UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA

**“Emprego de Computadores em
Elucidação Estrutural de Alcalóides”**

Alessandra Rodrigues Rufino

Tese de Doutorado

Prof. Dr. Vicente de Paulo Emerenciano

Orientador

São Paulo

Maio de 2005

“Emprego de computadores em elucidação estrutural de alcalóides”

ALESSANDRA RODRIGUES RUFINO

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

Aprovado(a) por:

Prof. Dr. VICENTE DE PAULO EMERENCIANO
IQ – USP
(Orientador e Presidente)

Prof. Dr. JOÃO PEDRO SIMON FARAH
IQ – USP

Profa. Dra. MARIA AUXILIADORA COELHO KAPLAN
UFRJ

Profa. Dra. EDNA TOMIKO MYIAKE KATO
FCF – USP

Prof. Dr. GILBERTO DO VALE RODRIGUES
UFMG

SÃO PAULO
12 DE MAIO 2005

Agradecimentos

A Deus, por tornar mais um sonho possível e estar sempre ao meu lado.

Ao Prof. Dr. Vicente de Paulo Emerenciano, pela orientação e amizade.

Ao Prof. Dr. Gilberto do Vale Rodrigues, pela criação dos programas e disponibilidade de resolver os problemas que surgiram durante a tese.

Ao Antônio, pela preciosa amizade e sabedoria que está sempre disponibilizada para ajudar alguém.

Ao Marcelo pela coorientação e disponibilidade de ensinar o funcionamento dos programas.

Ao Telmo, pela preciosa amizade construída ao longo destes anos de convivência.

Aos funcionários da secretaria de pós-graduação, pela grande disponibilidade, eficiência, agilidade e profissionalismo com que executam os seus trabalhos. Estes elogios são poucos, vocês merecem muito mais.

Aos funcionários da Biblioteca do Conjunto das Químicas, pela dedicação e disponibilidade.

Aos porteiros do Instituto de Química pelas conversas e amizade.

Aos meus colegas de instituto Biaggio, Sr. Luiz, Xico, Júlio, dona Rosa e muitos outros, pela amizade.

Ao meu irmão Carlindo, pela amizade que eu tanto prezo, pelas nossas conversas e pelos bons e maus momentos que passamos juntos.

À minha família que sempre me apoiou, em especial, a minha mãe, que é meu esteio. Obrigada pelas longas conversas, pelos conselhos e por nunca ter me deixado desistir.

Ao meu esposo, pela grande contribuição que deu nesta tese, além de todo o carinho e paciência.

À Fundação de Amparo a Pesquisa do Estado de São Paulo, pela bolsa e reserva técnica.

Finalmente, agradeço a todos que, de alguma forma, contribuíram para realização deste trabalho.

Ao meu marido, que nos últimos 14 anos esteve sempre ao meu lado, obrigada pelo amor, dedicação, carinho, amizade e companheirismo.

À minha mãe, pelo carinho, amizade, amor e dedicação. Obrigada por acreditar tanto em mim e por manter sempre acesa a chama da sua fé.

RESUMO

O Sistema Especialista SISTEMAT foi construído com o objetivo de auxiliar os pesquisadores da área de produtos naturais na tarefa de determinação estrutural, estendendo-se também ao químico orgânico sintético. Seus programas aplicativos fornecem propostas de esqueletos fazendo uso dos dados de diversas técnicas espectrométricas, sendo que a espectrometria de ressonância magnética nuclear de ^{13}C tem um papel de destaque entre as demais.

Este trabalho descreve a utilização do SISTEMAT como uma ferramenta auxiliar na determinação estrutural de substâncias pertencentes às subclasses dos alcalóides quinolínicos, quinolizidínicos, aporfínicos, benzilisoquinolínicos, isoquinolínicos, pirrolizidínicos, acridônicos e indólicos. Para a realização deste trabalho foi construído um banco de dados contendo 1182 alcalóides, sendo todos coletados da literatura. Nestes 1182 alcalóides, estão presentes 1156 espectros de RMN ^{13}C , 354 espectros de RMN ^1H , 320 espectros de massas e as substâncias de origem vegetal estão distribuídos em 49 Famílias, 164 Gêneros e 260 Espécies. Os testes realizados forneceram bons percentuais de acertos para o reconhecimento de esqueletos.

Outro programa utilizado neste trabalho foi o de redes neurais artificiais. As redes foram treinadas para auxiliar na determinação estrutural dos alcalóides aporfínicos, fornecendo a probabilidade de uma determinada substância pertencer ao esqueleto pesquisado. Para utilização das redes neurais foi construída uma planilha com os deslocamentos químicos de RMN ^{13}C , de 165 alcalóides aporfínicos, pertencentes a 12 esqueletos diferentes. A rede forneceu ótimos resultados, classificando os esqueletos com alto grau de confiabilidade.

ABSTRACT

The Expert System SISTEMAT was built with the objective of aiding the researchers of the area of natural products in the task of structural determination, also extending to the synthetic organic chemist. Their applications programs supply proposed skeletons making use of the data of several techniques spectrometrics, and the ^{13}C NMR has a main paper among the others.

This work describes the use of SISTEMAT as an auxiliary tool in the structural determination of substances belonging to the underclass of the alkaloids quinoline, quinolizidine, aporphine, benzyloquinoline, isoquinoline, pyrrolizidine, acridone and indoles. For the accomplishment of this work a database was built containing 1182 alkaloids, being all collected of the literature. In these 1182 alkaloids, are present 1156 spectra of ^{13}C NMR, 354 spectra of RMN 1:00, 320 spectra of masses and the substances of botanical origin are distributed in 49 Families, 164 Genders and 260 Species. They were accomplished around 100 tests, of which 30 are presented in this thesis. These tests supplied good percentile of the successes for the recognition of skeletons.

Another program used in this work the one of nets artificial neurais, in which the nets were trained to aid in the structural determination of the aporphine alkaloids was, supplying the probability of a certain substance to belong to the researched skeleton. For use of the nets neurais a spreadsheet was built with the chemical displacements of ^{13}C NMR, of 165 aporphine alkaloids, belonging to 12 different skeletons. The net supplied great results, classifying the skeletons with high reliability degree.

SUMÁRIO

I -	Introdução.....	1
II -	Alcalóides	
	II.1. Introdução.....	4
	II.2. Definições e classificação dos alcalóides.....	4
	II.3. Evolução das pesquisas com ênfase nos alcalóides derivados de plantas.....	7
	II.4. Biossíntese e atividades biológicas dos alcalóides.....	9
III -	Sistemas Especialistas em determinação estrutural	
	III.1. Sistema DENDRAL.....	35
	III.2. Sistema CHEMICS.....	38
	III.3. Sistema DARC/EPIOS.....	41
	III.4. Sistema SpecInfo/SpecSolv.....	42
	III.5. Sistema SENECA.....	47
IV -	O Sistema Especialista SISTEMAT	
	IV.1. Metodologia.....	52
	IV.1.1. Levantamento Bibliográfico	52
	IV.1.2. Organização dos bancos de dados.....	52
	IV.2. Programas de criação do banco de dados.....	54
	IV.2.1. Codificação Estrutural.....	54
	IV.2.2. Programa DATASIS.....	58
	IV.2.3. Programa ARQUISIS.....	62
	IV.3. Programas Aplicativos.....	64
	IV.3.1. Programa SISLIST.....	64
	IV.3.2. Programa MACRONO.....	75
	IV.3.3. Programa SISCONST.....	81
	IV.3.4. Programa C13MACH.....	84
	IV.3.5. Programa SISOCBOT.....	86
	IV.3.6. Programa MASMACH.....	87
	IV.3.7. Programa H1MACH.....	88

IV.3.8. Programa PESQUISA.....	90
IV.4. Resultados e Discussão.....	96
V- Redes Neurais Artificiais	
V.1. Introdução.....	134
V.2. Redes Neurais Artificiais x Sistemas Especialistas.....	134
V.3. Histórico	135
V.4. Definições	138
V.5. Benefícios das Redes Neurais Artificiais	141
V.6. Neurônios Artificiais.....	144
V.7. Tipos de Função de Ativação.....	146
V.8. Aprendizado da Rede.....	147
V.8.1. Aprendizado Supervisionado.....	148
V.8.1.1. Correção de Erros.....	149
V.8.2. Aprendizado não Supervisionado	150
V.8.2.1. Aprendizado Hebbiano.....	151
V.8.2.2. Aprendizado por Competição.....	153
V.8.3. Aprendizado por Reforço.....	154
V.9. Principais arquiteturas das Redes Neurais Artificiais.....	155
V.9.1. Número de camadas.....	155
V.9.2. Número de neurônios por camada intermediária.....	156
V.9.3. Tipo de conexão entre os neurônios.....	157
V.9.4. Conectividade da rede.....	158
V.10. Tipos de redes Neurais Artificiais.....	158
V.10.1. Perceptron.....	158
V.10.2. Perceptron de camadas múltiplas	159
V.10.3. Hopfield.....	159
V.10.4. Kohonen.....	160
V.11. Utilização das redes neurais artificiais	161
V.11.1. Utilização de redes em determinação estrutural e previsão de deslocamentos químicos.....	162

V.12. Identificação de esqueletos de alcalóides aporfínicos utilizando Redes Neurais Artificiais	165
V.12.1. Definição de uma topologia ideal para a rede Perceptron de camadas múltiplas	165
V.12.1.1. Taxa de Aprendizado e o termo momento.....	170
V.13. Metodologia.....	172
V.14. Resultados e discussão.....	179
VI. Conclusões.....	188
Referências Bibliográficas.....	190

Lista de Figuras

Figura 1	Vegetais superiores que apresentam alcaloides.....	5
Figura 2	Exemplos de alcalóides verdadeiros.....	6
Figura 3	Exemplos de protoalcalóides.....	6
Figura 4	Exemplos de pseudoalcalóides.....	7
Figura 5	Precursor comum dos alcalóides indol-monoterpênicos.....	11
Figura 6	Núcleos característicos dos alcalóides presentes no banco de dados.....	13
Figura 7	Esqueletos dos Alcalóides quinolínicos, aporfínicos, benzilisoquinolínicos, acridônicos, quinolizidínicos, pirrolizidínicos, isoquinolínicos e Indólicos; presentes na base de dados.....	14
Figura 8	Estrutura do Warburganal e a planta <i>Warburgia ugandensis</i>	36
Figura 9	Fragmentos inferidos por interpretação dos dados espectrais.....	36
Figura 10	Estrutura do monoterpene Genipina.....	42
Figura 11	Sobreposição dos átomos comuns aos dois fragmentos para formação de uma subestrutura maior.....	45
Figura 12	Fluxograma do processo de geração de estruturas do SpecSolv.....	45
Figura 13	Estrutura do alcalóide pirrolizidínico, Uspalatina.....	46
Figura 14	Estrutura do monochaetin.....	49
Figura 15	Ficha de codificação utilizada no levantamento bibliográfico.....	53
Figura 16	Quinolínico X.....	55
Figura 17	Direção das ligações do Vetor Passos.....	58
Figura 18	Codificação manual de um Alcalóide indólico.....	58
Figura 19	Tela do menu principal do programa DATASIS.....	59
Figura 20	Tela de desenho do programa DATASIS, onde foi codificado, automaticamente, o alcalóide indólico utilizado para explicação dos vetores nó e passo (Pg. 58, Fig. 18).....	60
Figura 21	Tela de codificação do programa DATASIS. Onde pode-se comparar os vetores obtidos de forma manual (Pg. 58) com os fornecidos automaticamente pelo programa.....	61
Figura 22	Tela de leitura dos dados do programa DATASIS.....	61
Figura 23	Esquema dos bancos arquitetados do SISTEMAT.....	62
Figura 24	Tela de entrada dos dados físico-químicos do programa DATASIS.....	63
Figura 25	Menu principal do programa MACRONO.....	79
Figura 26	Alcalóide pirrolizidínico.....	80
Figura 27	Tela de resultados do programa MACRONO.....	80
Figura 28	Substância utilizada para teste e esqueleto Quinolínico 21.....	82
Figura 29	Subestruturas fornecidas pelo programa SISCONST.....	83
Figura 30	Substância utilizada para teste e a estrutura do esqueleto	

	Benzilisoquinolínico 3.....	85
Figura 31	Resultados do índice de semelhança apresentados pelo programa C13MACH.....	85
Figura 32	Substância utilizada para teste e o resultado fornecido pelo programa.....	87
Figura 33	Tela do menu principal do programa H1MACH.....	89
Figura 34	Tela de resultados parciais do H1MACH.....	89
Figura 35	Tela do menu principal do programa PESQUISA.....	91
Figura 36	Tela de definição dos dados de entrada.....	92
Figura 37	Tela para entrada dos dados de RMN ¹³ C.....	92
Figura 38	Tela de acesso aos programas aplicativos.....	93
Figura 39	Proposta de subestrutura com as comparações dos respectivos deslocamentos químicos obtidos do banco de dados do SISTEMAT (em preto) com os do espectro problema (em vermelho).....	94
Figura 40	Propostas de esqueletos fornecidas pelo programa SISCONST.....	94
Figura 41	Tela de apresentação dos resultados de probabilidade de esqueletos.....	95
Figura 42	Fragmento correspondente ao pico 169.....	99
Figura 43	Desfragmentação de três alcalóides Isoquinolínicos, pertencentes a esqueletos diferentes que originam o mesmo fragmento.....	99
Figura 44	Subestruturas pertencentes a esqueletos diferentes, apresentando a mesma massa molecular, devido a diferença de substituintes.....	102
Figura 45	Substância codificada, encontrada no banco de dados. Esqueleto Quinolínico 24.....	115
Figura 46	Uma das muitas substâncias do banco de dados que apresenta uma pedaço semelhante ao da substância utilizada no teste 18.....	117
Figura 47	Representação de uma Rede com suas respectivas camadas e conexões....	138
Figura 48	Modelo de Transmissão de Impulso nervoso em Neurônios Biológicos.....	143
Figura 49	Modelo Básico do Neurônio Artificial.....	144
Figura 50	Função de ativação utilizada por MacCulloch e Pitts.....	145
Figura 51	Funções de ativação: (a) Linear; (b) Rampa; (c) Passo e (d) Sigmoidal.....	147
Figura 52	Algoritmo de aprendizado supervisionado.....	149
Figura 53	Exemplo do Processo de aprendizado por correção de erros.....	150
Figura 54	Algoritmo de Aprendizado Não Supervisionado.....	151
Figura 55	Aprendizado por Reforço.....	154
Figura 56	(a) Rede de camada única, (b) Rede de camadas múltiplas.....	156
Figura 57	(a) Redes com alimentação direta; (b) Redes com alimentação recorrente..	158
Figura 58	Rede parcialmente conectada.....	158
Figura 59	Perceptron simples.....	159
Figura 60	Rede Perceptron de Camadas múltiplas.....	159
Figura 61	Rede Hopfield.....	160

Figura 62	Rede Kohonen (mapa de característica).....	161
Figura 63	Fluxo de Processamento do algoritmo de retropropagação.....	166
Figura 64	Superfície do Erro para a Rede MLP – Considerando as coordenadas W1 e W2.....	168
Figura 65	Gráfico da superfície de erro, evidenciando os mínimos locais e globais.....	169
Figura 66	Influência do termo momento.....	171
Figura 67	Esqueletos dos alcalóides aporfínicos.....	172
Figura 68	Alcalóide aporfínico, 7-metildeidroglaucine.....	173
Figura 69	Redes de 3 e 4 camadas.....	175
Figura 70	Alcalóides utilizados para teste das Redes Neurais.....	176
Figura 71	Ciclos x Coeficiente de correlação do treinamento da rede Perceptron, com três camadas intermediárias de neurônios, para o Esq 01 dos Alcalóides aporfínicos.....	182
Figura 72	Ciclos x Coeficiente de correlação do treinamento da rede Perceptron, com quatro camadas intermediárias de neurônios, para o Esq 01 dos Alcalóides aporfínicos.....	182
Figura 73	(a) Erro médio para rede de 3 camadas. (b) Erro médio para a rede de 4 camadas.....	184
Figura 74	Erro individual da substância de Teste A (a) para redes de 3 camadas e (b) para redes de 4 camadas; em função da variação do número de ciclos.....	185
Figura 75	Variação do coeficiente de correlação em função da taxa de aprendizado.....	186

Lista de Tabelas

Tabela 1	Distribuição dos alcalóides em plantas (esqueletos/substâncias).....	8
Tabela 2	Principais famílias de plantas, contendo um número mínimo de 100 alcalóides.....	9
Tabela 3	Alcalóides derivados de plantas com utilização farmacêutica.....	12
Tabela 4	Pesos atribuídos aos átomos e ligações para construção da matriz topológica.	54
Tabela 5	Matriz topológica do alcalóide Quinolínico X.....	55
Tabela 6	Listagem do número de ocorrências de esqueletos de Alcalóides.....	65
Tabela 7	Listagem das ocorrências botânicas dos alcalóides na família Rutaceae.....	72
Tabela 8	Principais Macronós presentes nos alcalóides.....	76
Tabela 9	Dados de RMN ¹³ C da substância colocada em teste.....	82
Tabela 10	Probabilidade de esqueletos apresentada pelo programa SISCONST.....	83
Tabela 11	Sinais associados as subestruturas A, B e C.....	83
Tabela 12	Análise dos Esqueletos mais prováveis fornecidos pelo programa C13MACH...	86
Tabela 13	Percentuais de ocorrência de esqueletos de alcalóides na família Fabaceae...	87
Tabela 14	Propostas de esqueletos para o programa H1MACH.....	89
Tabela 15	Dados de entrada do Teste 01.....	98
Tabela 16	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 01.....	98
Tabela 17	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 01.....	99
Tabela 18	Dados de entrada do Teste 02.....	100
Tabela 19	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 02.....	100
Tabela 20	Dados de entrada do Teste 03.....	101
Tabela 21	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 03.....	101
Tabela 22	Dados de entrada do Teste 04.....	103
Tabela 23	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 04.....	103
Tabela 24	Dados de entrada do Teste 05.....	104
Tabela 25	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 05.....	104
Tabela 26	Dados de entrada do Teste 06.....	105
Tabela 27	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 06.....	105
Tabela 28	Dados de entrada do Teste 07.....	106
Tabela 29	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 07.....	106
Tabela 30	Dados de entrada do Teste 08.....	107
Tabela 31	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 08.....	107
Tabela 32	Dados de entrada do Teste 09.....	108
Tabela 33	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 09.....	108
Tabela 34	Dados de entrada do Teste 10.....	109

Tabela 35	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 10.....	109
Tabela 36	Dados de entrada do Teste 11.....	110
Tabela 37	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 11.....	110
Tabela 38	Dados de entrada do Teste 12.....	111
Tabela 39	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 12.....	111
Tabela 40	Dados de entrada do Teste 13.....	112
Tabela 41	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 13.....	112
Tabela 42	Dados de entrada do Teste 14.....	113
Tabela 43	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 14.....	113
Tabela 44	Dados de entrada do Teste 15.....	114
Tabela 45	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 15.....	114
Tabela 46	Dados de entrada do Teste 16.....	115
Tabela 47	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 16.....	115
Tabela 48	Dados de entrada do Teste 17.....	116
Tabela 49	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 17.....	116
Tabela 50	Dados de entrada do Teste 18.....	117
Tabela 51	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 18.....	117
Tabela 52	Dados de entrada do Teste 19.....	118
Tabela 53	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 19.....	118
Tabela 54	Dados de entrada do Teste 20.....	119
Tabela 55	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 20.....	119
Tabela 56	Dados de entrada do Teste 21.....	120
Tabela 57	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 21.....	120
Tabela 58	Dados de entrada do Teste 22.....	121
Tabela 59	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 22.....	121
Tabela 60	Dados de entrada do TESTE 23.....	122
Tabela 61	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 11.....	122
Tabela 62	Dados de entrada do Teste 24.....	123
Tabela 63	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 24.....	123
Tabela 64	Dados de entrada do Teste 25.....	124
Tabela 65	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 25.....	124
Tabela 66	Dados de entrada do Teste 26.....	125
Tabela 67	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 26.....	125
Tabela 68	Dados de entrada do Teste 27.....	126
Tabela 69	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 27.....	126
Tabela 70	Dados de entrada do Teste 28.....	127
Tabela 71	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 28.....	127
Tabela 72	Dados de entrada do Teste 29.....	128
Tabela 73	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 29.....	128

Tabela 74	Dados de entrada do Teste 30.....	129
Tabela 75	Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 30.....	129
Tabela 76	Linha da planilha que representa a substância 7-metilideidroglaucone.....	173
Tabela 77	Deslocamentos químicos dos Alcalóides aporfínicos que serão identificados pela rede.....	177
Tabela 78	Resultados dos treinamentos da rede para identificação dos esqueletos das substâncias testadas.....	180
Tabela 79	Resumo da identificação de Esqueletos.....	181
Tabela 80	Correlações estatísticas obtidas para o treinamento da Perceptron com 3 Camadas intermediárias para o Esqueleto Aporfínico 01.....	183
Tabela 81	Correlações estatísticas obtidas para o treinamento da Perceptron com 4 Camadas intermediárias para o Esqueleto Aporfínico 01.....	184
Tabela 82	Resumo dos resultados obtidos pelo treinamento dos Esqueletos 02, 05 e 06..	186

I - INTRODUÇÃO

O ser humano, desde a antigüidade, faz uso de plantas tanto para alimentação quanto para fins medicinais. No início, como não dispunha de muita tecnologia, limitava-se a estudar apenas o cultivo e armazenagem destas plantas para fins alimentícios. Com a evolução, o homem passou a fazer uso medicinal das plantas, estudando suas atividades biológicas e seus respectivos efeitos sobre o organismo.

O passo seguinte, foi a obtenção de extratos concentrados das plantas que apresentavam atividades biológicas, potencializando assim seus efeitos. Entretanto, nestes extratos estavam presentes diversas substâncias, que poderiam ou não contribuir com o efeito farmacológico apresentado pela planta. Surgiu então a necessidade de isolar as substâncias biologicamente ativas, das demais encontradas nos extratos.

A idéia de produzir sinteticamente as substâncias ativas, isoladas dos extratos, tornou necessário a elucidação estrutural das mesmas. Assim o desenvolvimento das técnicas de determinação estrutural passou a ser um estudo de grande interesse à comunidade científica.

As técnicas de determinação estrutural, inicialmente, consistiam na degradação das substâncias, para identificação dos elementos presentes; reações características de grupos funcionais seguidos de propostas estruturais, as quais nem sempre estavam corretas. As estruturas eram elucidadas baseando-se principalmente em suas propriedades químicas, visto que os dados físico-químicos disponíveis na época eram somente as temperaturas de fusão e ebulição e a solubilidade em solventes. Dependendo da complexidade da substância, sua elucidação estrutural levava meses e mesmo assim, os resultados não eram 100% confiáveis.

A determinação estrutural das substâncias orgânicas, se desenvolveu lentamente por muitos anos, até que na década de 50 começaram a surgir os equipamentos que revolucionariam a determinação estrutural; iniciou-se então a era da espectroscopia molecular.

As técnicas espectrométricas permitem uma melhor caracterização da amostra. Em muitos casos, a partir dos dados espectrométricos, é possível identificar uma substância desconhecida. Entretanto, para substâncias complexas, a elucidação estrutural torna-se uma tarefa difícil ou as vezes quase

impossível. A necessidade de obter respostas rápidas e elucidar estruturas complexas levou o químico orgânica a fazer uso do Sistemas Especialistas como uma ferramenta auxiliar na determinação estrutural.

“Os Sistemas Especialistas são aqueles que (com auxílio da inteligência artificial) lidam com problemas complexos do mundo real e exigem a interpretação de um especialista; solucionam estes problemas através do uso de um modelo computacional do raciocínio de um especialista humano, chegando as mesmas conclusões que este especialista humano chegaria caso se defrontasse com um problema comparável¹”.

Nos últimos anos, vários grupos de pesquisadores em **inteligência artificial** construíram sistemas especialistas contendo a perícia necessária para solucionar problemas de diagnósticos e tratamentos médicos², exploração geológica³, análise de estruturas químicas⁴⁻¹³, entre outros. A seguir serão apresentados alguns Sistemas Especialistas desenvolvidos na área da química orgânica.

- ❖ LHASA¹⁴ – Faz a previsão dos reagentes de partida utilizados para sintetizar uma determinada substância através das propostas de retrossíntese da mesma.
- ❖ CAMEO¹⁵ – Prevê, logicamente, os possíveis produtos gerados a partir de uma reação orgânica.
- ❖ MECHGEN¹⁶ – Utiliza cálculos cinéticos e dados termodinâmicos para propor mecanismos de reações orgânicas.
- ❖ Os sistemas desenvolvidos para determinação estrutural serão vistos com maiores detalhes.

O sistema especialista utilizado neste trabalho é o SISTEMAT¹⁷⁻²⁰. O SISTEMAT começou a ser desenvolvido em 1988 no Instituto de Química da Universidade de São Paulo; desde então novos módulos e programas aplicativos vem sendo criados e/ou aperfeiçoados, sempre buscando maiores facilidades para o usuário e o fornecimento de propostas de esqueletos cada vez mais confiáveis.

O SISTEMAT utiliza uma abordagem multiespectral aliada a dados botânicos e tem como principal objetivo auxiliar pesquisadores, da área de produtos naturais, no processo de determinação estrutural de substâncias e em estudos quimiotaxonômicos.

Este trabalho foi desenvolvido com o objetivo de auxiliar na elucidação estrutural de uma variedade de substâncias pertencentes às subclasses dos alcalóides quinolínicos, quinolizidínicos, aporfínicos, benzilisoquinolínicos, isoquinolínicos, pirrolizidínicos, acridônicos e indólicos; utilizando o Sistema Especialista SISTEMAT, apresentando os programas aplicativos criados para o reconhecimento de esqueletos e determinação estrutural das classes estudadas. Também foi utilizado um programa de redes neurais no qual as redes foram treinadas para auxiliar na determinação estrutural dos alcalóides aporfínicos, fornecendo a probabilidade de uma determinada substância pertencer ao esqueleto pesquisado.

Para realização deste trabalho foi construído um banco de dados contendo 1182 alcalóides, sendo todos coletados da literatura e obtidos por extração de produtos naturais ou síntese. Nestes banco, estão presentes 1156 espectros de RMN ^{13}C , 354 espectros de RMN ^1H , 320 espectros de massas e as substâncias de origem botânica estão distribuídos em 49 famílias, 164 gêneros e 260 espécies.

II - ALCALÓIDES

II.1 - Introdução

O grande interesse em alcalóides advém muito provavelmente das intensas ações farmacológicas pronunciadas em um largo espectro de ação²¹.

A história dos alcalóides está intimamente ligada a história do homem e da civilização. O registro do uso de chás, poções e remédios contendo alcalóides remota há mais de 4000 anos. Um exemplo histórico marcante é o suicídio compulsório de Sócrates pelo uso de extrato de cicuta em 399 a.C.

Somente a partir do século XIX, os farmacêuticos/químicos, contando com equipamentos, técnicas e uma fundamentação teórica mais avançada, conseguiram isolar e investigar essas substâncias em forma semi-pura. Um exemplo é a morfina, que foi isolada em forma pura em 1805 por Serturmer²². Por volta de 1820 vários outros alcalóides já haviam sido isolados de plantas por Pelletier e Caventou, tais como a estricnina, quinina, emetina e colchicina.

A primeira elucidação estrutural de um alcalóide ocorreu em 1870 com a determinação da coniina (Fig.1), princípio tóxico presente no extrato de cicuta (*Conium maculatum* L.). A coniina foi também o primeiro alcalóide a ser sintetizado, em 1886, por A. Ladenburg²³.

Os alcalóides são encontrados principalmente em angiospermas (vegetais superiores que apresentam flores (Fig. 1)). Entretanto, os alcalóides não estão limitados ao reino vegetal, eles também são encontrados em animais, insetos e microorganismos.

II.2- Definições e classificação dos alcalóides.

Como os alcalóides, inicialmente descobertos, eram derivados de plantas, começaram a surgir definições precoces de que os alcalóides eram substâncias que possuíam três características: caráter básico, presença de Nitrogênio e derivado de plantas. Com o desenvolvimento e isolamento de novas estruturas, as definições ficaram mais claras; o conceito de ser derivados também de aminoácidos foi somado à definição inicial, juntamente com a idéia da presença de um nitrogênio heterocíclico.

Figura 1 - vegetais superiores que apresentam alcalóides.

Na definição de R. Hagnaeur²⁴, “os alcalóides são substâncias mais ou menos tóxicas as quais atuam primariamente no sistema nervoso central. Eles têm um caráter básico, contêm Nitrogênio heterocíclico e são sintetizados por plantas a partir de aminoácidos ou seus derivados imediatos. Na maioria dos casos estão distribuídos no reino vegetal”.

Os alcalóides também são divididos em alcalóides Verdadeiros, Protoalcalóides e Pseudoalcalóides :

a) Alcalóides Verdadeiros (Fig. 2): são substâncias tóxicas apresentando uma grande gama de atividades fisiológicas; são invariavelmente básicos; contém nitrogênio em anel heterocíclico; são derivados de aminoácidos (AAs); são limitados quanto a distribuição taxonômica pois ocorrem somente em algumas famílias de vegetais .

Figura 2 – Exemplos de alcalóides verdadeiros.

b) Protoalcalóides (Fig. 3): são aminas simples, onde o nitrogênio não se encontra num heterociclo. São derivados de AAs e tem propriedades básicas. O termo “aminas biológicas” é freqüentemente empregado para definir estas substâncias .

Figura 3 – Exemplos de protoalcalóides .

c) Pseudoalcalóides: são substâncias básicas, mas não derivadas de AAs .

Figura 4 – Exemplos de pseudoalcalóides.

Embora ainda muito utilizada, a definição proposta por Hagnauer foi adaptada de acordo com a descoberta de novas substâncias nos últimos anos. Por exemplo, a definição original menciona unicamente plantas superiores (angiospermas) como os produtores de alcalóides. Entretanto, os alcalóides já foram isolados de animais (alcalóides esteroidais), insetos, organismos marinhos, microorganismos e plantas inferiores, o que invalida esta condição na definição de Hagnauer.

Um aspecto também ultrapassado da definição de alcalóides por Hagnauer é a atribuição obrigatória de atividade fisiológica sobre o sistema nervoso central (SNC). As atividades dos alcalóides nem sempre estão associadas ao SNC. Alguns pesquisadores (Peter G. Waterman²⁵, por exemplo) questionam inclusive a validade do conceito de atividade biológica como quesito relevante para a classificação dos alcalóides, ou mesmo para classificação taxonômicas em geral.

II.3 – Evolução das pesquisas com ênfase nos alcalóides derivados de plantas²⁶.

Seguindo uma ordem cronológica, o número de alcalóides derivados de plantas e caracterizados até 1950 eram aproximadamente 1000; em 1973 chegou a 3300 o número de estruturas elucidadas. Seguindo os extraordinários avanços das técnicas espectroscópicas dos últimos 30 anos, uma análise recente do banco de dados do NAPRALERT^{sm27} indicou 26900 estruturas conhecidas de alcalóides derivados de fontes naturais variadas, sendo que

21120 são derivadas de plantas e estão distribuídos em 7231 espécies, 1730 gêneros e em 186 famílias. Foram detectadas a presença de alcalóides em 35 novas famílias, os quais ainda não foram isolados e caracterizados; e 153 outras famílias sequer foram testadas para verificar se possuem ou não alcalóides em seus extratos.

Um total de 1872 esqueletos dos 21120 alcalóides descritos na literatura estão distribuídos em monocotiledônias, dicotiledônias e gimnospermas, conforme a sua subclasse (Tab.1).

Tabela 1 – Distribuição dos alcalóides em plantas (esqueletos/substâncias).

Classe dos alcalóides	Monocotiledônias	Dicotiledônias	Gimnosperma	Total
	Esquel / Subst.	Esquel / Subst.	Esquel / Subst.	Esquel / Subst.
Aminas Simples	17 / 98	48 / 1141	6 / 20	71 / 1213
Pirrolidínico/piperidínico	19 / 126	51 / 773	4 / 32	74 / 898
Piridínico	3 / 22	33 / 299	1 / 5	37 / 314
Tropânico	2 / 6	13 / 495	1 / 4	16 / 502
Pirrol / izidínico	11 / 40	20 / 855	0	31 / 884
Indolizidínico	4 / 5	21 / 240	0	25 / 245
Quinolizidínico	3 / 5	45 / 768	0	48 / 771
Quinolínico	3 / 7	125 / 1175	3 / 6	131 / 1184
Isoquinolínico	71 / 739	287 / 5040	14 / 175	372 / 5872
Indólico	30 / 142	539 / 5112	2 / 6	571 / 5191
Monoterpênico	1 / 1	23 / 132	0	24 / 132
Sesquiterpênico	5 / 12	30 / 214	0	35 / 226
Diterpênico	2 / 4	73 / 1302	3 / 3	78 / 1307
Triterpênico	1 / 1	29 / 122	0	30 / 123
Esteroidal	32 / 329	55 / 547	0	87 / 849
Peptídico	1 / 2	20 / 160	0	21 / 161
Espermínico	2 / 8	12 / 217	5 / 8	19 / 221
Flavonoides	1 / 1	5 / 18	0	6 / 19
Misto com 1 Nitrogênio	29 / 143	62 / 236	0	91 / 366
Misto com 2 N	16 / 52	55 / 386	3 / 4	73 / 411
Misto com 3 N	1 / 2	11 / 14	0	12 / 15
Misto com 4 N	4 / 29	15 / 203	1 / 12	20 / 216
Total	258 / 1774	1571 / 19449	43 / 275	1872 / 21120

As principais famílias de plantas, das quais estão isolados um número mínimo de 100 alcalóides, estão discriminadas na Tabela 2. Vale lembrar que nestas famílias e nas demais, cujo número de alcalóides são inferiores a 100,

existem ainda muitas espécies não estudadas (mais de 30000), o que indica que a pesquisa com alcalóides, embora antiga, ainda será fonte de inúmeros trabalhos e muitos avanços medicinais.

Tabela 2 - Principais famílias de plantas, contendo um número mínimo de 100 alcalóides.

Famílias De plantas	Número de Alcalóides	Famílias De plantas	Número de Alcalóides
Apocynaceae	2664	Buxaceae	243
Rutaceae	1730	Rhamnaceae	210
Ranunculaceae	1559	Erythroxylaceae	193
Fabaceae	1453	Euphorbiaceae	191
Papaveraceae	1309	Convolvulaceae	185
Menispermaceae	922	Asclepiadaceae	180
Annonaceae	849	Brassicaceae	180
Solanaceae	793	Simaroubaceae	171
Asteraceae	705	Aristolochiaceae	160
Rubiaceae	677	Hernandiaceae	146
Liliaceae	667	Cactaceae	135
Loganiaceae	563	Sterculiaceae	128
Amaryllidaceae	448	Magnoliaceae	123
La uraceae	425	Zygophyllaceae	117
Berberidaceae	373	Monimiaceae	115
Boraginaceae	287	Apiaceae	114
Piperaceae	258	Scrophulariaceae	106
Poaceae	256	Rosaceae	104
Celastraceae	252		

II.4 - Biossíntese e atividades biológicas dos alcalóides²⁸

Como mencionado anteriormente, alcalóides verdadeiros e protoalcalóides são substâncias derivadas de α -aminoácidos. Há portanto uma estreita correlação entre essas duas classes de substâncias. Os principais precursores presentes na biossíntese de alcalóides são a ornitina, lisina, fenilalanina, tirosina, ácido antranílico, triptofano, histidina e ácido-nicotínico (Esquema 1).

Esquema 1 – Biossíntese dos Alcalóides.

A biossíntese de alcalóides com estrutura mais complexa é também uma área de intensa investigação devido aos aspectos comerciais e medicinais associados a essas substâncias. Um bom exemplo de uma classe de alcalóides de grande interesse farmacêutico são os alcalóides indol-

monoterpênicos, uma classe de alcalóides com mais de 1800 substâncias. Muitos deles são agentes terapêuticos de eficácia comprovada e são obtidos normalmente de fontes naturais em larga escala para fins de comercialização. Os alcalóides indol-monoterpênicos - assim chamados porque são derivados do aminoácido triptofano (que contém o anel indólico) e do monoterpene secologanina - tem como precursor comum a 3 α (S)-strictosidina, como mostrado na Figura 5.

