Q_{N3}^{CHELP} , Q_{N4}^{CHELP} , Q_{C7}^{CHELP} , Q_{R2}^{CHELP} , α , vdWV, e R1vdWV principalmente (tab.15). A contribuição da carga derivada do potencial eletrostático sobre o

Nitrogênio 3 mostra que a classificação correta no grupo de baixa atividade acontece nos

compostos em que a carga na posição é menos negativa. Os compostos de alta atividade pior classificados foram 59, 63, e 82, que têm a carga nesta posição mais negativa, com valor semelhante aos pouco ativos. O mesmo se aplica a carga na posição do Nitrogênio 4, inclusive quanto as classificações erradas. As cargas sobre os carbonos 7 e 11 dos compostos pouco ativos bem classificados é menos positiva, e as más classificações correspondem às exceções a regra. As polarizabilidades e os volumes de van der Waals da molécula e do substituinte R1 são os mesmos utilizados no modelo anterior e sua influência já foi discutida. A segunda PC acumula 21% da variância total dos dados autoescalados, e as principais contribuições são Q_{C6}^{CHELP} , e Q_{C7}^{CHELP} . A carga na posição C6 tem contribuição de tal modo que valores mais negativos corresponde a um 'score' mais alto para o composto na segunda PC. A carga na posição C7 contribui com aumento diretamente proporcional no 'score' do composto na segunda PC. Esta PC apenas não tem nenhuma capacidade de modelar a separação entre grupos de atividade diferente, mais é importante principalmente na classificação correta dos compostos 54, 55, 60, e 74, muito ativos e que ficam bem distantes se considerada esta PC.

A terceira PC tem papel secundário, com 10% da variância acumulada, e contribuição principal das variáveis Q_{N4}^{CHELP} , Q_{R2}^{CHELP} , Q_{C6}^{CHELP} , α , e vdWV.

Tabela 15: Autovetores das componentes principais, e as variâncias acumuladas

↓Variável	PC#→	PC 1	PC 2	PC 3
Q_{N3}^{CHELP}		0,40323	-0,0148	0,07118
Q_{N4}^{CHELP}		0,40288	0,23128	0,31220
Q_{C6}^{CHELP}		0,37352	0,33421	0,29705
Q_{C7}^{CHELP}		0,41576	0,19675	0,19745
Q_{R2}^{CHELP}		0,37220	-0,0506	-0,5569
α		0,13649	-0,6661	0,32050

VdWV	-0,2573	0,58426	-0,1311
R1vdWV	-0,3733	0,09030	0,58325
Variância acumulada	57,65%	21,51%	10,68%

5.2.3. Modelo final da atividade *in vivo* dos 4-fenil-guanidino-tiazóis

O modelo obtido para os para relacionar as atividades *in vivo* com os parâmetros calculados é mostrado a seguir. Os dados de atividade foram medidos como a concentração μM necessária à inibição de 50% da atividade enzimática. Esta determinação é disponível para 23 compostos desta série.

A HCA foi realizada com os dados autoescalados, utilizando conexão simples entre as amostras. Um composto foi considerado com alta atividade quando sua $\text{IC}_{50} \leq 15 \mu\text{M}$.

O dendograma separa o grupo menos ativo com 14 compostos, e quatro classificações erradas, marcado em vermelho no centro do dendograma. Os compostos 71, 84, 88, e 90 são compostos com alta atividade classificados errado neste conjunto, com IC_{50} de 14, 15; 1,5; e 1,6 μM respectivamente. A similaridade entre os elementos é maior que 0,197UV,

distando 1,725UV da conexão com os demais compostos. Os compostos marcados com um ponto verde são os classificados no grupo de alta atividade em PCA (71, e 90 corretamente classificados com PCA).

Os compostos de alta atividade não estão agrupados propriamente. No canto superior esquerdo do dendograma, marcados em azul, foram classificados 7 compostos com alta

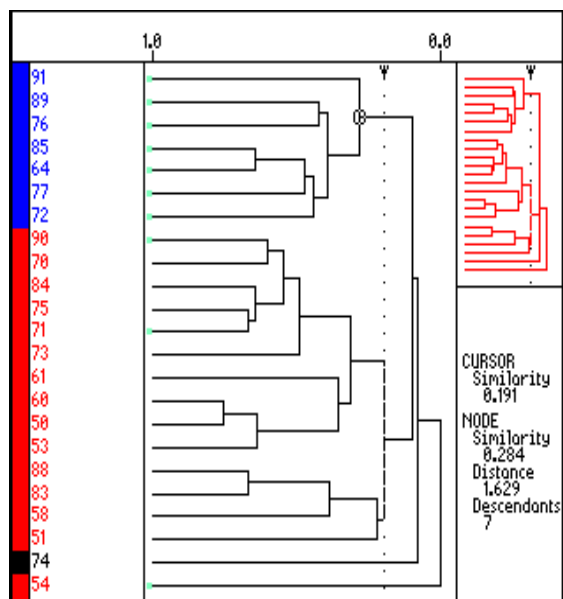
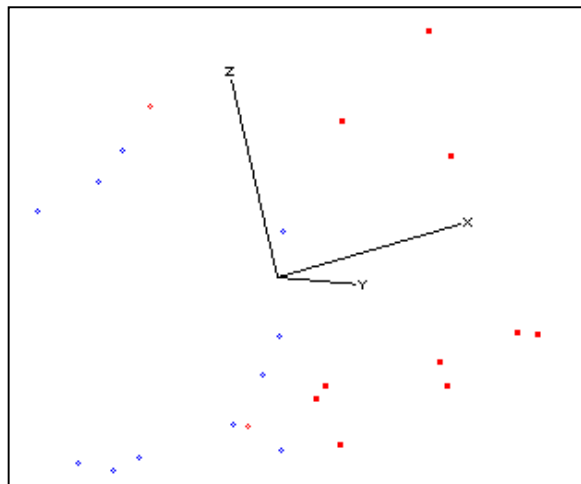


Figura 41: Dendograma obtido da HCA final para as atividades *in vivo* dos compostos guanidino-tiazóis.

atividade, agrupados com mais de 0,284UV de similaridade, a uma distância de 1,629UV da conexão com o grupo de baixa atividade. Aqui não houve classificação errada.

Ficam isolados dos grupos de alta ou baixa atividade o composto 74, de baixa atividade,



que aparece em preto no canto inferior esquerdo do dendograma. Está a 1,845UV de distância da conexão entre os grupos definidos de alta e baixa atividade. O composto 54, em vermelho, dista 1,925UV da conexão entre 74 e o centro de conexão com o grupo menos ativo.

Figura 42: Gráfico das 3 primeiras PCs para a PCA final com atividades *in vivo* dos guanidinothiazóis

A análise PCA foi feita com autoescalamento dos dados, em 5 PCs. Para este conjunto de 23 compostos 12 mais ativos e 11 menos ativos, foram usadas as 3 primeiras PCs (PC1 $\Rightarrow$ X, PC2 $\Rightarrow$ Y, PC3 $\Rightarrow$ Z), totalizando cerca de 85% da variância total dos dados autoescalados. A substituição da 3ª PC pela 4ª não muda o resultado, e pela 5ª aumenta uma classificação errada. O grupo mais ativo com 10 compostos, todos classificados certo com 3 PCs (fig.42) tem os compostos 71 e 90 classificados errado na primeira PC (fig.43). A primeira PC acumula 37% da variância total dos dados autoescalados (tab.16). É uma