A biossíntese da 3 α (S)-strictosidina é feita diretamente a partir da triptamina e da secologanina pela ação da strictosidina sintase, como mostrado no Esquema 2.

Figura 5 - Precursor comum dos alcalóides indol-monoterpênicos

Esquema 2 – Biossíntese da strictosidina.

Como mencionado anteriormente, o maior interesse em pesquisar alcalóides está na variada gama de atividades biológicas apresentadas pelos mesmos. Algumas destas atividades são mostradas na Tabela 3.

Tabela 3- Alcalóides derivados de plantas com utilização farmacêutica²⁶.

Alcaloide	Família	Gênero / espécie	Uso farmacêutico
Aconitina	Ranunculaceae	<i>Aconitum napellus</i>	Reumatismo, nevralgia
Aimalicina	Apocynaceae	<i>Catharanthus roseus</i>	Problemas circulatórios
Aimalina	Apocynaceae	<i>Rauvolfia serpentina</i>	Arritmia cardíaca
Anabasina	Fabaceae	<i>Anabasis aphylla</i>	Relaxante muscular
Arecolina	Palmae	<i>Areca catechu</i>	Vermífugo
Atropina	Solanaceae	<i>Atropa belladonna</i>	Antiespasmódico
Berberine	Berberidaceae	<i>Berberis vulgares</i>	Desintéria, infecções nos olhos
Cafeína	Rubiaceae	<i>Coffea arabica</i>	Estimulante SNC.
Captothecina	Icacinaeae	<i>Captotheca acuminata</i>	Antitumoral
Cocaína	Erythroxylaceae	<i>Erythroxylum coca</i>	Anestésico local
Deserpidina	Apocynaceae	<i>Rauvolfia canescens</i>	Anti-hipertensivo
L-dopa	Fabaceae	<i>Mucuna derringiana</i>	Tratamento do mal de Parkinson.
Efedrina	Ephedraceae	<i>Ephedra sinica</i>	Descongestionante e antiasmático
Glaucina	Papaveraceae	<i>Glaucium flavum</i>	Antitosse
Glaziovina	Lauraceae	<i>Ocotea glaziovii</i>	Antidepressivo
Huperzina A	Lycopodiaceae	<i>Huperzia serrata</i>	Demência senil (Senile dementia)
Lobeline	Campanulaceae	<i>Lobelia inflata</i>	Combate o tabagismo
Morfina	Papaveraceae	<i>Papaver somniferum</i>	Analgésico
Nicotina	Solanaceae	<i>Nicotiana tabacum</i>	Inseticida
Palmatine	Berberidaceae	<i>Coptis japonica</i>	Antipirético
Quinina	Rubiaceae	<i>Cinchona ledgeriana</i>	Antimalárico
Rescinnamina	Apocynaceae	<i>Rauvolfia serpentina</i>	Anti-hipertensivo, tranquilizante
Sanguinarina	Papaveraceae	<i>Sanguinaria canadensis</i>	Inibidor de placas dentais
Taxol	Taxaceae	<i>Taxus brevifolia</i>	Antitumoral
Theobromine	Sterculiaceae	<i>Theobroma cacao</i>	Diurético e vasodilatador
Yohimbine	Rubiaceae	<i>Pausinystalia yohimba</i>	Afrodisíaco

A partir desta breve revisão dos alcalóides, justificamos a importância do trabalho com a elucidação estrutural destas substâncias, juntamente com a construção de um banco de dados mais completo para utilização do Sistema Especialista SISTEMAT.

Como mencionado anteriormente, as subclasses de alcalóides utilizadas neste trabalho foram os alcalóides quinolínicos, quinolizidínicos, aporfínicos, benzilisoquinolínicos, isoquinolínicos, pirrolizidínicos, acridônicos e indólicos, de origem natural e/ou sintética; cujos respectivos núcleos característicos estão apresentados na Figura 6.

Figura 6 – Núcleos característicos dos alcalóides presentes no banco de dados.

A amostragem de alcalóides, inseridos no banco de dados do programa SISTEMAT, equivale apenas à 4,4% do total de alcalóides isolados apenas de produtos naturais²⁷, fora os alcalóides de origem sintética. Verificada a eficiência dos programas aplicativos do SISTEMAT na determinação estrutural das variadas subclasses dos alcalóides, pode-se então, começar a trabalhar com bancos completos das respectivas subclasses individualmente. Outra opção seria trabalhar com bancos completos, porém classificando os alcalóides por diferenciação de esqueletos encontrados em uma determinada família de plantas. Como a origem botânica é um dado já conhecido, mesmo antes da obtenção de qualquer extrato, pressupõe-se que um banco, construído com alcalóides de uma determinada família, teria mais utilidade para o químico de produtos naturais que estivesse estudando esta família, por exemplo, um banco da família Apocynaceae teria até 2664 alcalóides diferentes, como foi observado na Tabela 2.

Os 1183 alcalóides presentes no banco de dados estão distribuídos em 282 esqueletos diferentes (Fig. 7).

Figura 7- Esqueletos dos Alcalóides quinolínicos, aporfínicos, benzilisoquinolínicos, acridônicos, quinolizidínicos, pirrolizidínicos, isoquinolínicos e Indólicos; presentes na base de dados.

Figura 7- Continuação

Figura 7 – Continuação

Figura 7 – Continuação

QUINOLIZIDÍNICO 5

QUINOLIZIDÍNICO 6

QUINOLIZIDÍNICO 7

QUINOLIZIDÍNICO 8

QUINOLIZIDÍNICO 9

QUINOLIZIDÍNICO 10

QUINOLIZIDÍNICO 11

QUINOLIZIDÍNICO 12

QUINOLIZIDÍNICO 13

QUINOLIZIDÍNICO 14

QUINOLIZIDÍNICO 15

QUINOLIZIDÍNICO 16

QUINOLIZIDÍNICO 17

QUINOLIZIDÍNICO 18

QUINOLIZIDÍNICO 19

Figura 7 – Continuação

Figura 7 – Continuação

Figura 7 – Continuação

BENZILISOQUINOLÍNICO 17

BENZILISOQUINOLÍNICO 18

BENZILISOQUINOLÍNICO 19

BENZILISOQUINOLÍNICO 20

BENZILISOQUINOLÍNICO 21

BENZILISOQUINOLÍNICO 22

ACRIDÔNICO 1

ACRIDÔNICO 2

ACRIDÔNICO 3

ACRIDÔNICO 4

ACRIDÔNICO 5

ACRIDÔNICO 6

Figura 7 - Continuação

Figura 7 – continuação

Figura 7 – Continuação

Figura 7 – Continuação.

Figura 7 - Continuação

Figura 7 – Continuação

Figura 7 - Continuação

Figura 7 - Continuação

Figura 7 - Continuação

Figura 7 - Continuação

Figura 7 – Continuação

Figura 7 – Continuação

Figura 7 – Continuação

III – Sistemas Especialistas em determinação estrutural.

A evolução que ocorreu na espectroscopia a partir dos anos 60, com as técnicas avançadas dos aparelhos de ressonância magnética nuclear, possibilitou a aquisição de um grande número de dados espectrométricos de substâncias de origem natural e sintética. Entretanto, a velocidade de interpretação desses dados, muitas vezes, não consegue acompanhar a velocidade com que os mesmos são gerados, tornando o processo de determinação estrutural ineficiente. Um dos maiores desafios atuais, de quem trabalha com determinação estrutural (principalmente se tratando de produtos naturais), está na interpretação dos dados espectrométricos obtidos, principalmente quando se trata de ressonância magnética nuclear, que é hoje a técnica de onde se extraíam as mais valiosas informações em determinação estrutural. Para tentar agilizar esta fase do processo, nos últimos trinta anos, tem sido criados inúmeros sistemas especialistas em determinação estrutural de substâncias.

A rapidez com que os computadores trabalham, permite que uma resposta seja alcançada em tempo satisfatório se o número de combinações for pequeno, coisa que não ocorre quando o assunto é geração de estruturas.

Quando o problema em questão apresenta uma enorme variedade de caminhos a serem verificados, torna-se necessário a eliminação daqueles que possuem menor possibilidade de chegar a resposta desejada. Esta busca reduzida é denominada Busca Heurística. Ela é obtida pela imposição de regras as quais delimitam o espaço de busca, evitando assim a explosão combinatória.

Os programas computacionais, desenvolvidos para determinação estrutural, contêm uma grande base de dados que correlacionam os parâmetros espectrométricos, as respectivas estruturas químicas e o conhecimento dos especialistas. O mecanismo utilizado pela maioria dos sistemas especialistas pode ser resumido da seguinte forma:

- Obtenção do maior número de dados sobre o problema;
- Representação organizada destes dados para o sistema;

- Flexibilidade para aquisição de novos dados que venham a surgir posteriormente;
- Geração de propostas estruturais;
- Um grau de confiabilidade das propostas estruturais geradas;
- Facilidade e interação com o usuário e
- Acessibilidade do conhecimento humano especializado.

Atualmente existem vários Sistemas Especialistas em determinação estrutural. Como exemplos podemos citar os Sistemas Especialistas DENDRAL⁴, CHEMICS⁵, CASE⁶, ACCES⁷, DARC/EPIOS^{8,9}, SpecInfo¹⁰, X-PERT¹¹, ASSEMBLE¹² e SENECA¹³, entre outros. Alguns destes programas serão apresentados a seguir com mais detalhes .

III.1 – Sistema DENDRAL⁴

O DENDRAL foi o primeiro e mais famoso Sistema Especialista em determinação estrutural de substâncias orgânicas. Sua construção iniciou-se na década de 60 na Universidade de Stanford. Ele surgiu do interesse do Prof. Lederberg pela isomeria química, interesse este que o fez criar um algoritmo combinatorial capaz de gerar inúmeras estruturas isoméricas, a partir de uma determinada composição química. O Sistema DENDRAL resultou em muitas publicações e é citado por praticamente todos os livros que abordam o assunto de determinação estrutural através de Sistemas Especialistas.

O Sistema DENDRAL foi desenvolvido para uso em computadores de grande porte. Os dados espectrométricos mais utilizados são a Ressonância Magnética Nuclear de Carbono 13 (RMN ¹³C) e a espectrometria de massas.

Funcionamento do Sistema.

O funcionamento do sistema DENDRAL está dividido em três etapas: Planejamento, Geração de estruturas e Teste para avaliação das estruturas geradas. Diversos testes do sistema com substâncias naturais já foram realizados e publicados. Um exemplo de teste, importante e detalhado, da utilização do sistema DENDRAL foi realizado na identificação do sesquiterpeno *warburganal*²⁹, isolado em 1976 da *Warburgia ugandensis* (Fig. 8). As etapas desse teste são mostradas a seguir.

Figura 8 – Estrutura do Warburganal e a planta *Warburgia ugandensis*.

Planejamento –

O planejamento é a etapa crucial do processo. Envolve uma atuante participação do usuário, uma vez que este deverá inicialmente impor restrições ao programa, fornecendo a fórmula molecular e os fragmentos que ele julgar estarem presentes ou ausentes na substância em teste. Estes fragmentos são inferidos baseando-se na análise dos dados espectrais.

Para o Warburganal, foi necessário fornecer ao programa a fórmula molecular ($C_{12}H_{22}O_3$) e as seguintes informações estruturais apresentadas na Figura 9.

FRAGMENTOS PRESENTES

FRAGMENTOS AUSENTES

Figura 9 – Fragmentos inferidos por interpretação dos dados espectrais.

Geração de estruturas

Os módulos geradores do DENDRAL utilizam os dados fornecidos pelo usuário na fase do planejamento e geram todos os isômeros possíveis que satisfaçam as restrições impostas. Como o número de estruturas geradas pode ser muito grande, são utilizadas técnicas de GERAR-TESTAR para eliminar propostas contraditórias às regras de ligação ou às restrições impostas pelos fragmentos ausentes, fornecidos pelo usuário.

Nesta etapa, o usuário também tem um papel importante, pois o programa gerador de estruturas começa a criar novos fragmentos maiores e estruturas parciais que deverão ser novamente avaliadas e selecionadas. O usuário, então, pode eliminar ou manter as subestruturas que serão utilizadas para a geração de propostas estruturais. Com essa prévia seleção de subestruturas, o tempo computacional é reduzido e os grandes equívocos nas gerações de estruturas também diminuem. Cada subestrutura parcial apresentada é denominada CASO.

Para o Warburganal foram aprovados pelo usuário 17 casos, que, quando utilizados pelo programa gerador de estruturas forneceram 42 propostas estruturais compatíveis com todas as restrições impostas anteriormente.

Testes das estruturas geradas

Todas as estruturas propostas na fase de geração são posteriormente testadas. Os testes são feitos de duas formas diferentes:

- Confrontando as estruturas propostas com bancos de esqueletos de produtos naturais, verificando se estas estruturas incorporam esqueletos ou fragmentos de esqueletos de produtos naturais e
- Simulando os espectros de RMN ^{13}C e massas, comparando-os com os dados originais de modo a obter um índice de semelhança. O programa então fornece as estruturas com o melhores índices.

Das 42 estruturas geradas, apenas uma tinha o esqueleto completo de um produto natural e esta correspondia à estrutura do Warburganal. O mesmo resultado foi obtido quando se efetuou a confrontação dos dados

espectrométricos de RMN ^{13}C e massas, simulados pelo programa, com os dados reais, classificando o Warburganal como a estrutura que apresentou o maior índice de similaridade.

III.2 – Sistema CHEMICS⁵

O sistema CHEMICS (Combined Handling of Elucidation Methods for Interpretable Chemical Structures) foi projetado para determinação estrutural de substâncias orgânicas usando uma abordagem multiespectral. O CHEMICS teve seu desenvolvimento iniciado em 1967 na Universidade de Educação de Miyagi, Japão.

Existem duas abordagens distintas para a elucidação estrutural assistida por computador. A primeira delas consiste em montar um banco de dados espectrais o mais completo possível, inserindo as substâncias por meio de codificações. Uma outra alternativa consiste em construir um banco de fragmentos de estruturas. Com algumas poucas centenas de fragmentos pode-se construir praticamente qualquer molécula. O sistema CHEMICS trabalha unindo as duas técnicas citadas acima. Em uma primeira etapa, são geradas propostas de estruturas, que posteriormente têm seus dados confrontados com um banco de dados espectrais.

Funcionamento do Sistema

O sistema contém quatro módulos principais e mais de 30 sub-programas que compõem os módulos. A utilização do sistema segue a seguinte ordem de passos:

- Entrada dos dados disponíveis;
- Análise dos dados fornecidos;
- Uso do gerador de estruturas;
- Comparação dos dados analíticos com os dados contidos no banco.

Entrada dos dados disponíveis

Este módulo do sistema é o responsável pela entrada dos dados que caracterizam uma determinada substância, ou seja, fórmula molecular e dados espectrais, como: infravermelho, ultravioleta, espectro de massas, RMN ^1H e RMN ^{13}C . Algumas restrições também podem ser introduzidas, como a certeza

da presença ou ausência de um determinado grupo funcional. Esta fase é de fundamental importância e influi diretamente no resultado final da análise. Quanto mais completo for o conjunto de dados fornecidos ao sistema, menor será o número de estruturas propostas, podendo até mesmo ser única.

Análise dos dados fornecidos

O sistema CHEMICS contém mais de 570 fragmentos codificados em seus bancos de dados. A função deste módulo é analisar os dados da substância em teste e descartar o maior número possível de fragmentos. A fórmula molecular, caso exista, é analisada em primeiro lugar. São anotados o número de carbonos e de heteroátomos; também é calculado o índice de deficiência de hidrogênio. Os fragmentos são analisados e descartados, no caso de haver contradições entre os dados obtidos da análise da fórmula molecular e os fragmentos propostos. Por exemplo, se a fórmula molecular apresenta apenas um átomo de oxigênio, os fragmentos com um número superior de oxigênios serão excluídos. Fragmentos contendo ligações triplas serão excluídos se o cálculo do índice de deficiência de hidrogênio fornecer um valor igual a 1(um).

Os dados da espectrometria no infravermelho, caso existam, são analisados. O sistema contém caracterizações dos principais grupos funcionais existentes em substâncias orgânicas. A função de teste aplicada ao espectro no infravermelho fornece um resultado com três valores possíveis para cada grupo funcional. O valor -1 garante a ausência do grupo funcional em questão. Um valor + 1 garante a presença, enquanto que o valor 0 não permite afirmar nenhum dos dois extremos anteriores. Após a verificação da ausência de um determinado grupo funcional, todos os fragmentos que tenham este grupo são também descartados.

O espectro de ressonância magnética nuclear de hidrogênio, se existir, é analisado em terceiro lugar. Os sinais no espectro são agrupados a cada 20 Hz e o número de hidrogênios é calculado para cada grupo, a partir da área de cada sinal e da fórmula molecular. Fragmentos que não se adequarem aos dados assim calculados também serão descartados. Os demais métodos espectrométricos também fornecem informações que permitem o descarte de fragmentos.

Uso do gerador de estruturas

Os fragmentos que não foram descartados durante a fase de análise dos dados são fornecidos ao gerador de estruturas. O módulo de geração realiza três etapas distintas. A primeira etapa consiste em combinar todos os fragmentos selecionados, formando todos os conjuntos de fragmentos possíveis. Cada conjunto é confrontado com a fórmula molecular e descartado se esta não for satisfeita. Se o conjunto satisfizer à fórmula molecular, é feita uma verificação se as funções, descartadas pela análise do espectro no infravermelho, não surgiram pela união dos fragmentos. Em caso positivo, este conjunto também é descartado. Em uma segunda fase, os conjuntos aprovados serão submetidos à verificação de restrições impostas pelos demais dados espectrométricos. Finalmente, os conjuntos aprovados após passarem por todas as verificações descritas anteriormente são, em uma terceira etapa, usados pelo gerador propriamente dito. O gerador define as conexões entre os fragmentos de modo a preservar as regras de ligação, verificando se o número de ligações dos átomos está correto e a validade das mesmas.

Comparação dos dados analíticos com os dados contidos no banco

As estruturas propostas e que passaram por todos os testes das fases anteriores são submetidas à comparação de seus dados espectrais contra um banco de dados. Os dados da amostra são codificados de acordo com a codificação do banco de dados. A fórmula molecular, o conjunto de fragmentos e a conectividade entre estes fragmentos permitem localizar uma substância no banco de dados e recuperar seus dados espectrométricos codificados. A comparação entre os dois conjuntos de dados permite que o módulo de comparação forneça uma dentre três respostas possíveis:

- 1) A estrutura padrão não consta do banco de dados, portanto a comparação não é possível;
- 2) Os dois conjuntos de dados são claramente não equivalentes, portanto a estrutura proposta deve estar incorreta;
- 3) Os dois conjuntos de dados são equivalentes, portanto a estrutura proposta pode ser a correta.

Em qualquer uma das três situações anteriores, todas as estruturas geradas são apresentadas para a apreciação final do usuário. Este sistema é em vários pontos mais complexo do que o DENDRAL.

III.3 – Sistema DARC/EPIOS^{8,9a}

Este Sistema Especialista começou a ser desenvolvido, nos anos 60, na França. Atualmente utiliza principalmente a espectrometria de RMN ^{13}C e a fórmula molecular. Na Década de 70 alcançou um grande sucesso comercial para a determinação de estruturas pequenas. Não é orientado para química de produtos naturais e não utiliza uma abordagem multiespectral.

Funcionamento do Sistema

O usuário fornece ao programa a fórmula molecular e os dados do espectro de RMN ^{13}C . Ao contrário do DENDRAL, o Sistema Especialista DARC/EPIOS consegue deduzir automaticamente os fragmentos iniciais, tornando optativo a participação do usuário na etapa de planejamento, que mesmo sendo automático não o impede de fornecer ao programa as restrições que julgar importantes para a seleção dos fragmentos.

O programa dispõe de um banco de dados de ELCOs (Environment that is Limited, Concentric and Ordered, Ambiente Limitado Centrado Ordenado) originados de 11.000 espectros obtidos da literatura. Cada ELCO representa um carbono ressonante com a descrição de sua vizinhança química, ao qual está associado um deslocamento químico. Os 11.000 espectros foram fragmentados em milhares de ELCOs que formam o banco de dados do DARC/EPIOS.

O espectro-problema é fornecido ao programa que procura, no banco de dados, os ELCOs que satisfazem a todos os deslocamentos químicos deste espectro. Ao final da procura, obtém-se uma lista de fragmentos que são submetidos a uma rotina geradora de estruturas. O gerador focaliza um ELCO e atribui um deslocamento químico. A estrutura vai aumentando por sobreposição de novos ELCOs e o espectro vai sendo atribuído.

Para reduzir o número de estruturas propostas e diminuir o tempo de computação gasto, restrições podem ser impostas pelo usuário. Estas

restrições são fornecidas na forma de conectividades entre os átomos de carbono, normalmente obtidas do espectro INADEQUATE.

A Genipina^{9b} (Fig. 10) foi utilizada para testar o sistema DARC/EPIOS. A partir do espectro da substância foram selecionados 139 ELCOs e a estrutura correta foi obtida sem a participação do usuário, em um intervalo de tempo de 7,65 segundos. Fornecendo-se a conexão entre os carbonos **a** e **b**, a resposta foi obtida em 1,27 segundos.

Figura 10 – Estrutura do monoterpene Genipina

III.4 - Sistema SpecInfo/SpecSolv²⁰

O sistema SpecInfo de informação espectrométrica multidimensional^{20a} foi desenvolvido pela BASF em 1970, com o objetivo de criar uma ferramenta para auxiliar o espectroscopista analítico. Esse sistema encontra-se em sua terceira versão, utilizando uma interface gráfica do tipo X-Windows e sendo executado em estações de trabalho.

Em 1990, o SpecInfo já continha em sua base de dados, 16.000 espectros de massas, 110.000 espectros de RMN¹³C, 15.000 espectros de RMN ¹¹B, ¹⁵N, ¹⁷O, ¹⁹F e ³¹P e 20.000 espectros de infravermelho (IV). Para assegurar a confiabilidade do programa, cada espectro inserido na base de dados passa por um controle de qualidade, que realiza testes automatizados e também é inspecionado por um espectroscopista experiente, de modo que espectros de baixa qualidade são imediatamente descartados.

Funcionamento do Sistema

O SpecInfo tem como princípio a fundamentação teórica de que ambientes químicos de um determinado átomo em uma molécula influenciam as suas propriedades. Baseado nesse princípio, foi construído um sistema de codificação de subestruturas utilizando o código HOSE (Hierarchically Ordered

Spherical Description of Environment)³⁰⁻³¹. Esse código caracteriza o ambiente de cada átomo esfericamente ao redor de um centro específico de interesse, isto é, a partir de um nó central são descritos todos os átomos que o circundam até a terceira esfera, desta forma, os efeitos sobre acoplamento e deslocamento químico do átomo central fazem parte da informação calculada e armazenada para qualquer átomo na base de dados.

O resultado desse processo é um banco de dados contendo ambientes químicos de átomos, cuja aplicação não está limitada a classes específicas de substâncias. Uma vez que cada átomo da base de dados está correlacionado com um padrão espectral, é possível prever o espectro de uma substância, pois cada átomo será convertido automaticamente em uma lista de códigos HOSE. Através de uma busca na biblioteca de RMN ^{13}C , será fornecido um deslocamento químico para cada código HOSE da lista, portanto, um deslocamento químico para cada átomo de carbono da substância.

O sistema SpecSolv^{20b} é um novo módulo criado para o sistema SpecInfo. Esse módulo trabalha exclusivamente com RMN ^{13}C .

A elucidação estrutural com o SpecSolv é dividida em três etapas:

- a) Aquisição do espectro de RMN ^{13}C -DEPT experimental e extração dos deslocamentos químicos, multiplicidades e intensidades;
- b) A procura de subespectros em uma biblioteca de correlações subestrutura-subespectro (CSS) com posterior geração de uma lista com as melhores subestruturas; e
- c) União das subestruturas, através da sobreposição destas, seguido de um passo de validação das subestruturas intermediárias e do resultado final.

Base de Dados do SpecSolv

A base de dados do SpecSolv contém mais de 400.000 subestruturas geradas a partir dos códigos HOSE de terceira esfera e mais de 100.000 a partir de códigos HOSE de segunda esfera. Essas CSS foram originadas a partir da base de dados do SpecInfo que, atualmente, apresenta mais de 200.000 espectros de RMN ^{13}C . Cada CSS apresenta uma informação estrutural completa (subespectros, matriz de conectividade e código HOSE para todos os átomos) juntamente com todos parâmetros de RMN ^{13}C dos átomos de carbono (deslocamento químico, multiplicidade e intensidade).

Procura de Subestruturas

Consiste na comparação do espectro problema com todos os subespectros da base de dados. Durante a busca, são selecionados somente as subestruturas que apresentarem os átomos de carbonos com os deslocamentos químicos, multiplicidades e intensidades, dentro da faixa definida pelo usuário. As subestruturas selecionadas são classificadas de acordo com um fator de comparação que indica o desvio médio, em ppm, dos deslocamentos químicos do espectro problema e os encontrados nos subespectros, que geraram as subestruturas.

O SpecSolv não requer o conhecimento dos tipos de átomos do espectro questionado, mas restrições de heteroátomos multivalentes, como por exemplo, N, P e S, reduzem o número de CSSs propostos e, consequentemente, o tempo computacional.

Gerador de Estruturas

As subestruturas geradas são interligadas pela sobreposição dos átomos comuns as duas subestruturas (Fig. 11). Uma subestrutura maior é gerada e posteriormente validada com uma nova previsão espectral, baseado nos CSSs. Se o espectro previsto for compatível com o espectro questionado, a subestrutura gerada é aceita e utilizada como novo ponto de partida, até que todas as subestruturas sejam sobrepostas com a subestrutura de partida e/ou todos os deslocamentos químicos do espectro problema estejam completamente atribuídos. Caso o espectro previsto não seja compatível com o questionado, a subestrutura gerada é rejeitada, e o processo inicia-se com a escolha da próxima subestrutura melhor classificada. O processo de geração de estruturas do SpecSolv é descrito no Fluxograma da Figura 12.

Figura 11 – Sobreposição dos átomos comuns aos dois fragmentos para formação de uma subestrutura maior

Figura 12 – Fluxograma do processo de geração de estruturas do SpecSolv.

Observa-se que durante o processo de sobreposição das subestruturas individuais, os deslocamentos químicos da subestrutura resultante ficam disponíveis na memória central do computador, assim um novo subespectro

podem ser gerado pela simples combinação dos subespectros originais de uma forma rápida, pois não é necessário o acesso a uma base de dados.

A validação instantânea das estruturas intermediárias é uma das principais diferenças entre o SpecSolv e outros programas geradores de estruturas, que realizam a validação de estruturas somente após a geração da molécula inteira. Esse processo no SpecSolv proporciona uma importante redução do tempo computacional quando comparado a outros programas convencionais.

Para verificar a eficiência do SpecSolv como uma ferramenta na área da determinação estrutural, foi realizado um teste com o alcalóide pirrolizidínico **Uspalatina**³² (Fig. 13).

A elucidação estrutural da uspalatina foi realizada automaticamente pelo SpecSolv, em menos de três minutos. Os dados de entrada, fornecidos ao programa, foram os deslocamentos químicos, multiplicidades e intensidades dos sinais do espectro de RMN ¹³C e foi atribuído um valor máximo do tipo C_xH_xO_xN_x para a fórmula molecular. O número de átomos de nitrogênio também foi estimado próximo do valor real, com o objetivo de minimizar o tempo computacional. Na busca de subespectros foram fornecidas 150 subestruturas, e dentre elas o gerador apresentou apenas uma substância, como resultado final, cujo o espectro simulado pelo programa era compatível com o experimental, apresentando um fator de comparação de 0,448. A estrutura da substância gerada correspondeu a estrutura correta do alcalóide Uspalatina.

Figura 13. Estrutura do alcalóide pirrolizidínico, Uspalatina.

Limitações do SpecSolv

O módulo SpecSolv é capaz de gerar estruturas e prever espectros para substâncias com massas até aproximadamente 1000Da e que apresentem somente átomos de C, H, N, O, S, P e X. Moléculas maiores podem requerer

um tempo computacional muito grande devido o número de possibilidades combinatoriais para a geração de subestruturas obtidas por sobreposições. Além disso, o sistema somente é capaz de elucidar estruturas que possam ser descritas pela sobreposição de subestruturas, portanto, o sistema não possibilita a resolução de componentes individuais contidos em uma mistura.

A previsão de deslocamentos químicos do sistema é afetada negativamente pelo efeito do solvente ou influências estereoquímicas, uma vez que o sistema não apresenta informações estereoquímicas e não descreve efeitos configuracionais. Como qualquer outro sistema que utilize esse tipo de dado, o SpecSolv depende crucialmente da qualidade e diversidade da base de dados de correlações subestrutura-subespectro.

III.5 – Sistema Especialista SENECA¹³

Com a chegada da espectroscopia 2D nos anos 70 e seu desenvolvimento nas rotinas laboratoriais, tornou-se mais claro que problemas de elucidação estrutural assistida por computadores poderiam ser resolvidos utilizando-se a fórmula molecular e a conectividade dos dados obtidos por espectros 2D. Um sinal no espectro de hidrogênio 1D, numa frequência particular, pode fornecer diferentes ambientes químicos a partir de diferentes tipos de prótons. Um sinal cruzado num espectro 2D claramente liga os dois núcleos envolvidos de um modo positivo/negativo, dependendo do tipo de experimento que é executado. Um sinal cruzado nos diz que o próton ressonante em uma frequência F_1 é também separado por 2, 3, ou algumas vezes 4 ligações a partir do átomo de carbono ressonante a uma frequência F_2 (então chamados acoplamentos $^2J_{CH}$, $^3J_{CH}$, e $^4J_{CH}$). Algumas destas informações foram utilizadas anteriormente no programa LUCY³³. Este programa foi criado, baseado no conhecimento da composição molecular da substância em questão e numa série de experimentos de RMN (RMN 1D: RMN ^{13}C , DEPT-90, DEPT-135; RMN 2D: HH-COSY, HSQC, HMBC), com ênfase particular na ambigüidade da informação HMBC. O programa LUCY era capacitado para determinar estruturas com até 30 átomos de carbonos, desde que as informações extraídas dos espectros de RMN estivessem absolutamente livre de erros. Isto era alcançado por geração exaustiva de todas as estruturas que eram compatíveis com os dados fornecidos na entrada

do programa. Este programa utilizava informações do HMBC em uma etapa inicial (antes de gerar a estrutura final), o que fazia reduzir o tempo de busca. A partir deste ponto, vários programas começaram a surgir utilizando o mesmo princípio de funcionamento.

O LUCY apresentou algumas limitações. Primeiramente, a etapa indispensável do pré-processamento usando informações do HMBC ($^2J_{CH}$, $^3J_{CH}$), não considerando que existe um pequeno, mas significativo conjunto de sinais causados pelo acoplamento $^4J_{CH}$. Isto levou o programa a gerar resultados falso-positivo e falso-negativo. A Segunda limitação era em função da geração exaustiva de todos os constituintes que concordassem com os dados de entrada. Isso fazia com que o limite atual de átomos na molécula nunca pudesse se expandir, visto que ocasionaria uma explosão combinatorial. O SENECA foi projetado focalizando estes assuntos para superar as limitações do LUCY.

O SENECA é escrito em linguagem JAVA, apresenta um módulo que trabalha exclusivamente com a fórmula molecular da substância desconhecida e as correlações obtidas através de experimentos de RMN. Esse módulo utiliza os dados do DEPT-90 e DEPT-135 e os dados do espectro C,H COSY (HMQC e HSQC), HN HMBC e os dados do espectro H,H COSY.

Baseado nos dados de entrada da formula molecular de uma substância desconhecida e de um conjunto de restrições obtidas a partir dos dados espectrais de RMN, o processo se inicia com estruturas geradas randômicamente (S_1). A Energia E_1 é a medida de concordância da estrutura com as restrições determinadas. A estrutura é então transformada em um isômero (S_2), com a formula molecular determinada. Um novo valor de energia (E_2) é fornecido para a estrutura recentemente gerada; se $E_2 < E_1$ significa que a nova (S_2) está mais coerente com os dados espectroscópicos do que a primeira estrutura (S_1), então esta estrutura é aceita imediatamente. Se $E_2 > E_1$, é gerado um número aleatório R (com $0 < R < 1$), e a estrutura só é aceita se :

$$R < e^{(E_2 - E_1) / kT}$$

Este processo, quando simulado em uma temperatura constante converge para uma distribuição de Boltzmann. Kirkpatrick et al³⁴. mostraram

que, utilizando Metropolis³⁵ em temperaturas sucessivamente baixas, somente o ótimo poderia ser suficientemente populado a essas temperaturas e o sistema pára quando atinge um estado ótimo. No SENECA foi empregado este princípio para determinar a constituição de uma substância desconhecida baseado em um conjunto de dados derivados de métodos espectroscópicos. Em outras palavras, o SENECA tenta otimizar sistemas multivariáveis, não lineares e discretas.

A monochaetin³⁶ (Fig. 14) foi uma das substâncias utilizadas para testar a eficiência do SENECA.

Figura 14 – Estrutura do monochaetin

O primeiro passo do procedimento de elucidação estrutural foi o fornecimento da formula molecular da substância testada ($C_{18}H_{20}O_5$). A Segunda etapa do processo consistiu em entrar com os dados espectroscópicos: 18 valores de deslocamentos químicos de RMN ^{13}C e as correspondentes informações fornecidas pelo DEPT para 12 átomos de carbonos (*d*, *t*, *q*); 12 correlações $^1J_{CH}$ obtidas a partir do CH COSY; 6 correlações $^3J_{HH}$ obtidas a partir do HH COSY e por último, 28 correlações ambíguas obtidas a partir do HMBC as quais podem ser interpretadas como $^2J_{CH}$, $^3J_{CH}$ e algumas com $^4J_{CH}$ e até distâncias mais longas. Todos estes dados foram inseridos manualmente no programa. O próximo passo foi importar todos os dados espectrométricos disponíveis. Posteriormente todos os átomos de carbono na molécula são classificados de acordo com os respectivos deslocamentos químicos. Se houver menos sinais de átomos de carbono no espectro do que na molécula, deslocamento de sinais com dupla, tripla ou intensidade alta são usados múltiplas vezes, de maneira adequada. Durante este processo automatizado é atribuído o numero de átomos de

hidrogênio que estão ligados a cada átomo de carbono, baseado na inspeção dos dados espectrais do DEPT. A intervenção do usuário somente é requerida em casos de ambigüidade, por exemplo se houverem múltiplas possibilidades para atribuição de hidrogênios ligados em heteroátomos. A última ação antes de iniciar o processo final para elucidação estrutural é acionar os programas denominados “**JUÍZES**”, para que estes verifiquem se a estrutura gerada concorda com todas as restrições impostas a partir da interpretação dos dados como HH COSY ou HMBC. Depois de passar por todos os testes, que possam indicar qualquer tipo de informação confrontada, são feitas as propostas estruturais e finalmente as estruturas selecionadas ficam disponíveis para visualização. O programa forneceu 8 proposta estruturais compatíveis com as restrições espectrométricas fornecidas e, dentre elas, estava a estrutura do monochaetin. Então foi utilizado uma regra, a qual introduziu uma restrição adicional para ligações entre heteroátomos. De posse desses novos dados, o programa forneceu uma única proposta estrutural, que se tratava da estrutura correta do monochaetin.

IV – O SISTEMA ESPECIALISTA SISTEMAT

O sistema especialista SISTEMAT¹⁷⁻²⁰ foi desenvolvido no Instituto de Química da Universidade de São Paulo, na década de 80. Ele faz uso de diversas técnicas, envolvendo várias espectrometrias, para a determinação de classe e esqueletos de produtos naturais. O principal objetivo deste programa é auxiliar os pesquisadores, da área de produtos naturais, no processo de determinação estrutural das substâncias e em seus estudos quimiotaxonômicos, sendo que a quimiotaxonomia ainda foi pouco explorada pelo sistema.

Uma das grandes e principais vantagens do SISTEMAT, em relação a outros sistemas especialistas em determinação estrutural, é que ele possui um sistema de codificação estrutural e compactação de dados extremamente eficiente, possibilitando assim a utilização de microcomputadores, enquanto que os demais sistemas são executados em computadores de grande e médio porte. Esse método de codificação também permite que, a partir da estrutura codificada, o próprio sistema descubra informações químicas das substâncias como fórmula molecular, número de oxidação e grupos funcionais.

O programa SISTEMAT é composto basicamente de dois grandes pacotes de programas:

- Programas de criação de bancos de dados;
- Programas aplicativos.

Os primeiros programas, desenvolvidos no SISTEMAT, foram escritos em linguagem FORTRAN e os mais recentes em PASCAL. Ambos já foram testados com sucesso em outros trabalhos e teses³⁷⁻⁴² e serão atualmente aplicados à base de dados de alcalóides.

IV.1 – Metodologia

IV.1.1 - Levantamento Bibliográfico

Os bancos de dados foram construídos a partir de um levantamento bibliográfico dos dados espectrais e botânicos das substâncias pertencentes às subclasses dos alcalóides quinolínicos, quinolizidínicos, aporfinicos, benzilisoquinolínicos, isoquinolínicos, pirrolizidínicos, acridônicos e indólicos. Os levantamentos dos dados, arquivados no sistema, foram realizados na biblioteca do Instituto de Química da USP, onde foram consultados periódicos como *Phytochemistry*, *Planta Médica*, entre outros e na biblioteca do Instituto de Química da UNICAMP, onde foram consultados os periódicos *Heterocycles*, *Journal Natural Products*, entre outros periódicos, não encontrados ou incompletos na biblioteca do IQ- USP.

IV.1.2- Organização dos Bancos de dados

Para o bom funcionamento de qualquer sistema especialista é fundamental que os dados contidos em seus bancos sejam apresentados para o sistema de forma organizada e precisa. Sendo assim, os dados presentes no sistema especialista SISTEMAT são organizados da seguinte forma:

- Representação da molécula;
- Nome trivial;
- Número da ficha;
- Classe química;
- Esqueleto;
- Referências;
- Dados botânicos;
 - Família;
 - Gênero e
 - Espécie.
- Dados físico químicos
 - RMN ^{13}C
 - RMN ^1H
 - Massas.

Para armazenar os dados extraídos da literatura e facilitar o trabalho de digitação para entrada destes dados no programa, é utilizado uma ficha de codificação (Fig. 15), onde estão presentes todas as informações importantes da substância em questão.

FICHA: 341	NOME: PULCHINE	BANCO: ALCA 007	SOLV: CDCl ₃	
ESQUELETO: Aporfínico 5		REFERÊNCIA: J. Nat. Prod., 1988, pg. 389		
FAMÍLIA: Lauraceae	GÊNERO: Ocotea	ESPÉCIE: pulchella		
<div>Desenho da estrutura</div>		RMN ¹³ C	RMN ¹ H	Massas
		1.	1.	1.
		2.	2.	2.
		3.	3.	3.
		4.	4.	4.
		5.	5.	5.
		6.	6.	6.
		7.	7.	7.
		8.	8.	8.
		9.	9.	9.
		10.	10.	10.
		11.	11.	11.
		12.	12.	12.
		13.	13.	13.
		14.	14.	14.
				

Figura 15 – Ficha de codificação utilizada no levantamento bibliográfico.

Para atribuir os deslocamentos químicos característicos aos seus respectivos átomos de carbono, foi necessário criar uma numeração seqüencial, padronizada, para cada tipo de esqueleto. Esta numeração, denominada de **vetor biogenético**, nem sempre é idêntica a numeração biogenética encontrada na química de produtos naturais. Isto deve-se ao fato de que, na literatura, diferentes numerações podem ser encontradas para um mesmo tipo de esqueleto de alcalóide. Neste trabalho foram utilizadas aquelas mais comumente encontradas. Como exemplo observa-se o alcalóide (Fig. 15) pertencente ao esqueleto Aporfínico 5, onde a numeração segue de 1 a 17.

Pode-se observar, pelo levantamento realizado, a grande diversificação estrutural dos alcalóides (Fig. 7), o que torna esta classe de substâncias um grande desafio para os espectroscopistas.

IV.2 – Programas de criação do banco de dados.

O primeiro passo para construção do SISTEMAT, foi a elaboração de um programa responsável pelo armazenamento de dados, ou seja, um programa para criação da base de dados, onde as substâncias eram inseridas ao banco através de um processo de codificação estrutural, processo este que inicialmente era feito de forma manual, o que o tornava muito lento e passível de erros. Em 1992, com a criação do programa CODISIS¹⁹, a codificação das substâncias passou a ser semi-automática e o trabalho do usuário se resumiu somente à definição da seqüência numérica da cadeia carbônica da substância a ser codificada. Atualmente, esse programa foi aperfeiçoado e recebeu o nome de DATASIS, o qual realiza a codificação totalmente automática. O reflexo imediato desta automatização foi o aumento na velocidade de arquivamento dos dados e a diminuição do número de erros.

IV.2.1 – Codificação Estrutural

O primeiro passo na construção de qualquer sistema especialista na área de Química Orgânica é a representação de uma substância em uma linguagem matemática que possa ser entendida pelo computador. Tal processo é chamado de codificação estrutural.

A representação matemática da estrutura de uma substância no SISTEMAT é feita usando-se uma matriz topológica. Esta matriz é construída a partir do desenho da formula estrutural da substância, com seus respectivos átomos numerados. O SISTEMAT foi projetado para trabalhar com 11 “tipos” de átomos e 4 “tipos” de ligações (Tab.4).

Tabela 4 – Pesos atribuídos aos átomos e ligações para construção da matriz topológica.

ÁTOMO		LIGACÃO	
TIPO	PESO	TIPO	PESO
Carbono	1	Não Ligado	0
Carbono Aromático	2	Simples	1
Oxigênio	3	Aromática	1
Nitrogênio	4	Dupla	2
Nitrogênio Aromático	5	Tripla	3
Flúor	6		
Cloro	7		
Bromo	8		
Iodo	9		
Enxofre	10		
Fósforo	11		

A diagonal da matriz topológica contém os pesos dos átomos, na ordem da numeração dos mesmos. Os elementos fora da diagonal representam os pesos das ligações entre os átomos das coordenadas do elemento.

Pode-se observar um exemplo de matriz topológica (Tabela 5), construída para o alcalóide **Quinolínico X** (Fig.16):

Figura 16 - Quinolínico X

Tabela 5 - Matriz topológica do alcalóide Quinolínico X.

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

Analisando a matriz topológica (Tab.5), o programa pode encontrar todas as informações estruturais da molécula, exceto a representação das ligações alfa e beta, ligações transanulares de configuração *cis* e os substituintes comuns encontrados em produtos naturais, por exemplo metoxila, acetato e outros. Estas informações são codificadas em quatro vetores distintos que são associados à matriz. Entretanto, o armazenamento da matriz topológica e da matriz dos vetores ocuparia um grande volume de memória do computador, contrariando um dos princípios do SISTEMAT, que é a maior compactação possível dos dados. Para solucionar este problema, fez-se uso da técnica do vetor reduzido, cuja aplicação só foi possível por que a maioria

dos elementos da matriz topológica são nulos. Este vetor consiste numa sequência das posições e valores dos elementos não nulos da matriz.

Para que o programa transformasse os números da matriz numa linguagem familiar ao químico, “desenho da molécula”, foi necessário um outro método de codificação. Então foi desenvolvido um método de codificação arbitrário, que conduzissem a um vetor com menor tamanho possível, contendo todas as informações necessárias. Este vetor foi denominado **vetor nó**.

Para exemplificar a codificação de uma estrutura, temos que imaginá-la desenhada em papel quadriculado, onde cada átomo ocupa um quadrado e cada ligação ocupa 2 ou mais quadrados. As ligações que ocupam mais do que 2 quadrados são denominadas ligações longas, estas ligações têm sempre que ocupar um número ímpar de quadrados, para que o programa as reconheçam.

A primeira etapa da codificação é a numeração da substância, que respeita as seguintes regras:

- a) Dois átomos ligados por uma ligação longa não podem apresentar números consecutivos.

- b) Os nós monoatômicos (átomos ligados por uma única ligação a um átomo pertencente a cadeia principal) devem ser numerados em primeiro lugar. Um átomo deixa de ser considerado como nó monoatômico quando ele faz parte da cadeia, como átomo inicial ou final da mesma. Observa-se na Figura 18 que os átomos **1** e **2** (oxigênio hidroxílico e carbono metílico) foram considerados **nós** monoatômicos, e por isso, eles não fazem parte da cadeia principal, que é numerada do átomo **3** ao **11**.
- c) A numeração da estrutura deve ser feita de modo que o programa possa reconhecer quais são as ligações alfa e beta. Portanto, foi estabelecido que

se um átomo estiver em posição alfa, todas as suas ligações com átomos de maior numeração serão consideradas alfa, o mesmo sendo aplicada às ligações beta.

A próxima etapa da codificação é a elaboração do **vetor nó**, que deve conter a seguinte sequência de informações:

1. Os números iniciais e finais de cada cadeia. Precedidos por **-1**, o número dos átomos aos quais os nós monoatômicos estão ligados, seguidos pelos números dos átomos não incluídos na cadeia principal, isto é, dos átomos com numeração não sequencial que estão interligados.
2. Precedido por **-2**, os números atômicos dos heteroátomos e as posições que os mesmos ocupam.
3. Precedidas por **-3**, as posições dos átomos das ligações duplas.
4. Precedida por **-4**, a posição de um átomo do ciclo aromático que faça parte de apenas um anel.
5. Precedidas por **-5**, as posições dos átomos das ligações triplas.
6. Precedidos por **-6 -1**, os átomos beta; por **-6-2**, os átomos alfa; por **-6-3**, os átomos com carga positiva e por **-6-4**, aqueles com carga negativa.
7. Precedidos por **-7**, os átomos das ligações transanulares de conformação *cis*.
8. Precedidos por **-8**, os códigos e as posições dos macronós.
9. O sinal **-9**, para fechar este primeiro vetor.

O Vetor nó permite ao programa reconhecer o tipo de ligação, os tipos de átomos e as cadeias existentes na molécula, mas faltam ainda outras informações para que o computador possa transformar a matriz topológica numa representação gráfica, ou seja, no desenho da substância. Portanto é necessário indicar as direções das ligações, estas informações estão armazenadas no **vetor passo**. Este contém as coordenadas do átomo inicial da cadeia e as orientações sucessivas das ligações que a compõe, codificadas

de acordo com a Figura 17. O vetor final será o resultado da concatenação dos vetores nó e passo.

Figura 17. Direção das ligações do vetor passos

O alcalóide indólico da Figura 18 será, portanto, representado pelo seguinte vetor nó e passo:

Figura 18 – Codificação manual de um Alcalóide indólico.

IV.2.2 – Programa DATASIS

O programa DATASIS³⁸ permite a codificação automática dos vetores **nó** e **passo**, a partir do desenho da molécula na tela do computador. Estes desenhos podem ser armazenados em arquivos de esqueletos e/ou substâncias funcionalizadas para agilizar a entrada de novos dados no banco. Desta forma, para inserir uma substância com esqueleto semelhante ao armazenado, basta escolher o esqueleto e/ou substância nos respectivos arquivos e fazer as alterações necessárias. Este programa, como mencionado anteriormente, permite que, além dos esqueletos identificados com seus respectivos dados espectrométricos, informações como a classe química da substância, o tipo de esqueleto carbônico e a origem botânica da substância (família, gênero e espécie) sejam inseridos no banco para melhor servir ao pesquisador da área de produtos naturais.

Iniciado o programa DATASIS é apresentada uma tela onde o usuário deve fornecer o nome do arquivo de Banco Fonte, ou seja, o banco no qual

serão armazenados os dados pesquisados na literatura, posteriormente são inseridos os nomes dos arquivos de esqueletos e substâncias. Logo em seguida, aparece a tela do menu principal (Fig. 19).

Figura 19 – Tela do menu principal do programa DATASIS

As opções **0** e **8** são para voltar ao sistema operacional, sendo que pela opção **0** a volta é definitiva, de forma que para iniciar o programa novamente é necessário definir os arquivos de Banco Fonte, Substância e Esqueleto; enquanto que a opção **8** indica que a saída do sistema será temporária, não exigindo a definição dos arquivos para que a tela volte ao menu principal.

A opção **1** é utilizada para redefinir arquivos de Banco Fonte, Substância e Esqueleto, no caso em que a definição inicial tenha sido feita de forma errônea.

A opção **2**, como o próprio nome informa, é utilizada para criar bancos arquitetutados, informando ao programa a quantidade máxima de substâncias que serão inseridas no banco.

A opção **3** leva a outro menu, contendo 3 novas opções que permite ao usuário voltar ao menu anterior (opção 0), somente desenhar (opção 1), somente obter o código (opção 2) e por último e mais importante codificar uma estrutura (opção 3). A entrada de uma nova substância no banco é realizada através desta última opção. Então, logo que a opção 3 é selecionada, o programa pede o nome trivial da substância a ser inserida no banco. A partir

do nome, o programa pode reconhecer se esta substância já se encontra arquivada no banco de dados ou não. Na ausência de um nome trivial, o programa usa um identificador matemático para fazer esse reconhecimento.

Confirmado que se trata de uma substância inédita, o usuário entra com o número da ficha (Fig.15), na qual estão presentes os dados pesquisados na literatura e posteriormente aparece uma tela (Fig.20) onde a substância será desenhada (com o auxílio dos símbolos utilizados para representação dos átomos e ligações) e codificada automaticamente.

Após o desenho da substância, o programa numera todos os átomos da estrutura e pede a confirmação da mesma. Em caso afirmativo, ele apresenta o vetor completo, o número de bytes usados na compactação, número de dados encontrados no vetor de compactação (não podendo ultrapassar 99 dados), o nox e o tempo (Fig. 21).

Figura 20 – Tela de desenho do programa DATASIS, onde foi codificado, automaticamente, o alcalóide indólico utilizado para explicação dos vetores nó e passo (Pg. 58, Fig. 18).

Figura 21 – Tela de codificação do programa DATASIS. Onde pode-se comparar os vetores obtidos de forma manual (Pg. 58) com os fornecidos automaticamente pelo programa.

Depois de fornecer os dados codificados, o programa apresenta a tela onde é realizada a leitura dos dados (Fig.22). Nesta tela o usuário deve entrar com os dados obrigatórios (classe e esqueleto da substância, código da revista, ano e página) e se existirem, os dados optativos (família, gênero e espécie da planta e índice de evolução morfológica - IEM). Quando a substância for um derivado sintético e não possuir os dados botânicos, o item **IEM** é preenchido com o valor [-1] e quando houver dados botânicos utiliza-se o valor [1].

LEITURA DOS DADOS

CLASSE	:	[ALCALOIDE	1
ESQUELETO	:	[QUINOLINICOM	1
CODIGO DA REVISTA	:	[51]	
ANO	:	[28]	
PAGINA	:	[3456]	
FAMILIA	:	[RUTACEAE	1
GENERO	:	[MEDICOSMA	1
ESPECIE	:	[FAREANA	
IEM	:	[1]	

OBSERVAÇÕES:

1- HAVENDO DADOS BOTANICOS, IEM > 0
 2- NAO HAVENDO DADOS BOTANICOS, IEM= -1
 3- SEM FAMILIA E COM GENERO, FAMILIA= IGNORADA

Figura 22- Tela de leitura dos dados do programa DATASIS.

Ao término da leitura dos dados, o programa retorna à tela de entrada do nome trivial e uma nova substância poderá ser codificada.

No caso da entrada de algum dado errado, estes podem ser corrigidos através da opção **6** do menu principal do programa DATASIS (Fig.19) e caso necessário, pode-se eliminar definitivamente uma substância através da opção **7**. Isto é necessário quando ocorre algum erro no desenho da estrutura, que não foi percebido durante a entrada dos dados.

O próximo passo para entrada de dados é realizado através da opção **4** do menu principal do programa DATASIS (Fig.19), ou seja, entrada de dados físico-químicos. Entretanto esta etapa só é possível após a arquitetura do Banco Fonte realizada pelo programa **ARQUISIS**.

IV.2.3 – Programa ARQUISIS³⁸

O ARQUISIS é um programa de arquitetura do Banco Fonte. Entende-se por arquitetura, a distribuição dos dados do Banco Fonte em bancos interligados, que contêm apenas um tipo de dado específico (Fig.23).

Figura 23 – Esquema dos bancos arquitetados do SISTEMAT.

- **BNUM** – Número matemático criado para cada substância.
- **BTIP** – Classe das substâncias e tipos de dados
- **BESQ** – Nome do esqueleto
- **BREF** – Referências bibliográficas
- **BREV** – Nome das revistas
- **BESP** – Espécies

estudadas as seguintes classes químicas de substâncias: Alcalóides (apresentados nesta tese), monoterpenos e iridóides³⁷, sesquiterpenos lactonizados e não-lactonizados³⁸, diterpenos⁴⁰, triterpenos³⁹, esteróides⁴¹ e flavonóides⁴².

Uma grande vantagem do programa SISTEMAT é que, como os dados inseridos no banco, através do programa DATASIS, são classificados pela diferenciação dos esqueletos, a busca de estruturas a partir destes esqueletos, previamente definidos, elimina o problema de explosão combinatória e, conseqüentemente, fornece um número menor de estruturas equivocadas.

IV.3 – Programas Aplicativos

Os programas aplicativos do SISTEMAT possibilitam a análise dos dados presentes no sistema e a obtenção de informações importantes para o processo de elucidação estrutural. Através dos dados de RMN ¹³C, Espectros de Massas, RMN ¹H e dados botânicos, estes programas podem indicar o tipo de esqueleto de uma substância desconhecida.

IV.3.1 – Programa SISLIST³⁸

O programa SISLIST fornece dois tipos de listagens contidas no Banco de dados. As listagens podem ser realizadas por número de substâncias e tipos de esqueletos ou por dados botânicos. No primeiro caso, o programa fornece uma lista do total de substâncias da classe química estudada e a distribuição destas por tipo de esqueleto. No segundo caso, o programa apresenta os dados botânicos onde são listados o número de ocorrências, por gênero e espécie estudados, para cada família de plantas presentes no banco de dados.

A Tabela 6 apresenta uma listagem do número de ocorrência de cada esqueleto, presente no banco de dados, pertencente a classe dos alcalóides estudados neste trabalho, por exemplo, o alcalóide Quinolizidínico 1 é representado, no banco de dados, por 55 substâncias diferentes.

Tabela 6: Listagem do número de ocorrências de esqueletos de Alcalóides .

Classe : ALCALÓIDE		Ocorrência por esqueleto
Esqueletos		Total : 1183
1.	Quinolínico 1	1
2.	Quinolínico 2	2
3.	Quinolínico 3	1
4.	Quinolínico 4	2
5.	Quinolínico 5	2
6.	Quinolínico 6	13
7.	Quinolínico 7	1
8.	Quinolínico 8	2
9.	Quinolínico 9	1
10.	Quinolínico 10	8
11.	Quinolínico 11	3
12.	Quinolínico 12	1
13.	Quinolínico 13	1
14.	Quinolínico 14	2
15.	Quinolínico 15	2
16.	Quinolínico 16	1
17.	Quinolínico 17	1
18.	Quinolínico 18	2
19.	Quinolínico 19	2
20.	Quinolínico 20	1
21.	Quinolínico 21	4
22.	Quinolínico 22	2
23.	Quinolínico 23	5
24.	Quinolínico 24	1
25.	Quinolínico 25	1
26.	Quinolínico 26	1
27.	Quinolínico 27	1
28.	Quinolínico 28	1
29.	Quinolínico 30	4
30.	Quinolínico 31	2
31.	Quinolínico 32	2
32.	Quinolínico 33	2
33.	Quinolínico 35	1
34.	Quinolínico 36	1
35.	Quinolínico 37	1
36.	Quinolínico 38	1
37.	Quinolínico 39	1

Classe : ALCALÓIDE		Total de ocorrência
Esqueletos		Por esqueleto
38.	Quinolínico 40	2
39.	Quinolínico 41	20
40.	Quinolínico 42	6
41.	Quinolínico 43	1
42.	Quinolínico 44	1
43.	Quinolínico 45	2
44.	Quinolínico 46	1
45.	Quinolínico 47	1
46.	Quinolínico 48	1
47.	Quinolínico 49	1
48.	Quinolizidínico 1	55
49.	Quinolizidínico 2	4
50.	Quinolizidínico 3	2
51.	Quinolizidínico 4	1
52.	Quinolizidínico 5	1
53.	Quinolizidínico 6	1
54.	Quinolizidínico 7	1
55.	Quinolizidínico 8	1
56.	Quinolizidínico 9	1
57.	Quinolizidínico 10	1
58.	Quinolizidínico 11	1
59.	Quinolizidínico 12	2
60.	Quinolizidínico 13	1
61.	Quinolizidínico 14	1
62.	Quinolizidínico 16	1
63.	Quinolizidínico 17	2
64.	Quinolizidínico 18	1
65.	Quinolizidínico 19	1
66.	Quinolizidínico 20	1
67.	Quinolizidínico 21	4
68.	Quinolizidínico 22	2
69.	Quinolizidínico 23	1
70.	Quinolizidínico 24	2
71.	Quinolizidínico 25	2
72.	Quinolizidínico 26	1
73.	Quinolizidínico 27	2
74.	Quinolizidínico 28	2
75.	Quinolizidínico 29	1

Classe : ALCALÓIDE		Total de ocorrência
Esqueletos		Por esqueleto
76.	Quinolizidínico 30	1
77.	Quinolizidínico 31	1
78.	Quinolizidínico 32	1
79.	Quinolizidínico 33	1
80.	Benzilisoquinolínico 1	21
81.	Benzilisoquinolínico 2	17
82.	Benzilisoquinolínico 3	17
83.	Benzilisoquinolínico 4	1
84.	Benzilisoquinolínico 5	3
85.	Benzilisoquinolínico 6	13
86.	Benzilisoquinolínico 7	1
87.	Benzilisoquinolínico 8	2
88.	Benzilisoquinolínico 9	5
89.	Benzilisoquinolínico 10	2
90.	Benzilisoquinolínico 11	3
91.	Benzilisoquinolínico 12	1
92.	Benzilisoquinolínico 13	1
93.	Benzilisoquinolínico 14	3
94.	Benzilisoquinolínico 15	1
95.	Benzilisoquinolínico 16	1
96.	Benzilisoquinolínico 17	1
97.	Benzilisoquinolínico 18	3
98.	Benzilisoquinolínico 19	1
99.	Benzilisoquinolínico 20	1
100.	Benzilisoquinolínico 21	1
101.	Benzilisoquinolínico 22	1
102.	Aporfínico 1	42
103.	Aporfínico 2	5
104.	Aporfínico 3	2
105.	Aporfínico 4	15
106.	Aporfínico 5	24
107.	Aporfínico 6	1
108.	Aporfínico 7	3
109.	Aporfínico 8	1
110.	Aporfínico 9	5
111.	Aporfínico 10	1
112.	Aporfínico 11	2
113.	Aporfínico 12	1

Classe ▾, ALCALÓIDE	Total de ocorrência
Esqueletos	Por esqueleto
114. Acridônico 1	5
115. Acridônico 2	2
116. Acridônico 3	1
117. Acridônico 4	7
118. Acridônico 5	2
119. Acridônico 6	1
120. Acridônico 7	39
121. Acridônico 8	1
122. Acridônico 9	1
123. Acridônico 10	1
124. Acridônico 11	1
125. Acridônico 12	16
126. Acridônico 13	2
127. Acridônico 14	1
128. Acridônico 15	1
129. Acridônico 16	1
130. Acridônico 17	40
131. Acridônico 18	7
132. Acridônico 19	8
133. Acridônico 20	1
134. Acridônico 21	1
135. Acridônico 22	1
136. Acridônico 23	1
137. Acridônico 24	1
138. Acridônico 25	1
139. Acridônico 27	1
140. Acridônico 28	3
141. Acridônico 29	3
142. Acridônico 30	2
143. Acridônico 31	1
144. Acridônico 32	1
145. Acridônico 34	3
146. Acridônico 35	1
147. Acridônico 36	1
148. Acridônico 37	3
149. Acridônico 38	2
150. Acridônico 39	1
151. Acridônico 40	1

Classe : ALCALÓIDE		Total de ocorrência
Esqueletos		Por esqueleto
152.	Acridônico 41	5
153.	Acridônico 42	1
154.	Acridônico 43	1
155.	Acridônico 44	6
156.	Acridônico 45	1
157.	Acridônico 46	1
158.	Acridônico 47	3
159.	Acridônico 49	2
160.	Acridônico 50	2
161.	Acridônico 51	1
162.	Acridônico 52	1
163.	Acridônico 53	1
164.	Acridônico 55	1
165.	Pirrolizidínico 1	94
166.	Pirrolizidínico 2	2
167.	Isoquinolínico 1	49
168.	Isoquinolínico 2	32
169.	Isoquinolínico 3	5
170.	Isoquinolínico 4	20
171.	Isoquinolínico 5	2
172.	Isoquinolínico 6	8
173.	Isoquinolínico 7	1
174.	Isoquinolínico 8	2
175.	Isoquinolínico 9	2
176.	Isoquinolínico 10	2
177.	Isoquinolínico 11	1
178.	Isoquinolínico 12	3
179.	Isoquinolínico 13	1
180.	Isoquinolínico 14	10
181.	Isoquinolínico 15	2
182.	Isoquinolínico 17	2
183.	Isoquinolínico 18	2
184.	Isoquinolínico 19	1
185.	Isoquinolínico 20	2
186.	Isoquinolínico 21	2
187.	Isoquinolínico 22	6
188.	Isoquinolínico 23	2
189.	Isoquinolínico 24	2

Classe : ALCALÓIDE		Total de ocorrência
Esqueletos		Por esqueleto
190.	Isoquinolínico 25	2
191.	Isoquinolínico 26	1
192.	Isoquinolínico 27	1
193.	Isoquinolínico 28	1
194.	Isoquinolínico 30	1
195.	Isoquinolínico 31	1
196.	Isoquinolínico 32	6
197.	Isoquinolínico 33	1
198.	Isoquinolínico 34	2
199.	Isoquinolínico 35	2
200.	Isoquinolínico 36	1
201.	Isoquinolínico 37	2
202.	Isoquinolínico 38	1
203.	Isoquinolínico 39	2
204.	Isoquinolínico 40	4
205.	Isoquinolínico 41	2
206.	Isoquinolínico 42	1
207.	Indol 1	3
208.	Indol 2	1
209.	Indol 3	12
210.	Indol 4	4
211.	Indol 5	10
212.	Indol 6	3
213.	Indol 7	2
214.	Indol 8	1
215.	Indol 9	1
216.	Indol 10	2
217.	Indol 11	4
218.	Indol 12	1
219.	Indol 13	1
220.	Indol 14	1
221.	Indol 15	1
222.	Indol 16	1
223.	Indol 17	10
224.	Indol 18	1
225.	Indol 19	1
226.	Indol 20	2
227.	Indol 21	3

Classe : ALCALÓIDE		Total de ocorrência
Esqueletos		Por esqueleto
228.	Indol 22	2
229.	Indol 23	6
230.	Indol 24	16
231.	Indol 25	30
232.	Indol 26	4
233.	Indol 27	2
234.	Indol 28	2
235.	Indol 29	2
236.	Indol 30	1
237.	Indol 31	2
238.	Indol 32	3
239.	Indol 33	4
240.	Indol 34	1
241.	Indol 35	7
242.	Indol 36	2
243.	Indol 37	1
244.	Indol 38	7
245.	Indol 39	1
246.	Indol 40	7
247.	Indol 41	3
248.	Indol 42	10
249.	Indol 43	10
250.	Indol 44	2
251.	Indol 45	1
252.	Indol 46	2
253.	Indol 47	1
254.	Indol 48	5
255.	Indol 49	3
256.	Indol 50	1
257.	Indol 51	2
258.	Indol 53	2
259.	Indol 54	1
260.	Indol 55	2
261.	Indol 56	1
262.	Indol 57	3
263.	Indol 58	1
264.	Indol 59	1
265.	Indol 60	2

Classe : ALCALÓIDE	Total de ocorrência
Esqueletos	Por esqueleto
	1
266. Indol 61	1
267. Indol 62	12
268. Indol 63	2
269. Indol 64	1
270. Indol 65	5
271. Indol 66	6
272. Indol 67	1
273. Indol 68	1
274. Indol 69	12
275. Indol 70	5
276. Indol 71	1
277. Indol 72	1
278. Indol 73	6
279. Indol 74	3
280. Indol 75	1
281. Indol 76	2
282. Indol 77	

Na Tabela 7 são apresentados o número de ocorrências botânicas dos alcalóides, dados por gêneros e espécies, presentes na família Rutaceae.

Tabela 7- Listagem das ocorrências botânicas dos alcalóides na família Rutaceae.

Família : Rutaceae	
Gênero: <i>Galipea</i>	
Espécie : <i>bracteata</i>	Total = 6
Total do gênero : <i>Galipea</i>	6
Gênero: <i>Zanthoxylum</i>	
Espécie : <i>acutifolium</i>	Total = 8
Espécie : <i>simulans</i>	Total = 4
Espécie : <i>arborescens</i>	Total = 1
Total do gênero : <i>Zanthoxylum</i>	13
Gênero: <i>Orixa</i>	
Espécie : <i>japonica</i>	Total = 3
Total do gênero : <i>Orixa</i>	3

Família: Rutaceae		
Gênero: <i>Melicope</i>		Total = 1
Espécie : <i>semecarpifolia</i>		Total = 1
Espécie : <i>leratti</i>		2
Total do gênero : <i>Melicope</i>		
Gênero: <i>Dictyoloma</i>		
Espécie : <i>peruviana</i>		Total : 2
Total do gênero : <i>Dictyoloma</i>		2
Gênero: <i>Esenbeckia</i>		
Espécie : <i>almawillia</i>		Total = 2
Espécie : <i>litoralis</i>		Total = 1
Espécie : <i>leiocarpa</i>		Total = 2
Total do gênero : <i>Esenbeckia</i>		5
Gênero: <i>Evodia</i>		
Espécie : <i>rutaecarpa</i>		Total = 7
Espécie : <i>roxburghiana</i>		Total = 3
Total do gênero : <i>Evodia</i>		10
Gênero: <i>Lyngbya</i>		
Espécie : <i>majuscula</i>		Total = 2
Total do gênero : <i>Lyngbya</i>		2
Gênero: <i>Vepris</i>		
Espécie : <i>louisii</i>		Total : 4
Total do gênero : <i>Vepris</i>		4
Gênero: <i>Oricia</i>		
Espécie : <i>renieri</i>		Total : 3
Total do gênero : <i>Oricia</i>		3
Gênero: <i>Eriosteman</i>		
Espécie : <i>gardneri</i>		Total = 1
Total do gênero : <i>Eriosteman</i>		1
Gênero: <i>Haplophyllum</i>		
Espécie : <i>tuberculatum</i>		Total = 2
Espécie : <i>glabrinum</i>		Total = 2
Total do gênero : <i>Haplophyllum</i>		2
Gênero: <i>Skimmia</i>		
Espécie : <i>reevesiana</i>		Total : 4
Total do gênero : <i>Skimmia</i>		4
Gênero: <i>Teclea</i>		
Espécie : <i>simplicifolia</i>		Total = 1
Total do gênero : <i>Teclea</i>		1

Família : Rutaceae		
Gênero: <i>Clausena</i>	Total = 1	
Espécie : <i>anisata</i>	1	
Total do gênero : <i>Clausena</i>		
Gênero: <i>Agathosma</i>	Total = 8	
Espécie : <i>Desconhecida</i>	8	
Total do gênero : <i>Agathosma</i>	8	
Gênero: <i>Euxylophora</i>	Total = 7	
Espécie : <i>paraensis</i>	7	
Total do gênero : <i>Euxylophora</i>	7	
Gênero: <i>Acronychia</i>	Total = 1	
Espécie : <i>laurifolia</i>	1	
Espécie : <i>bauerii</i>	23	
Total do gênero : <i>Acronychia</i>	24	
Gênero: <i>Glycosmis</i>	Total = 1	
Espécie : <i>arborea</i>	1	
Espécie : <i>citrifolia</i>	6	
Total do gênero : <i>Glycosmis</i>	7	
Gênero: <i>Medicosma</i>	Total = 1	
Espécie : <i>fareana</i>	1	
Total do gênero : <i>Medicosma</i>	1	
Gênero: <i>Bosistoa</i>	Total = 1	
Espécie : <i>transversa</i>	1	
Total do gênero : <i>Bosistoa</i>	1	
Gênero: <i>Angostura</i>	Total = 1	
Espécie : <i>paniculata</i>	1	
Total do gênero : <i>Angostura</i>	1	
Gênero: <i>Severinia</i>	Total = 9	
Espécie : <i>buxifolia</i>	9	
Total do gênero :	9	
Gênero: <i>Ruta</i>	Total = 12	
Espécie : <i>graveolens</i>	12	
Total do gênero : <i>Ruta</i>	12	
Total da Família Rutaceae :	151	

IV.3.2 – Programa MACRONO*

Os produtos naturais, em geral, apresentam grupos substituintes, tais como, metoxila, acetila, tigloíla entre outros, ligados ao esqueleto carbônico da substância por meio de um heteroátomo. Estes substituintes são normalmente provenientes da esterificação das hidroxilas presentes no esqueleto da substância. Estes grupos, no SISTEMAT, são denominados de **macronós**.

Durante a análise espectral realizada pelos programas aplicativos do SISTEMAT, a presença dos sinais pertencentes aos macronós pode confundir os algoritmos de identificação de esqueletos e subestruturas, conduzindo a propostas incorretas. Para solucionar este problema, foi desenvolvido o programa MACRONO³⁸, que fornece os possíveis substituintes ligados ao esqueleto carbônico de uma substância, a partir da análise dos dados do espectro de RMN ¹³C. Assim, antes que os dados do espectro-problema sejam submetidos aos programas que realizam a identificação de esqueletos e fornecem as propostas estruturais, é realizada uma análise prévia para identificação e remoção dos sinais dos substituintes.

Para o desenvolvimento deste programa, foi criado um banco de dados com **58** tipos de substituintes mais comuns encontrados em lactonas sesquiterpênicas. No entanto, quando o programa MACRONO foi utilizado para a identificação dos substituintes presentes em alcalóides, este apresentou um baixo nível de desempenho quando comparado ao observado para os sesquiterpenos lactonizados. A causa deste problema estava na base de dados do programa:

- durante o arquivamento dos substituintes, haviam sido utilizadas as faixas de deslocamentos químicos dos ácidos livres, que não correspondem as faixas dos respectivos ésteres;
- a variedade de substituintes arquivada era muito pequena quando comparada a encontrada em alcalóides.

A solução encontrada para este problema foi completar a base de dados inserindo os tipos de substituintes freqüentemente encontrados nos alcalóides.

As estruturas dos principais macronós, presentes no banco de dados dos alcalóides, são apresentadas na Tabela 8.

* Embora escreva-se MACRONO, esta palavra é pronunciada como macronó (nós macros ou nós atômicos).

Tabela 8 ; Principais Macronós presentes nos alcalóides.

Nome	Estrutura do Macronó	Sigla:	Código :	δ (RMN ^{13}C) ± 2
Metoxila		ME	180	56,0 (q)
Etoxila		Et	22	67,0(t) ; 16,0(q)
Pentadecanoíla (Palmitoíla)		PAL	23	177,0(s); 35,0(t); 33,0(t); 30,0(t) (x6); 29,0(t); 25,0(t); 23,0(t); 14,0(q).
Glicosila		GLY	12	63,0(t); 78,0(d); 71,0(d); 78,0(d); 75,0(d); 105,0(d)
Benzoíla		BZT	83	165,0(s); 133,0(s); 131,0(d); 129,0(d); 130,0(d); 130,0(d); 129,0(d).
Angeloíla		ANG	199	166,2(s); 139,6(d); 126,4(s); 19,0(q); 15,0(q).
Tigloíla		TIG	128	168,0(s); 129,0(s); 137,0(d); 14,0(q); 12,0(q).
4-hidroxi- 3,5-dimetoxi – tetraidropiranox-2-ila		MTH	52	95,0(d); 84,0(d); 76,0(d); 80,0(d); 63,0(t); 58,0(q); 58,0(q).
Acetoíla		ACT	80	170,0(S); 21,0(Q).
3-metoxycarbonil- butanoxila		MBM	18	176,0(s); 52,0(q); 65,0(t); 33,0(t); 36,0(d); 17,4(q);
3-metoxycarbonil- butenox-2-ila		MBE	24	167,6(s); 136,4(d); 130,0(s); 13,0(q); 65,0(t); 53,0(q).