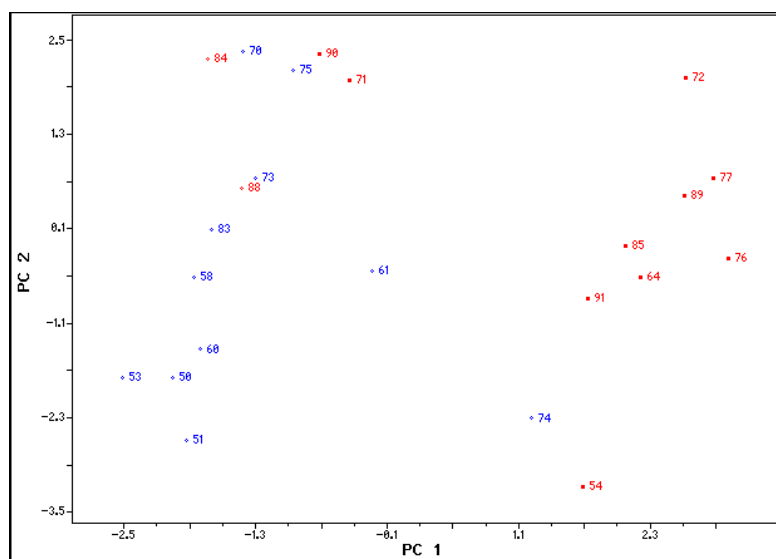


Figura 43: Gráfico das duas primeiras PCs do modelo PCA final com atividades *in vivo*

combinação dos volumes de van der Waals das moléculas, do substituinte R1, da área disponível para o solvente na superfície de van der Waals, e da polarizabilidade calculada, principalmente. O valor da polarizabilidade dos compostos bem classificados nesta 1ª PC é alto, maior que 32,000, enquanto os mau classificados têm valores

ate 26,600 aproximadamente (composto 71). Os volumes dos compostos bem classificados são maiores que 295 Å³, e a área acessível para o solvente maior que 324 Å² (composto 85). O volume do substituinte R1 dos bem classificados corresponde ao grupo benzila nesta posição.

A 2ª PC acumula 28% da variância dos dados, e não cria 2 classes para os compostos mais ou menos ativos. Sua maior contribuição é auxiliar na separação dos compostos 71 e 90 que têm alta atividade e não foram bem classificados somente com a primeira PC, e dos compostos 61 e 74 que têm baixa atividade e também não foram bem classificados

somente com a primeira PC. Sua composição tem grande influência dos índices de densidade de orbitais de fronteira, normalizados pela energia do orbital correspondente (HOMO ou LUMO), nas posições C5, C9, e C10; O momento de dipolo calculado com as orbitais híbridas é outro fator importante, seguido pelas cargas calculadas na posição N3, e R2. A densidade orbital em LUMO sobre o carbono 5, em HOMO sobre o carbono 9 e em HOMO sobre o carbono 10 é menos negativa nos compostos com alta atividade bem classificados na segunda PC. Um valor baixo na componente nos orbitais híbridos do momento de dipolo aproxima os compostos do grupo de alta atividade. Uma carga mais positiva na posição R2, e mais negativa na posição N3 também ajudam na classificação correta dos compostos mais ativos pela 2ª PC. Na 3ª PC as principais contribuições são as cargas nas posições R2 e N3, da mesma maneira que na 2ª PC. As densidades de elétrons em HOMO e LUMO atuam de maneira inversa na 2ª e na 3ª PCs, ainda que sua atuação na 3ª PC seja pequena.

Tabela 16: Autovetores e variâncias acumuladas por cada PC no modelo final para atividades *in vivo*.

↓Variável	PC#→	PC 1	PC 2	PC 3
F_{C5}^{LUMO}		0,17492	0,46177	0,17318
F_{C9}^{HOMO}		0,07369	0,48690	0,18637
F_{C10}^{HOMO}		-0,0229	0,48564	0,17117
Q_{N3}^{CHELP}		0,04165	-0,1689	0,67662
Q_{R2}^{CHELP}		-0,0087	0,20221	-0,6617
$\mu^{híbrido}$		-0,0034	-0,4737	0,02820
α		0,49379	-0,0109	-0,0846
Volume vdW		0,50654	-0,0085	-0,0060
Área acessível ao solvente de vdW		0,48803	-0,1303	-0,0451
Volume vdW de R1		0,47227	-0,0529	-0,0150
Variância acumulada		37%	28%	20%

5.3. Modelos finais de regressão PLS com todas as variáveis selecionadas com PCA.

5.3.1. Modelo de regressão PLS para atividades *in vitro*.

Para os compostos 4-fenil 2-guanidinothiazóis foi construído um modelo de correlação quantitativa com as atividades *in vitro*. As variáveis usadas neste modelo foram Q_{N3}^{Mulliken} , Q_{N4}^{Mulliken} , Q_{C9}^{Mulliken} , α , vdWV , e R1vdWV , respectivamente (nas colunas). A seguir são

mostrados os comandos executados no MATLAB para análise dos dados:

```
» load am1_fim.mat
» am1_fim0
am1_fim0 =
```

•	0,3300	-0,3800	-0,1400	21,8195	183,9000	18,3200
•	0,2700	-0,3800	-0,1400	31,1784	273,4000	131,8800
•	0,3300	-0,3800	-0,1400	22,6049	197,7000	18,3200
•	0,3300	-0,3900	-0,1300	23,1656	199,0000	18,3200
•	0,3300	-0,3800	-0,1400	23,4521	199,0000	18,3200
•	0,3300	-0,3800	-0,1300	22,4181	190,8000	18,3200
•	0,2900	-0,3800	-0,1300	32,8028	290,8000	131,8800
•	0,3300	-0,3400	-0,1300	33,2900	295,1000	131,8800
•	0,2700	-0,3800	-0,1400	32,2269	293,6000	131,8800
•	0,2900	-0,3800	-0,1300	32,2844	279,0000	131,8800
•	0,3300	-0,3800	-0,1400	22,8810	202,2000	18,3200
•	0,3200	-0,3800	-0,1500	26,3259	244,6000	18,3200
•	0,3200	-0,3800	-0,1500	23,2545	199,9000	18,3200
•	0,3300	-0,3800	-0,1400	26,6694	244,7000	18,3200
•	0,3300	-0,3800	-0,1400	23,5759	204,0000	18,3200
•	0,3300	-0,3800	-0,1400	31,6972	255,5000	18,3200
•	0,3600	-0,2800	-0,1500	32,0625	304,3000	131,8800
•	0,3400	-0,3400	-0,1300	32,8602	288,7000	131,8800
•	0,2700	-0,3800	-0,1400	33,0913	310,2000	131,8800
•	0,2900	-0,3800	-0,1300	42,1141	353,5000	131,8800
•	0,3300	-0,3800	-0,1400	23,2381	195,0000	18,3200
•	0,3300	-0,3800	-0,1400	26,2517	228,3200	18,3200
•	0,3900	-0,3300	-0,1500	24,0035	220,8000	18,3200
•	0,3800	-0,3200	-0,1500	26,6246	252,9000	18,3200
•	0,3700	-0,2800	-0,1400	34,2347	301,8000	131,8800
•	0,3800	-0,3300	-0,1600	24,3254	241,4000	18,3200
•	0,3100	-0,3800	-0,1300	30,8774	228,8200	18,3200
•	0,3300	-0,3800	-0,1400	24,9693	231,9000	18,3200
•	0,3800	-0,2700	-0,1500	34,1975	323,6000	131,8800
•	0,3800	-0,2700	-0,1400	34,6882	316,3000	131,8800
•	0,3100	-0,3800	-0,1300	30,8774	306,1000	141,6200
•	0,2900	-0,3800	-0,1400	34,4674	311,9000	131,8800
•	0,3300	-0,3800	-0,1400	27,3243	257,5000	18,3200
•	0,2700	-0,3800	-0,1400	37,5912	335,2000	131,8800

- 0,3300 -0,3800 -0,1400 22,6912 206,2000 18,3200
- 0,3300 -0,3800 -0,1400 22,8992 207,2000 18,3200
- 0,3300 -0,3800 -0,1400 23,2203 205,2000 18,3200
- 0,3800 -0,2900 -0,1400 32,8822 295,2000 131,8800
- 0,2700 -0,3800 -0,1400 32,9490 298,1000 131,8800
- 0,3800 -0,2800 -0,1400 33,3815 300,9000 131,8800
- 0,3800 -0,3300 -0,1500 23,4393 210,6000 18,3200
- 0,2700 -0,3800 -0,1400 34,1889 298,1000 131,8800
- 0,3300 -0,3800 -0,1400 25,1892 232,2000 18,3200

```

» [fim0_ax,fim0_mx,fim0_stdv]=auto(am1_fim0);
» [vitro_y,vitro_my,vitro_stdy]=auto(vitro');
» [prs,cp,mlv,b,w,p,q,t21,tv]=plsconvrnf(fim0_ax,vitro_y,50,30,6,0,1);