Tabela 8 ; Principais Macronós presentes nos alcalóides (cont.)

Nome	Estrutura do Macronó	Sigla:	Código :	δ (RMN ^{13}C) ± 2
3-hidroxi-butoxila		3HB	25	66,0(t); 38,0(t); 66,0(d); 23,5(q).
Prenoxila		PRE	32	71,0(t); 119,7(d); 137,0(s); 18,1(q); 25,9(q).
2-metil-butirola		2MB	26	167,2(s); 41,0(d); 26,6(t); 11,3(q); 16,7(q).
1H-Pirrole-metanoíla		CNE	54	161,0(s); 123,5(s); 116,1(d); 110,3(d); 122,9(d).
Veratroíla		VET	69	171,0(s); 124,0(s); 111,8(d); 123,6(d); 152,9(s); 148,4(s); 110,4(d); 56,0(q); 56,0(q).
Isovaniloíla		IVN	129	161,0(s); 123,0(s); 123,7(d); 110,4(d); 153,0(s); 148,5(s); 111,9(d); 56,2(q).
Fenilmetoxila		BNZ	43	70,9(t); 137,0(s); 127,5(d); 127,5(d); 128,5(d); 128,5(d); 127,9(d).
9-hidroxi-8,8-dimetil-2-oxo- 9,10-diidro-2H,8H-piranox-10-ila		CUM	42	158,8(s); 112,9(d); 142,9(d); 112,4(s); 129,5(d); 114,7(d); 156,4(s); 108,2(s); 154,5(s); 77,2(d); 70,7(d); 78,9(s); 25,8(q); 21,7(q).
Metoxi-metanoxila		MME	182	103,3(t); 53,7(t).

Tabela 8 ; Principais Macronós presentes nos alcalóides (Cont.)

Nome	Estrutura do Macronó	Sigla:	Código :	δ (RMN ^{13}C) ± 2
Tercbutil-dimetil silício		IPS	53	25,9(q) (x3); 18,55(s); -5,4(q); -5,4(q).
Tercbutoxila		TBU	105	81,1(s); 28,3(q)
Isovaleroíla		IVA	122	173,5(s); 43,3(t); 25,7(d); 22,2(q); 22,2(q).
2,2-dietoxi-etanoxila		DET	35	68,4(t); 100,2(d); 62,7(t); 62,7(t); 15,3(q); 15,3(q).
4-bromo-3-metil-butenoxila-2		BIP	46	65,2(t); 125,3(d); 137,4(s); 34,1(t); 15,5(q).
2-acetilglicosila		2AG	31	96,2(d); 74,1(d); 75,3(d); 71,5(d); 78,4(d); 62,4(t); 170,0(s); 21,4(q).
2-cloro-fenilaminila		2CB	143	127,9(s); 138,1(s); 133,3(d); 130,8(d); 126,1(d); 126,1(d).
Terc-butoximetanoíla		MTB	148	159,8(s); 84,9(s); 27,5(q) (x3).
2,3,4,6-tetra-acetil-glicosila		TAG	94	95,3(d); 70,5(d); 73,5(d); 67,7(d); 72,3(d); 61,7(t); 170,0(s)(x3); 20,8(q)(x3).
2-cloro-etoxila		2CE	154	72,5(t); 41,6(t).

O programa MACRONO, ao ser executado pela primeira vez, apresenta a tela do menu principal (Fig. 25).

Figura 25 – Menu principal do programa MACRONO.

A partir da opção 5, os deslocamentos químicos de RMN ^{13}C e as respectivas multiplicidades de um espectro problema são fornecidos ao programa e através da opção 6, o programa pesquisa quais os possíveis macronós presentes no espectro da substância. O programa fornece os nomes dos prováveis macronós, os sinais da amostra à ele atribuídos, o número de carbonos do substituinte e a faixa de erro médio. O número de carbonos do substituinte é um dado importante pois, se por exemplo, o espectro problema apresentar 16 sinais e for de um alcalóide acridônico pertencente ao esqueleto **ACD01** (14 carbonos), conclui-se que **dois sinais** não fazem parte do esqueleto carbônico. Analisando as subestruturas propostas pelo programa, os macronós que apresentarem mais que dois carbonos podem ser automaticamente descartados. A faixa de erro médio é a média dos desvios entre os deslocamentos químicos do espectro problema e os da base de dados. Portanto, o substituinte mais provável será aquele que apresentar a menor faixa de erro médio aliado a um número de carbonos compatível ao espectro problema.

Para exemplificar o uso do programa MACRONO foram utilizados os dados de RMN ^{13}C de um alcalóide pirrolizidínico (Fig.26), isolado da *Senecio nemorensis*⁴³ (Asteraceae) e que não consta da base de dados.

Figura 26 – Alcalóide pirrolizidínico

Dados do espectro de RMN ^{13}C (CDCl_3) :

173.5 (s)	172.2 (s)	132.4 (s)	129.7 (d)	81.6 (s)	78.2 (d)	71.5 (d)	70.9 (d)
63.0 (t)	62.5 (t)	53.7 (t)	43.3 (t)	35.9 (t)	32.3 (d)	25.7 (d)	22.2 (q)
22.2 (q)	17.5 (q)	16.3 (q)	14.5 (q)				

A substância analisada possui um esqueleto com quinze átomos de carbono. O número total de deslocamentos químicos foi de vinte átomos, o que significa que o macronó deverá ser uma subestrutura contendo cinco átomos de carbono.

A tela de resultados do programa apresentou duas propostas, as quais são mostradas na Figura 27. Isovalerato (5 carbonos) e Etóxido (2 carbonos). Baseando-se no número de átomos de carbono, pode-se descartar a possibilidade de o Etóxido ser o macronó em questão.

POSSIVEIS MACRONOS							
CODIGO	MACRONO					ERRO MEDIO	SINAIS
40)	ISOVALERATO					0.890	5
SINAIS:		173.20	43.30	25.70	22.20	22.20	
64)	ETOXIDO					0.375	2
SINAIS:		62.50	16.30				

Figura 27 – Tela de resultados do programa MACRONO.

Deste modo, os sinais referentes ao isovalerato podem ser retirados, restando apenas os quinze sinais (correspondem ao esqueleto carbônico do alcalóide) que serão utilizados pelos programas aplicativos na busca de esqueletos.

IV.3.3 – Programa Aplicativo SISCONST³⁸

O programa SISCONST utiliza os dados de RMN ^{13}C (deslocamentos químicos e multiplicidade) para determinar os esqueletos mais prováveis e fornecer grandes fragmentos estruturais (com os respectivos sinais de RMN ^{13}C atribuídos) cujos dados são compatíveis com os do espectro problema. O programa confronta os sinais do espectro questionado com todos os espectros pertencentes à base de dados. Se um sinal do espectro questionado está presente, por exemplo, no espectro da primeira substância do banco e tem a mesma multiplicidade, os deslocamentos químicos dos carbonos vizinhos interligados são então confrontados com os do espectro-problema. Este processo se repete até que seja fornecido o maior fragmento, da substância do banco, que apresenta deslocamentos químicos compatíveis com os dados do espectro-problema.

Após verificar todas as substâncias do banco de dados, o programa apresenta como resultados os esqueletos mais prováveis e os fragmentos aos quais os sinais do espectro-problema podem ser atribuídos. Para cada esqueleto proposto é atribuído um *percentual de reconhecimento (PR)* que fornece ao químico a confiabilidade da proposição deste esqueleto. Quanto maior for o valor do *PR*, mais confiável será a resposta do programa.

O usuário, além dos deslocamentos químicos e das multiplicidades da substância testada, deve fornecer ao programa as seguintes informações :

- número mínimo de átomos para proposição de subestruturas;
- a faixa de erro admitida pelo programa, geralmente, δ 1.0;
- gradiente, que permite o aumento automático da faixa de erro, recomendado um valor entre 0.5-1.0;
- espaço de busca da pesquisa, ou seja, se esta deve ser realizada sobre todas as substâncias da classe ou apenas sobre as substâncias de um esqueleto específico.

Neste último item, a escolha depende das informações que o pesquisador possui sobre a substância a ser elucidada, ou seja, se ele tiver certeza do tipo de esqueleto a que ela pertence, a busca será restrita apenas ao banco desse esqueleto. Quando se trata de uma substância totalmente desconhecida, é realizada uma busca utilizando todas as substâncias existentes no banco de dados.

Na busca do provável esqueleto, o programa somente escolhe substâncias que apresentem subestruturas com pelo menos metade do número total de carbonos. A partir das substâncias selecionadas é calculado a probabilidade de esqueleto do espectro-problema. Para a realização das pesquisas foi fixado como intervalo de erro, no deslocamento químico de RMN ^{13}C , 1,0δ para subestruturas e 5,0δ para esqueletos. Se a busca não for bem sucedida, os intervalos são automaticamente aumentados de acordo com um gradiente de 0,5 e 1,0δ para subestruturas e esqueletos, respectivamente. Caso este intervalo alcance valor maior que 3,0δ para subestruturas e 10,0δ para esqueletos, o programa cessa a busca, não apresenta previsão de esqueletos e também não fornece subestruturas.

Para exemplificar o funcionamento do programa SISCONST foram utilizados os sinais de um espectro de RMN ^{13}C de um alcalóide quinolínic, pertencente ao grupo do esqueleto Quinolínico21 (Fig.28).

Figura 28 – Substância utilizada para teste e esqueleto Quinolínico 21

Para a busca do esqueleto e suas respectivas subestruturas, o usuário inicialmente fornece o número total de sinais significativos da substância analisada (9 sinais) e seus respectivos deslocamentos químicos e multiplicidades (Tab.9).

Tabela 9 – Dados de RMN ^{13}C da substância colocada em teste.

SINAL	Deslocamento Químico	Multiplicidade
1	150,5	dubl
2	118,9	dubl
3	135,6	dubl
4	128,8	dubl
5	119,8	dubl
6	160,7	sing
7	107,3	dubl
8	123,5	sing
9	150,0	sing

Posteriormente o usuário estipula um número mínimo de cinco sinais coincidentes (subestruturas com menos de cinco sinais não serão consideradas neste teste) e um gradiente de δ 1,0.

O programa apresentou a probabilidade da substância pertencer a um determinado tipo de esqueleto (Tab.10) e listou duas subestruturas A, B (Fig.29) com atribuição dos deslocamentos químicos (Tab.11).

Tabela 10. Probabilidade de esqueletos apresentada pelo programa SISCONST

Esqueleto	Probabilidade (%)
Quinolínico 21	73,5%
Quinolínico 7	2,0%
Quinolínico 9	2,0%
Quinolínico 23	6,1%
Quinolínico 38	2,0%
Outros	14,4%

Figura 29 – Subestruturas fornecidas pelo programa SISCONST.

Tabela 11 – Sinais associados as subestruturas A, B e C .

Átomo	A	B	C
1	148,5	---	---
2	119,8	---	---
3	136,1	---	---
4	---	---	127,5
5	---	118,8	121,4
6	---	161,6	---
7	---	106,4	---
8	121,3	---	125,3
9	150,0	149,5	---
10	Heteroátomo	---	---
11	---	Heteroátomo	---

As estruturas A e B foram fornecidas pelo programa SISCONST, entrando com todos os dados de deslocamentos químicos contidos no espectro problema. Entretanto, para a obtenção da subestrutura C, o programa foi simulado novamente, desta vez utilizando o deslocamento 128,8(d) (correspondente ao deslocamento químico do carbono 4, não atribuído anteriormente nas subestruturas A e B), e três deslocamentos dos átomos vizinhos a ele (135,6 d; 123,5 s; 119,8 d – correspondentes aos átomos C3, C5 e C8 da Figura 28). Desta forma, obteve-se a estrutura completa da substância testada pela sobreposição dos 3 fragmentos fornecidos pelo programa.

IV.3.4 - Programa C13MACH³⁸

O C13MACH, assim como o SISCONST, é um sistema que utiliza os dados dos espectros experimentais de RMN ¹³C para prever o tipo de esqueleto de um determinado produto natural, trabalhando com um algoritmo de comparação e fornecendo também as estruturas existentes no banco de dados que têm espectros semelhantes ao da amostra analisada. Este programa confronta os sinais do espectro teste com os dados do banco e verifica o índice de semelhança entre eles, o qual é calculado a partir do número de sinais do espectro analisado que estão presentes em cada substância do banco, dentro da faixa de erro estabelecida. O programa seleciona um número de substâncias (que pode variar de 1 a 50) que apresentam maior similaridade. Quanto maior o índice de semelhança, maior será a probabilidade do espectro analisado pertencer a uma substância com o mesmo esqueleto daquela presente no banco.

O programa fornece, como dados de saída, as estruturas das substâncias onde ocorre alguma coincidência espectral (Fig. 31), o Índice de Semelhança (IS) e a porcentagem de ocorrência desses deslocamentos químicos nos vários esqueletos do banco de dados.

O usuário deve fornecer as seguintes informações para execução do programa.

- A faixa de erro admitida pelo programa, ou seja, de quanto pode ser a diferença entre o deslocamento químico fornecido e o deslocamento químico das substâncias presentes no banco;
- O número de substâncias com melhores índices de semelhanças que devem ser considerados para o cálculo da probabilidade de esqueleto. Este valor pode variar de acordo com a classe de substâncias estudada;
- Os dados dos deslocamentos químicos do espectro problema com as respectivas multiplicidades.

Este programa foi testado com o espectro de RMN ^{13}C de uma substância desconhecida pelo banco, cuja estrutura pertence ao esqueleto Benzilisoquinolínico 3 (Fig.30).

Figura 30 – Substância utilizada para teste e a estrutura do esqueleto Benzilisoquinolínico 3.

Figura 31 – Resultados do índice de semelhança apresentados pelo programa C13MACH.

Observa-se que o alto Índice de Semelhança (Fig. 31) encontrado na substância **A** é coerente, visto que a única diferença entre ela e a substância teste é a ligação entre C10 e o oxigênio heterocíclico, que na substância teste é uma ligação beta e na substância proposta pelo programa é uma ligação alfa.

O programa fornece então, o resultado da análise de Classe e o resultado da probabilidade de esqueleto (Tab.12) .

Resultado da Análise das Classes .

Classe : ALCALOIDE 100%

Tabela 12: Análise dos Esqueletos mais prováveis fornecidos pelo programa C13MACH :

Esqueleto	Probabilidade
Benzilisoquinolinico 3	92,0%
Aporfínico 4	8,0%

O programa C13MACH não exige a condição de que os átomos de carbonos, que tiverem seu dados espectrais confrontados, estejam interligados, ou seja, ele não considera o ambiente químico dos carbonos para efeito de confrontamento de dados. Isto pode conduzir a um resultado impreciso, tornando-o menos eficiente que o programa SISCONST.

4.3.5 – Programa SISOCBOT

O programa SISOCBOT³⁸ permite a análise dos dados botânicos utilizando como dado de entrada obrigatório, o nome da família na qual se quer realizar a pesquisa, sendo o gênero e a espécie dados optativos. Como resultado da análise, o programa fornece o percentual de ocorrência dos esqueletos na família e/ou gênero escolhidos. Estas informações podem ser utilizadas como restrições no processo de determinação estrutural.

Para exemplificar a utilização do programa SISOCBOT, foi realizada uma busca para determinar os percentuais de ocorrência de esqueletos dos alcalóides na Família Fabaceae (Tab.13)

Tabela 13. Percentuais de ocorrência de esqueletos de alcalóides na família Fabaceae.

Família Fabaceae	
<i>Esqueleto</i>	<i>Probabilidade</i>
Quinolizidínico 1	80%
Quinolizidínico 31	20%

IV.3.6 – O Programa MASMACH³⁸

Neste programa os dados de espectrometria de massas de uma substância em teste são confrontados com os dados contidos no banco de dados, da mesma forma como no C13MACH.

O banco de espectrometria de massas, assim como os dados de carbono e os dados botânicos, é criado através do programa DATASIS, utilizando-se a opção 18 da tela de entrada dos dados físico-químicos (Fig.24),

O programa MASMACH pode ser executado apenas sobre uma classe de substâncias, sobre um esqueleto específico ou sobre todas as classes. Como o objetivo é a identificação de uma substância (alcalóide) desconhecida, a pesquisa é, geralmente, realizada sobre todas as classes.

Uma substância pertencente ao esqueleto do alcalóide **Indol 25**⁴⁴ foi submetida ao teste (Fig.32) para demonstrar a utilização do programa.

Esqueleto	Probabilidade (%)
	INDOL25	92,6
	INDOL63	3,7
	ISOQ02	1,9
	ISOQ04	1,8

$m/z = 452 (M^+), 352, 251(100), 169.$

Figura 32 – Substância utilizada para teste e o resultado fornecido pelo programa.

IV.3.7 – O Programa H1MACH³⁸

O programa aplicativo H1MACH foi construído baseado na espectrometria de RMN ^1H , com o objetivo de ampliar o conjunto de dados armazenados no sistema e contribuir como uma ferramenta a mais na determinação estrutural de substâncias orgânicas.

O espectro de RMN ^1H fornece muitas informações, tais como, deslocamentos químicos, multiplicidades, integração dos sinais e constantes de acoplamentos. A interpretação destas informações não é simples (como ocorre com a técnica de RMN ^{13}C). Esta técnica é muito sensível a frequência dos aparelhos, interferindo na qualidade dos dados obtidos pela interpretação dos resultados. Para construir um banco, confiável, contendo todas as informações fornecidas pelos espectros de RMN ^1H , seria necessário que todos os dados (coletados da literatura) fossem obtidos em aparelhos utilizando a mesma frequência e os mesmos solventes. Visto que isso não é possível, os únicos dados inseridos nos bancos de RMN ^1H são os deslocamentos químicos dos hidrogênios presentes na substância.

A entrada dos deslocamentos químicos no banco de dados do programa é realizada obedecendo a mesma seqüência numérica estabelecida para os átomos de carbono, ou seja, o primeiro dado de RMN ^1H representará os hidrogênios localizados no primeiro átomo de carbono da codificação. No caso de um metileno, deverão ser inseridos dois deslocamentos químicos, independente de serem magneticamente equivalentes. Para os hidrogênios metílicos, apenas um valor será inserido no banco.

Da mesma forma que o C13MACH, o programa H1MACH trabalha por confrontamento de dados espectrais, fornecendo uma previsão de esqueleto no final da análise e apresentando as estruturas das substâncias que possuem um maior grau de similaridade com o espectro questionado.

O programa H1MACH é iniciado com a tela do menu principal (Fig. 33), onde são apresentadas 4 opções:

- Opção 1 – Separa os dados de RMN ^1H dos banco dos dados arquitetados, para serem utilizados pelo H1MACH;
- Opção 2 – Possibilita a entrada dos dados do espectro-problema;
- Opção 3 – Efetua a pesquisa de esqueleto;
- Opção 4 – Saída do sistema.

Figura 33 – Tela do menu principal do programa H1MACH

Para exemplificar a utilização do programa, foi efetuado um teste com o alcalóide Indólico 25 (Fig.32), para o qual o programa forneceu estruturas com os respectivos índices de similaridades (Fig.34) e as propostas de esqueletos (Tab.14) correspondentes aos dados de RMN ^1H do espectro-problema.

Dados de entrada δ (RMN ^1H): 4,00; 7,29; 7,05; 7,17; 7,51; 8,12; 1,63; 5,20; 3,34; 3,55; 7,89.

Figura 34 – Tela de resultados parciais do H1MACH.

Tabela 14 – Propostas de esqueletos para o programa H1MACH.

Esqueleto	Probabilidade (%)	Esqueleto	Probabilidade (%)
1) INDOL25	31,82	5) INDOL23	6,82
2) INDOL05	27,27	6) INDOL53	6,82
3) INDOL34	6,82	7) INDOL54	6,82
4) INDOL21	6,82	8) INDOL70	6,82

IV.3.8 – O programa PESQUISA

Inicialmente, as buscas por esqueletos exigia a entrada dos dados individualmente para cada programa aplicativo, por exemplo, os dados de RMN ^{13}C eram digitados para o programa SISCONST e novamente digitados para o programa C13MACH e cada correção realizada, por exemplo, nos dados de deslocamentos químicos (da substância em teste); era exigido que todos os deslocamentos fossem digitados novamente com suas respectivas multiplicidades. Com desenvolvimento do programa PESQUISA este trabalho tornou-se desnecessário.

O programa PESQUISA foi desenvolvido durante este trabalho para facilitar o processo de elucidação estrutural dos esqueletos de produtos naturais utilizando dados espectrométricos de RMN ^{13}C , RMN ^1H , massas e dados botânicos. Através do programa PESQUISA, o usuário tem acesso a todos os programas aplicativos descritos anteriormente, com a enorme vantagem de possuir um banco para a entrada dos dados da substância-teste, no qual os dados ficam disponíveis para serem utilizados em mais de um programa aplicativo. O usuário digita apenas uma vez os dados espectrais da substância teste e os dados botânicos (se a mesma for de origem natural), para que este conjunto de dados possa ser analisado por todos os programas aplicativos e posteriormente estes dados ficam arquivados em um banco do programa tornando-se disponível quando necessário.

O programa Pesquisa fornece os resultados individuais de probabilidade de esqueleto para cada um dos programas aplicativos. Terminada a análise, o programa PESQUISA gera um resultado global, considerando todos os resultados individuais de cada programa e os respectivos pesos atribuídos a cada um deles.

O programa PESQUISA é orientado por menus explicativos que facilitam a interface com o usuário. Ele se inicia com a geração de um arquivo, responsável pela identificação da substância colocada em teste. Os dados espectrométricos e botânicos desta substância são adicionados ao banco de dados do PESQUISA, podem ser recuperados e utilizados nos testes quantas vezes forem desejados. Dado o nome do arquivo, o programa abre a tela do menu principal (Fig. 35)

Figura 35 – Tela do menu principal do programa PESQUISA.

0 – Voltar para o Sistema. O programa retorna para a tela inicial .

1 – Entrar com os dados: Pode-se completar um arquivo inserindo os dados que estão faltando ou definir um novo arquivo e entrar com os dados para fazer a pesquisa em uma nova substância(Fig. 36)

2 – Gerenciar Macronós – Esta opção substitui o programa aplicativo MACRONO, apresentado anteriormente; nela os macronós são introduzidos, codificados, pesquisados e posteriormente eliminados para que não interfira na previsão de esqueletos realizadas posteriormente.

3 – Executar o PESQUISA – Escolhe-se o tipo de teste para busca de esqueleto, por exemplo: SISCONST, C13MACH, H1MACH, MASMACH ou SISOCBOT (Fig.38).

4– Combinar resultados – Esta opção fornece um rendimento global do programa.

5 – Alterar pesos dos resultados – Para cada programa aplicativo é estipulado um peso. Este peso é utilizado para calcular a probabilidade global e varia em função da eficiência que cada programa apresenta em suas propostas de esqueletos. Esta opção permite a modificação destes pesos

6 – DOSSHEL – Finalmente, esta opção retorna para o DOS.

Figura 36– Tela de definição dos dados de entrada .

A tela de entrada dos dados (Fig.37) é auto explicativa, nela encontra-se a opção de **Adicionar um novo dado(1)**, **Excluir(2)** e **Modificar(3)** os dados existentes, ou ainda, **entrar com todos os dados (4)**. Na parte superior da tela está indicada a especificação do dado de entrada (neste caso RMN ^{13}C), após a entrada dos dados especificados na tela seleciona-se a opção **0** para retornar à tela de definição dos dados para que os demais dados (RMN ^1H , massas e dados botânicos) possam ser inseridos no banco de dados do programa.

Figura 37 – Tela para entrada dos dados de RMN ^{13}C .

Após a entrada dos dados, retorna-se ao menu principal (Fig.35), seleciona-se a opção 2 para identificar os possíveis macronós existentes na substância utilizada para teste, posteriormente descarta-se os deslocamentos químicos correspondentes aos macronós identificados, para que sejam executados os programas aplicativos de pesquisa de esqueletos, os quais são

apresentados separadamente em uma tela (Fig.38) acessada através da opção 3 do menu principal do programa PESQUISA.

Figura 38 – Tela de acesso aos programas aplicativos.

Nesta tela seleciona-se a pesquisa a ser realizada e o programa roda automaticamente, buscando todos os dados necessários criados anteriormente, fornecendo a probabilidade de esqueleto e suas possíveis subestruturas. Como exemplo, é mostrado o funcionamento do programa SISCONST, que foi simulado com os dados de um alcalóide pertencente ao esqueleto Indol 25, cujo os deslocamentos químicos de RMN ^{13}C são àqueles mostrados na Figura 37. O usuário segue as seguintes etapas:

- O sistema pergunta se existe dados para rodar o programa;
- Em seguida pede um número mínimo de átomos por subestruturas;
- a margem de erro e o gradiente por subestrutura.

De posse de todos estes dados, o programa fornece as subestruturas encontradas através do confronto dos dados (Fig.39), pergunta se a proposta da subestrutura é boa e o usuário escolhe se quer ver as outras subestruturas ou obter diretamente o resultado da busca de esqueleto (Fig. 40).

Figura 39 – Proposta de subestrutura com as comparações dos respectivos deslocamentos químicos obtidos do banco de dados do SISTEMAT (em preto) com os do espectro problema (em vermelho).

SKELETONS PROPOSED BY SISCONST		
SKELETON		PROBABILITY (%)
1) INDOL03		1.45
2) INDOL05		5.93
3) INDOL20		1.16
4) INDOL21		2.89
5) INDOL24		16.12
6) INDOL17		2.19
7) INDOL25		41.43
8) INDOL27		1.16
9) INDOL06		1.81
10) INDOL29		1.81
11) INDOL30		1.16
12) INDOL31		2.31
13) INDOL32		1.88
14) INDOL33		2.31
15) INDOL53		1.81
16) INDOL63		3.11
17) INDOL54		1.81
18) INDOL48		1.45
19) OTHER		8.32

ERROR RANGE= 5.0
PROGRAM PESQUISA (GIA) - 01/04/2005 - 17:58

Figura 40 – Propostas de esqueletos fornecidas pelo programa SISCONST.

Após rodar todos os programas aplicativos, o módulo PESQUISA fornece uma tela com os resultados individuais de cada programa aplicativo e um resultado global (Fig.41).

SKELETONS PROBABILITY (%)							
SKELETON	1	2	3	4	5	6	7
1) INDOL25	41.4	62.5	0.0	90.9	31.0	0.0	5.0
2) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
3) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
4) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
5) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
6) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
7) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
8) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
9) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
10) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
11) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
12) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
13) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
14) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
15) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
16) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
17) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
18) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
19) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
20) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
21) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
22) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
23) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
24) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
25) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
26) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
27) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
28) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
29) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
30) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
31) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
32) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
33) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
34) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
35) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
36) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
37) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
38) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
39) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
40) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
41) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
42) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
43) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
44) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
45) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
46) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
47) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
48) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
49) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
50) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
51) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
52) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
53) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
54) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
55) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
56) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
57) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
58) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
59) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
60) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
61) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
62) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
63) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
64) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
65) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
66) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
67) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
68) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
69) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
70) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
71) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
72) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
73) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
74) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
75) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
76) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
77) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
78) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
79) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
80) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
81) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
82) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
83) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
84) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
85) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
86) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
87) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
88) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
89) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
90) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
91) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
92) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
93) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
94) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
95) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
96) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
97) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
98) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
99) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0
100) INDOL25	16.1	55.5	0.0	0.0	22.0	0.0	0.0

Figura 41 – Tela de apresentação dos resultados de probabilidade de esqueletos.

As substâncias que não possuem dados para rodar o programa, apresentam como resultados valores iguais a zero, que são desconsiderados para o cálculo de probabilidade global. Por exemplo, como esta substância teste é um derivado sintético, os dados botânicos ficaram em branco e o programa SISOCBOT (coluna n.º 3 – Fig. 41) não foi utilizado na busca de esqueletos. O mesmo ocorreu para o programa REGRAS, que é um programa de desfuncionalização (coluna n.º 6 – Fig. 41), cujos dados que permitem sua análise não foram arquivados no banco de dados do sistema e, portanto, ele não foi abordado nesta tese. Os demais programas terão seus resultados discutidos no próximo capítulo.

IV.4 – Resultados e Discussão

Neste capítulo serão apresentados e discutidos os resultados dos testes de determinação estrutural e identificação de esqueletos realizados nos programas aplicativos SISCONST, C13MACH, H1MACH, MASMACH e SISOCBOT.

Foram realizados 30 testes cujos resultados serão discutidos nesta tese. Nos testes, com substâncias pertencentes às subclasses dos alcalóides quinolínicos, aporfínicos, acridônicos, benzilisoquinolínicos, quinolizidínicos e pirrolizidínicos serão apresentados somente os resultados dos programas aplicativos SISCONST, C13MACH, MACRONO e SISOCBOT; visto que, no início não existia a intenção de trabalhar com os programas MASMACH e H1MACH, e portanto, os respectivos dados não foram arquivados na base de dados do programa. Nas substâncias pertencentes a subclasse dos alcalóides isoquinolínicos e indólicos foram efetuados testes utilizando, quando possível, todos os programas aplicativos. Em ambos os casos, somente as substâncias ditas inéditas foram usadas para os testes, ou seja, as que não estão presentes na base de dados. Apenas as cinco primeiras propostas de esqueletos, fornecidas para cada teste, serão apresentadas nas tabelas de resultados. No resultado global apresentado pelo programa Pesquisa (média ponderada das probabilidades fornecidas pelos programas aplicativos), foram atribuídos os seguintes pesos para cada programa: SISCONST:1,0; C13MCH:0,5; H1MACH:0,5; MASMACH:0,5 e SISOCBOT:0,5. Como o programa regras não foi utilizado neste trabalho, foi atribuído a ele um peso igual a zero.

MACRONO

Os prováveis macronós presentes nas substâncias utilizadas para teste serão apresentados com seus respectivos deslocamentos químicos (fornecidos pelo programa), encontrados pela comparação entre os deslocamentos químicos do substituinte proposto com àqueles presentes na substância teste.

SISCONST

As subestruturas fornecidas pelo programa SISCONST foram apresentadas com os respectivos deslocamentos químicos atribuídos pelo programa através do confrontação dos dados espectrais da substância teste com aquelas presentes no banco de dados. Para realização dos testes do programa SISCONST foi fornecido um número mínimo de átomos de carbono, correspondente a metade dos átomos presentes na substância utilizada para teste e uma faixa de erro, o qual foi estipulado em 2.0 ppm. Nos casos em que o programa não forneceu uma proposta de esqueleto, esta faixa foi estendida em até 4,0 ppm, sempre variando em um gradiente de 1,0 ppm.

C13MACH

No programa C13MACH, a probabilidade de uma substância pertencer a um determinado tipo de esqueleto foi obtida com base nas dez primeiras substâncias listadas pelo programa, com maior índice de semelhança (IS) e dentro de uma faixa de erro de 2,0 ppm.

SISOCBOT

O programa SISOCBOT forneceu os percentuais de ocorrência de esqueletos, na família e gênero de onde a substância teste foi isolada. Nos casos onde a mesma era um derivado sintético, o teste não foi realizado.

MASMACH

Para o programa MASMACH foi estipulado um número máximo de dez substâncias que apresentassem uma maior semelhança, quando comparados os dados espectrométricos da substância teste com os dados presentes no banco.

H1MACH

No programa H1MACH a margem de tolerância para os deslocamentos químicos foi de 0,5 ppm e foram estipulados um número máximo de dez substâncias que apresentasse um maior Índice de Semelhança, quando comparados os dados espectrométricos da substância teste com os dados presentes no banco.

As tabelas com todos os resultados mostram apenas as cinco primeiras propostas de esqueletos, descartando as demais. Isso faz com que a soma das probabilidades geralmente não fechem em 100%.

Teste 01

Derivado sintético
Heterocycles, 1997, pg.1529

Macronós Propostos:

Metoxila –

Sinal : 56,90

Tercbutoximetanoila –

Sinais : 159,80; 84,90; 27,50; 27,50; 27,50

Esqueleto Indol 25

Tabela 15 – Dados de entrada do Teste 01.

Dados de RMN ¹³ C (CDCl ₃)						
134,3(s)	55,1(d)	51,4(t)	19,6(t)	108,1(s)	127,5(s)	118,1(d)
119,3(d)	121,3(d)	110,8(d)	136,0(s)	32,9(t)	36,9(d)	117,6(s)
146,4(d)	12,8(q)	118,9(d)	133,6(s)	51,7(t)	167,7(s)	149,8(s)
84,9(s)	27,5(q)	27,5(q)	27,5(q)	56,9(q)		
Dados de RMN ¹ H (CDCl ₃)						
4,00	7,29	7,05	7,17	7,51	8,12	1,63
5,20	3,34	3,55	7,89			
Dados de massa (m/z)						
452 (M ⁺)	352	251 (100)	169 (80)			

Subestruturas propostas pelo SISCONST

Tabela 16- Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 01

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL25	41,9	62,5	0,0	0,0	48,1	38,7
2) INDOL05	5,9	0,0	0,0	56,3	20,5	17,7
3) INDOL24	16,1	9,4	0,0	0,0	0,0	8,3

* Valor atribuído para o programa MASMACH utilizando o pico em m/z 169.

Ao executar os programas aplicativos para obtenção de propostas de esqueletos (Tab.16), utilizando os dados espectrais da substância apresentada no TESTE 01(Tab15), foi observado que, embora no resultado global o programa tenha proposto o esqueleto correto, o programa MASMACH

não forneceu um bom resultado. Isso ocorreu por que na desfragmentação deste alcalóide é gerado um pico intenso em $m/z = 169$ (Fig.42) que corresponde ao fragmento do pico base do alcalóide pertencente ao esqueleto Indol 05, e este fragmento também está presente em diversos outros esqueletos (Fig 7) que possuem este núcleo em comum.

Figura 42 – Fragmento correspondente ao pico 169

A partir do momento que este pico foi retirado, e o teste realizado com os demais picos presentes, o resultado fornecido pelo programa MASMACH, além de acertar a proposta de esqueleto, este acerto veio seguido de um alto grau de confiança (Tab.17).

Tabela 17 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 01

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MAFMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL25	41,9	62,5	0,0	92,6	48,1	57,4
2) INDOL24	16,1	9,4	0,0	3,7	0,0	9,1
3)INDOL 63	3,1	9,3	0,0	0,0	18,0	6,7
4) INDOL05	6,0	0,0	0,0	0,0	20,5	6,5
5) INDOL10	0,0	9,5	0,0	0,0	0,0	1,9

* Valor atribuído para o programa MASMACH SEM o pico em m/z 169.

OBS: Os picos de fragmentos de massa só podem ser retirados se já existir um indício do esqueleto, proposto pelos demais programas, que identifique pelo menos a subclasse do alcalóide pesquisado, para que possa ser identificado um núcleo comum a vários esqueletos e então o pico correspondente a este núcleo pode ser removido para não interferir nos resultados (um exemplo é mostrado na Figura 43).

Figura 43 – Desfragmentação de três alcalóides Isoquinolínicos, pertencentes a esqueletos diferentes que originam o mesmo fragmento.

Teste 02

Rubiaceae – *Mitragyna speciosa*

J. Med.Chem., 2002, pg.1953

Macronós Propostos:

Metoxila –

Sinal : 57,9 e 61,6

Acetoíla –

Sinais : 170,1; 21,1

Esqueleto Indol 25

Tabela 18 – Dados de entrada do Teste 02.

Dados de RMN ¹³ C (CDCl ₃)						
170,1 (s)	169,3 (s)	160,7 (d)	143,7 (s)	137,9 (s)	136,2 (s)	121,3 (d)
120,4 (t)	112,0 (d)	111,4 (s)	108,9 (d)	106,3 (s)	61,1 (d)	57,3 (t)
53,4 (t)	40,7 (d)	40,0 (d)	29,9 (t)	22,9 (t)	21,1 (q)	19,1 (t)
12,9 (q)	61,6 (q)	57,5 (q)				
Dados de RMN ¹ H (CDCl ₃)						
7,95	7,43	7,15	7,05	6,75	3,15	3,04
3,04	3,04	2,92	2,71	2,55	2,55	2,46
2,21	1,80	1,75	1,63	0,86	0,86	0,86
Dados de massa (m/z)						
426	411	297	242			

Subestruturas propostas pelo SISCONST

Tabela 19 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 02

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL25	80,2	56,0	100,0	100,0	69,3	80,9
2) INDOL55	8,5	0,0	0,0	0,0	0,0	2,8
3) INDOL33	7,6	0,0	0,0	0,0	0,0	2,5
4) INDOL03	0,0	0,0	0,0	0,0	14,9	2,5
5) INDOL63	3,8	0,0	0,0	0,0	4,0	1,9

Teste 03

Ancistrocladaceae – *Ancistrocladus*

tanzaniensis

J. Nat. Prod., 2003, pg.1159

Macronós Propostos

Metoxila :

Sinais : 57,8; 60,5; 55,7; 56,5

Esqueleto Isoquinolínico 41

Tabela 20 – Dados de entrada do Teste 03.

Dados de RMN ^{13}C (CDCl_3)						
175,6(s)	165,9(s)	163,5(s)	157,1(s)	157,5(s)	140,2(s)	137,1(1)
135,8(s)	128,8(d)	123,4(d)	123,3(s)	122,7(s)	116,1(s)	108,9(d)
108,4(s)	105,0(d)	94,0(d)	57,8(q)	60,5(q)	55,7(q)	56,5(q)
17,2(q)	47,6(d)	31,5(t)	24,5(t)	22,1(q)		

Dados de RMN ^1H (CDCl_3)						
7,06	6,84	6,70	6,54	6,51	3,79	2,86
2,86	2,86	2,60	2,33	2,33	2,33	2,21
1,23	1,23	1,23				

Dados de massa (m/z)						
419	407	406	392	388	196	

Subestruturas propostas pelo SISCONST

Tabela 21 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 03

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) ISOQ41	87,8	0,0	28,6	50,0	8,2	43,7
2) ISOQ40	0,0	20,9	57,1	50,0	32,9	26,8
3) ISOQ16	0,0	0,0	14,3	0,0	0,0	2,4
4) INDOL 35	0,0	0,0	0,0	0,0	26,6	4,4
5) QUINOL 04	0,0	11,1	0,0	0,0	0,0	1,9

Os resultados dos programas aplicativos indicaram dois esqueletos, os Isoquinolínicos 40 e 41, como as melhores propostas. O resultado global indicou com maior probabilidade o esqueleto Isoquinolínico 41 (resultado gerado em função da proposta de esqueleto obtido a partir do programa SISCONST). Entretanto os demais programas aplicativos forneceram uma probabilidade maior ou igual para o esqueleto Isoquinolínico 40. Isso ocorreu devido a grande semelhança existente entre os dois esqueletos (Fig.7) e a maior amostragem de substâncias, contidas no banco de dados, pertencentes ao esqueleto Isoquinolínico 40 (Tab. 6). O resultado apresentado pelo programa MASMACH, (50:50) pode ser explicado pela variação comum que ocorre com os substituintes de produtos naturais. Pode-se observar (Fig.44) que apenas a modificação de um substituinte tornam iguais as massas dos esqueletos Isoquinolínicos 40 e 41. Portanto, apesar da grande semelhança entre os esqueletos Isoquinolínicos 40 e 41, o programa foi capaz de identificar o esqueleto correto. Este exemplo mostra a grande aplicabilidade e versatilidade do SISTEMAT.

Figura 44 – Subestruturas pertencentes a esqueletos diferentes, apresentando a mesma massa molecular, devido a diferença de substituintes.

Teste 04

Derivado Sintético

Heterocycles, 1999, Pg. 243

Macronós Propostos

Metoxila :

Sinail : 52,7

Esqueleto Indol 17

Tabela 22 – Dados de entrada do Teste 04.

Dados de RMN ¹³ C (CDCl ₃)						
178,1(s)	148,6(s)	145,5(s)	140,6(s)	136,7(s)	123,5(s)	112,9(d)
112,4(s)	111,6(d)	59,7(d)	55,1(t)	52,7(q)	51,5(3)	46,4(d)
30,4(t)	24,9(t)	23,7(t)				
Dados de RMN ¹ H (CDCl ₃)						
3,90	2,67	8,78	8,49	9,62		
Dados de massa (m/z)						
374	357	327	288	287	243	

Subestruturas propostas pelo SISCONST

Tabela 23 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 04

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL 17	26,7	75,6	0,0	0,0	0,2	29,5
2) ISOQ 01	0,0	0,0	0,0	0,0	52,5	9,7
3) INDOL 03	0,0	0,0	0,0	50,0	0,0	9,3
4) INDOL 04	0,0	0,0	0,0	50,0	0,0	9,3
5) INDOL630	11,8	0,0	0,0	0,0	0,0	4,4

Teste 05

Apocinaceae – *Kopsia fruticosa*

Phytochemistry, 2004, Pg. 2119

Macronós Propostos

Metoxila :

Sinail : 55,2

Esqueleto Indol 36

Tabela 24 – Dados de entrada do Teste 05.

Dados de RMN ¹³ C (CDCl ₃)						
164,7(s)	145,9(s)	145,2(d)	138,8(s)	136,7(s)	134,1(s)	129,1(d)
126,5(d)	120,3(d)	115,1(d)	109,4(d)	70,3(s)	67,9(d)	62,3(s)
55,2(q)	53,1(t)	47,2(t)	37,5(s)	37,1(t)	32,2(t)	28,1(t)
Dados de RMN ¹ H (CDCl ₃)						
6,90	6,84	6,78	6,69	5,94	5,80	3,77
3,36	3,29	3,00	2,48	2,19	1,91	1,65
1,50	1,32	1,25				
Dados de massa (m/z)						
364	336	281	249	199		

Subestrutura proposta pelo SISCONST

Tabela 25 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 05

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL43	49,3	0,0	22,2	16,7	5,7	23,9
2) INDOL17	0,0	0,0	0,0	66,7	0,0	11,1
3) INDOL35	12,1	0,0	38,9	0,0	0,0	10,5
4) INDOL36	16,8	0,0	11,1	0,0	0,0	7,4
5) INDOL03	8,4	0,0	11,1	0,0	5,7	5,6

Os esqueletos denominados Indol 43 e 36 (Fig.7) se diferenciam apenas pela substituição do hidrogênio, ligado ao nitrogênio heterocíclico (Indol 36), por uma metila (Indol 43) . Embora o resultado não identifique o esqueleto Indol 36, como predominante, ele aparece entre os cinco primeiros esqueletos. Um especialista seria capaz de perceber o equívoco do programa, visto que, nos dados de RMN ¹³C não consta nenhum deslocamento químico característico deste átomo de carbono (quadrupeto $\delta \cong 35,0 - 48,0$). A subestrutura fornecida é coerente e muito útil para elucidação da substância teste.

Teste 06

Rubiaceae – *Neonalclea sessilifolia*

Phytochemistry, 2003, Pg. 359

Macronós Propostos

Metoxila :

Sinai : 55,2

Glicose :

Sinais : 100,4; 78,7; 78,1; 74,7; 71,7; 63,0.

Esqueleto Indol 49

Tabela 26 – Dados de entrada do Teste 06.

Dados de RMN ¹³ C (CDCl ₃)						
169,3(s)	167,4(s)	154,8(d)	144,0(s)	143,0(s)	137,8(s)	137,1(s)
135,2(d)	132,2(s)	131,2(d)	123,4(d)	122,6(s)	122,4(d)	119,9(t)
116,4(d)	113,5(d)	110,1(s)	100,4(d)	97,4(d)	74,7(d)	78,1(d)
71,7(d)	78,7(d)	63,(t)	45,7(d)	34,9(d)	34,0(t)	
Dados de RMN ¹ H (CDCl ₃)						
8,81	8,29	7,67	7,67	7,55	7,38	5,90
5,84	5,12	5,08	3,61	3,56	3,56	2,69

Subestrutura proposta pelo SISCONST

Tabela 27 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 06

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL49	81,3	20,3	11,1	0,0	13,0	41,4
2) INDOL25	10,6	0,0	22,2	0,0	0,0	8,7
3) INDOL42	3,3	0,0	22,2	0,0	0,0	5,7
4) INDOL04	0,0	11,8	11,1	0,0	0,0	4,6
5) ISOQ 01	0,0	20,0	0,0	0,0	0,0	4,0

Teste 07

Rubiaceae – *Psychotria leiocarpa*

Phytochemistry, 2004, Pg. 449

Macronós Propostos

Glicose :

Sinais : 99,6; 74,9; 78,0; 71,6; 78,3; 62,7.

Esqueleto Indol 42

Tabela 28 – Dados de entrada do Teste 07.

Dados de RMN ¹³ C (CDCl ₃)						
166,3(s)	149,2(d)	137,7(s)	136,1(s)	133,4(d)	129,6(s)	122,9(d)
121,3(d)	120,7(t)	119,3(d)	114,8(d)	111,5(s)	109,1(s)	99,63(d)
97,5(d)	78,3(d)	78,0(d)	74,9(d)	71,6(d)	62,7(t)	54,5(d)
44,1(d)	41,6(t)	35,6(t)	27,9(d)	22,3(t)		
Dados de RMN ¹ H (CDCl ₃)						
7,63	7,48	7,44	7,12	7,06	5,51	5,48
5,26	5,14	5,12	5,05	3,40	2,84	2,78
2,75	2,65	2,30	1,38			
Dados de massa (m/z)						
661	659	515	499	353	187	185

Subestrutura proposta pelo SISCONST

Tabela 29 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 07

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL42	37,2	10,6	22,2	80,0	76,1	43,9
2) INDOL25	6,1	9,4	22,2	6,7	0,0	8,4
3) INDOL49	16,5	0,0	11,1	0,0	0,0	7,3
4) INDOL04	9,2	0,0	11,1	0,0	0,0	4,9
5) INDOL 43	9,2	0,0	0,0	0,0	5,2	3,9

Teste 08

Rubiaceae – *Chimarrhis turbinata*

J. Nat. Prod., 2003, Pg. 1017

Macronós Propostos

Glicose :

Sinais :100,0; 74,2; 77,5; 71,3; 78,0; 62,6.

Esqueleto Indol 25

Tabela 30– Dados de entrada do Teste 08.

Dados de RMN ¹³ C (CDCl ₃)						
171,1(s)	156,5(d)	138,6(s)	138,0(s)	134,8(d)	127,2(s)	122,8(d)
120,0(d)	119,3(t)	118,6(d)	111,7(d)	108,5(s)	108,1(s)	100,0(d)
97,0(d)	78,0(d)	77,5(d)	74,2(d)	71,3(d)	62,6(t)	52,3(q)
52,6(d)	48,5(t)	45,0(d)	34,4(t)	32,4(t)	23,8(t)	
Dados de RMN ¹ H (CDCl ₃)						
7,80	7,47	7,29	7,11	7,03	5,90	5,85
5,37	2,25	4,60	3,90	3,42	3,26	3,10
3,00	2,77	2,40	2,20			
Dados de massa (m/z)						
532	452	352				

Subestrutura proposta pelo SISCONST

Tabela31 – Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 08

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL25	26,0	9,4	22,2	96,2	5,6	30,9
2) INDOL49	23,5	12,1	11,1	0,0	0,0	11,7
3) INDOL23	5,7	0,0	7,4	0,0	50,0	11,5
4) INDOL04	11,4	22,3	11,1	0,0	0,0	9,4
5) INDOL20	13,8	9,4	0,0	0,0	5,6	7,1

**Cactaceae – *Dolichothele*
*longimamma***

J. Nat. Prod. 2002, Pg. 748

Teste 09

Macronós Propostos :

Substância sem macronó .

O Esqueleto **Inédito**- não consta no Banco de dados

OBS: Vale lembrar que existe uma diferença entre substância inédita e esqueleto inédito, ou seja, todas as substâncias testadas são inéditas. Entretanto, quase todas possuem um esqueleto básico no banco de dados que lhes servem de referência. No caso de compostos como os do teste 09 este esqueleto não existe, no banco, para efeito de comparação.

Tabela 32 – Dados de entrada do Teste 09.

Dados de RMN ¹³ C (CDCl ₃)						
136,1(s)	126,0(s)	122,9(d)	120,9(d)	117,9(s)	115,3(s)	110,8(d)
91,3(s)	32,4(q)					
Dados de RMN ¹ H (CDCl ₃)						
7,57	7,41	7,27	7,18	3,82	3,82	3,82

Subestruturas propostas pelo SISCONST

Tabela 33 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 09

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL70	38,5	10,6	0,0	0,0	0,0	20,4
2) INDOL76	34,9	9,6	0,0	0,0	0,0	18,5
3) ISOQ01	2,4	0,0	0,0	0,0	49,2	12,3
4) ISOQ02	0,0	0,0	0,0	0,0	49,2	12,3
5) INDOL71	10,8	0,0	0,0	0,0	0,0	5,0

O programa forneceu como propostas as estruturas que melhor representam o esqueleto da substância teste. Através da sobreposição das subestruturas, o esqueleto em questão foi obtido por completo. Os altos valores das propostas de esqueletos atribuídos pelo programa H1MACH, para os alcalóides Isoquinolínicos, foram dadas em função da grande semelhança entre os deslocamentos químicos dos hidrogênios presentes nos anéis aromáticos destas duas subclasses de alcalóides.

Teste 10

Esponja – *Hyrtios erectus*

J. Nat. Prod. 2002, Pg. 1173

Macronós Propostos :

Substância sem macronó .

Esqueleto Indol 70

Tabela 34 – Dados de entrada do Teste 10.

Dados de RMN ¹³ C (CDCl ₃)						
151,3(s)	133,4(s)	129,9(s)	124,7(d)	112,9(d)	112,5(d)	112,2(s)
103,8(d)	63,9(t)	30,2(t)				
Dados de RMN ¹ H (CDCl ₃)						
7,15	7,00	6,91	6,64	3,77	3,77	2,89
2,89						
Dados de massa (m/z)						
177	146	117	91			

Subestruturas propostas pelo SISCONST

Tabela 35 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 10

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL70	51,6	19,5	66,7	1,9	0,8	32,0
2) ISOQ01	6,3	0,0	0,0	48,5	58,1	19,9
3) ISOQ05	6,3	0,0	0,0	48,5	0,0	10,2
4) ISOQ02	4,7	9,8	0,0	0,0	40,4	9,9
5) INDOL67	1,6	9,4	33,3	0,0	0,0	7,6

Teste 11

Haliclonidae, *Haliclona* sp

J. Nat. Prod. 2001, Pg. 1249

Macronós Propostos :

Metoxila : 56,2

Esqueleto Isoquinolínico 06

Tabela 36 – Dados de entrada do Teste 11.

Dados de RMN ¹³ C (CDCl ₃)						
156,0(s)	154,6(s)	151,8(s)	135,8(s)	129,3(d)	120,9(d)	119,0(s)
109,2(d)	104,7(d)	58,6(t)	56,2(q)			
Dados de RMN ¹ H (CDCl ₃)						
8,19	8,01	7,58	7,47	5,34		
Dados de massa (m/z)						
206	---	---	---	---	---	---

Subestruturas propostas pelo SISCONST

Tabela 37 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 11

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) ISOQ06	28,1	26,9	0,0	33,3	89,2	41,1
2) ISOQ04	0,0	0,0	0,0	66,7	0,0	13,3
3) ISOQ02	24,0	0,0	0,0	0,0	0,0	9,6
4) ISOQ01	15,6	8,6	0,0	0,0	0,0	8,0
5) ISOQ39	16,7	0,0	0,0	0,0	0,0	6,7

Teste 12

Menispermaceae – *Menispermum*
dauricum

Phytochemistry. 2004, Pg. 929

Esqueleto Isoquinolínico 01

Macronós Propostos :

Metoxila

Sinais : 55,8

Tabela 38 – Dados de entrada do Teste 12.

Dados de RMN ^{13}C (CDCl_3)						
165,4(s)	161,2(s)	152,5(s)	146,3(s)	132,2(s)	120,0(s)	110,6(d)
107,0(d)	104,8(d)	55,8(q)				
Dados de RMN ^1H (CDCl_3)						
7,47	7,08	6,95	6,42			
Dados de massa (m/z)						
192	135	76				

Subestruturas propostas pelo SISCONST

Tabela 39 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 12

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) ISOQ 01	35,4	40,4	0,0	74,8	45,0	46,2
2) ISOQ 02	29,8	9,4	0,0	24,9	45,0	27,8
3) INDOL74	20,5	30,5	0,0	0,0	0,0	14,3
4) ISOQ 06	8,1	9,4	0,0	0,0	0,0	5,1
5) ISOQ 04	1,9	10,3	0,0	0,0	0,0	2,8

Teste 13

Cactaceae – *Dolichothele longimamma*

J. Org. Chem. 1976, Pg. 319

Macronós Propostos :

Substância sem macronó .

Esqueleto Isoquinolínico02

Tabela 40 – Dados de entrada do Teste 13.

Dados de RMN ^{13}C (CDCl_3)						
156,7(s)	142,1(s)	130,3(d)	128,3(s)	114,8(d)	112,6(d)	60,7(t)
57,6(t)	38,4(q)	31,2(t)				
Dados de RMN ^1H (CDCl_3)						
7,10	6,82	6,77	4,33	4,33	3,47	3,47
3,47	3,47	3,05	3,05	3,05		
Dados de massa (m/z)						
163	162	120	91	44		

Subestruturas propostas pelo SISCONST

Tabela 41 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 13

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) ISOQ 02	93,1	38,9	75,0	50,0	61,0	68,5
2) ISOQ 01	3,5	17,0	25,0	30,0	25,5	17,4
3) ISOQ 04	0,0	0,0	0,0	10,0	12,8	3,8
4) ISOQ 06	0,0	0,0	0,0	10,0	0,0	1,7
5) ISOQ 39	0,0	8,8	0,0	0,0	0,0	1,5

Teste 14

Fabaceae – *Lupinus argenteus*

J. Org. Chem. 1990, Pg. 1204

Macronós Propostos :

Acetoila :

Sinais : 169,5; 21,5

Esqueleto Quinolizidínico 01

Tabela 42 – Dados de entrada do Teste 14.

Dados de RMN ¹³ C (CDCl ₃)						
169,6(s)	169,5(s)	169,5(s)	137,6(s)	101,7(d)	78,2(s)	73,3(d)
61,6(d)	53,9(t)	52,3(t)	36,9(d)	25,1(t)	25,0(t)	24,9(t)
21,5(q)	21,5(q)	18,9(t)	16,7(t)	16,6(t)		

Subestruturas propostas pelo SISCONST

Tabela 43 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 14

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) QUINOLIZID 01	88,3	74,4	83,3	83,6
2) QUINOLIZID 31	0,0	0,0	16,7	4,2
3) INDOL 48	1,2	8,4	0,0	2,7
4) QUINOLIZID 18	0,0	8,6	0,0	2,2
5) QUINOLIZID 3	0,0	8,5	0,0	2,1

Teste 15

Mitrephoreae – *Popowia*
pisocarpa

J. Nat. Prod. 1986, Pg. 1018

Macronós Propostos :

Metoxila :

Sinais : 58,5; 60,1

Esqueleto Benzilisoquinolínico 01

Tabela 44 – Dados de entrada do Teste 15.

Dados de RMN ¹³ C (CDCl ₃)						
155,8(s)	144,4(s)	144,1(s)	133,4(s)	130,5(d)	130,5(d)	127,6(s)
125,2(s)	121,9(d)	114,6(d)	112,3(d)	110,6(d)	64,4(d)	46,2(t)
41,8(q)	39,9(t)	24,4(t)				

Subestruturas propostas pelo SISCONST

Tabela 45 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 15

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) BENZILISOQ 01	58,7	62,8	0,0	60,0
2) APORFINICO 05	13,8	0,0	0,0	9,2
3) BENZILISOQ 03	9,5	0,0	0,0	6,3
4) BENZILISOQ 08	0,0	9,8	0,0	3,3
5) ISOQ 04	0,0	9,6	0,0	3,2

Teste 16

Derivado Sintético

J. Heteroc.Chem 1993, Pg. 977

Macronós Propostos :

Etoxila –

Sinais : 65,6 – 16,7

Esqueleto **Inédito** – Não consta no banco de dados

Tabela 46 – Dados de entrada do Teste 16.

Dados de RMN ^{13}C (CDCl_3)						
192,1(s)	128,5(s)	166,9(s)	144,0(s)	137,8(s)	137,3(d)	128,4(d)
123,8(d)	122,9(d)	115,7(d)	115,5(s)	67,7(s)	67,0(t)	40,2(t)
35,9(t)	29,9(q)	16,1(q)	14,9(q)			

Subestruturas propostas pelo SISCONST

Tabela 47 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 16

Esqueleto	SISCONST(%)	C13MACH(%)	GLOBAL(%)
1) QUINOLÍNICO 24	83,1	0,0	55,4
2) ACRIDÔNICO 17	12,8	9,7	11,8
3) APORFÍNICO 01	0,0	19,4	6,5
4) QUINOLÍNICO 03	5,1	0,0	3,4
5) QUINOLÍNICO 06	5,1	0,0	3,4

A substância utilizada no teste 16 não possui seu esqueleto definido no banco de dados. Entretanto, mesmo se tratando de um esqueleto inédito, o programa forneceu a estrutura de um esqueleto semelhante (Fig.45).

Figura 45 – Substância codificada, encontrada no banco de dados. Esqueleto **Quínolínico 24**

Teste 17

Derivado Sintético

J. Heteroc. Chem 1991, Pg. 167

Macronós Propostos :

Metoxila : 58,1

Esqueleto – Quinolínico 10

Tabela 48 – Dados de entrada do Teste 17.

Dados de RMN ¹³ C (CDCl ₃)						
162,9(s)	161,7(s)	156,4(s)	144,9(s)	128,9(d)	126,9(d)	123,2(d)
121,7(d)	118,2(s)	103,5(s)	100,8(d)	58,1(q)	52,0(s)	24,8(q)
24,8(q)						

Subestruturas propostas pelo SISCONST

Tabela 49 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 17

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) QUINOLÍNICO 10	35,1	19,2	0,0	29,8
2) QUINOLÍNICO 23	17,6	0,0	0,0	11,7
3) BENZILISOQ 23	13,5	0,0	0,0	9,0
4) QUINOLÍNICO 06	9,5	0,0	0,0	6,3
5) ISOQ 03	6,8	0,0	0,0	4,5

Teste 18

Derivado Sintético
Chem. Ber. 1975, Pg. 1043

Macronós Propostos :
Substância sem macronó

Esqueleto **Inédito**– Não consta no banco de dados

Tabela 50 – Dados de entrada do Teste 18.

Dados de RMN ^{13}C (CDCl ₃)						
169,9(s)	56,8(d)	42,0(t)	34,1(t)	33,1(t)	30,7(t)	25,6(t)
24,7(t)	19,5(t)					

Subestruturas propostas pelo SISCONST

Tabela 51 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 18

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) QUINOLIZID 01	0,0	60,3	0,0	60,2
2) QUINOLIZID 22	0,0	10,4	0,0	10,4
3) QUINOLIZID 04	0,0	10,1	0,0	10,1
4) QUINOLIZID 02	0,0	9,9	0,0	9,9
5) QUINOLIZID 27	0,0	9,3	0,0	9,3

No teste 18, foi utilizada uma substância com esqueleto inédito. O programa SISCONST não apresentou propostas de esqueletos, embora tenha sugerido boas propostas de subestruturas, que, por sobreposição dos dois fragmentos, obteve-se a estrutura completa da substância submetida ao teste. A proposta fornecida pelo programa C13MACH foi baseada no fato de que existem, no banco de dados, varias substâncias pertencentes ao esqueleto Quinolizidínico 01 (Fig46), cujos deslocamentos químicos de RMN ^{13}C de um pedaço da estrutura coincidem com os da substância testada.

Leguminosae - *Calpurnia aurea*
Quinolizidínico 01
Mag. Res. Chem., 1999, 779

Figura 46 – Uma das muitas substâncias do banco de dados que apresenta uma pedaço semelhante ao da substância utilizada no teste 18.

Teste 19

Rutaceae – *Acronychia bauerii*

Chem. Pharm.Bull., 1983, Pg. 3084

Esqueleto – Acridônico 18

Macronós Propostos :

Metoxila : 59,8

Tabela 52 – Dados de entrada do Teste 19.

Dados de RMN ¹³ C (CDCl ₃)						
181,9(s)	161,5(s)	159,1(s)	150,5(s)	146,0(s)	138,2(s)	131,0(s)
125,0(s)	123,6(d)	122,7(d)	122,6(d)	121,5(d)	117,8(d)	115,2(d)
110,3(s)	107,1(s)	102,2(s)	76,4(s)	49,0(q)	27,1(q)	27,1(q)
25,9(q)	24,5(t)	18,0(q)				

Subestruturas propostas pelo SISCONST

Tabela 53 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 19

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) ACRIDÔNICO 18	56,5	51,3	20,0	46,1
2) ACRIDÔNICO 19	37,5	9,4	12,0	24,1
3) ACRIDÔNICO 07	1,5	30,3	36,0	17,3
4) ACRIDÔNICO 04	0,0	9,0	4,0	3,2
5) ACRIDÔNICO 23	4,2	0,0	4,0	3,1

Neste teste, além do programa fornecer a proposta de esqueleto correta, ele também indicou, como segundo maior valor, o esqueleto Acridônico 19, que se diferencia do Acridônico 18 somente pela ausência da metila ligada ao Nitrogênio heterocíclico.

Teste 20

Rutaceae – *Glycosmis citrifolia*

J. Nat. Prod., 1995, Pg. 1629

Esqueleto – Acridônico 17

Tabela 54 – Dados de entrada do Teste 20.

Dados de RMN ¹³ C (CDCl ₃)						
180,8(s)	159,4(s)	159,3(s)	156,5(s)	142,6(s)	141,3(s)	136,3(s)
129,8(s)	121,8(d)	116,2(s)	113,1(d)	104,7(s)	94,0(d)	46,4(q)

Subestruturas propostas pelo SISCONST

Tabela 55- Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 20

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) ACRIDÔNICO 17	45,2	31,9	41,7	41,0
2) ACRIDÔNICO 07	0,0	19,5	41,7	15,3
3) ACRIDÔNICO 12	10,8	9,4	0,0	7,8
4) ACRIDÔNICO 47	10,2	0,0	0,0	5,1
5) ACRIDÔNICO 31	5,1	0,0	8,3	4,6

Teste 21

Rutaceae – *Glycosmis citrifolia*

Chem. Pharm. Bull., 1983, Pg. 3084

Esqueleto – Acridônico 07

Tabela 56 – Dados de entrada do Teste 21.

Dados de RMN ¹³ C (CDCl ₃)						
181,4(s)	164,5(s)	161,1(s)	157,5(s)	147,7(s)	142,3(s)	138,4(s)
124,1(d)	122,4(d)	120,8(d)	118,7(s)	108,2(d)	106,7(s)	102,5(s)
98,2(d)	76,6(s)	49,0(q)	27,1(q)	27,1(q)		

Subestruturas propostas pelo SISCONST

Tabela 57-Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 21

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) ACRIDÔNICO 07	66,7	74,7	36,0	61,0
2) ACRIDÔNICO 19	5,5	8,1	12,0	7,8
3) ACRIDÔNICO 18	0,0	9,1	20,0	7,3
4) ACRIDÔNICO 01	6,5	0,0	8,0	5,3
5) ACRIDÔNICO 04	7,4	0,0	4,0	4,7

Teste 22

Boraginaceae – *Symphytum*
asperum

Phytochemistry, 1990, Pg. 11

Esqueleto – Pirrolizidínico 01

Tabela 58– Dados de entrada do Teste 22.

Dados de RMN ^{13}C (CDCl_3)						
174,1(s)	166,9(s)	139,1(d)	132,9(s)	128,6(d)	127,4(s)	82,7(s)
77,1(s)	77,1(s)	73,1(d)	69,8(d)	62,8(t)	62,8(t)	53,7(t)
34,5(t)	25,8(q)	24,8(q)	20,4(q)	18,5(q)	15,7(q)	

Subestruturas propostas pelo SISCONST

Tabela 59 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 22

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) PIRROLIZIDÍNICO 01	100,0	100,0	100,0	100,0

Teste 23

Esponja – *Cribrochalina vasculum*

J. Nat. Prod., 2000, Pg. 793

Macronós Propostos:

Angeloila-

Sinais : 168,0 - 137,9 - 127,9 - 20,6 - 15,7

Metoxila –

Sinal : 57,8

Esqueleto – Isoquinolínico 14

Tabela 60 – Dados de entrada do TESTE 23.

Dados de RMN ¹³ C (CDCl ₃)						
184,2(s)	181,4(s)	167,2(s)	158,2(s)	156,6(s)	153,7(2)	139,6(d)
138,6(s)	130,1(s)	126,9(s)	122,4(s)	118,9(d)	65,0(t)	20,5(q)
15,6(q)	8,7(q)					
Dados de RMN ¹ H (CDCl ₃)						
8,91	7,87	5,84	5,84	2,09	2,09	2,09
Dados de massa (m/z)						
315	232	216	82	55		

Subestruturas propostas pelo SISCONST

Tabela 61 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 11

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) ISOQ 14	95,4	74,6	100,0	66,7	12,6	74,1
2) ISOQ 02	0,0	0,0	0,0	0,0	43,1	7,2
3) ISOQ 03	0,0	0,0	0,0	0,0	43,1	7,2
4) ISOQ 27	2,3	0,0	0,0	16,7	0,6	3,6
5) INDOL 55	0,0	0,0	0,0	16,7	0,0	2,8

Teste 24

Dioncophyllaceae - *Triphyophyllum*
peltatum

Phytochemistry., 1998, Pg. 1151

Esqueleto – Isoquinolinico 15

Macronós Propostos:

Metoxila –

Sinal : 58,5

Tabela 62 – Dados de entrada do Teste 24.

Dados de RMN ¹³ C (CDCl ₃)						
156,2(s)	156,0(s)	150,9(s)	136,5(s)	136,3(s)	135,9(s)	130,9(d)
130,3(d)	129,7(s)	127,2(s)	123,7(d)	120,8(d)	118,0(d)	118,0(s)
113,3(s)	106,6(d)	50,9(d)	49,05(d)	37,0(t)	21,9(q)	21,9(q)
20,3(q)						
Dados de RMN ¹ H (CDCl ₃)						
7,36	7,25	7,22	7,21	6,83	6,57	4,36
2,91	2,69	2,69	2,40	2,40	2,40	1,53
1,53	1,53	1,22	1,22	1,22		
Dados de massa (m/z)						
378	377	362				

Subestruturas propostas pelo SISCONST

Tabela 63 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 24

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) ISOQ 15	67,4	11,6	100,0	0,0	33,1	55,9
2) ISOQ 41	16,7	9,4	0,0	0,0	0,0	8,6
3) INDOL 25	0,0	0,0	0,0	0,0	23,8	4,8
4) ISOQ 01	0,0	19,3	0,0	0,0	0,0	3,9
5) QUINOL26	7,4	0,0	0,0	0,0	0,0	3,0

Teste 25

Apocynaceae – *Aspidosperma ramiflorum*

Phytochemistry., 1996, Pg. 963

Substância sem macronós

Esqueleto – Indol 40

Tabela 64 – Dados de entrada do Teste 25.

Dados de RMN ¹³ C (CDCl ₃)						
137,5(s)	136,6(s)	136,5(s)	136,1(s)	134,5(s)	127,4(t)	127,4(s)
121,4(d)	121,4(d)	121,4(d)	119,2(d)	119,2(d)	117,8(d)	117,8(d)
111,3(d)	111,3(d)	108,2(s)	106,9(s)	54,0(t)	53,3(d)	51,3(t)
51,3(d)	42,1(t)	37,8(t)	33,3(t)	30,9(d)	22,4(t)	18,3(t)
13,0(q)						
Dados de RMN ¹ H (CDCl ₃)						
8,40	8,40	8,30	7,39	7,12	7,10	6,95
6,90	6,80	6,70	5,55	4,21	3,76	3,66
3,30	3,20	3,15	3,10	3,08	3,05	3,00
2,92	2,50	2,50	2,25	2,07	1,67	1,66
1,66	1,66	1,46				
Dados de massa (m/z)						
436	252	185	171			

Subestruturas propostas pelo SISCONST

Tabela 65 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 25.

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL 40	50,0	22,1	29,2	33,3	85,2	45,0
2) INDOL 41	50,0	0,0	12,5	0,0	0,0	18,8
3) INDOL 24	0,0	39,8	8,3	0,0	0,0	8,0
4) INDOL 23	0,0	0,0	4,2	33,3	0,0	6,3
5) INDOL 63	0,0	0,0	33,3	0,0	3,1	6,1

Teste 26

Derivado sintético

Heterocycles., 1996, Pg. 2377

Substância sem macronós
(metoxila utilizado como cadeia)

Esqueleto – Indol 03

Tabela 66 – Dados de entrada do Teste 26.

Dados de RMN ¹³ C (CDCl ₃)						
155,7(s)	135,0(s)	130,4(s)	127,6(s)	120,7(d)	119,4(d)	118,2(d)
108,9(d)	104,2(s)	62,1(q)	54,4(d)	51,4(t)	43,7(t)	41,9(s)
39,0(t)	29,7(t)	23,6(t)	20,7(t)	16,9(t)	9,0(q)	
Dados de RMN ¹ H (CDCl ₃)						
7,51	7,30	7,18	7,12	5,05	4,74	4,20
3,38	3,18	3,02	2,62	2,51	2,50	2,24
2,10	1,88	1,52	1,44	1,27	1,06	1,06
1,06						
Dados de massa (m/z)						
323						

Subestruturas propostas pelo SISCONST

Tabela 67 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 26

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL 03	43,5	54,6	0,0	0,0	33,5	43,7
2) INDOL 25	15,2	8,9	0,0	0,0	29,6	17,2
3) INDOL 21	0,0	18,6	0,0	0,0	4,7	5,8
4) INDOL23	4,2	0,0	0,0	0,0	7,8	4,0
5) INDOL 24	7,4	0,0	0,0	0,0	0,0	3,7

Teste 27

Fungo - *Rhizopus arrhizus*

J. Nat. Prod., 1999, Pg. 988

Substância sem macronós

Esqueleto – Aporfínico 07

Tabela 68 – Dados de entrada do Teste 27.

Dados de RMN ¹³ C (CDCl ₃)						
181,5(1)	150,6(1)	148,1(2)	147,4(1)	147,1(2)	138,3(1)	135,0(1)
134,3(2)	131,9(1)	131,1(2)	128,1(2)	125,1(2)	123,2(2)	119,2(1)
118,9(2)						

Subestruturas propostas pelo SISCONST

Tabela 69 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 27

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) APORFINICO 07	27,1	37,6	0,0	30,6
2) APORFINICO 01	10,4	8,6	0,0	9,8
3) ISOQ 01	0,0	27,0	0,0	9,0
4) ACRIDÔNICO 17	11,4	0,0	0,0	7,6
5) INDOL 66	5,9	0,0	0,0	4,0

Teste 28

Derivado Sintético

Heterocycles, 1999, Pg. 1827

Macronós Propostos :

Metoxila : 55,6

Metoxila: 60,5

Esqueleto –Indol 11

Tabela 70 – Dados de entrada do Teste 28.