```

O PRESS mínimo cumulativo é com 3 VLs. O modelo com 6 variáveis obtidas na seleção com PCA/HCA teve desempenho apenas razoável na predição dos valores de atividade biológica dos compostos. Os coeficientes de regressão mostram (fig.44) que há contribuição significativa de todas as variáveis selecionadas para obtenção do modelo. O percentual da variância capturada com o modelo PLS é mostrado na tabela 17. A variância acumulada para as variáveis dependentes, o bloco

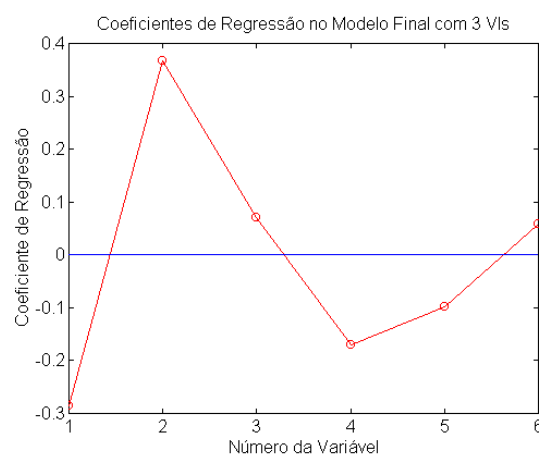


Figura 44: Valor do coeficiente de cada variável na matriz de regressão do modelo PLS.

Y, é pequena nestas três variáveis latentes.

Tabela 17: Variâncias acumuladas nas VLs

-----Bloco-X-----			-----Bloco-Y-----	
VL #	Esta VL	Total	Esta VL	Total
----	-----	-----	-----	-----
1	37.00	37.00	31.68	31.68
2	43.29	80.28	1.97	33.65
3	15.49	95.77	2.17	35.82

O limite de Q para 95 por cento de confiança é 0.963766. Os compostos 63, 67, e 74 têm o valor de Q obtido maior ou igual ao valor limite, e o composto 49 se aproxima bastante do limite (fig.45). Os compostos 63 e 67 têm em comum o fato de serem os únicos da série que têm o substituinte *p*-fenil na posição R3, além de pertencerem ao grupo de baixa atividade. A diferença entre eles é o substituinte R1, que é benzila no composto 67 e hidrogênio no composto 63. O composto 67 é tem o segundo menor valor de atividade *in vitro* medido para esta série de compostos, e o 49 o menor.

O composto 74 pertence ao grupo de alta atividade, situada na faixa intermediária dos valores para o seu grupo. Tem substituição *m*-metoxi na posição R3, e hidrogênio nas demais.

O limite de T^2 para 95 por cento de confiança é 8,94226. Os valores obtidos na exclusão de cada um dos compostos podem ser vistos na

figura 46.

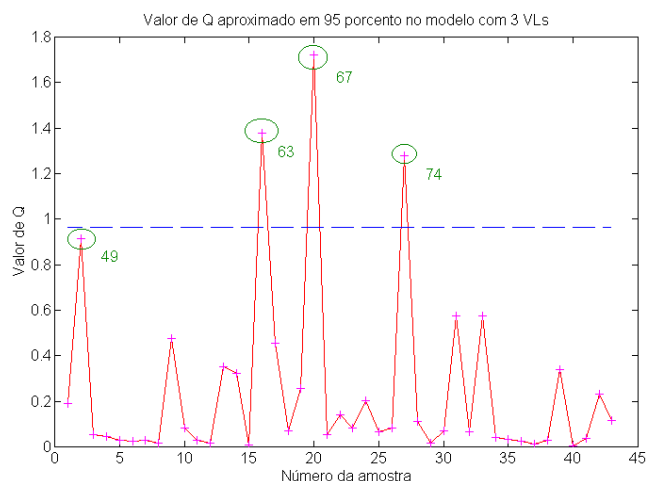


Figura 45: Valor de Q para a exclusão de cada um dos compostos no modelo PLS, e valor para limite de confiança de 95% no resultado

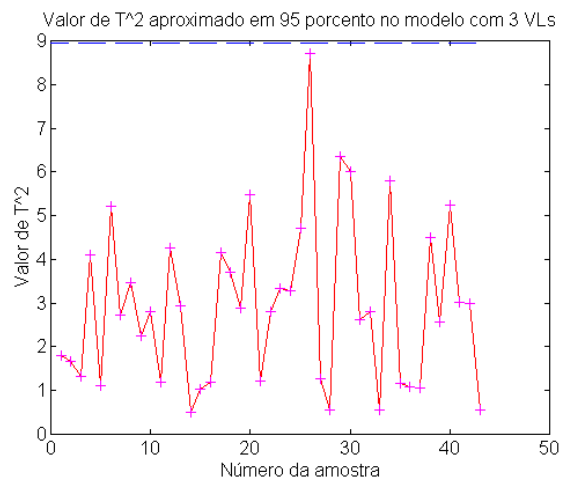


Figura 46: Valor de T^2 para a exclusão de cada um dos compostos no modelo PLS, e valor para limite de confiança de 95 no resultado

5.3.2. Regressão PLS para atividades *in vivo*.

O modelo PLS foi obtido com as variáveis selecionadas

com PCA. Neste caso F_{c5}^{LUMO} , F_{c9}^{HOMO} , F_{c10}^{HOMO} ,

Q_{N3}^{CHELP} , Q_{R2}^{CHELP} , $\mu^{hibrido}$, α , $vdWV$, $vdWSAV$, e

$vdWVR1$, na ordem em que aparecem no gráfico da figura

47.

O coeficiente de regressão mostra que as variáveis que

têm menor importância no ajuste do modelo aos dados

experimentais são a carga no substituinte R2, a

polarizabilidade e o volume de van der Waals acessível ao

solvente, variáveis 5, 7, e 9, e foram mantidas com o

intuito de comparar o desempenho dos modelos (PCA,

PLS, Neural) com o mesmo conjunto de variáveis

independentes. PRESS mínimo cumulativo é com 4

variáveis latentes, e o percentual da variância capturada

com o modelo PLS é mostrado na tabela 18:

Tabela 18: Variâncias acumuladas nas VLs

-----Bloco-X----- Bloco-Y-----

VL # Esta VL Total Esta VL Total

-----	-----	-----	-----	-----
1	34.70	34.70	46.11	46.11
2	20.21	54.90	10.58	56.70
3	12.38	67.28	9.85	66.55
4	21.12	88.40	2.35	68.90

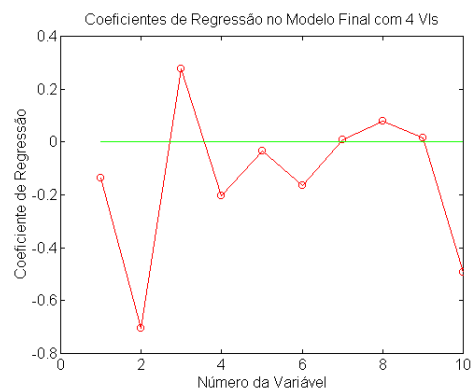


Figura 47: Valor do coeficiente de cada variável na matriz de regressão do modelo PLS.