Dados de RMN ¹³ C (CDCl ₃)						
169,1(s)	168,6(s)	151,2(s)	138,3(d)	137,0(s)	135,9(s)	129,3(s)
124,0(d)	122,7(d)	118,0(d)	115,6(d)	115,3(s)	96,1(d)	58,2(d)
52,4(q)	52,3(q)	50,7(d)	47,8(t)	32,3(t)	31,7(d)	22,3(t)
Dados de RMN ¹ H (CDCl ₃)						
7,93	7,41	7,29	7,20	6,17	4,57	4,29
3,20	2,83					
Dados de massa (m/z)						
454	398	284	267	214	170	169

Subestruturas propostas pelo SISCONST

Tabela 71 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 28

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	MASMACH(%)	H1MACH(%)	GLOBAL(%)
1) INDOL 11	24,1	37,4	0,0	22,2	35,0	28,6
2) INDOL 05	0,0	0,0	0,0	50,0	50,0	11,0
3) INDOL03	6,6	0,0	0,0	30,0	0,0	8,7
4) INDOL06	14,9	0,0	0,0	5,6	0,0	7,1
5) INDOL14	2,4	11,0	0,0	5,6	6,3	5,5

Teste 29

Derivado Sintético

Banco de dados SADLER, 1999.

Biblioteca – Unicamp

Esqueleto – Isoquinolínico 32

Tabela 72 – Dados de entrada do Teste 29.

Dados de RMN ¹³ C (CDCl ₃)						
163,8(1)	163,4(1)	158,0(1)	132,9(1)	132,2(2)	128,2(2)	127,1(2)
123,7(1)	121,9(1)	121,9(2)	121,9(1)	112,5(2)	56,9(3)	38,1(3)
Dados de RMN ¹ H (CDCl ₃)						
7,88	7,87	7,71	7,56	7,24	3,46	3,46
2,95	2,95					

Subestruturas propostas pelo SISCONST

Tabela 73 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 29.

Esqueleto	SISCONST(%)	C13MACH(%)	H1MACH(%)	GLOBAL(%)
1) ISOQ 32	44,8	22,8	29,5	35,5
2) ISOQ 31	23,4	0,0	7,4	13,5
3) ISOQ 01	0,0	48,0	0,0	12,0
4) ISOQ 22	20,1	0,0	0,0	10,1
5) INDOL03	0,0	0,0	26,2	6,6

Teste 30

Rutaceae – *Severinia buxifolia*

Chem. Pharm. Bull., 2000, Pg. 85

Esqueleto – Acridônico 18

Tabela 74 – Dados de entrada do Teste 30.

Dados de RMN ¹³ C (CDCl ₃)						
184,2(s)	165,6(s)	160,8(s)	149,5(s)	149,2(s)	138,2(s)	132,1(s)
131,4(s)	125,8(s)	124,4(d)	124,0(d)	123,7(d)	120,5(d)	117,2(d)
116,1(s)	115,2(s)	105,3(s)	47,9(q)	27,1(t)	25,8(q)	25,7(q)
23,2(t)	18,1(q)	17,9(q)				

Subestruturas propostas pelo SISCONST

Tabela 75 - Tabela de resultados do programa Pesquisa, para o Alcalóide do Teste 30.

Esqueleto	SISCONST(%)	C13MACH(%)	SISOCBOT(%)	GLOBAL(%)
1) ACRIDÔNICO 18	58,8	22,1	0,0	34,9
2) ACRIDÔNICO 19	30,6	21,3	44,4	31,7
3) ACRIDÔNICO 22	10,0	0,0	11,1	7,8
4) ACRIDÔNICO 17	0,0	0,0	22,2	5,6
5) ACRIDÔNICO 07	0,0	19,2	0,0	4,8

A segunda subestrutura foi fornecida pelo programa SISCONST, após serem descartados os sinais dos deslocamentos químicos atribuídos aos átomos da primeira subestrutura proposta. Considerando-se o número de átomos existentes na substância teste, observa-se que, através da união das duas subestruturas propostas pelo programa SISCONST, foi obtido um grande fragmento com um total de 23 átomos de carbonos e que corresponde a 95,8% da substância submetida ao teste.

Analisando os resultados fornecidos pelos programas aplicativos do Sistema Especialista SISTEMAT, pode-se observar sua eficiência na tarefa de determinação estrutural de substâncias orgânicas, sejam elas de origem natural ou sintética.

Foram realizados testes com 30 substâncias inéditas ao banco de dados, pertencentes às subclasses dos alcalóides estudados neste trabalho. Destas 30 substâncias submetidas aos testes, 3 pertenciam à esqueletos inéditos (sem representação no banco de dados).

O programa SISCONST obteve um índice de acerto de 96%. Nos casos de substâncias com esqueletos inéditos, o programa apresentou como proposta um esqueleto (em grande parte) semelhante ao pesquisado, como ocorreram nos casos dos testes 09, 16 e 18 (pgs. 108, 115 e 117). Na ausência de previsão por falta de parâmetros de comparação, o programa forneceu subestruturas que possibilitaram a previsão do esqueleto por meio de sobreposição ou união das mesmas, conduzindo a uma subestrutura maior ou até mesmo ao esqueleto completo da substância, como ocorreram nos casos dos testes 09 e 18. Em todas as propostas estruturais fornecidas pelo SISCONST, o esqueleto em questão, quando não foi apresentado como melhor proposta, foi sugerido entre os cinco primeiros esqueletos com maior probabilidade de ocorrência, como aconteceu no Teste 05 (pg.104), onde a substância submetida ao teste pertencia ao esqueleto Indol 36 e foi proposto, com maior probabilidade, o esqueleto Indol 43, em função da grande semelhança entre os dois (Fig.7 – pgs. 30 e 31) e a maior representatividade do esqueleto Indol 43, que aparece no banco de dados representado por 10 substâncias, com relação ao esqueleto Indol 36, que é representado por apenas 2 substâncias no banco de dados (Tab. 6 – pg.71).

O programa C13MACH apresentou um índice de acerto de 74%, inferior ao do programa SISCONST. Este resultado já era esperado, visto que, o funcionamento do programa C13MACH baseia-se apenas na confrontação entre os deslocamentos químicos da substância teste com os deslocamentos das substâncias presentes no banco de dados. O C13MACH não considera o ambiente químico dos átomos de carbonos confrontados; isso significa que, se for realizado um teste com uma substância pertencente ao esqueleto **A**, este esqueleto possuir 2 representantes no banco e neste mesmo banco existir um

esqueleto **B**, com aproximadamente 50% dos deslocamentos químicos semelhantes aos do esqueleto **A** e tiver 10 representantes no banco, o programa fornecerá como principal proposta, o esqueleto **B**. Este fato ocorreu nos Testes 03, 08 e 25 (pgs.101, 107 e 124) onde todas as propostas foram sugeridas em função da maior representação do esqueleto no banco de dados, causando um equívoco nos resultados que forneceram propostas de esqueletos erradas.

O programa H1MACH e o programa SISOCBOT apresentaram um percentual de acerto de 67% e 69% respectivamente, pelos mesmos motivos justificados para o programa C13MACH.

O programa MASMACH teve um acerto de 62,5%, sendo que, em certas ocasiões, como no Teste 01(pg.98), o descarte de um dos picos, que representava um fragmento comum a outros esqueletos, possibilitou o fornecimento de uma proposta de esqueleto correta e com alto grau de confiabilidade. Entretanto, na maioria dos casos, este artifício não pode ser utilizado, visto que, nos alcalóides existem inúmeros fragmentos comuns a vários esqueletos (alguns exemplos são dados nas Figs. 42 e 43 do Teste 01 – pg.99).

Estes resultados, embora satisfatórios, tendem a melhorar com o crescimento do banco de dados, que atualmente possui uma quantidade muito variada de esqueletos proporcionalmente a quantidade de substâncias que os representam. A idéia de construir um banco de dados de alcalóides, por família de plantas ao invés de tipos de esqueletos, possibilitaria uma análise mais eficiente do programa e, provavelmente, aumentaria a quantidade de substâncias por esqueleto e conseqüentemente o índice de acerto para propostas de esqueleto.

Comparando-se os resultados dos testes realizados com outras classes, estudadas anteriormente^{37,38}, os resultados obtidos dos testes realizados com a classe dos alcalóides apresentaram um índice de acerto menor, isso pode ser atribuído a dois fatores principais:

- Uma variedade de esqueletos muito grande, que consiste em 282 esqueletos diferentes para representar apenas 1183 substâncias, o que daria o equivalente a 4 substâncias por esqueleto, isso se as

mesmas fossem divididas igualmente, o que não ocorre. Neste banco de dados tem esqueleto representado por 94 substâncias diferentes, como ocorre no caso dos alcalóides pertencentes ao esqueleto Pirrolizidínico 01, o que explica o excelente resultado obtido no Teste 22 (pg.121). Por outro lado, existem inúmeros esqueletos que são representados por somente uma ou duas substâncias, como ocorreu no caso do Teste 05, comentado anteriormente.

- Outro fator que influencia nos resultados, é que muitos esqueletos possuem uma grande semelhança entre suas estruturas, se diferenciando apenas pela adição de uma metila (por exemplo os esqueletos Isoquinolínicos 40 e 41, mostrados na Fig.44 como justificativa para o resultado fornecido pelo programa C13MACH no Teste 03) ou uma ligação intramolecular entre os átomos de carbonos dos esqueletos (como ocorre com os esqueletos de alcalóides Benzilisoquinolínico 01 e Aporfínico 05).

Em todos os casos o Sistema Especialista SISTEMAT identificou corretamente a subclasse do alcalóide testado e, de um modo geral, os programas aplicativos apresentaram resultados satisfatórios que podem ser utilizados como uma importante ferramenta na elucidação estrutural de alcalóides.

Limitações do programa –

Em nenhuma ocasião, o programa de criação do banco de dados DATASIS, aceitou nitrogênios com cargas positivas (alcalóides sob a forma de sais quaternários), o que impediu a entrada de uma grande quantidade de substâncias no banco de dados.

Muitos alcalóides descritos na literatura possuem um número maior que 30 ou até mesmo 40 átomos de carbono, o que torna impossível sua entrada no banco de dados em função de extrapolar o limite de compactação do vetor que é de 99 dados.

Em alguns casos, de alcalóides com estruturas mais complexas, as ligações longas também não foram aceitas na codificação da substância.

Os macronós que possuíam um carbono quaternário, formado por ligações simples, ligado ao átomo de oxigênio (por exemplo terc-butoxila), não foram codificados.

Em função destas limitações foi estudado um método alternativo para a elucidação estrutural dos alcalóides, método este que utiliza as Redes Neurais Artificiais, como uma ferramenta na elucidação estrutural. Essa metodologia foi aplicada na determinação de esqueletos dos alcalóides aporfínicos e será discutida no próximo capítulo.

V - REDES NEURAIS ARTIFICIAIS (RNAs)

V.1 Introdução

A solução de problemas através das RNAs é bastante atrativa, pois o paralelismo constitui-se na característica principal das RNAs, onde esta cria a possibilidade de um desempenho superior em relação a solução de problemas baseados nos modelos convencionais. Nas RNAs, o procedimento usual na solução de problemas passa inicialmente por uma fase de aprendizagem, onde um conjunto de exemplos é apresentado para a rede, que extrai automaticamente as características necessárias para representar a informação fornecida. Essas características são utilizadas posteriormente para gerar respostas para o problema.

A capacidade de aprender através de exemplos e de generalizar a informação aprendida são os atrativos principais da solução de problemas através das RNAs. A generalização está associada à capacidade da rede em aprender através de um conjunto reduzido de exemplos e, posteriormente, dar respostas coerentes a dados não apresentados à rede. Isto é uma demonstração que a capacidade das RNAs vai, além de, simplesmente mapear relações de entrada e saída. As RNAs são capazes de extrair informações, não apresentadas de forma explícita, através de exemplos. Outras características importantes são a capacidade de auto-organização e processamento temporal que, aliadas àquelas citadas anteriormente, fazem das RNAs uma ferramenta computacional extremamente poderosa e atrativa para a solução de problemas complexos.

Embora a programação e a matemática envolvidas na tecnologia de redes neurais sejam complexas, a utilização dos programas de redes neurais pode ser bastante simples e os resultados são freqüentemente extraordinários.

V.2 – Redes Neurais Artificiais x Sistemas Especialistas⁴⁵

As **redes neurais** e os **sistemas especialistas** foram considerados os maiores avanços tecnológicos desde o transistor.

Há tempos atrás, as redes neurais e os sistemas especialistas competiam entre si para disputar financiamentos de pesquisas, hoje cada um é

aceito individualmente com suas forças e restrições, ficando mais esclarecida a diferença entre eles e a necessidade dos dois.

Os sistemas especialistas são capazes de raciocínio dedutivo usando regras impostas pelo programador. As redes neurais criam suas próprias regras, aprendendo através de exemplos. Os sistemas especialistas apresentam bons resultados passo a passo em operações lógicas, enquanto que as redes neurais são boas no reconhecimento de padrões imediatos e previsões. Juntos eles podem resolver problemas além da extensão de uma solução computacional.

V.3 – Histórico

As simulações em redes neurais parecem ser um desenvolvimento recente. Entretanto, este campo foi estabelecido antes da invenção dos computadores, sobrevivendo a várias décadas.

A tentativa inicial de trabalhar com redes neurais foi a criação do primeiro neurônio artificial, produzido em 1943 pelo neurofisiologista Warren McCulloch e o matemático Walter Pitts⁴⁶, que inicialmente fizeram simulações usando lógica formal. Eles desenvolveram modelos de redes neurais baseados nos seus conhecimentos de neurologia.

Muitos outros artigos e livros surgiram desde então, porém, por um longo período de tempo, pouco resultado foi obtido. Em 1949 Donald Hebb⁴⁷ escreveu um livro intitulado "The Organization of Behavior" (A Organização do Comportamento) que apresenta uma idéia de que o condicionamento psicológico clássico está presente em qualquer animal pelo fato de ser uma propriedade de neurônios individuais. Suas idéias não eram novas, mas Hebb foi o primeiro a propor uma lei de aprendizagem específica para as sinapses dos neurônios. Este primeiro e corajoso passo serviu de inspiração para que muitos outros pesquisadores perseguissem a mesma idéia e embora muito tenha sido estudado e publicado nos anos que se seguiram (1940-1950), estes serviram mais como base para desenvolvimento posterior que para o próprio desenvolvimento.

Em 1956 no "Dartmouth College" nasceram os dois paradigmas da Inteligência Artificial, a simbólica e o conexionista. A Inteligência Artificial Simbólica tenta simular o comportamento inteligente humano desconsiderando

os mecanismos responsáveis por tal. Já a Inteligência Artificial Conexionista acredita que construindo-se um sistema que simule a estrutura do cérebro, este sistema apresentará inteligência, ou seja, será capaz de aprender, assimilar, errar e aprender com seus erros.

Dois grupos, Farley e Clark, 1954 em Rochester, Haibit e Duda, 1956 na Holanda utilizaram simulações de computador³. No primeiro grupo, pesquisadores da IBM, mantiveram contato fechado com neurocientistas. Então, sempre que os modelos não trabalhavam direito, eles os consultavam. Esta interação estabeleceu uma tendência de multidisciplinaridade que continua até os dias de hoje.

Os neurocientistas não eram os únicos influentes no desenvolvimento de redes neurais, os psicólogos e engenheiros também contribuíram para o progresso na área de simulações em redes neurais. Frank Rosenblatt⁴⁸ (1958) despertou um interesse considerável neste campo, quando desenvolveu o **Mark I Perceptron**. Devido a profundidade de seus estudos, suas contribuições técnicas e de sua maneira moderna de pensar, muitos o vêem como o fundador da neurocomputação na forma em que a temos hoje. Seu interesse inicial para a criação do Perceptron era o reconhecimento de padrões.

O Perceptron era formado por três camadas, com a camada mediana conhecida como camada de associação. Este sistema poderia aprender, conectar ou associar uma determinada contribuição a uma unidade de produção fortuita.

Outro sistema era o **ADALINE** (Elemento Linear Adaptável) que foi desenvolvido em 1960 por Widrow e Hoff⁴⁹ (da Universidade de Stanford). O ADALINE era um dispositivo eletrônico análogo ao Perceptron, construído a partir de componentes simples. O método usado para aprendizagem foi diferente do que era utilizado para o Perceptron; ele utilizava o Least-Mean-Squares (LMS) para aprendizagem de regras. A principal contribuição do ADALINE foi o princípio de treinamento desenvolvido, denominado *regra delta*. Widrow também fundou a primeira companhia de hardware de neurocomputadores e componentes.

Em 1969 Minsky e Papert⁵⁰ escreveram um livro no qual generalizaram as limitações do Perceptron de camada simples para os sistemas de camadas

múltiplas. Os autores disseram : "...nosso julgamento intuitivo é que a extensão (para sistemas de camadas múltiplas) é estéril". De fato, eles provaram que o *perceptron* podia representar somente funções linearmente separáveis, não conseguindo representar funções booleanas simples, limitando o seu uso.

A obra de Minsky e Papert desestimulou o apoio financeiro de empresas e agências de fomento à pesquisa e, como consequência, reduziu o desenvolvimento de novos trabalhos. Com isso, as pesquisas com redes neurais foram praticamente abandonadas até meados da década de 80.

Nos anos 80, muitos dos pesquisadores passaram a publicar diversas propostas para a exploração de desenvolvimento de redes neurais e suas aplicações. Porém talvez o fato mais importante deste período tenha ocorrido quando, um administrador de programas da DARPA⁵¹ (Defense Advanced Research Projects Agency) decidiu ouvir os argumentos da neurocomputação e seus projetistas, e divergindo dos caminhos tradicionais dos conhecimentos convencionais, fundou em 1983 grupos de pesquisas neste campo. Esta iniciativa não só abriu as portas para a neuro computação, como também deu à DARPA o status de uma das líderes mundiais em se tratando de "moda" tecnológica.

Outro especialista que emergiu neste período foi John Hopfield⁵², físico mundialmente conhecido, se interessou pela neurocomputação e escreveu artigos que persuadiram centenas de cientistas, matemáticos, e tecnólogos altamente qualificados a se unirem nesta nova área emergente. Apesar de um terço dos pesquisadores da área terem aderido à neurocomputação pela influência de Hopfield, foi em 1986 que este campo de pesquisa "explodiu" com a publicação do livro "Parallel Distributed Processing" (Processamento Distribuído Paralelo) editado por David Rumelhart e James McClelland⁵³. Neste livro foi apresentado um algoritmo de treinamento de redes neurais de camadas múltiplas, partindo da *regra delta*, com um modo eficiente de alteração dos pesos da rede com camadas intermediárias entre a camada de entrada e saída, baseado no erro conhecido na camada de saída. Esse algoritmo foi chamado de *regra delta generalizada* ou *algoritmo de retro-propagação*⁵⁴ (backpropagation).

Em 1987 ocorreu em São Francisco a primeira conferência de redes neurais em tempos modernos, a IEEE International Conference on Neural

Networks, e também foi formada a International Neural Networks Society (INNS). A partir destes acontecimentos decorreram a fundação do INNS journal em 1989, seguido do Neural Computation e do IEEE Transactions on Neural Networks em 1990.

Desde 1987, muitas universidades anunciaram a formação de institutos de pesquisa e programas de educação em neurocomputação. De 1990 até 2005 aproximadamente 28.541 artigos foram publicados nesta área*.

V.4 – Definições

Redes Neurais Artificiais (RNA) são sistemas inspirados nos neurônios biológicos que tentam simular seu processo de aprendizagem. Elas possuem a capacidade de adquirir, armazenar e utilizar conhecimento experimental. São sistemas paralelos distribuídos (conexionismo) compostos por unidades de processamentos simples (nodos ou neurônios) que calculam funções matemáticas (normalmente não lineares). Estas unidades estão dispostas em uma ou mais camadas, e interligadas por um grande número de conexões (Fig.47). Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede.

Figura 47 – Representação de uma Rede com suas respectivas camadas e conexões

O trabalho em redes neurais artificiais tem sido motivado pelo reconhecimento de que o cérebro humano processa informações de uma forma inteiramente diferente do computador digital convencional. O cérebro é um *computador* (sistema de processamento de informação) altamente *complexo*,

* Informação obtida da base de dados Web Of Science.

não-linear e paralelo. Ele tem a capacidade de organizar seus constituintes estruturais, conhecidos por *neurônios*, de forma a realizar certos processamentos (por exemplo, reconhecimento de padrões, percepção e controle motor) a uma velocidade muito superior a velocidade do mais rápido computador digital hoje existente. Considere, por exemplo, a *visão humana*, que é uma tarefa de processamento de informação⁵⁵⁻⁵⁷. A função do sistema visual é fornecer uma *representação* do ambiente à nossa volta e, mais importante que isso, fornecer a informação de que precisamos para *interagir* com o ambiente. Para sermos específicos, o cérebro realiza rotineiramente tarefas de reconhecimento perceptivo (p. ex., reconhecendo um rosto familiar inserido em uma cena não-familiar) em aproximadamente 100-200 ms, ao passo que tarefas de complexidade muito menor podem levar dias para serem executadas em um computador convencional.

Como outro exemplo, considere o *sonar* de um morcego. O sonar é um sistema ativo de localização por eco. Além de fornecer informações sobre a distância até um alvo (p. ex., um inseto voador), o sonar de um morcego transmite também informação sobre a velocidade relativa do alvo, o tamanho do alvo, o tamanho de várias características do alvo e a elevação do mesmo^{58,59}. A complexa computação neural necessária para extrair toda essa informação ocorre no interior de um cérebro do tamanho de uma ameixa. De fato, um morcego guiado por eco pode perseguir e capturar seu alvo com uma facilidade e taxa de sucesso que são de causar inveja a um engenheiro de radar ou sonar.

Como, então, um cérebro humano ou o cérebro de um morcego faz isso?

O cérebro tem uma grande estrutura e a habilidade de desenvolver suas próprias regras através do que usualmente denominamos "experiência". Na verdade, a experiência vai sendo acumulada com o tempo, sendo que o mais dramático desenvolvimento (i. e., por ligações físicas) do cérebro humano acontece durante os dois primeiros anos de vida; mas o desenvolvimento continua para muito além desse estágio.

Um neurônio em "desenvolvimento" é sinônimo de um cérebro plástico: a *plasticidade* permite que o sistema nervoso em desenvolvimento se adapte ao seu meio ambiente. Assim como a plasticidade parece ser essencial para o

funcionamento dos neurônios como unidades de processamento de informação do cérebro humano, também ela o é com relação às redes neurais construídas com neurônios artificiais. Na sua forma mais geral, uma *rede neural* é uma máquina que é projetada para *modelar* a maneira como o cérebro realiza uma tarefa particular ou função de interesse; a rede é normalmente implementada utilizando-se componentes eletrônicos ou é simulada por programação em um computador digital.

Para alcançarem bom desempenho, as redes neurais empregam uma interligação maciça de células computacionais simples denominadas "neurônios" ou "unidades de processamento". Pode-se então sugerir a seguinte definição de uma rede neural vista como uma máquina adaptativa⁴⁵:

"Uma rede neural é um processador maciçamente e paralelamente distribuído, constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso". Ela se assemelha ao cérebro em dois aspectos:

- *O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.*
- *Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.*

O procedimento utilizado para realizar o processo de aprendizagem é chamado de *algoritmo de aprendizagem*, cuja função é modificar os pesos sinápticos da rede de uma forma ordenada para alcançar um objetivo de projeto desejado.

A modificação dos pesos sinápticos é o método tradicional para o projeto de redes neurais. Entretanto, é possível também para uma rede neural modificar sua própria topologia, o que é motivado pelo fato de os neurônios no cérebro humano poderem morrer e que novas conexões sinápticas possam crescer.

As redes neurais são também referidas na literatura como *neurocomputadores*, *redes connexionistas*, *processadores paralelamente distribuídos*, etc.

V.5 – Benefícios das Redes Neurais⁴⁵

É evidente que uma rede neural extrai seu poder computacional através, primeiro, de sua estrutura maciçamente paralelamente distribuída e segundo de sua habilidade de aprender e portanto de generalizar. A *generalização* se refere ao fato de a rede neural produzir saídas adequadas para entradas que não estavam presentes durante o treinamento (aprendizagem). Estas duas capacidades de processamento de informação tornam possível a resolução de problemas complexos (de grande escala) que são atualmente intratáveis. Entretanto, é importante reconhecer que existe um longo caminho a percorrer antes de ser construído (se porventura for possível) uma arquitetura computacional que se assemelhe ao cérebro humano.

As redes neurais possuem importantes propriedades que as tornam úteis e capacitadas no desenvolvimento de suas tarefas, são elas:

- *Não-linearidade.* É uma propriedade muito importante, particularmente se o mecanismo físico responsável pela geração do sinal de entrada (p. ex., sinal de voz) for inerentemente não-linear.
- *Mapeamento de Entrada-Saída.* Um paradigma popular de aprendizagem chamado *aprendizagem com um professor ou aprendizagem supervisionada* envolve a modificação dos pesos sinápticos de uma rede neural pela aplicação de um conjunto de *amostras de treinamento* rotuladas ou *exemplos da tarefa*. Cada exemplo consiste de um *sinal de entrada* único e de uma *resposta desejada* correspondente. Apresenta-se para a rede um exemplo escolhido ao acaso do conjunto e os pesos sinápticos (parâmetros livres) da rede são modificados para minimizar a diferença entre a resposta desejada e a resposta real da rede, produzida pelo sinal de entrada, de acordo com um critério estatístico apropriado. O treinamento da rede é repetido para muitos exemplos do conjunto até que a rede alcance um estado estável onde não haja mais modificações significativas nos pesos sinápticos. Os exemplos de treinamento previamente aplicados podem ser reaplicados durante a sessão de treinamento, mas em uma ordem diferente. Assim, a rede aprende dos

exemplos ao construir um *mapeamento de entrada-saída* para o problema considerado.

- *Adaptabilidade.* As redes neurais têm uma grande capacidade de *adaptar* seus pesos sinápticos à modificações do meio ambiente. Em particular, uma rede neural treinada para operar em um ambiente específico pode ser facilmente *retreinada* para lidar com pequenas modificações nas condições operativas do ambiente. Além disso, trabalhando em um ambiente *não-estacionário* (i. e., onde as estatísticas mudam com o tempo), uma rede neural pode ser projetada para modificar seus pesos sinápticos em tempo real. A arquitetura natural de uma rede neural para classificação de padrões, processamento de sinais e aplicações de controle, aliada à capacidade de adaptação da rede, a torna uma ferramenta muito útil para classificação adaptativa de padrões, processamento adaptativo de sinais e controle adaptativo.
- *Resposta a Evidências.* No contexto de classificação de padrões, uma rede neural pode ser projetada para fornecer informação não somente sobre qual padrão particular *selecionar*, mas também sobre a confiança ou *crença* na decisão tomada. Esta última informação pode ser utilizada para rejeitar padrões ambíguos, caso eles estejam presentes, e com isso melhorar o desempenho de classificação da rede.
- *Informação Contextual.* O conhecimento é representado pela própria estrutura e estado de ativação de uma rede neural. Cada neurônio da rede é potencialmente afetado pela atividade de todos os outros neurônios na rede. Consequentemente, a informação contextual é tratada naturalmente pela rede neural.
- *Tolerância a Falhas.* Uma rede neural, implementada na forma física (em *hardware*), tem o potencial de ser inerentemente *tolerante a falhas*, ou capaz de realizar computação robusta, no sentido de que seu desempenho se degrada suavemente sob condições de operação adversas. Se um neurônio ou suas conexões são danificados, por exemplo, a recuperação de um padrão armazenado é prejudicada em qualidade. Contudo, devido à natureza distribuída da informação armazenada na rede, o dano deve ser extenso para que a resposta global da rede seja degradada seriamente.

Assim, a princípio, uma rede neural exhibe uma degradação suave do desempenho em vez de apresentar uma falha catastrófica.

- *Implementação em VLSI.* Uma virtude benéfica particular da tecnologia VLSI (*very-large-scale-integration*) é que ela fornece um meio de capturar comportamentos realmente complexos de uma forma altamente hierárquica.
- *Uniformidade de Análise e Projeto.* Basicamente, as redes neurais desfrutam de universalidade como processadores de informação. Dizemos isso no sentido de que a mesma notação é utilizada em todos os domínios envolvendo a aplicação de redes neurais. Esta característica se manifesta de diferentes modos:
 - Os neurônios, de uma forma ou de outra, representam um ingrediente *comum* a todas as redes neurais.
 - Esta uniformidade torna possível *compartilhar* teorias e algoritmos de aprendizagem em aplicações diferentes de redes neurais.
- *Analogia Neurobiológica.* O projeto de uma rede neural é motivado pela analogia com o modelo básico de uma célula nervosa (Fig.48), também chamada de neurônio. Este modelo de neurônio é formado por *dendritos*, ou seja, conexões pelas quais os sinais de outras células chegam até o neurônio. O corpo da célula é chamado de soma e o sinal de saída do neurônio é transmitido por uma ramificação chamada axônio. A região entre o axônio de um neurônio e o *dendrito* de outro é chamado de *sinapse*.

Figura 48 - Modelo de Transmissão de Impulso nervoso em Neurônios Biológicos

O modelo de funcionamento dos neurônios é relativamente simples: os impulsos que chegam ao neurônio através de seus *dendritos* são “somados” e, caso a soma seja maior que um determinado limiar, o neurônio é ativado e dispara um impulso, através do *axônio*, para outras células.

V.6 - Neurônios artificiais -

O neurônio artificial (Fig.49) – ou unidade de processamento – foi inspirado no neurônio biológico (Fig.48). As *sinapses* são representadas por pesos w_{ij} , os quais são responsáveis pela soma ponderada das entradas x_i . Os pesos sinápticos são números, positivos ou negativos, do acoplamento sináptico entre o neurônio que envia o sinal e o outro que recebe. Quando o peso w de uma sinapse é positivo, ela é dita excitatória; e quando ele é negativo, ela é dita inibitória.

O núcleo da célula é representado por uma função, chamada de *função de ativação*, usada para restringir a amplitude da saída de um neurônio. A função de ativação também é referida como *função restritiva* já que restringe (limita) o intervalo permissível de amplitude do sinal de saída a um valor finito. O bias é aplicado externamente e tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo, respectivamente. O axônio é representado pela saída y_j do neurônio.

Figura 49 - Modelo Básico do Neurônio Artificial.

A estrutura do neurônio artificial, descrita acima, foi inicialmente apresentada por McCulloch e Pitts⁴⁶, que utilizaram a função de ativação

degrau (Fig.50) para representar o núcleo de um neurônio. O modelo foi expresso por:

$$y_j = f(u_j) = f\left(\sum_{i=1}^p w_{ij}x_i - \theta_j\right) \quad (5.1)$$

em que $f(u_j)$ é a função de ativação do neurônio, w_{ij} são os pesos sinápticos, x_i ($i=1,2,\dots,p$) são as p entradas, θ_j é o limiar de ativação do neurônio (viés ou bias) e y_i representa a saída do neurônio.

Figura 50 – Função de ativação utilizada por MacCulloch e Pitts

A primeira etapa do processamento do neurônio artificial é calcular o *nível de ativação* u , que é a combinação linear usada para representar o potencial resultante na soma de um neurônio biológico. A esta combinação é somado um número θ .

Considerando uma entrada x_0 , igual a -1 e o correspondente w_{0j} igual ao limiar θ_j , a equação 5.1 pode ser escrita como:

$$y_j = f\left(\sum_{i=0}^p w_{ij}x_i\right) \quad (5.2)$$

em que $w_{0j} = \theta_j$, $x_0 = -1$.

Desse modo, considerando a função degrau, a saída do neurônio é calculada por:

$$\begin{aligned} y_j &= 0 & \text{se} & \sum_{i=1}^p w_{ij}x_i < \theta_j \\ y_j &= 1 & \text{se} & \sum_{i=1}^p w_{ij}x_i \geq \theta_j \end{aligned} \quad (5.3)$$

Esse modelo de neurônio artificial é a base de praticamente todas as redes neurais desenvolvidas a partir de então. As principais inovações no modelo de neurônios vieram com a utilização de novos tipos de função de ativação.

V.7 – Tipos de Função de ativação⁶⁰.

A partir do modelo proposto por McCulloch e Pitts foram derivados vários outros modelos que permitem a produção de uma saída qualquer, não necessariamente zero ou um, e com diferentes funções de ativação. A Figura 51 ilustra, graficamente, quatro funções de ativação diferentes: a função linear, a função rampa, a função degrau (passo) e a função sigmoideal.

A função de ativação linear, mostrada na Figura 51a, é definida pela equação.

$$y = \alpha x \quad (5.4)$$

onde α é um número real que define a saída linear para os valores de entrada, y é a saída e x é a entrada.

A função linear pode ser restringida para produzir valores constantes em uma faixa $[-\gamma, +\gamma]$, e neste caso, passa a ser a função rampa (função linear por partes) mostrada, graficamente, na Figura 51b definida pela notação abaixo.

$$y = \begin{cases} +\gamma & \text{se } x \geq \gamma \\ x & \text{se } x = \gamma \\ -\gamma & \text{se } x \leq -\gamma \end{cases} \quad (5.5)$$

Os valores máximo e mínimo da saída são $-\gamma, +\gamma$. A função rampa é usada, geralmente, como uma função linear simplificada.

A função passo (função de limiar), ilustrada na Figura 51c, é similar a uma função rampa, pois a função produz a saída $+\gamma$ para valores de x maiores que zero, caso contrário a função produz o valor de $-\gamma$. A função passo é definida por.

$$y = \begin{cases} +\gamma & \text{se } x > 0 \\ -\gamma & \text{se } x \leq 0 \end{cases} \quad (5.6)$$

A função sigmoidal, conhecida também como *S-shape*, ilustrada na Figura 51d, é uma função semi-linear. É possível definir várias funções sigmoidais. Uma das funções sigmoidais mais importantes é a função logística definida por.

$$y = \frac{1}{1 + e^{-x/T}} \quad (5.7)$$

onde parâmetro T determina a suavidade de curva.

Figura 51 - Funções de ativação: (a) Linear; (b) Rampa; (c) Passo e (d) Sigmoidal

V.8 – Aprendizado da rede

As RNAs possuem a capacidade de aprender através de exemplos e fazer interpolações e extrapolações do que aprenderam. No aprendizado conexionista não se procura obter regras como na abordagem simbólica da Inteligência Artificial (IA), mas determinar a intensidade das conexões entre os neurônios. Um conjunto de procedimentos bem definidos para adaptar os parâmetros de uma RNA, afim de que a mesma possa *aprender* uma determinada função é chamado de *algoritmo de aprendizado*. Como era de se esperar, não há um único algoritmo de aprendizado, mas sim um conjunto de ferramentas representadas por diversos algoritmos, cada um com suas vantagens e desvantagens. Estes algoritmos, basicamente, diferem pela maneira através da qual o ajuste dos pesos é feito.

A utilização de uma RNA na solução de uma tarefa passa, inicialmente, por uma fase de aprendizagem, onde a rede extrai características relevantes de padrões de informações apresentados para a mesma, criando assim uma representação própria para o problema. A etapa de aprendizagem consiste em um processo iterativo de ajuste de parâmetros da rede, os pesos das conexões entre as unidades de processamento, que guardam o conhecimento que a rede adquiriu do ambiente em que está operando. Uma definição geral do que vem a ser aprendizagem pode ser expressa da seguinte forma⁶¹.

“Aprendizagem é o processo pelo qual os parâmetros de uma rede neural são ajustados através de uma forma continuada de estímulos fornecidos pelo ambiente no qual a rede está operando, sendo que o tipo específico de aprendizagem realizado, é definido pela maneira particular como ocorrem os ajustes realizados nos parâmetros”.

Diversos métodos para treinamento de redes foram desenvolvidos e podem ser agrupados em dois paradigmas principais: Aprendizado Supervisionado e Aprendizado Não Supervisionado. Outros dois paradigmas bastante conhecidos são os de Aprendizado por Reforço (que é um caso particular de Aprendizado Supervisionado) e Aprendizado por Competição (que é um caso particular de Aprendizado Não Supervisionado).