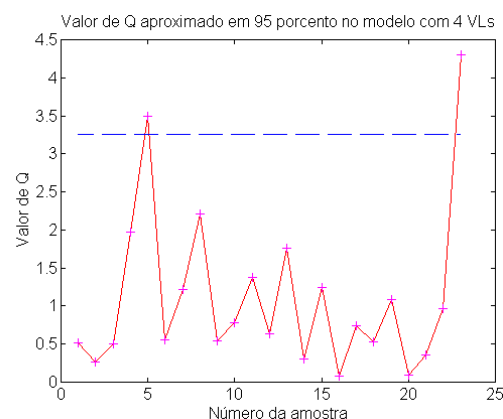


Figura 48: Valor de Q para a exclusão de cada um dos compostos no modelo PLS, e valor para limite de confiança de 95 no resultado

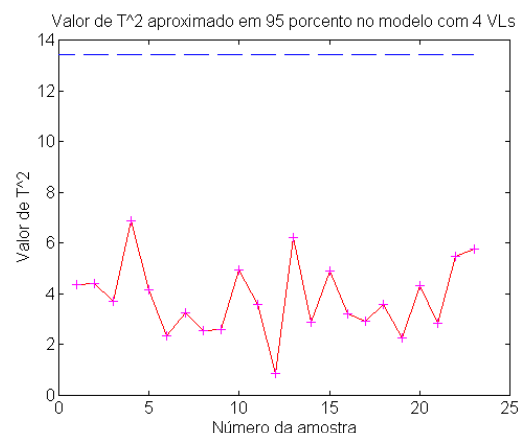


Figura 49: Valor de T^2 para a exclusão de cada um dos compostos no modelo PLS, e valor para limite de confiança de 95 no resultado

Com quatro variáveis latentes o modelo acumula 88% da variância total das variáveis independentes autoescaladas, e 69% da variância total do valor autoescalado de $\log IC_{50}$, que foi usado como variável dependente. No caso da regressão PLS pode haver mudança no método de escalamento das variáveis dependentes e independentes. O uso de quatro variáveis latentes é possível num conjunto de 23 amostras, ainda que no limite (5 a 6 amostras para cada variável independente correlacionada).

O limite de Q para 95 por cento de confiança é 3,26013. Apenas dois compostos ultrapassam este limite de erro na predição de seus valores de atividade, 54, e 91, a quinta e a vigésima terceira amostras respectivamente (fig.48). Têm em comum a presença de um grupo benzila na posição R1. Estes compostos estão situados no extremo superior da atividade biológica, e ainda que a predição dos seus valores não seja obtida por extrapolação, a retirada destes compostos do conjunto afeta bastante a média e dificulta a previsão.

O limite de T^2 para 95 por cento é 13,4075 e nenhum composto ultrapassa este limite (fig.49).

5.3. Redes neurais usando os parâmetros selecionados com PCA/HCA. Para comparação entre as metodologias foram realizados testes de treinamento de redes neurais para os dados obtidos. A partir da seleção de variáveis feito por PCA e HCA para cada tipo de variável, foram treinadas redes neurais para reproduzir o padrão encontrado nesta série de compostos. Os resultados obtidos para os modelos com rede neural para as variáveis consideradas e a atividade *in vivo* e *in vitro* serão mostrados a seguir..

5.3.1. Modelos de classificação das atividades *in vitro*, usando cargas de Mulliken.

O modelo inicialmente criado com o programa PSDD obtido do QCPE, foi obtido com uma rede de 6 neurônios na primeira camada, 10 neurônios na segunda, e dois na terceira, por retropropagação, sem utilização de bias, com os outros parâmetros definidos iguais aos

padrões internos do programa, e utilizados para todos os modelos de correlação com atividades *in vitro*. As variáveis utilizadas na primeira camada foram Q_{N3}^{Mulliken} , Q_{N4}^{Mulliken} , Q_{C9}^{Mulliken} , α , vdWV , e R1vdWV . O composto 82 foi classificado errado já no treinamento da rede neural como muito ativo (tab.19, classificação resultante).

Para previsão sucessiva de cada composto foi treinada uma rede com 20 neurônios na segunda camada, e limite na função de erro de 10^{-6} para cada iteração. Este modelo neural não foi capaz de prever acuradamente 10 compostos entre os 43 do conjunto todo, ou 23%, usando estas variáveis para treinamento da rede. Os compostos marcados em vermelho, 48, 54, 60, 69, e 89, são os que pertencem ao grupo de alta atividade e foram previstos no grupo de baixa atividade. Os compostos marcados em azul, 57, 59, 63, 78, e 82, pertencem ao grupo de baixa atividade e foram previstos no grupo de alta atividade (tab.19, classificação prevista 1).

Foi criado um outro modelo com 20 neurônios na segunda camada, e limite de convergência fixado em 0,01 para as mesmas variáveis do caso anterior, e os demais parâmetros de criação da rede também iguais.

Neste caso a diminuição do limite de convergência aplicado a cada ciclo melhorou a capacidade de predição do modelo. Foram classificados errado nove compostos do conjunto de 43, ou seja 21% dos compostos foram previstos erradamente (tab.19, classificação prevista 2).

Tabela 19: Resultados obtidos com redes neurais para as atividades *in vitro* dos 4-fenil 2-guanidinothiazóis com os dados selecionados com PCA/HCA, e cargas de Mulliken. Os resultados mostrados em vermelho são para os compostos muito ativos (classe 1) erradamente classificados. Os mostrados em azul são os pouco ativos (classe 2) classificados errado.

Padrões de Treinamento			Classificação o Resultante		Classificação o Prevista 1		Classificação Prevista 2	
Num.	Classe	Padrão						
48	1	1 0	,996	,004	,252	,748	,927	,074
49	2	0 1	,000	1,000	,000	1,000	,000	1,000

50	1	1 0	1,000	,000	,691	,309	1,000	,000
51	1	1 0	1,000	,000	,899	,101	1,000	,000
52	1	1 0	1,000	,000	,953	,047	1,000	,000
53	1	1 0	1,000	,000	,985	,015	1,000	,000
54	1	1 0	,883	,116	,001	,999	,000	1,000
55	1	1 0	1,000	,000	,980	,020	1,000	,000
56	2	0 1	,004	,996	,009	,991	,008	,992
57	2	0 1	,114	,886	,950	,050	,989	,011
58	1	1 0	1,000	,000	,668	,332	1,000	,000
59	2	0 1	,034	,966	1,000	,000	,951	,048
60	1	1 0	,983	,017	,050	,950	,021	,980
61	1	1 0	1,000	,000	1,000	,000	1,000	,000
62	1	1 0	1,000	,000	,945	,055	1,000	,000
63	2	0 1	,040	,960	,994	,006	1,000	,000
64	1	1 0	,985	,015	,898	,102	,962	,043
65	1	1 0	1,000	,000	,995	,005	1,000	,000
66	2	0 1	,056	,944	,139	,861	,122	,878
67	2	0 1	,002	,998	,011	,989	,027	,974
68	1	1 0	1,000	,000	,917	,083	1,000	,000
69	1	1 0	1,000	,000	,000	1,000	1,000	,000
70	1	1 0	1,000	,000	,997	,003	1,000	,000
71	1	1 0	,997	,004	,948	,052	1,000	,000
72	1	1 0	1,000	,000	1,000	,000	1,000	,000
73	1	1 0	,989	,012	,991	,008	,889	,110
74	1	1 0	1,000	,000	,756	,245	1,000	,000
75	1	1 0	1,000	,000	,999	,001	1,000	,000
76	1	1 0	,990	,010	,714	,285	,980	,022
77	1	1 0	,998	,002	1,000	,000	1,000	,000
78	2	0 1	,003	,997	,787	,212	,011	,990
79	2	0 1	,064	,936	,311	,689	,641	,365
80	1	1 0	,969	,031	1,000	,000	,182	,814
81	2	0 1	,003	,997	,010	,990	,010	,990
82	2	0 1	1,000	,000	,999	,001	1,000	,000
83	1	1 0	1,000	,000	,586	,414	1,000	,000
84	1	1 0	1,000	,000	,805	,195	1,000	,000
85	1	1 0	1,000	,000	1,000	,000	1,000	,000
86	2	0 1	,086	,914	,095	,905	,240	,758
87	1	1 0	1,000	,000	1,000	,000	1,000	,000
88	1	1 0	1,000	,000	1,000	,000	1,000	,000
89	1	1 0	,891	,109	,005	,995	,012	,988
90	1	1 0	1,000	,000	1,000	,000	1,000	,000

Os compostos que foram erradamente previstos no modelo de “Previsão 1” e que foram corretamente classificados no modelo de “Previsão 2” fora 48, 69, e 78. Os que foram bem previstos em “Previsão 1” e erradamente em “Previsão 2” foram 79, e 80.

5.3.2. Modelo quantitativo para as atividades *in vitro*, com cargas de Mulliken.

Esta rede foi treinada com os mesmas variáveis utilizadas para classificação e previsão.

Foram usados 20 neurônios na segunda camada, e limite de erro de 10^{-5} para cada iteração. Testes com outros limites de erro foram feitos, sem nenhuma melhora significativa no modelo.

Tabela 20: Padrões de treinamento e valores ajustados pelo modelo.

>	48	49	50	51	52	53	54	55	56
Valor:	52,000	-1,000	71,000	48,000	48,000	61,000	99,000	80,000	38,000
Calc:	51,582	4,729	54,932	51,045	56,529	63,108	106,912	78,871	33,288
>	57	58	59	60	61	62	63	64	65
Valor:	35,000	78,000	27,000	60,000	56,000	61,000	25,00	111,000	53,000
Calc:	41,021	58,703	27,738	58,753	60,984	62,513	26,600	109,69	51,496
>	66	67	68	69	70	71	72	73	74
Valor:	16,000	5,000	45,000	52,000	63,000	75,000	77,000	70,000	48,000
Calc:	13,2383	3,900	53,391	49,391	64,324	75,066	71,594	70,053	52,277
>	75	76	77	78	79	80	81	82	83
Valor:	54,000	99,000	85,000	25,000	24,000	41,000	25,000	26,000	69,000
Calc:	81,942	101,540	77,138	27,157	32,194	48,097	29,672	62,66	64,598
>	84	85	86	87	88	89	90		
Valor:	75,000	85,000	8,000	49,000	100,000	111,000	97,000		
Calc:	62,989	72,497	39,225	49,280	105,673	133,534	81,649		

Tabela 21: Regressão linear para os valores calculados:

$$Y = A + B * X$$

	Parâmetro	Valor
A	8,99457	3,64264
B	0,87998	0,05766

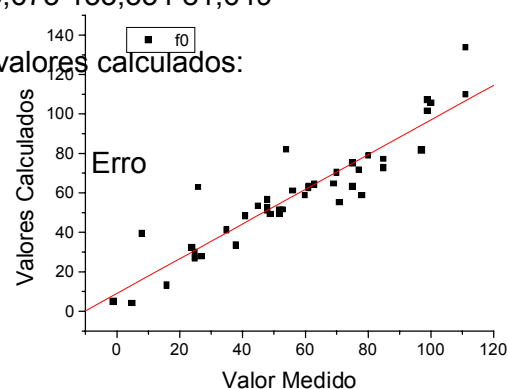


Figura 50: Gráfico da regressão entre os valores

R	SD	N	P
0,92214	10,75039	43	<0,0001

5.3.3. Modelo com as atividades *in vitro* com cargas CHELP.

Para as variáveis selecionadas no com PCA/HCA, utilizando-se as cargas derivadas do potencial eletrostático, foram treinadas redes neurais que relacionam-se com as atividades medidas *in vitro*, e estudada a influência da variação do número de neurônios na segunda camada. As variáveis utilizadas são Q_{N3}^{CHELP} , Q_{N4}^{CHELP} , Q_{C6}^{CHELP} , Q_{C7}^{CHELP} , Q_{R2}^{CHELP} , α , $vdWV$, e $vdWVR1$, totalizando 8. Os resultados são mostrados na tabela 22.

Os padrões de treinamento foram acuradamente reproduzidos pelo modelo neural, e estão mostrados como “classificação resultante” na tabela 22. A predição um a um com uma rede treinada com limite de erro de 10^{-5} e o mesmo número de neurônios na primeira e segunda camada (8) é mostrada como “previsão 1” na tabela 22, e o mesmo resultado para uma rede treinada com o dobro de neurônios na segunda camada (16) é mostrado como “previsão 2” na tabela 22. Os compostos pouco ativos classificados errado são marcados em azul, e os muito ativos classificados errado são marcados em vermelho.

O modelo com mais neurônios na segunda camada foi mais eficiente na previsão da classe dos compostos retirados no treinamento do modelo, diminuindo de 17 para 15 as predições erradas. Significa que o melhor modelo errou 35% das predições com este conjunto de variáveis e padrões de treinamento da rede neural.

Tabela 22: Resultados dos modelos com redes neurais para as atividades *in vitro* dos 4-fenil 2-guanidinothiazóis, usando as cargas derivadas do potencial eletrostático

Padrão de Treinamento.			Classificação Resultante		Classificação Prevista 1		Classificação Prevista 2	
Num.	Classe	Padrão	1	2	1	2	1	2
48	1	1 0	1,000	,000	1,000	,000	1,000	,000
49	2	0 1	,009	,991	,998	,002	,996	,004

50	1	1 0	1,000	,000	,000	1,000	,001	,999
51	1	1 0	,999	,001	1,000	,000	1,000	,000
52	1	1 0	1,000	,000	1,000	,000	1,000	,000
53	1	1 0	,999	,001	1,000	,000	,999	,001
54	1	1 0	,983	,017	,000	1,000	,000	1,000
55	1	1 0	,991	,008	,936	,065	1,000	,000
56	2	0 1	,007	,993	,387	,614	,168	,833
57	2	0 1	,017	,984	,000	1,000	,000	1,000
58	1	1 0	,999	,001	1,000	,000	,995	,005
59	2	0 1	,020	,980	1,000	,000	1,000	,000
60	1	1 0	1,000	,000	1,000	,000	1,000	,000
61	1	1 0	1,000	,000	,829	,171	,976	,024
62	1	1 0	1,000	,000	1,000	,000	1,000	,000
63	2	0 1	,020	,979	1,000	,000	,058	,947
64	1	1 0	,995	,005	1,000	,000	1,000	,000
65	1	1 0	,992	,008	1,000	,000	1,000	,000
66	2	0 1	,021	,979	1,000	,000	,999	,001
67	2	0 1	,018	,982	,000	1,000	,000	1,000
68	1	1 0	1,000	,000	1,000	,000	,905	,094
69	1	1 0	1,000	,000	1,000	,000	1,000	,000
70	1	1 0	,971	,029	,270	,730	,592	,409
71	1	1 0	,997	,003	,998	,002	,999	,001
72	1	1 0	,996	,004	1,000	,000	1,000	,000
73	1	1 0	,979	,021	,003	,997	,009	,991
74	1	1 0	1,000	,000	,000	1,000	,000	1,000
75	1	1 0	1,000	,000	1,000	,000	1,000	,000
76	1	1 0	,993	,007	,836	,162	,986	,014
77	1	1 0	,972	,028	,005	,995	,015	,985
78	2	0 1	,004	,996	,999	,001	,997	,002
79	2	0 1	,000	1,000	,002	,998	,004	,996
80	1	1 0	1,000	,000	,913	,087	,936	,065
81	2	0 1	,020	,980	1,000	,000	1,000	,000
82	2	0 1	,026	,974	1,000	,000	1,000	,000
83	1	1 0	1,000	,000	,713	,281	1,000	,000
84	1	1 0	,991	,009	1,000	,000	,998	,002
85	1	1 0	,995	,005	1,000	,000	,992	,008
86	2	0 1	,028	,972	,026	,974	,266	,738
87	1	1 0	,980	,020	,000	1,000	,000	1,000
88	1	1 0	1,000	,000	,000	1,000	,000	1,000
89	1	1 0	,974	,025	,431	,560	,430	,565
90	1	1 0	1,000	,000	1,000	,000	1,000	,000

5.3.4. Modelo quantitativo para as atividades *in vitro*, com cargas CHELP.

O modelo foi ajustado para reproduzir os dados quantitativos de atividade biológica para os dados selecionados. Trata-se do treinamento de uma rede neural, e não de previsão com exclusão sucessiva de cada amostra. Os resultados reproduzem fielmente as atividades medidas experimentalmente, como pode ser visto no gráfico da figura 51 e nas

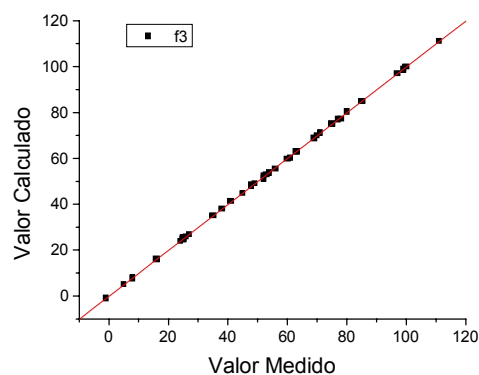
tabelas 23 e 24, a regressão linear entre os dados medidos experimentalmente e os ajustados pelo modelo.