V.8.1 – Aprendizado Supervisionado

Este método de aprendizado é o mais comum no treinamento das RNAs, sendo chamado *Aprendizado Supervisionado* porque a saída desejada que a rede deve resultar é fornecida por um supervisor (professor) externo (Fig.52). O objetivo é ajustar os parâmetros da rede, de forma a encontrar uma ligação entre os pares de entrada e saída fornecidos. O supervisor indica, explicitamente, um comportamento bom ou ruim para a rede, visando direcionar o processo de treinamento. A rede tem sua saída corrente (calculada) comparada com a saída desejada, recebendo informações do supervisor sobre o erro da resposta atual. Os pesos das conexões são ajustados para minimizar o erro. A minimização da diferença é feita com

incrementos”, já que pequenos ajustes são feitos nos pesos à cada etapa de treinamento, de tal forma que estes caminhem, se possível, para uma solução.

Figura 52 - Algoritmo de aprendizado supervisionado.

O Aprendizado Supervisionado pode ser implementado, basicamente, de duas formas: *off-line* e *on-line*.

Para treinamento *off-line*, uma única correção é feita por ciclo. Todos os exemplos do conjunto de treinamento são apresentados à rede, seu erro médio é calculado e a partir deste erro fazem-se as correções dos pesos, os dados do conjunto de treinamento não mudam, sendo que uma vez obtida uma solução para a rede, esta deve permanecer fixa.

No aprendizado *on-line* a correção de pesos ocorre toda vez que um conjunto de treinamento é apresentado à rede. Cada correção dos pesos baseia-se somente no erro do exemplo presente naquela iteração. Assim, em cada ciclo ocorrem N correções e o conjunto de dados muda, continuamente, visto que a rede deve estar em contínuo processo de adaptação.

A desvantagem do Aprendizado Supervisionado é que, na ausência do professor, a rede não conseguirá aprender novas estratégias para situações não cobertas pelos exemplos do treinamento da rede. Os exemplos mais conhecidos de algoritmos de Aprendizado Supervisionado são a regra delta ou correção de erros⁴⁹ e a sua generalização para redes de camadas múltiplas, o algoritmo de *back-propagation*⁵⁴ (retro-propagação).

V.8.1.1 – Correção de Erros

O processo de aprendizado por correção de erros procura minimizar a diferença entre a soma ponderada das entradas pelos pesos (saída calculada pela rede) e a saída desejada, ou seja, o erro da resposta atual da rede

(Fig.53). O termo $e(t)$ denota o erro e deve ser calculado através da seguinte expressão: $e(t) = d(t) - y(t)$, onde $d(t)$ é a saída desejada e $y(t)$ é a resposta atual (calculada) no instante de tempo t . A forma genérica para alteração dos pesos por correção de erros é definida por:

$$w_i(t+1) = w_i(t) + \eta e(t) x_i(t) \quad (5.8)$$

onde η é a taxa de aprendizado e $x_i(t)$ é a entrada para o neurônio i no tempo t .

Segundo a equação (5.8), o ajuste dos pesos deve ser proporcional ao produto do erro pelo valor de entrada da sinapse naquele instante de tempo. Esta expressão aparece tanto no algoritmo de treinamento do perceptron⁵³ quanto no algoritmo para treinamento do ADALINE e posterior generalização para o algoritmo *back-propagation*⁵⁴.

Figura 53 – Exemplo do Processo de aprendizado por correção de erros.

V.8.2 – Aprendizado Não Supervisionado

No Aprendizado Não Supervisionado, como o próprio nome sugere, não há um professor ou supervisor para acompanhar o processo de aprendizado

(Fig.54). Para este tipo de aprendizado, somente os padrões de entrada estão disponíveis para a rede, ao contrário do Aprendizado Supervisionado, cujo conjunto de treinamento possui pares de entrada e saída. A partir do momento em que a rede estabelece uma harmonia com as regularidades estatísticas da entrada de dados, desenvolve-se nela uma habilidade de formar representações internas para codificar características da entrada e criar novas classes ou grupos automaticamente. Este tipo de aprendizado, só se torna possível, quando existe redundância nos dados de entrada. Sem redundância seria impossível encontrar quaisquer padrões ou características dos dados de entrada.

Figura 54 – Algoritmo de Aprendizado Não Supervisionado

Uma das vantagens do Aprendizado Não Supervisionado pode ser visualizada na ausência do supervisor ou professor (responsável pela indicação da saída desejada). A desvantagem está intimamente ligada ao conjunto de dados de entrada, que deve ser redundante para que a rede consiga abstrair características em seu treinamento.

V.8.2.1 – Aprendizado Hebbiano

Os resultados obtidos por Hebb⁴⁷ motivaram os primeiros métodos de aprendizado em RNAs. Como originalmente postulado, a regra de Hebb propõe que o peso de uma conexão sináptica deva ser ajustado, caso exista sincronismo entre os “níveis de atividade” das entrada e saídas. Caso dois neurônios, em lados distintos da sinapse, sejam ativados sincronamente, tem-se um fortalecimento desta sinapse. Entretanto, se os neurônios forem ativados assincronamente, a sinapse será enfraquecida ou mesmo eliminada. Em outras palavras, se o neurônio pré-sináptico tiver grande influência na ativação do neurônio pós-sináptico, a conexão entre eles dever ser reforçada.

A sinapse Hebbiana possui quatro características principais:

- *Mecanismo iterativo*: no aprendizado Hebbiano não é permitido analisar atividades pré-sinápticas de forma isolada; qualquer

modificação na sinapse Hebbiana depende (estatisticamente ou deterministicamente) da interação entre os dois tipos de atividades (pré e pós-sinápticas).

- *Mecanismo local*: uma sinapse Hebbiana é uma transmissão com sinais contínuos que produz modificações sinápticas locais que são entradas específicas. É a partir do mecanismo local que sinapses Hebbianas efetuam Aprendizado Não Supervisionado.
- *Mecanismo dependente do tempo*: as modificações em uma sinapse Hebbiana dependem do momento exato de ocorrência das atividades pré e pós-sinápticas.
- *Mecanismo correlacionado ou conjunto*: a sinapse Hebbiana pode ser chamada de sinapse conjuncional pelo fato de que a ocorrência conjunta de atividades pré e pós-sinápticas é suficiente para que haja uma modificação. Além disso, pode também ser chamada de sinapse correlacional, porque uma correlação entre estas mesmas atividades é suficiente para gerar mudanças.

Pode-se expressar o postulado de Hebb em termos matemáticos a partir da seguinte equação, que consiste em uma regra para a mudança do peso sináptico w_{ij} (regra do produto de atividade):

$$\Delta w_{ij}(t) = \eta \cdot y_i(t) x_j(t) \quad (5.9)$$

onde η é uma constante positiva que determina a taxa de aprendizado, y_i é a saída desejada no instante de tempo t e x_j é a entrada do neurônio j no instante de tempo t .

Apesar dos vetores de entrada e saída serem fornecidos (como no Aprendizado Supervisionado), a regra de Hebb é classificada como Aprendizado Não Supervisionado, já que não existe supervisor externo para verificar a qualidade da resposta da rede para prover um ajustes de pesos. Neste caso, o treinamento da rede é feito, independentemente, da resposta atual, através de um mecanismo local à sinapse.

V.8.2.2 – Aprendizado por Competição

Aprendizado por Competição é um caso particular de Aprendizado Não Supervisionado⁶²⁻⁶⁵. A idéia neste caso é, dado um padrão de entrada, fazer com que as unidades de saída disputem entre si para serem ativadas. Existe, portanto, uma competição entre as unidades de saída para decidir qual delas será a vencedora e, conseqüentemente, terá a sua saída ativada e seus pesos atualizados no treinamento. As unidades de entrada são diretamente conectadas às unidades de saída, sendo que estas últimas também podem estar ligadas entre si via conexões laterais inibitórias, ou negativas. A unidade de saída com maior ativação inicial terá mais chance de vencer a disputa com as outras unidades, que perderão o poder de inibição ao longo do tempo sobre as unidades de maior ativação. A unidade mais forte fica ainda mais forte e seu efeito inibidor sobre as outras unidades de saída torna-se dominante. Com o tempo, todas as outras unidades de saída ficarão completamente inativas, exceto a vencedora. Um algoritmo simples de aprendizado competitivo é:

- Apresentar um vetor de entrada;
- Calcular a ativação inicial de cada unidade de saída;
- Deixar as unidades de saída competirem até que apenas uma permaneça ativa;
- Aumentar os pesos sobre as conexões entre as unidades de saída ativa e as unidades de entrada ativa. Com isso, a unidade de saída terá maior probabilidade de ficar ativa na próxima repetição do padrão.

Um problema neste algoritmo é que uma unidade de saída pode se tornar dominante e ficar ativa todo tempo, podendo captar para si todo o espaço de entrada. A solução para tal problema está em racionar os pesos, de tal forma que a soma dos pesos sobre as linhas de entrada de uma unidade seja limitada a 1. Para aumentar o peso de uma conexão, é preciso diminuir o peso de outra, tal que:

$$\Delta w_j = \eta x_j / M - \eta w_j \quad (5.10)$$

para todo $j = 1, 2, \dots, n$, onde w_j é o peso da conexão da unidade de entrada j com a unidade ativa, x_j é o valor do j -ésimo bit de entrada, M é o número de unidades de entrada ativas no vetor de entrada e η é a taxa de aprendizagem.

O fato de um único neurônio está ativo em um determinado instante, torna a aprendizagem competitiva muito adequada para descobrir características salientes que podem ser utilizadas na classificação de padrões.

V.8.3 – Aprendizado por Reforço⁶⁶

Aprendizado por Reforço pode ser visto como um caso particular de Aprendizado Supervisionado. A principal diferença entre o Aprendizado Supervisionado clássico e o Aprendizado por Reforço é a medida de desempenho usada em cada um dos sistemas. No Aprendizado Supervisionado, a medida de desempenho é baseada no conjunto de respostas desejadas usando um critério de erro conhecido, enquanto que, no Aprendizado por Reforço o desempenho é baseado em qualquer medida que possa ser fornecida ao sistema. No Aprendizado por Reforço, a única informação de realimentação fornecida à rede é se uma determinada saída está correta ou não, isto é, não é fornecida a rede a resposta exata para o padrão de entrada.

Figura 55 - Aprendizado por Reforço

Aprendizado por Reforço é uma forma de aprendizado *on-line* obtido por um mapeamento de entrada-saída, através de um processo de triagem de erro desenvolvido para maximizar o índice de desempenho escalar chamada de *signal de reforço*. O termo Aprendizagem por Reforço foi, inicialmente, citado por Minsky⁶⁶ em seus estudos iniciais de IA. A idéia básica que está por trás do termo “reforço” tem sua origem em estudos experimentais sobre aprendizado dos animais⁶⁷. Neste contexto, é interessante lembrar a *Lei do Efeito*⁶⁸ que diz quanto maior a satisfação obtida com uma certa experiência em um animal,

maiores as chances dele aprender. Sutton⁶⁹ descreve o Aprendizado por Reforço como:

“Se uma ação tomada pelo sistema de aprendizagem é seguida de estados satisfatórios, então a tendência do sistema de produzir esta ação particular é reforçada. Se não for seguida de estados satisfatórios, a tendência do sistema de produzir esta ação é enfraquecida”.

V.9 – Principais arquiteturas das Redes Neurais

A conexão entre os neurônios que formam as redes neurais artificiais foi motivado pela estrutura na qual os neurônio biológico estão conectados dentro do cérebro. O cérebro humano possui de 10^{10} a 10^{11} neurônios, cada um com 10^3 a 10^4 dendritos, ou seja, aproximadamente 10 bilhões de neurônios e 60 trilhões de conexões⁷⁰.

Com o objetivo de simular algumas características do cérebro, como aprendizagem e capacidade de generalização, os neurônios artificiais foram interligados de várias formas, formando redes neurais artificiais de diferentes arquiteturas que podem variar em função do número de camadas da rede, número de neurônios por camada intermediária, tipo de conexão entre os neurônios e a conectividade da rede. A definição da arquitetura de uma rede neural artificial restringe o tipo de problema que pode ser tratado pela rede.

V.9.1 – Número de camadas.

Redes de camada única - existe somente uma camada de neurônio entre qualquer entrada e saída da rede (Fig.56a). Estas redes estão capacitadas a resolver somente problemas linearmente separáveis.

Redes de camadas múltiplas possuem uma ou mais camadas de neurônios entre a entrada e saída da rede. Estão hábeis a solucionar problemas não linearmente separáveis, podendo implementar qualquer função contínua. A utilização de duas camadas intermediárias permite a aproximação de qualquer função⁷¹.

Figura 56 – (a) Rede de camada única, (b) Rede de camadas múltiplas.

V.9.2 - Número de neurônios por camada intermediária

A topologia de uma rede de camadas múltiplas (como a Perceptron de camadas múltiplas) está intimamente ligada com a definição dos neurônios de camada intermediária; este número é, geralmente, definido empiricamente e depende, fortemente, da distribuição dos padrões de treinamento e validação da rede. Alguns métodos, no entanto, têm sido propostos. Os mais utilizados são:

- Definir o número de unidades em função do número de entradas e saídas;
- Utilizar um número de conexões dez vezes menor que o número de casos.

Para solução de problemas práticos de reconhecimento de padrões, destina-se para a rede um número de unidades intermediárias suficiente para a solução do problema. Deve-se ter cuidado para não utilizar unidades em excesso, o que pode levar a rede a memorizar padrões de treinamento ao invés de extrair as características gerais que permitirão a generalização ou reconhecimento de padrões não vistos durante o treinamento (este problema é chamado de *overfitting*). Também não deve ser utilizado um número muito pequeno de conexões, forçando a rede a gastar tempo em excesso tentando encontrar uma representação ótima e fazendo com que as unidades utilizadas possam ficar sobrecarregadas, tendo que lidar com um elevado número de restrições, ocasionando a não convergência durante o seu treinamento (este problema é chamado de *underfitting*). Uma forma de evitar o *overfitting* é estimar o erro de generalização durante o processo de treinamento. Para isto, o conjunto de dados deve ser dividido em *conjunto de treinamento* e *conjunto de validação*. O conjunto de treinamento é utilizado na modificação dos pesos,

e o conjunto de validação é utilizado para estimar a capacidade de generalização da rede durante o processo de aprendizagem⁶⁰.

O treinamento da rede deve ser interrompido quando o erro do conjunto de validação começar a subir, ou seja, quando a rede começar a incorporar o ruído presente nos dados causando degradação na sua capacidade de generalização.

V.9.3 – Tipo de conexão entre os neurônios

Redes com alimentação direta ou feedforward (Fig.57a) - Os neurônios de uma camada são ligados aos neurônios da camada seguinte, através dos pesos sinápticos. Não existem conexões entre os neurônios de uma mesma camada e não é permitido “pular” camadas (por exemplo, conexão entre os neurônios da primeira camada com os da terceira). As redes Perceptron de Camadas múltiplas são os tipos de *redes com alimentação direta* mais utilizadas, possuindo a habilidade de resolver operações lógicas complexas (excelente para aproximar funções) e classificação de padrões.

Segundo Westead⁷², as redes neurais *com alimentação direta* correspondem a mais de 90% das aplicações com redes existentes.

Redes com alimentação recorrente ou feedback (Fig.57b) – Na rede recorrente as saídas dos neurônios são utilizadas para alimentar as entradas. A rede *Hopfield* é o tipo mais conhecido de redes recorrentes e é utilizada em problemas de otimização, memória associativa e para produzir conversores analógicos-digitais⁷³.

Figura 57 – (a) Redes com alimentação direta; (b) Redes com alimentação recorrente.

V.9.4 – Conectividade da rede – Uma rede é considerada *totalmente conectada*, quando cada um dos nós de uma camada da rede está conectada a todos os nós da camada adjacente seguinte (Fig. 56 e 57). Entretanto, se alguns dos elos de comunicação (conexões sinápticas) estiverem faltando na rede, esta é considerada *parcialmente conectada* (Fig. 58)

Figura 58 – Rede parcialmente conectada.

V.10 – Tipos de Redes Neurais Artificiais

V.10.1 – Perceptron

O Perceptron⁴⁸ (Fig.59) foi primeiramente introduzido por F. Roseblatt em 1958. Este é um tipo simples de rede neural com duas camadas de neurônios; aceita somente dados de entrada *binários* e valores do tipo (0 ou 1) e utiliza o método de correção de erros no seu treinamento. O processo de

aprendizagem é **supervisionado**. A rede tem habilidade de resolver operações lógicas básicas como **E** ou **OU** e reconhecimento de padrões.

Figura 59 – Perceptron simples

V.10.2 – Perceptron de Camadas Múltiplas (Extensão da rede Perceptron)

As redes Perceptron de camadas múltiplas⁵⁴ (Fig.60) possuem uma ou mais camadas intermediárias (ocultas) entre suas camadas de entrada e saída, opera com aprendizado supervisionado e utiliza o algoritmo de retropropagação de erros. A Perceptron de camadas múltiplas está habilitada a resolver operações lógicas complexas, incluindo os problemas XOR (operação lógica com valores de X e Y contraditórios).

Figura 60 – Rede Perceptron de Camadas múltiplas

V.10.3 – Rede Hopfield

A rede Hopfield⁵² (Fig.61) foi primeiramente introduzida pelo físico J.J. Hopfield em 1982 e corresponde a um tipo de rede neural que são denominados “*modelos termodinâmicos*”.

Esta rede é composta de um jogo de neurônios, totalmente conectados. Não existe uma diferenciação entre os dados dos neurônios de entrada e os dados de saída. É uma rede recorrente, com tipo de aprendizado não supervisionado. As principais aplicações da rede Hopfield são o

armazenamento e reconhecimento de padrões, por exemplo, arquivos de imagens e otimização de problemas.

Figura 61 – Rede Hopfield

V.10.4 – Redes Kohonen – Mapa de características.

A rede Kohonen Feature Map⁶³ foi introduzida pelo professor finlandês Teuvo Kohonen (University of Helsinki) em 1982.

Este é provavelmente o tipo de rede neural mais vantajoso, se o processo de aprendizagem do cérebro humano for simulado. O coração deste tipo de rede é o mapa de características, uma camada onde os neurônios são organizados de acordo com certos valores de entrada.

Este tipo de rede neural consiste em ambas *alimentação direta* (camada de entrada de dados para caracterizar o mapa) e *recorrente* (mapa de características). Utiliza um método de aprendizado não supervisionado e possui uma organização própria. Este tipo de rede é muito utilizado para classificação de padrão, otimização de problemas e simulação.

Figura 62 – Rede Kohonen (mapa de característica)

V.11 - Utilização das redes neurais –

As RNAs são sistemas muito flexíveis com capacidade de resolução de problemas variados, em áreas distintas. Tais problemas normalmente se resumem em otimização, simulação ou classificação.

- ✓ Classificação de padrões e reconhecimento de padrões⁷⁴⁻⁷⁸;
- ✓ Controladores de processos^{79,80};
- ✓ Simulações^{81,82};
- ✓ Processamento de imagens⁸³ e análise de fala⁸⁴;
- ✓ Modelagem de sistemas complexos e otimização de processos⁸⁵;
- ✓ Mercado Financeiro (análise de créditos)⁸⁶;
- ✓ Medicina⁸⁷⁻⁸⁹, etc.

Uma habilidade das RNAs que deve ser mencionada explicitamente é a sua tolerância a erros. Isso significa que, se uma rede neural for treinada para um problema específico, poderá reconhecer os resultados corretos, mesmo se o problema a ser resolvido não estiver exatamente igual aos já treinados anteriormente.

Por exemplo, suponha que uma rede neural tenha sido treinada para reconhecer a fala humana. Durante o processo de aprendizagem, uma certa pessoa tem que pronunciar algumas palavras que são instruídas pela rede. Então, se a rede foi corretamente treinada, ela deverá reconhecer essas palavras mesmo quando forem ditas por outras pessoas.

V.11.1 – Utilização de redes neurais artificiais em determinação estrutural e previsão de deslocamentos químicos.

Com a introdução das redes neurais artificiais na química, seu potencial foi imensamente explorado na classificação de espectros de produtos naturais⁹⁰, determinação estrutural de compostos orgânicos⁹¹ e previsão de deslocamentos químicos de RMN ^1H e ^{13}C ⁹²⁻¹⁰⁰. Recentemente, foi mostrado a descrição numérica do código HOSE (Utilizado no SpecInfo - seção III.4) para redes neurais¹⁰¹.

A computação de deslocamentos químicos com uma rede neural artificial requer uma descrição numérica do ambiente químico do átomo de carbono de interesse. Esta descrição deve reunir certas condições para serem utilizadas como dados de entrada de uma rede neural. As características dos ambientes químicos descritos para cada átomo de carbono devem ser apresentadas de forma bem detalhada para a rede, para que a influência de uma respectiva característica estrutural possa ser compreendida por uma unidade de entrada de dados.

Meiler¹⁰² desenvolveu o programa C-Shift utilizando redes neurais artificiais para previsão de deslocamentos químicos de RMN ^{13}C de substâncias orgânicas, comparando-as com outros métodos tradicionais. Foi utilizada uma descrição numérica para caracterizar o ambiente químico dos átomos de carbonos utilizados como dados de entrada da rede. As redes neurais artificiais são capazes de combinar as vantagens do método de incrementos¹⁰³ e o código HOSE para previsão de deslocamentos químicos, funcionando de forma mais rápida e precisa.

Após serem treinadas, as redes neurais (utilizadas por Meiler) executam uma previsão aproximadamente 1000 vezes mais rápida que os programas tradicionais, e elas independem do acesso direto à base de dados. São, geralmente mais precisas, que os métodos de incrementos e que o código HOSE, na estimativa de deslocamentos químicos de moléculas recém sintetizadas que ainda não tem boa representatividade no banco de dados.

Todas as redes de Meiler foram criadas, treinadas e analisadas utilizando o Software SMART¹⁰⁴. As redes de três camadas foram treinadas utilizando o algoritmo de retropropagação de erros. A velocidade de

aprendizado η foi decrescente, variando de 0,01 à 0,0001 durante o processo de treinamento da rede. O treinamento foi realizado com número de ciclos entre 300 e 7000, dependendo do tamanho e do conjunto de dados utilizado no treino.

Para verificar a eficiência das redes neurais, Meiler realizou um teste com o agente antitumoral TAXOL, onde foram comparados os resultados obtidos com redes neurais com os resultados de outros programas que utilizam o método de incrementos, código HOSE e Química Quântica.

- Utilizando o programa C-Shifft (redes neurais), foi obtido um desvio padrão de 1,3 ppm.
- Os resultados obtidos utilizando o código HOSE até a quinta esfera forneceu um desvio padrão de 2,6 ppm para o programa SpecInfo; 3,7 ppm para o PredictIt NMR e 2,9 ppm para programa ACD CNMR. Considerando que, o resultado do programa ACD foi satisfatório em função da estrutura do taxol ser bem conhecida e já estar codificada em seu banco de dados juntamente com 14 outras estruturas similares a ela.
- Os programas que utilizam métodos de Incremento apresentaram os piores resultados, fornecendo desvios de 5,6 ppm e 5,3 ppm para os programas ChemDraw e SpecTool, respectivamente.
- As mais extensivas rotinas de química quântica dos programas Cosmos e Gaussian levaram também a grandes desvios (4,6 ppm e 5,1 ppm respectivamente) causados principalmente pela representação incorreta da estrutura tridimensional.

Além de apresentar ótimos resultados, as redes neurais podem ser retreinadas dinamicamente para serem ajustadas, caso seja inserida uma nova substância ou até mesmo uma nova classe de substâncias.

Nesta tese, as redes neurais artificiais foram utilizadas na identificação de esqueletos dos alcalóides Aporfínicos utilizando deslocamentos químicos de RMN ^{13}C como dados de entrada da rede.

V.12 - Identificação de esqueletos dos alcalóides Aporfínicos utilizando Redes Neurais Artificiais.

O primeiro passo para trabalhar com redes neurais, é analisar o problema a ser resolvido e o tipo de dados disponíveis, para escolher a melhor rede e as condições de treinamento que se adeque ao caso. No processo de identificação de esqueletos dos alcalóides Aporfínicos foi utilizada a rede Perceptron de Camadas múltiplas, com o algoritmo de retropropagação de erros.

Embora já tenha sido brevemente definida no item V.10.2 (Pg.160), é necessário que sejam introduzidos alguns conceitos para um melhor entendimento dos resultados apresentados posteriormente.

V.12.1 - Definição de uma Topologia Ideal para a Rede Perceptron de Camadas múltiplas.

Uma das maiores dificuldades em se definir a estrutura de uma RNA é o fiel dimensionamento de sua topologia. Normalmente, o número de camadas e o número de neurônios em cada camada é definido em função de uma inspeção prévia nos dados e da complexidade do problema.

Em uma rede de camadas múltiplas, o processo realizado por cada neurônio é definido pela combinação dos processamentos realizados pelos neurônios da camada anterior que estão conectados a ele. Quando se segue da primeira camada intermediária em direção à camada de saída, as funções de implemento se tornam cada vez mais complexas. A utilização de mais de duas camadas intermediárias não é recomendada, visto que, cada vez que o erro medido durante o treinamento é propagado para a camada anterior, ele se torna menos útil ou preciso. A única camada que tem uma noção precisa do erro cometido pela rede é a camada de saída.

Com relação ao número de neurônios nas camadas intermediárias, este é em geral definido empiricamente, como mencionado na seção V.9.2 (Pg.158).

Uma vez definida a topologia inicial, a estrutura final mais adequada para o modelamento é, normalmente, obtida através de refinamentos sucessivos, que pode levar a um tempo de dimensionamento alto, já que este tem um grande componente empírico.

O objetivo desta etapa de ajuste é a obtenção de uma topologia de rede que modele com precisão os dados do conjunto de treinamento, mas também resulte em uma aproximação com boa capacidade de generalização. Como, neste caso, o conjunto de treinamento de uma RNA é composto de dados experimentais, estes contêm, implicitamente, erros inerentes aos processos de amostragem. Desta forma, a aproximação através de RNAs deve ser feita visando a obtenção de uma estrutura que seja capaz de modelar os dados sem modelar o ruído contido neles. Este é um problema no projeto de RNAs conhecida na literatura como *variação do bias**, que envolve a obtenção de um modelo que não seja muito rígido a ponto de não modelar fielmente os dados, mas que também não seja, excessivamente, flexível a ponto de modelar também o ruído.

O algoritmo de aprendizado mais conhecido para treinamento destas redes e utilizado nesta tese é o *método de retropropagação de erro (back-propagation)*. A maioria dos métodos de aprendizado para RNAs do tipo Perceptron de camadas múltiplas utiliza variações deste algoritmo.

O *back-propagation* é um algoritmo supervisionado, que utiliza pares (entrada, saída desejada) para ajustar os pesos da rede através de um mecanismo de correção de erros. O treinamento ocorre em duas fases, onde cada fase percorre a rede em um sentido. Estas duas fases são chamadas de *fase de alimentação direta* e *fase de alimentação recorrente*.

Figura 63 – Fluxo de Processamento do algoritmo de retropropagação.

* o bias é aplicado externamente e tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo, respectivamente.

A fase *forward* envolve os seguintes passos:

- (a) Os dados são apresentados à camada de entrada da rede (C^0),
- (b) Os neurônios das segunda camada (C^1) utilizam os dados de entrada para calcular (através da função de ativação) os sinais de saída que serão enviados, como dados de entrada, para a camada C^{j+1} .
- (c) As saídas produzidas pelos neurônios da última cada são comparadas às saídas desejadas

Começando da última camada, até chegar na camada de entrada, a fase *backward* envolve as etapas:

- (a) Os nodos da camada atual ajustam seus pesos de forma a reduzir seus erros
- (b) O erro de um nodo da camada intermediária é calculado, utilizando os erros dos nodos da camada seguinte conectados a ele, ponderado pelos pesos das conexões entre eles

O algoritmo *back-propagation*, que faz uso destas duas fases, é apresentado a seguir:

- 1 – Estabelecer os pesos e parâmetros
- 2 – Repetir a operação até o erro ser mínimo ou a realização de um dado número de ciclos:
 - 2.1 – Para cada padrão de treinamento X
 - 2.1.1 – Definir saída da rede através de fase *forward*
 - 2.1.2 – Comparar saídas produzidas com as saídas desejadas
 - 2.1.3 – Atualizar pesos do nodos através da fase *backward*

O *back-propagation* é baseado na regra delta, sendo então chamada de regra delta generalizada. Este algoritmo propõe uma forma de definir o erro dos nodos das camadas intermediárias, possibilitando o ajuste de seus pesos. Os ajustes dos pesos são realizados utilizando o método do gradiente descendente (seção V.8.1.1 – equação 5.8). A derivação da regra delta generalizada é simples, a função a ser minimizada é uma função erro, definida pela soma dos erros quadráticos e representada por:

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (5.12)$$

onde E é a medida de erro total, p é o número de padrões, k é o número de unidades de saída, d_i é a i -ésima saída desejada e y_i é a i -ésima saída gerada pela rede. Esta equação define o erro total cometido pela rede, ou a diferença encontrada entre a saída calculada pela rede e a saída desejada.

A regra delta generalizada requer que as funções de ativação utilizadas pelos nodos sejam contínuas, diferenciáveis, geralmente, não decrescente da entrada total recebida pelo nodo. Estas funções são chamadas de funções semi-lineares. A Equação abaixo ilustra o cálculo do valor de ativação.

$$y_j^p = f_j(net_j^p) \quad (5.13)$$

onde:

$$net_j^p = \sum_{i=1}^n x_i^p w_{ji} \quad (5.14)$$

A constante n represente o número de conexões de entrada do nodo j e w_{ij} o peso da conexão entre a entrada x_i^p e o nodo j .

Há garantia de convergência se a superfície do erro for simples. A visualização da influência do treinamento da rede no valor do erro está ilustrada pela superfície de erro da Figura 64, onde está representado um caso simples de uma rede com apenas um neurônio e duas conexões. Os valores dos pesos da rede definem a coordenada de um ponto da superfície do erro. O erro produzido pela rede, para cada combinação de valores de pesos, é dado pela altura da superfície naquele ponto. Assim, quanto mais alto estiver o ponto, maior é o erro produzido pela rede.

Figura 64 – Superfície do Erro para a Rede MLP – Considerando as coordenadas $W1$ e $W2$.

As três equações anteriores mostram como encontrar o erro global da rede e como encontrar o erro local de um determinado padrão de entrada, processadas as duas equações anteriores deve-se aplicar a equação seguinte que ajustará os pesos das diversas camadas:

$$\Delta w_{ji} = \eta \delta_j x_i \quad (5.15)$$

$$\text{ou} \quad w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j(t) x_i(t) \quad (5.16)$$

onde o erro $\delta_j = (d_j - y_j)$ é utilizado para o ajuste dos pesos da camada de saída; e o erro $\delta_j = \sum w_{ij} \delta_i$ é utilizado para o ajuste dos pesos das camadas intermediárias, i unidade em questão, j camada em questão.

Problemas Enfrentados no Treinamento de Redes Perceptron de camadas múltiplas.

A frequência de ajustes dos pesos influencia o desempenho obtido durante o treinamento.

O principal problema diz respeito a lentidão do algoritmo para superfícies complexas. Uma forma de minimizar este problema é considerar efeitos de segunda ordem para o gradiente descendente. Não é raro o algoritmo convergir para mínimos locais.

Figura 65 – Gráfico da superfície de erro, evidenciando os mínimos locais e globais.

Mínimos locais são pontos na superfície de erro que apresentam uma solução estável, embora não seja correta. Algumas técnicas são utilizadas tanto para acelerar o algoritmo de *retropropagação* como para reduzir a incidência de mínimos locais:

- Utilizar taxa de aprendizado (η) decrescente;
- Utilizar um termo *momento*;

V.12.1.1 - Taxa de aprendizado e o termo momento.

O algoritmo de retropropagação fornece uma "aproximação" para a trajetória no espaço de pesos calculada pelo método da descida mais íngreme. Quanto menor for o parâmetro da taxa de aprendizado (η), menores serão as variações dos pesos sinápticos da rede (de uma iteração para a outra), mais suave será a trajetória no espaço de pesos e mais demorado será o processo. Por outro lado, se for utilizada uma taxa de aprendizado muito alta, as grandes modificações nos pesos sinápticos resultantes podem tornar a rede instável (i. e., oscilatória).

Entre as várias técnicas utilizadas para acelerar o processo de treinamento e evitar mínimos locais, a adição de um termo *momentum* é uma das mais freqüentes. O termo *momentum* é representado por:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j(t) x_i(t) + \alpha (w_{ij}(t) - w_{ij}(t-1)) \quad (5.17)$$

onde α é taxa de *momentum* a ser utilizada no processamento do aprendizado da rede. A inclusão da taxa de *momentum* na fórmula de ajuste dos pesos aumenta a velocidade de aprendizado (aceleração), reduzindo o perigo de instabilidade. A taxa *momentum* pode acelerar o treinamento em regiões muito planas para a superfície de erro. Além disso, ele suprime a oscilação de pesos em vales e ravinas*, ajudando a rede a sair dos mínimos locais.

* Cavidades da superfície de erro onde se localizam os mínimos locais.

Figura 66 – Influência do termo momento.

Terminada a apresentação dos conceitos necessários para compreensão dos parâmetros, envolvidos no treinamento da rede e utilizados neste trabalho, o próximo passo foi aplicar a teoria no desenvolvimento de uma metodologia para treinar a rede Perceptron de camadas múltiplas, em determinação de esqueletos dos alcalóides Aporfínicos.

V.13 – Metodologia

Como mencionado nas seções anteriores as redes neurais artificiais utilizam métodos de aprendizagem, ou seja, o seu treinamento deve ser feito com a utilização de dados bem detalhados e corretos, para que ela não aprenda de forma errônea. Para isso foi construída uma planilha em Excel (Anexo 1) com a codificação de 162 alcalóides aporfínicos pertencentes a 12 esqueletos diferentes (Fig.67) com seus respectivos deslocamentos químicos de RMN ^{13}C . Os Aporfínicos foram escolhidos em função de não possuir esqueletos com estruturas tão variadas como os outros alcalóides do banco de dados). A construção desta planilha é o passo principal para a obtenção de bons resultados.

Figura 67 – Esqueletos dos alcalóides aporfínicos.

Na primeira coluna da planilha foram colocados, sob a forma de códigos (APF001-APF162), os alcalóides presentes no banco de dados. Nas 12 colunas seguintes são identificados(0 ou 1) os respectivos esqueletos de cada um dos 162 alcalóides. Nas últimas 22 colunas estão os deslocamentos químicos de RMN ^{13}C de cada tipo de átomo de carbono encontrado nos

esqueletos dos alcalóides, seguindo uma numeração biogenética criada para esta planilha, numeração esta que vai de C-1 a C-18, e 4 colunas que diferenciam quais e quantos substituintes estão ligados ao nitrogênio dos alcalóides aporfínicos. A codificação (Tab.76) da substância 7-metildehidroglaucine¹⁰⁵ (Fig.68), alcalóide pertencente ao esqueleto Aporfínico 05, mostra melhor a construção da planilha

Figura 68 – Alcalóide aporfínico, 7-metildehidroglaucine.

Tabela 76- Linha da planilha que representa a substância 7-metildehidroglaucine .

	Esq. 1	Esq. 2	Esq.3	Esq.4	Esq.5	Esq.6
Identificação dos esqueletos	0	0	0	0	1	0
	Esq. 7	Esq. 8	Esq.9	Esq.10	Esq.11	Esq.12
	0	0	0	0	0	0
Deslocamentos químicos de RMN ¹³ C	C-1	C-2	C-3	C-4	C-5	C-6
	49.6	24.9	111.8	149.8	144.6	124.1
	C-7	C-8	C-9	C-10	C-11	C-12
	140.7	129.7	121.7	120.8	129.6	121.1
	C-13	C-14	C-15	C-16	C-17	C-18
	109.1	147.1	148.8	104.6	14.1	?
	N-Me1*	N-Me2*	N-Et(CH ₂)*	N-Et(CH ₃)*		
	42.5	?	?	?		

*É bom lembrar que os grupos etila e metila, ligados ao átomo de nitrogênio, estão desfuncionalizados em suas representações, o que não significa que o mesmo ocorra em todas as moléculas da planilha, e o N-Me2 representa a segunda metila ligada ao Nitrogênio, esteja a substância em sua forma lônica (Esqueleto 6) ou não (Esqueleto 4).