Tabela 23: Padrões de treinamento e valores ajustados pela rede neural.

	>	48	49	50	51	52	53	54
55	56							
		Valor:52,000	-1,000	71,000	48,000	48,000	61,000	99,000
					80,000	38,000		
		Calc.:52,3707	-,9659	71,0572	48,7035	47,6150	60,3092	98,8016
							80,3220	38,008
	>	57	58	59	60	61	62	63
				64	65			
		Valor:35,000	78,000	27,000	60,000	56,000	61,000	25,000
		111,00	53,000					
		Calc.:35,0814	77,1342	26,8901	59,6186	55,3201	60,2966	25,058
							110,970	53,066
	>	66	67	68	69	70	71	72
					73	74		
		Valor:16,000	5,000	45,000	52,000	63,000	75,000	77,000
							70,000	48,000
		Calc.:16,0344	5,0538	44,7238	51,2071	62,9838	74,8256	76,9366
							69,901	47,688
	>	75	76	77	78	79	80	81
					82	83		
		Valor:54,000	99,000	85,000	25,000	24,000	41,000	25,000
							26,000	69,000
		Calc.:53,5758	98,8540	84,9420	24,6617	24,0263	41,228	24,973
							25,720	68,768
	>	84	85	86	87	88	89	90
		Valor:75,000	85,000	8,000	49,000	100,000	111,000	97,000
		Calc.:75,2749	84,8516	7,7559	48,9539	99,9774	110,950	97,0102

Tabela 24: Regressão Linear para os valores

calculados:			
$Y = A + B * X$			
Parâmetro	Valor	Erro	
A	-0,09443	0,10712	
B	0,99942	0,0017	
R	SD	N	P
0,99994	0,31613	43	<0,0001

**Figura 51:** Gráfico da regressão entre os valores experimentais e os valores obtidos nos

5.3.5. Modelos de classificação *in vivo*.

Para relacionar as atividades *in vivo* reportadas para os 23 compostos da série, foram treinadas redes neurais. Foi também estudada a influência da variação do limite de erro aplicado em cada iteração. Os resultados obtidos são mostrados na tabela 25. As 10 variáveis utilizadas na primeira camada de neurônios foram selecionadas por PCA/HCA.

São: F_{C5}^{LUMO} , F_{C9}^{HOMO} , F_{C10}^{HOMO} , Q_{N3}^{CHELP} , Q_{R2}^{CHELP} , $\mu^{hibrido}$, α , VvdW, vdWSA, e R1VvdW

Tabela 25: Resultados dos modelos com redes neurais para as atividades *in vivo* dos 4-fenil 2-guanidino-tiazóis, usando as variáveis selecionadas com PCA/HCA.

Padrões de treinamento			Padrão Calculado		Padrão Previsto 1		Padrão Previsto 2	
Num.	Classe	Padrão	1	2	1	2	1	2
50	1	1 0	1,000		1,000		1,000	
			,000		,000		,000	
51	1	1 0	,997		,996		,991	
			,003		,004		,008	
53	1	1 0	1,000		1,000		1,000	
			,000		,000		,000	
54	2	0 1	,000		,000		,000	
			1,000		1,000		1,000	
58	1	1 0	,997		,996		,988	
			,004		,004		,012	
60	1	1 0	1,000		1,000		1,000	
			,000		,000		,000	
61	1	1 0	,991		,906		,390	

				,010	,094	,610
64	2	0	1	,007	1,000	1,000
				,992	,000	,001
70	1	1	0	,982	,004	,026
				,017	,996	,975
71	2	0	1	,018	1,000	,996
				,983	,000	,004
72	2	0	1	,000	,000	,000
				1,000	1,000	1,000
73	1	1	0	1,000	1,000	1,000
				,000	,000	,000
74	1	1	0	1,000	1,000	1,000
				,000	,000	,000
75	1	1	0	,975	,000	,000
				,025	1,000	1,000
76	2	0	1	,000	,000	,000
				1,000	1,000	1,000
77	2	0	1	,000	,000	,000
				1,000	1,000	1,000
83	1	1	0	,982	,000	,002
				,019	1,000	,998
84	2	0	1	,020	1,000	,999
				,980	,000	,001
85	2	0	1	,000	,000	,000
				1,000	1,000	1,000
88	2	0	1	,022	1,000	1,000
				,978	,000	,000
89	2	0	1	,000	,000	,000
				1,000	1,000	1,000
90	2	0	1	,013	,823	,915
				,987	,173	,088
91	2	0	1	,000	,000	,000
				1,000	1,000	1,000

Os padrões de atividade do conjunto de treinamento foram fielmente reproduzidos pelo modelo neural, não havendo nenhuma classificação errada nos padrões calculados pelo modelo.

O modelo neural inicial foi treinado para reproduzir as classificações com 22 neurônios na segunda camada, sem bias, por retropropagação, com limite de erro fixado em 10^{-5} . Os parâmetros para estabelecer os limites das variáveis internas na formação da rede neural utilizados foram os padrões internos do programa.

Para as mesmas variáveis foi testado o modelo de previsão sucessiva, que retira seqüencialmente cada um dos compostos do conjunto e prevê a sua posição com o modelo criado a partir dos restantes. Os parâmetros para a criação da rede foram os mesmos que no caso anterior.

Para o caso da previsão dos compostos por exclusão sucessiva de cada um deles o modelo errou 8 classificações, no conjunto de 23 compostos, totalizando 35% de erro. As previsões erradas foram os compostos 64, 71, 84, 88, e 90 previstos como pouco ativos no modelo, e os compostos 70, 75, e 83 previstos como muito ativos no modelo. Não há relação clara entre a classificação e alguma estrutura química presente nos compostos. Outro modelo foi construído para as mesmas variáveis, e os mesmos parâmetros para a criação da rede, exceto o limite de erro, fixado em 0,01. O resultado é mostrado nos padrões previstos 2. Houveram 9 classificações erradas, uma a mais que no caso anterior, o composto 61 que é pouco ativo foi erradamente classificado no grupo muito ativo. No mais, as mesmas classificações erradas.

5.3.6. Ajuste de rede quantitativa para as atividades *in vivo*.

O modelo quantitativo foi obtido com as mesmas variáveis que o modelo classificatório para as atividades *in vivo*. O programa foi instruído a realizar regressão não linear múltipla, com a palavra chave QSAR. O limite de erro fixado em 10^{-4} , com 22 neurônios na segunda camada (tab.26). O ajuste da rede aos dados experimentais é muito bom (tab.27), como pode ser observado no gráfico (fig.52) dos valores calculados versus valores experimentais.

Tabela 26: Padrões de treinamento e resultados da rede treinada.

>	50	51	53	54	58	60	61	64	
				70					
Valor:	2,890	3,970	3,457	1,792	2,996	3,466	3,611	2,303	3,466
Calc:	2,886	3,966	3,446	1,784	2,989	3,454	3,602	2,321	3,466

> 71 72 73 74 75 76 77 83

84

Valor: 2,639 1,649 3,296 4,094 3,555 ,833 2,303 3,135 2,708

Calc: 2,654 1,686 3,327 4,109 3,558 ,838 2,307 3,142 2,709

> 85 88 89 90 91

Valor: 2,398 ,406 -1,204 ,470 ,262

Calc: 2,375 ,397 -1,216 ,451 ,254

Tabela 27: Regressão linear dos valores
calculados:

$$Y = A + B * X$$

Parâmetro

Valor

Erro

A

-0,00524

0,00625

B

1,00237

0,00229

R

SD

N

P

0,99995

0,01489

23

<0,0001

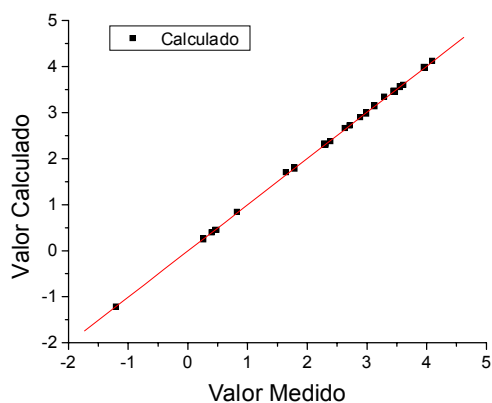


Figura 52: Gráfico da regressão
entre os valores experimentais
e os valores obtidos nos

5.3.7. Modelo de previsão quantitativa das atividades *in vivo*.

O modelo de previsão quantitativa foi obtido com os mesmos dados do ajuste da rede, porém com limite de erro fixado em 0,01. O maior valor do limite de convergência produziu melhor resultado na predição dos valores de atividade (tab.28). Este modelo tem desempenho razoável na predição quantitativa dos valores. O valor de R é obtido na previsão da atividade dos compostos pelo modelo (tab.29), traçando-se o gráfico do valor previsto versus o valor medido experimentalmente (fig.53), pode ser comparado ao valor denominado q^2 , utilizado em modelos de CoMFA para testar o ajuste de uma regressão PLS^{lvii}. Para predição de compostos, com a mesma metodologia de “deixar um de fora”, é

considerado bom se $q^2 > 0,3$ (no caso $R^2 = 0,305$). Para este tipo de regressão não deve-se esperar resultados como os obtidos na ajuste simples de uma função aos dados experimentais, quer seja função linear ou polinomial.

Tabela 28: Padrões de treinamento e resultados da rede treinada.

>	50	51	53	54	58	60	61	64	
70									
Valor:	2,890	3,970	3,457	1,792	2,996	3,466	3,611	2,303	3,466
Previ:	2,874	3,722	4,354	1,982	2,475	3,304	2,661	3,832	2,577
>	71	72	73	74	75	76	77	83	
84									
Valor:	2,639	1,649	3,296	4,094	3,555	,833	2,303	3,135	2,708
Previ:	4,298	1,699	2,767	4,569	2,149	-2,388	1,075	,461	4,676
>	85	88	89	90	91				
Valor:	2,398	,406	-1,204	,470	,262				
Previ:	1,034	2,570	2,540	2,706	-1,127				

revistos

Previsto2

Categoria	Previsto2
85	1.034
88	2.570
89	2.540
90	2.706
91	-1.127

Tabela 29: Regressão linear entre os valores previstos e os experimentais

$$Y = A + B * X$$

Parâmetro	Valor	Erro	
A	0,69322	0,52112	
	0,57993	0,19091	
R	SD	N	P
0,55252	1,24108	23	0,00626

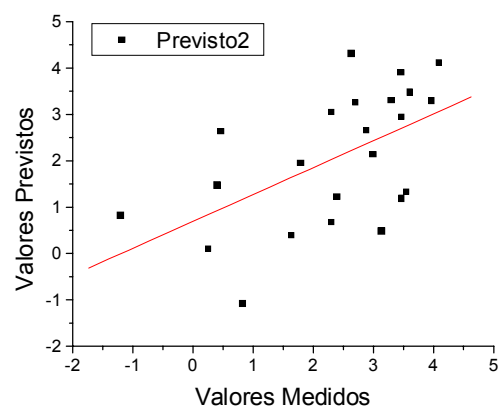


Figura 53:
Gráfico da

Conclusões

- ⇒ As ferramentas estatísticas utilizadas na análise exploratória dos dados (HCA, PCA) foram muito importantes para obtenção de relações entre atividades biológicas e parâmetros físico-químicos calculados por métodos semi-empíricos, em função das possibilidades oferecidas para seleção de variáveis com capacidade de modelagem.
- ⇒ Os Hamiltonianos MINDO3 e MNDO são capazes de calcular polarizabilidades e momentos de dipolo e cargas nas posições dos átomos C1, C2, C19 e C20 (ver figura 3.5) que correlacionam-se com as atividades biológicas dos compostos diterpenos, melhores que os obtidos com AM1 ou PM3 para estes compostos.
- ⇒ A utilização do Hamiltoniano AM1 modificado para incluir efeitos de solvatação em água permitiu que fossem calculadas cargas que relacionam-se melhor com os dados de atividade biológica que os obtidos com o AM1.
- ⇒ As cargas derivadas do potencial eletrostático e de Mulliken calculadas com Hamiltoniano AM1 parametrizado para reproduzir solvatação em água, criam modelos que são equivalentes nos resultados de correlação com as atividades biológicas dos compostos Diterpenos.
- ⇒ Os resultados obtidos com Hamiltoniano AM1 para os guanidinothiazóis criam um modelo de classificação das atividades biológicas *in vitro* e *in vivo* melhor que os obtidos com MINDO3, MNDO, ou PM3.
- ⇒ As variáveis selecionadas que relacionam-se com as atividades *in vitro* são a polarizabilidade, e o volume de van der Waals das moléculas e dos substituintes na posição R1. As cargas calculadas com método CHELP ou de Mulliken sobre os nitrogênios N3 e N4 (ver figura 5.20) dos guanidinothiazóis também são importantes.
- ⇒ As variáveis que correlacionam-se com as atividades *in vivo* são as polarizabilidades, o momento de dipolo calculado com os orbitais híbridos, os volumes de van der Waals das

moléculas e dos substituintes na posição R1, o volume acessível ao solvente, as densidades eletrônicas em HOMO nos carbonos C9 e C10, a densidade em LUMO sobre o carbono C5, e as cargas CHELP sobre o nitrogênio N3 e o grupo R2.

⇒ Os parâmetros calculados estão ligados às estruturas apontadas em trabalhos anteriores de relacionamento entre estrutura química e atividade biológica, como responsáveis pela atividade mostrada pelos compostos guanidino-tiazóis.

⇒ A estrutura da guanidina foi apontada como importante pelas cargas calculadas nas posições deste grupo, e pelo volume da região correspondente ao grupo substituinte próximo à guanidina

⇒ Os modelos de classificação obtidos com PCA na classificação dos compostos pela sua atividade biológica *in vitro* são melhores que os publicados anteriormente, baseados em parâmetros de conectividade.

⇒ No caso das classificações *in vivo* os resultados são equivalentes aos publicados.

Porém os modelos aqui propostos são capazes de separar os grupos baseados em parâmetros moleculares que relacionam-se com as estruturas químicas dos compostos, e portanto podem ser mais úteis na tentativa de propor modificações que aumentem a potência dos compostos.

⇒ Os modelos quantitativos obtidos não mostram grande capacidade de predição dos valores de atividade biológica para os compostos, tanto no caso dos modelos de regressão PLS multivariada, como no caso de regressão com redes neurais.

Bibliografia

- ⁱ Relatório elaborado durante o seminário “Desafios e Oportunidades para a Indústria Farmacêutica” realizado em 22/10/**1997**.
- ⁱⁱ Yuji Takahata, Projeto Técnico Científico Elaborado pelo Grupo de pesquisa, Instituto de Química da UNICAMP.
- ⁱⁱⁱ Cartier,A.;Rivail,J.L; *Chemom. Intell. Lab. Sys.*, **1987**, 1, 335.
- ^{iv} Gilbert,G.; Chan,C.H.; Thomas,E.; *Postgraduate Medicine*; **89**(4): 91-98, (1991).
- ^v Larson,D.E.; “Clinic Family Health Book”. New York: William Morrow and Company, Inc., **1990**.
- ^{vi} Fedotin,M.S.; *Postgraduate Medicine*,; **94**(3): 38-45, (1993).
- ^{vii} Goodman & Gilman’s; “The Pharmacological Basis of Therapeutics”; Editor Joel G. Hardman; **1996**, pp-901-917.
- ^{viii} Maloney,P.C.; *J.Membrane Biol.* **67**. 1-12 (1982).
- ^{ix} Gustavson, S.; Adam, H. O.; Loof, L.; Nyberg, A.; Nyren, O.; *Lancet* ; **124** ,8342, (1983).
- ^x Holt, W.F.; Murphy, W.R.; LaMattina,J.L; *Gastroenterology*, **92**, 1439, (1987).
- ^{xi} Madan,A.K.; Goel,A.; *J.Chem.Inf.Comput.Sci.*;, **35**, 504-509, (1995).
- ^{xii} Eiichi Kitazawwa, Aiya Sato, Shuji Takahashi, Harumitsu Kumano, and Akira Ogiso, *Chem.Pharm.Bull.* **28**(1):227-234 (1980).
- ^{xiii} *Gastroenterology*. **5**(43) (1945)
- ^{xiv} José C.Simões, Antônio J.Ribeiro da Silva, Harry Serruya, Maria Helena da Silva Bentes, *Ciência e Cultura* **31**(10):1140-1143 (1979).
- ^{xv} Kubo,I.; Asaka,Y.; and Shibata,K.; *Phytochemistry* **30**(8):2545-2546 (1991).
- ^{xvi} Yoshitatsu Ichihara, Koichi Takeya, Yukio Hitotsuyanagi, Hiroshi, *Planta Med.* **58**(6):549-551 (1992).
- ^{xvii} Morita, S.O.; Masami S.; Hirota F.; Motidome,M.; and Itakawa,H.; *Planta Med.* **55**(2):127-133 (1987).
- ^{xviii} Åsbrink , L ; Fridh , C ; Lindholm , E . ; *Chemical Physics Letters*;, **52** , 63-75, (1977).
- ^{xix} Guadagnin, P.H, Bruns, R.E, *Química Nova* , , **19**, 148-151, (1996).

-
- ^{xx} Reed, J., *J. Chem. Ed.*, , **69**(10), 785, (1992).
- ^{xxi} Zhou, Z.; Parr, R.G.; *J.Am.Chem.Soc.*, , **112**, 5720, (1990).
- ^{xxii} Osmialowski, K.; Halkiewics,J.; Radecki,A.; Kaliszan, R.; *J.Chromatogr.*, , **346**, 53, (1985).
- ^{xxiii} Pearson, R.G.; *J. Org. Chem.*, , **54**, 1423, (1989).
- ^{xxiv} Zhang, Y.; Zhao,C.Y.; Fang,W.H; You,X.Z.; *Theor. Chem. Acc.*, , **96**, 129-134, (1997).
- ^{xxv} Stewart, J.J.P.; Kenneth,M.D.; Kurtz, H.A.; *Theoretical Chemistry*, , **11**, 82-87, (1989).
- ^{xxvi} Mulliken, R.S., *J. Chem. Phys.*, , **23**, 1833, (1955).
- ^{xxvii} Levine, I. A. "Quantum Chemistry" 4th Ed., Prentice-Hall, Inc., New York, (**1991**); Szabo, A., Ostlund., N. S., "Modern Quantum Chemistry", 1th Ed., Dover Publications, Inc., Mioneola, (**1996**).
- ^{xxviii} Reed, A.E., Weinstock, R. B., Weinhold, F.^a; *J.Chem. Phys.*, , **83**, 735-746, (1985).
- ^{xxix} Wiberg, K. B, Rablen, P. R; *J. Comp. Chem.*, , **14**, pp.1504-1518, (1993).
- ^{xxx} Chirlian, L. E., Francl, M.M.; *J. Comp. Chem.*, , **8**, pp. 894-905, (1987).
- ^{xxxi} Breneman, C.M., Wiberg, K.B.; *J. Comp. Chem.*, **11**, pp. 361-373, (1990).
- ^{xxxii} Lowe, J.P., "Quantum Chemistry" Academic Press, Inc., San Diego, **1978**.
- ^{xxxiii} Clark, T., "A Handbook of Computacional Chemistry, John Willey & sons. **1995**
- ^{xxxiv} Grier, D.D., Streitweiser, A.; *J.Am. Chem. Soc*, **104**, pp. 3556-3564, (1982).
- ^{xxxv} Christoffersen, R.E., Baker, K. A. *Chem. Phys. Lett.*, , **8**, 4, (1971).
- ^{xxxvi} Pollak, M. Rein, R.; *J. Chem Phys.*, **47**, 2045-2052, (1967).
- ^{xxxvii} Löwdin; *J.Chem. Phys*, , **21**, 374-375, (1953).
- ^{xxxviii} Davison, E. R., *J. Chem. Phys.*, **46**, 3320, (1953).
- ^{xxxix} Yanes; M., Stewart, R.F., Pople, J., *Acta Crystallogr. A*, **34**, 641, (1978).
- ^{xl} Ferreira, M.M.C., *J. Mol. Struct. (Teochem)*, **266**, 223, (1992).

-
- ^{xli} Politzer, P. Harris, R.R.; *J. Am. Chem. Soc.*, **92**, pp. 6451-6454, (1970).
- ^{xlii} Cornell, W. D., Cieplak, P., Bayly, C. I., Kollman, P. A., *J. Am. Chem. Soc.*, **115**, 9620, (1993).
- ^{xliii} Cox, S.R., Williams, D.E., *J. Comp. Chem.*, **2**, 304, (1981).
- ^{xliv} Breneman, C.M., Wiberg, K.B. *J. Comp. Chem.*, **11**, 361, (1990).
- ^{xl} (a) Pauling, L. "The Nature of Chemical Bond", Cornell University Press, Ithaca, N.Y., **1960**, pp. 260-2.(b) Valvani, S.C. et al *J. Phys. Chem.* **1976**, *80*(8), 829-835.
- ^{xlv} Dixon, R.W.; Leonard, J.M.; Hehre, W.J.; *Israel Journal of Chemistry*, **33**, pp.427-434, (1993)
- ^{xlvii} "Pirouette Multivariate Data Analysis for IBM System"; V.2.0; Infometrix, Seattle, WA, (1996).
- ^{xlviii} Sharaf, M.A.; Illman, D.L.; Kowalski, B.R.; "Chemometrics"; Wiley (1986).
- ^{xlix} Martens, H.; Naes, T.; "Multivariate Calibration"; Wiley, New York, 1989.
- ^l Ichikawa, H.; "QCPE 615. PSDD: Perceptron-type Neural Network Simulator"; Hoshi College of Pharmacy, 2-4-41 Ebara, Shinagawa, Tokyo 142, Japan.
- ^{li} Wise, B.M.; Gallagher, N.B.; "PLS_Toolbox Version 1.5"; Eigenvector Technologies; P.O. Box 483, 196 Hyacinth Avenue, Manson, WA, 98831. MATLAB é uma marca registrada de The Mathworks, Inc.
- ^{lii} Chirlian, L.E.; Franci, M.M.; *J. Comp. Chem.*; **8**, pp 894-905, (1987).
- ^{liii} Beil, R.; *Br. J. Pharmacol.*; **82**, 651, (1984).
- ^{liiii} Agarwal, A.; Pearson, P.P.; Taylor, E.W.; Li, H.B.; Dahlgren, T.; Herslöf, M.; Yang, Y.; Lambert, G.; Nelson, D.L.; Regan, J.W.; Martin, A.R.; *J. Med. Chem.*, **36**, 4006-4014, (1993).
- ^{lv} 'Guidebook on Molecular Modeling in Drug Design'; Edited by Cohen, N.C.; Academic Press, Inc; (1996)
- ^{lvi} Beil, R.; *Br. J. Pharmacol.*; **82**, 651, (1984).
- ^{lvii} Agarwal, A.; Pearson, P.P.; Taylor, E.W.; Li, H.B.; Dahlgren, T.; Herslöf, M.; Yang, Y.; Lambert, G.; Nelson, D.L.; Regan, J.W.; Martin, A.R.; "Three-Dimensional quantitative Structure-Activity Relationships of 5-HT Receptor Binding Data for Tetrahydropyridylindole Derivatives: A comparison of the Hansch and CoMFA Methods"; *J. Med. Chem.* **1993**, *36*, 4006-4014.