Na planilha contendo os deslocamentos químicos, o carbono C-1 pode aparecer de cinco formas diferentes. Dependendo da natureza do alcalóide representado, C-1 pode ser encontrado sob a forma de uma carbonila, um

metileno (CH_2), pode ser um carbono aromático, estar ou não ligado a um grupo O-R qualquer, ou ainda estar ligado a um átomo de nitrogênio positivo. Independente da natureza do C-1, todos os deslocamentos são colocados numa mesma coluna, pois a rede consegue, através do deslocamento do C-1 e dos átomos vizinhos, identificar a natureza deste átomo de carbono. Da mesma forma são adicionados todos os outros deslocamentos químicos correspondentes do C-2 até o C-18.

O número de átomos de carbono pode variar de um esqueleto para o outro (por exemplo nos aporfínicos 1,6,7,e,12 da figura 67), quando o esqueleto não possuir um determinado átomo (por exemplo C-17, C-18 e/ou carbonos dos grupos metila e etila ligados ao nitrogênio), no lugar do deslocamento químico é colocado um sinal de interrogação (?).

Os sinais de interrogação são tratados pelo programa como valores de dados “ausentes”, uma vez que os mesmos não tenham sido armazenado no banco dados. Estes sinais são atribuídos por dois motivos distintos: *difficuldade de interpretação dos espectros* ou simplesmente por que *o dado se refere a um átomo de carbono que não está presente na substância* (Ex. carbono C-17 em uma substância pertencente ao esqueleto Aporfínico 1).

Terminada a construção da planilha, o próximo passo consistiu em transportá-la para o programa de redes neurais artificiais*, definindo as condições de aprendizado.

Cada tipo de esqueleto pertencente aos alcalóides aporfínicos foi utilizado, como dado de Saída do programa, para cada série de teste.

Os deslocamentos químicos de RMN ^{13}C foram utilizados como dados de Entrada da rede. No programa Statistica Neural Networks – StatSoft, a divisão dos casos para o primeiro aprendizado da rede é realizada de forma aleatória nas proporções de 2:1:1 para *treino*, *verificação* e *teste* da rede, respectivamente. Após o treinamento, dependendo dos resultados apresentados pela rede, tem-se a opção de fixar estes dados (se os resultados forem bons) ou selecionar uma nova distribuição aleatória até que a rede “entenda” que existem exemplos de casos, bem distribuídos, que lhe proporcionará segurança para execução dos treinos, verificações e testes.

* Neste trabalho foi utilizado o programa Statistica Neural Networks Versão 4 – StatSoft¹⁰⁶.

Os tipos de redes utilizadas neste trabalho foram as Perceptron Multicamadas com *três* e *quatro* camadas intermediárias de neurônios (Fig.69)

Figura 69 – Redes de 3 e 4 camadas.

A complexidade da rede, i.e., a unidade de neurônios por camada intermediária, foi determinada automaticamente pelo programa utilizando a Solução de Problemas Inteligentes (IPS – Intelligent Problem Solver). Geralmente, aumentando o número de neurônios intermediários, aumenta-se o poder de modelagem da rede neural artificial. Entretanto deve-se cuidar para que o *overffiting* (definido na seção V.8.2) seja evitado. O *overffiting* é observado quando a rede começa a fornecer boas correlação para *treino* e correlações ruins para *verificações*.

No IPS, foi selecionado um tempo de treinamento médio para a duração do processo, conduzindo a uma procura rápida por uma rede ótima. Também foi selecionado para manter as redes com as melhores performance e salvar um número máximo de dez redes.

Para pesquisar o provável esqueleto de uma substância desconhecida, basta que, seja adicionada na planilha uma linha com seus respectivos deslocamentos químicos de RMN ^{13}C , e posteriormente utilize esta linha como teste.

Os deslocamentos químicos, das substâncias submetidas aos testes, são inseridos em ordem crescente nas linhas da planilha, uma vez que, se tratando de um esqueleto desconhecido, não faz sentido entrar com os dados na seqüência biogenética da substância.

Terminado o treinamento da rede, pode-se avaliar sua eficiência, a confiabilidade dos seus resultados e a possibilidade de uma substância pertencer ou não ao esqueleto pesquisado. Isto é feito pela análise das correlações, obtidas através das regressões estatística, e do erro individual calculado para cada substância, obtido a partir da probabilidade da mesma pertencer ao esqueleto pesquisado, ambos fornecidos pelo programa.

Foram adicionadas na planilha os deslocamentos químicos de RMN ^{13}C (Tab.77) de 12 substâncias, desconhecidas pela rede (Fig.70), pertencentes aos esqueletos dos alcalóides aporfínicos.

Figura 70 – Alcalóides utilizados para teste das Redes Neurais.

Tabela 77 - Deslocamentos químicos dos Alcalóides aporfínicos que serão identificados pela rede.

Carbono	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10	C-11	C-12	C-13	C-14	C-15	C-16	C-17	N-		Esqueletos ^a
																		-Me1	-Me2	
Teste A	42.8	29.2	108.3	146.9	142.6	112.3	54.0	129.4	126.4	37.4	129.1	122.8	148.5	145.6	114.5	123.4	?	?	?	ESQ 01
Teste B	156.6	176.1	112.8	153.3	155.0	124.7	132.7	124.9	120.0	114.3	133.3	127.2	128.1	127.7	128.4	129.6	?	33.5	?	ESQ 02
Teste C	52.8	23.5	111.0	149.1	149.2	130.5	62.8	123.5	129.9	33.7	123.0	122.9	114.1	144.2	144.8	145.5	?	40.3	?	ESQ 02
Teste D	55.3	36.0	116.5	145.8	144.3	126.3	140.0	132.8	127.3	106.3	133.1	128.3	126.9	127.0	127.1	131.9	47.8	152.1	?	ESQ 03
Teste E	60.1	31.1	106.6	146.6	155.6	124.3	120.9	131.5	125.6	121.5	126.6	131.7	115.7	119.6	126.9	151.3	?	44.6	44.6	ESQ 04
Teste F	45.6	19.8	147.1	143.4	145.2	116.5	172.0	122.7	117.4	73.1	145.2	121.1	129.5	113.6	159.5	109.5	33.7	?	?	ESQ 05
Teste G	45.8	26.4	105.6	151.2	142.4	115.6	169.7	132.8	113.6	72.5	145.2	106.6	157.6	98.4	161.3	101.1	32.4	?	?	ESQ 06
Teste H	61.5	23.5	108.0	148.0	142.2	118.2	69.7	119.8	118.3	28.6	123.7	122.5	112.1	146.2	145.5	114.3	?	42.6	53.4	ESQ 06
Teste I	60.3	23.8	110.6	152.9	142.2	126.0	69.1	125.2	118.4	28.9	124.3	120.2	143.5	149.7	111.5	119.6	?	42.9	53.5	ESQ 06
Teste J	59.4	32.5	116.0	150.5	145.5	124.2	127.3	129.3	125.4	128.2	129.3	132.4	121.6	126.7	126.7	125.7	?	49.1	?	ESQ 08
Teste K	148.1	123.2	118.9	147.1	?	150.7	147.5	138.3	119.3	181.5	132.0	135.0	125.1	134.4	131.1	128.1	?	?	?	ESQ 09
Teste L	42.7	24.2	140.9	134.8	144.0	108.1	54.1	119.1	130.8	37.6	128.9	122.5	148.3	144.5	113.6	123.4	?	?	?	ESQ 01

a – Cada um dos alcalóides testados está identificado de acordo com os esqueletos aos quais pertencem.

As substâncias utilizadas para os testes foram escolhidas de acordo com a representatividade dos seus respectivos esqueletos na planilha, ou seja, os esqueletos que estão representados na planilha por menos de 4 substâncias não foram utilizados para teste. Como vimos anteriormente, as Redes Neurais aprendem por meio de exemplos, então deve haver um número suficiente de exemplos para sua aprendizagem.

Para identificar os esqueletos dos alcalóides submetidos aos testes, foi simulada uma situação real, na qual o usuário ignora totalmente a identidade das substâncias testadas com relação a classificação dos seus respectivos esqueletos. Portanto, para cada uma das doze substâncias, foram atribuídos “sinais de ausência” (?) nas suas colunas de identificação de esqueletos.

Após o treinamento das redes Perceptron de 3 e 4 camadas intermediárias de neurônios, a eficiência de ambas foram comparadas, com posterior validação da melhor arquitetura de rede encontrada.

V.14 – Resultados e discussão

Foram realizados treinamentos individuais para cada um dos 12 esqueletos presentes na planilha (Esq 01 – Esq 12), obtendo-se como resultado a probabilidade (de 0-1) dos alcalóides submetidos aos teste pertencerem ou não ao esqueleto utilizado como saída do programa de redes neurais e também as correlações obtidas para os *treinos* e *testes* realizados pela rede (Tab. 78).

Para o treinamento da rede, utilizando como saída o Esqueleto 01, foi observado que os compostos **A** e **L** têm, respectivamente, 98,78, 99.05 % de probabilidade de pertencerem a este esqueleto. Também, foi observado que nenhuma das substâncias utilizados para o teste pertence aos esqueletos 10,11 e 12, como era de se esperar.

Os resultados mostraram que não existe a possibilidade de uma mesma substância pertencer a mais do que um esqueleto.

Da mesma forma que o Esqueleto 01, todos os outros esqueletos foram treinados. Analisando os resultados, pode-se concluir que a rede não somente identifica um esqueleto positivamente como também não fornece um falso positivo como resultado, e isso é constatado ao observar-se, por exemplo, a probabilidade de o teste **A** pertencer ao ESQ03 que é de 0.53%, um valor praticamente insignificante.

Tabela 78 – Resultados dos treinamentos da rede para identificação dos esqueletos das substâncias testadas.

Esqueletos	Probabilidade (%) das Substâncias testadas pertencerem aos esqueletos usados como saída da rede.												Correlações			Erro RMS	
	A	B	C	D	E	F	G	H	I	J	K	L	Treino	Teste	Treino	Teste	
testados																	
ESQ 01	98.78	14.31	1.12	1.71	1.10	0.14	1.68	1.39	1.93	10.95	0.17	99.05	0.99178	0.86230	0.06158	0.25040	
ESQ 02	3.69	99.69	97.84	5.78	17.4	3.76	5.44	11.95	3.69	8.79	1.91	24.45	0.99987	0.87208	0.00659	0.21450	
ESQ 03	0.53	0.50	0.15	99.72	7.82	1.10	0.23	1.46	0.35	2.04	1.02	2.70	0.99873	0.94651	0.00861	0.13010	
ESQ 04	0.06	0.16	0.59	0.23	99.45	1.25	1.85	2.42	1.93	1.87	1.22	0.48	0.99212	0.78715	0.01529	0.13650	
ESQ 05	0.22	0.31	0.33	0.57	0.17	99.75	99.40	0.04	0.02	0.52	0.61	0.07	0.99421	0.78789	0.02343	0.17350	
ESQ 06	0.87	2.38	6.71	1.66	1.88	0.36	1.43	98.79	99.96	0.34	0.59	0.49	0.98983	0.79407	0.04327	0.19130	
ESQ 07	0.00	0.03	0.01	0.02	0.02	0.01	0.00	0.02	0.01	0.00	0.00	0.00	0.99999	0.99990	0.00041	0.01535	
ESQ 08	4.92	4.35	4.29	5.29	4.57	4.22	4.66	4.19	4.56	98.65	4.67	5.54	0.99068	0.74252	0.04575	0.12870	
ESQ 09	0.04	1.45	0.47	1.63	0.04	0.78	0.23	0.43	1.32	0.85	98.17	0.79	0.99885	*	0.00720	0.12900	
ESQ 10	2.78	2.72	0.89	1.95	1.25	2.49	2.53	1.83	1.13	0.90	2.65	0.87	*	*	0.02478	0.07648	
ESQ 11	0.01	0.01	0.01	0	0.00	0.01	0.01	0.01	0	0.01	0.00	0.02	*	*	0.13360	0.00780	
ESQ 12	1.45	1.85	1.92	1.98	0.34	0.05	0.56	1.38	2.27	0.69	1.32	1.95	*	*	0.01035	0.01294	

*Esqueletos que não possuem um número suficiente de casos para fornecer correlações.

OBS: As somas dos resultados obtidas para cada teste (A, B ...) ultrapassam o valor igual a 100. Isto ocorre por que cada linha de resultados corresponde a um treinamento individual realizado para cada um dos 12 esqueletos presentes na planilha de dados.

Os resultados da identificação de esqueletos estão resumidos na Tabela 79 para melhor visualização.

Tabela 79 – Resumo da identificação de Esqueletos.

Substância teste	Esqueleto	Probabilidade %
Teste A	ESQ 01	98,78
Teste B	ESQ 02	99,69
Teste C	ESQ 02	97,84
Teste D	ESQ 03	99,72
Teste E	ESQ 04	99,45
Teste F	ESQ 05	99,75
Teste G	ESQ 05	99,40
Teste H	ESQ 06	98,89
Teste I	ESQ 06	99,96
Teste J	ESQ 08	98,65
Teste K	ESQ 09	98,17
Teste L	ESQ 01	99,05

Analisando os resultados pode-se afirmar, com alto grau de confiança, a qual esqueleto cada uma destas substâncias pertence.

Partindo-se das propostas de esqueletos mostrada na Tabela 79, os espaços da planilha, previamente ocupados por valores ausentes (?), agora podem ser preenchidos com suas respectivas identificações de esqueletos.

A rede foi retreinada utilizando uma nova planilha onde todas as substâncias estavam devidamente identificadas, possibilitando assim a variação e análise de alguns parâmetros tais como o tipo de algoritmo utilizado, número de ciclos, velocidade de aprendizado, ruídos, entre outros. A variação destes parâmetros contribuíram para o aperfeiçoamento do treinamento da rede, aumentando sua eficiência com a obtenção de resultados mais confiáveis.

No treinamento das redes de três e quatro camadas, para o ajuste de parâmetros, o número de ciclos (épocas) foi aumentado gradativamente, numa faixa que variou de 100-1000 ciclos (com um passo de 100) e depois de 1000-10000 ciclos (com um passo de 1000). Foi fixado uma taxa de aprendizagem igual a 0,1 e uma taxa de momento igual a 0,3. Através dos resultados apresentados, foi possível comparar a eficiência entre as duas redes e observar a relação entre o coeficiente de correlação (dos *treinos*, *verificações* e *testes*) e o número de ciclos, representados graficamente nas Figuras 71 e 72.

Figura 71 – Ciclos x Coeficiente de correlação do treinamento da rede Perceptron, com três camadas intermediárias de neurônios, para o Esq 01 dos Alcalóides aporfínicos.

Figura 72 – Ciclos x Coeficiente de correlação do treinamento da rede Perceptron, com quatro camadas intermediárias de neurônios, para o Esq 01 dos Alcalóides aporfínicos.

Nestes gráficos observa-se que o aumento do número de ciclos resultou em comportamentos diferentes para o treino e teste das Redes Neurais. Os coeficientes de correlação dos treinos, permaneceram praticamente constantes e sempre apresentando um bom desempenho para ambas as redes de 3 e 4 camadas intermediárias de neurônios. Nas redes de três camadas foi observado que o coeficiente de correlação obtido nos testes aumentou de 0,93 para 0,99 quando o número de ciclos foi variado de 1 para 6000. Os coeficientes de correlação permaneceram quase constante entre 6000 e 8000 ciclos, decaindo após 8000 ciclos. O melhor coeficiente de correlação observado

para os testes, no treinamento da rede de 3 camadas, foi de **0,98854**, treinada com 8000 ciclos (Tab.80).

Tabela 80 - Correlações estatísticas obtidas para o treinamento da Perceptron com 3 Camadas intermediárias para o Esqueleto Aporfínico 01

Ciclos	Treino	Verificação	Teste
0	0,99012	0,96093	0,93469
100	0,99804	0,95391	0,92827
200	0,99879	0,96389	0,93831
300	0,99912	0,96741	0,94335
400	0,99935	0,9716	0,94752
500	0,99949	0,97307	0,94814
600	0,99964	0,97606	0,94818
700	0,99968	0,97736	0,94832
800	0,99975	0,9771	0,94679
900	0,9998	0,97805	0,94605
1000	0,99981	0,9779	0,9671
2000	0,99983	0,97821	0,96642
3000	0,99986	0,97793	0,96654
4000	0,99987	0,98414	0,97289
5000	0,99979	0,98529	0,97592
6000	0,99989	0,98831	0,98819
7000	0,99994	0,98905	0,98838
8000	0,99996	0,98946	0,98854
9000	0,99986	0,99071	0,98313
10000	0,99994	0,9899	0,98029

Nas redes com 4 camadas intermediárias de neurônios, o coeficiente de correlação para os testes praticamente não variou (aproximadamente 0,95) entre 600 e 10000 ciclos (Tab.81). Estes resultados mostraram as redes de 3 camadas são melhores que os da rede de 4 camadas para identificação dos esqueletos aporfínicos 01. Os resultados de ambas as redes foram piores quando o treinamento foi efetuado com um número de ciclos inferior a 1000.

A melhor arquitetura de rede encontrada, para o teste do Esqueleto 01, foi a com 3 camadas intermediárias de neurônios do tipo 22:22-5-1:1. Esta notação indica que a rede possui 22 entradas (deslocamentos químicos), 3 camadas intermediárias contendo 22, 5 e 1 neurônios respectivamente e uma saída (Esqueleto).

Tabela 81 - Correlações estatísticas obtidas para o treinamento da Perceptron com 4 Camadas intermediárias para o Esqueleto Aporfínico 01

Ciclos	Treino	Verificação	Teste
0	0,9683	0,85905	0,82735
100	0,98129	0,95896	0,90365
200	0,98991	0,97612	0,9246
300	0,99591	0,97938	0,94364
400	0,99726	0,98301	0,94617
500	0,99805	0,98418	0,95156
600	0,99859	0,98409	0,94765
700	0,99892	0,98649	0,94397
800	0,99918	0,98801	0,94572
900	0,99937	0,98765	0,94370
1000	0,99951	0,98858	0,9450
2000	0,99959	0,9893	0,94527
3000	0,99976	0,98975	0,94934
4000	0,99979	0,99044	0,95187
5000	0,99988	0,99014	0,94935
6000	0,9999	0,9905	0,94977
7000	0,99994	0,99011	0,94868
8000	0,99997	0,98968	0,94952
9000	0,99998	0,98905	0,94942
10000	0,99999	0,98846	0,95032

Neste treinamento também foi observado a relação entre o número de ciclos e o erro médio (0-1) apresentado para o conjunto de casos, no treino do Esq01, para as redes de 3 e 4 camadas intermediárias (Figs.73 a e 73 b).

(a)

(b)

Figura 73 – (a) Erro médio para rede de 3 camadas. (b) Erro médio para a rede de 4 camadas.

Observa-se que, o erro médio para as redes de 3 camadas foi menor do que o erro para a rede de 4 Camadas.

Também foi observada a relação entre a variação do número de Ciclos e os erros (0-1) apresentados pela Substância do teste **A**, no treino do Esq01 (Fig.74), para as duas redes de 3 e 4 camadas respectivamente.

Figura 74 - Erro individual da substância de Teste A (a) para redes de 3 camadas e (b) para redes de 4 camadas; em função da variação do número de ciclos*.

Pode-se observar que de 100 à 800 Ciclos o gráfico não mostra uma relação direta entre número de ciclos e o Erro individual apresentado na identificação da substância teste **A**. Entretanto, acima de 2000 ciclos o gráfico mostra uma constante redução do erro para ambas as redes de 3 e 4 camadas, sendo que os melhores resultados são também apresentados pela rede de 3 camadas.

Definida a melhor rede para o treinamento do Esqueleto 01 (3 camadas, algoritmo de retropropagação de erro e número de ciclos igual a 8000), seus parâmetros foram fixados e foi avaliada a influência da variação da taxa de aprendizado da rede, verificando sua relação com o coeficiente de correlação. A taxa de aprendizado foi variada de 0,1 a 0,6 (Fig.75).

* Observa-se que as escalas dos gráficos são diferentes para melhor visualização dos valores.

Figura 75 - Variação do coeficiente de correlação em função da taxa de aprendizado

A taxa de aprendizado foi variada de 0,1 a 0,2 com um passo de 0,01 e de 0,2 a 0,8 com um aumentando gradualmente de 0,05.

O coeficiente de correlação não foi muito influenciado pela variação da taxa de aprendizado, a melhor correlação foi obtida com uma taxa de aprendizado igual a 0,6 (0.99836). Quando a taxa de aprendizado ultrapassou este valor, o treinamento do algoritmo parou prematuramente porque seus pesos começaram a divergir de forma muito rápida.

Após a definição dos melhores parâmetros para o treinamento da rede (ciclos, velocidade, algoritmo de aprendizado, número de camadas etc.), foi efetuada a validação da rede utilizando o programa Stat-Soft versão 6. Esta validação é realizada automaticamente, quando na escolha dos parâmetros de treinamento é selecionada a opção *cross-validation* (validação cruzada) a qual realiza o treinamento de forma que, a cada novo ciclo um caso é deixado de fora do treinamento. O erro médio obtido para o treinamento na validação da rede foi de 0,09025.

O mesmo treinamento realizado para o Esqueleto 01, foi aplicado para os Esqueletos 02, 05 e 06. Os resultados mostraram boas correlações para os respectivos testes (Tab.82).

Tabela 82 – Resumo dos resultados obtidos pelo treinamento dos Esqueletos 02, 05 e 06.

Esqueletos utilizados no Treinamento da rede	Condições de treinamento (TA =Taxa de aprendizado, m = momento)	Melhor correlação para o conjunto de teste	Validação da rede (Erro médio)
Treinamento do Esqueleto 02	9000 ciclos; TA=0,5; m=0,3	0.95451	
Treinamento do Esqueleto 05	8000 ciclos; TA=0,6; m=0,3	0.93716	
Treinamento do Esqueleto 06	8000 ciclos; TA=0,6; m=0,3	0.99853	

As redes neurais artificiais, exploradas neste trabalho, identificaram todos os esqueletos das substâncias submetidas aos testes, mostrando resultados de probabilidades altamente confiáveis (acima de 97%), até mesmo nos casos onde as substâncias pertencentes ao mesmo esqueleto possuíam grandes diferenças nos valores de deslocamentos químicos, como ocorreu nos casos dos deslocamentos químicos dos átomos de carbonos C-1 e C-2 para os testes **B** (156,6δ; 176,1δ) e **C** (52,8δ; 23,5δ) respectivamente.

Outro mérito das redes neurais foi que, além de identificar esqueletos semelhantes utilizando deslocamentos químicos diferentes, como observado acima, as redes conseguiram diferenciar esqueletos nos quais a grande maioria dos deslocamentos químicos eram semelhantes, diferenciando-se apenas pela presença ou ausência de um grupo metila, como foi observado nas substâncias teste **A** e **C**.

As redes neurais mostraram ser uma ferramenta simples e eficiente para solucionar problemas de elucidação estrutural utilizando dados de deslocamentos químicos de RMN ^{13}C .

VI – Conclusões

Os resultados dos testes, executados no programa SISTEMAT, mostraram um bom desempenho de seus programas aplicativos, caracterizando-o como uma valiosa ferramenta que pode ser utilizada para auxiliar o químico de produtos naturais no processo de determinação de estruturas e atribuição de sinais de deslocamentos químicos de RMN ^{13}C .

Nos testes realizados com diferentes substâncias, pertencentes a variadas subclasses de alcalóides, o programa SISCONST, além das propostas de esqueletos, forneceu subestruturas coerentes que, sobrepostas ou unidas, conduziram a subestruturas maiores e, algumas vezes, a estruturas completas das substâncias submetidas aos testes, mostrando ser muito útil na tarefa de determinação estrutural.

O SISTEMAT é o único sistema especialista que, através do programa SISOCBOT, correlaciona os dados botânicos de uma determinada planta aos esqueletos das substâncias isoladas da mesma, e utiliza esta correlação como um parâmetro a mais para auxiliar na determinação estrutural do composto isolado.

Os testes realizados no programa MACRONO, demonstraram a eficiência do programa para identificar os substituintes, cujos sinais poderiam trazer problemas, na execução dos programas SISCONST e C13MACH, se não fossem previamente identificados e retirados.

Comparando os resultados dos programas aplicativos SISCONST e C13MACH, pode-se dizer que o programa mais eficiente na tarefa de proposição de esqueletos é o SISCONST. Para fornecer uma proposta de esqueleto ou uma subestrutura, o SISCONST seleciona no banco de dados somente as substâncias que contêm um número mínimo de átomos de carbonos interligados, com os deslocamentos químicos e as multiplicidades desejadas, enquanto que para o C13MACH basta que existam, no banco de dados, átomos de carbonos e hidrogênios com os respectivos deslocamentos e multiplicidades, sem a restrição de que estejam interligados, para que uma proposta de esqueleto seja fornecida.

Os Programas H1MACH e MASMACH, embora tenham apresentado um desempenho regular, deram uma valiosa contribuição, visto que, quando não acertavam a proposta de esqueleto, indicavam um esqueleto semelhante, o que já é muito útil quando tem-se uma substância cuja a estrutura é totalmente ignorada.

Para todos os treinamentos realizados com as redes neurais artificiais foi observado que as redes Perceptron de camadas múltiplas, com três camadas intermediárias de neurônios, mostraram resultados melhores do que as redes de quatro camadas.

A variação de valores para a velocidade de aprendizado e para o ruído tiveram uma influência pequena nos resultados. Entretanto, pôde-se observar que o aumento no número de ciclos melhorou significativamente os resultados fornecidos pela rede.

As redes neurais artificiais mostraram ser uma ferramenta simples e eficiente para resolver problemas de elucidação estrutural fazendo uso de dados de deslocamentos químicos de RMN ^{13}C , até mesmo quando ocorre uma grande similaridade entre os esqueletos pesquisados. Embora, neste trabalho as RNAs tenham sido treinadas para identificar esqueletos de alcalóides aporfínicos, utilizando dados de RMN, tanto a classe das substâncias como o tipo de dados espectrométricos podem ser alterados, desde que sejam feitas pequenas adaptações relativas a entrada dos dados e o treinamento das redes.

Referências Bibliográficas

1. Weiss, S.M. e Kulikowski, C.A.; "Guia Prático para Projetar Sistemas Especialistas", Rio de Janeiro, Livros Técnicos e Científicos Editora S.A., **1988**.
2. Shortliffe, E.H. "Computer-based medical consultations: MYCIN"; New York, Elsevier, **1976**.
3. Duda, R.O.; Hart, P.E.; Konolige, K. e Reboh, R., "A computer based consultant for mineral exploration"; New York, Technical Report, SRI International, **1979**.
4. Lindsay, R.K.; Buchanan, B.G.; Feigenbaum, E.A.; Lederberg, J., "Applications of Artificial Intelligence for Organic Chemistry: the DENDRAL project"; New York, McGraw-Hill, **1980**.
5. a) Sasaki, S.I. e Kudo, Y.J., "Principle for Exhaustive Enumeration of Unique Structures Consistent with Structural Information" *Journal of Chemical Information e Computer Sciences*, **1976**, 16(1), 43-49. b) Sasaki, S.I. e Kudo, Y.J., "Structure elucidation system using structural information from multisources: CHEMICS" *Journal of Chemical Information e Computer Sciences*, **1985**, 25(3), 252-257.
6. Christie, B.D. e Munk, M.E., "The role of two-dimensional nuclear magnetic resonance spectroscopy in computer-enhanced structure elucidation", *Journal of the American Chemistry Society*, **1991**, 113(10), 3750-3757.
7. Bremser, W., Fachinger W. "Multidimensional spectroscopy"; *Magnetic Resonance in Chemistry*, **1985**, 23, 1056-1071.
8. Attias, R., "DARC substructure search system: a new approach to chemical information"; *Journal of Chemical Information e Computer Sciences*, **1983**, 23, 102-108.
9. a) Carrier, G.; Panaye, A.; Dubois, J.E., "Topological structural information in the CAS file: statistical occurrences of DARC concentric fragments. Basic carbon substructures", *Journal of Chemical Information e Computer Sciences*, **1990**, 30, 110-121. b) Frederiksen S.M., Stermitz F.R., "Pyridine monoterpene alkaloid formation from iridoid glycosides. A novel PMTA dimer from geniposide" *Journal of Natural Products*, **1996**, 59, 41-46.
10. a) Bremser, W.; Grzonka, M., "SpecInfo – "A multidimensional spectroscopic

- interpretation system"; *Mikrochimica Acta*, **1991**, *11*, 483-491. b) Neudert, R. ; Penk, M. , "Enhanced structure elucidation"; *Journal of Chemical Information e Computer Sciences*, **1996**, *36*, 244-248 .
11. Elyashberg, M.E.; Martirosian, E.R.; Karasev, Y.Z.; Thiele, H.; Somberg, H., "X-PERT: a user-friendly expert system for molecular structure elucidation by spectral methods"; *Analytica Chimica Acta*, **1997**, *337*, 265-286.
 12. Badertscher, M.; Korytko, A.; Schulz, K.P.; Madison, M.; Munk, M.E.; Portmann, P.; lunghans, M.; Fontana, P.; Pretsch, E., "Assemble 2.0: a structure generator"; *Chemometrics e Intelligent Laboratory Systems*, **2000**, *51*, 73-79 .
 13. Steinbeck, C., "SENECA: A Platform-Independent, Distributed, e Parallel System for Computer-Assisted Structure Elucidation in Organic Chemistry" *Chemical Information e Computer Sciences*, **2001**, *41*, 1500 – 1507.
 14. Corey, E.J., Long A. K.; Rubenstein, S.D., "Computer-assisted analysis in organic synthesis"; *Science*, **1985** *228*, 408-418.
 15. Jorgensen, W.L.; Laird, E.R.; Gushurst, A.J.; Fleischer, J.M.; Gothe, S.A; Helson, H.E.; Paderes, G.D.; Sinclair, S., "CAMEO: a program for the logical prediction of the products of organic reactions"; *Pure & Applied Chemistry*, **1990**, *62*, 1921-1932.
 16. Németh, A.; Vidóczy, T.; Héberger, K.; Kúti, Z.; Wágner, L., "MECHGEN: Computer aided generation e reduction of reaction mechanisms"; *Journal of Chemical Information e Computer Sciences*, **2002**, *42*, 208 -214.
 17. Gastmans, J.P.; Furlan, M.; Lopes, M.N.; Borges, I.H.G.; Emerenciano, V.P., "A inteligência artificial aplicada à química de produtos naturais. O programa SISTEMAT. Parte I - Bases teóricas"; *Química Nova*, **1990**, *13*, 10-16 .
 18. Gastmans, J.P.; Furlan, M.; Lopes, M.N.; Borges, I.H.G.; Emerenciano, V.P., "A inteligência artificial aplicada à química de produtos naturais. O programa SISTEMAT. Parte II - Organização do programa e aplicativos"; *Química Nova*, **1990**, *13*, 75-80.
 19. Emerenciano, V.P.; Rodrigues, G.V.; Gastmans, J.P., "Um método de codificação semi-automático assistido por computador para o SISTEMAT"; *Química Nova*, **1993**, *16*, 431- 434.

20. Ferreira, M.J.P.; Rodrigues, G.V.; Emerenciano, V.P., "MONOREG - An expert system for structural elucidation of monoterpenes"; *Canadian Journal of Chemistry*, 2001, 79, 1915 - 1925.
21. Roberts, M. F.; Wink, M., "Alkaloids : Biochemistry Ecology e Medicinal Applications", Ed. Plenum; New York, **1998**.
22. Correia, C.R.D., "Síntese Estereoseletiva de alcalóides e N-Heterocíclicos" – Ed. UFSCar, **2001**.
23. Cordell, G.A., "Introduction to Alkaloids: A Biogenetic Approach", Ed. John-Wiley & Sons: New York, **1981**.
24. Manske, R.H.F., "The Alkaloids Chemistry e Biology", Vol.1, Academic Press: New York, **1950**.
25. Waterman, P.G., "The Alkaloids", Cordell, G.A. Ed.; Vol 50, Academic Press: New York; **1998**, pp 537-565.
26. Cordell GA, Quinn-Beattie ML, Farnsworth NR "The potential of alkaloids in drug discovery", *Phytotherapy Research* **2001**, 15, 183-205.
27. **NAPRALERTsm** – Um acrônimo para **NA**tural, **PR**oducts **ALERT** – Banco de dados de produtos naturais.
28. Mann, J. "Chemical Aspects of Biosynthesis" – Oxford University Press Inc., New York, **1996**.
29. Kubo, I.; Lee, Y.M.; Pettei, M.; Pilkiewicz, F. e Nakanishi, K. "Potent army worm antifeedants from the east african warburgia plants"; *Journal of the Chemical Society - Chemical Communications*, **1976**, 1013-1014.
30. Bremser, W. "HOSE - A novel substructure code"; *Analytica Chimica Acta*, **1978**, 103, 355-365.
31. Will, M.; Fachinger, W. e Richert, I.R. "Fully automated structure elucidation – A spectroscopist's dream comes true"; *Journal of Chemical Information e Computer Sciences*, **1996**, 36, 221-227.
32. Pestchanker, M.J.; Ascheri, M.S. e Giordano, O.S., "Uspallatine, a pyrrolizidine

- alkaloid from *Senecio uspallatensis*"; *Phytochemistry*, **1985**, 24, 1622-1624.
33. Steinbeck, C. , "Lucy – A Program for Structure Elucidation from NMR Correlation Experiments", *Angew .Chem.Int. Ed. Engl*, **1996**, 35, 1984-1986.
 34. Kirkpatrick, S.; Gerlatt, C. D. J.;Vecchi, M. P., "Optimization by Simulated Annealing". *Science*, **1983**, 220, 671-680.
 35. Diaconis, P. e Ram,A., " Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques", *Technical report*, Stanford Univ., Stanford, C.A. **2000**.
 36. Isaza, J.H.; Yoshida,T., " Monochaetin, From *Monochaetum multiflorum*" *Chem. Pharm. Bull.* **1999**, 47(10), 1510-1511.
 37. Ferreira, M.J.P., "Um sistema especialista em determinação estrutural de monoterpenos e iridóides", *Tese de Mestrado*, IQ-USP, **1999**.
 38. Rodrigues, G.V., "Um sistema especialista em determinação estrutural de sesquiterpenos lactonizados. Um enfoque multiespectral aliado a dados botânicos", *Tese de Doutorado*, IQ-USP, **1996**.
 39. Oliveira, F.C.; Ferreira, M.J.P.; Núñez, C.V.; Rodrigues,G.V. e Emerenciano, V.P., ¹³C NMR spectroscopy of eudesmane sesquiterpenes"; *Progress in Nuclear Magnetic Resonance Spectroscopy*, **2000**, 37, 1-45.
 40. Macari, P.A.T.; Gastmans, I.P.; Rodrigues, G.V. e Emerenciano, V.P., "An expert system for structure elucidation of triterpenes"; *Spectroscopy -1n International Journal*, **1994**, 12, 139-166.
 41. Ferreira, M.J.P., "Esteróides: Análise espectral, geração de estruturas e simulação de dados de RMN ¹³C", *Tese de Doutorado*, IQ-USP, **2003**.
 42. Emerenciano, V.P.; Melo,L.D.; Rodrigues,G.V. e Gastmans, I.P., "Application of Artificial Intelligence in Organic Chemistry. Part XIX. Pattern recognition e structural determination of flavonoids using 13C-NMR spectra"; *Spectroscopy –* **1997**, 13, 181.
 43. Roeder, E., " Carbon-13 NMR spectroscopy of Pyrrolizidine alkaloids" *Phytochemistry*, **1990**, 29(1), 11-29.
 44. Lounasmaa, M. e Hanhinen, P., " Formation of 3-iso-19-epicathenamine, isolated as its cyano adduct 21 α -cyano-3-isorauniticine", *Heterocycles*, **1997**, 45(8), 1529-1536.
 45. Haykin, S., "Redes neurais - Princípios e prática", ARTMED Editora S.A., **2001**, Pg.59.

46. McCulloch, W.S. e Pitts, W., "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, **1943**, 5, 115-133.
47. Hebb, D. O., *The Organization of Behavior : A neuropsychological Theory*, Wiley & Sons, New York, **1949**.
48. Rosenblatt, F. "The Perceptron: A Probabilistic Model for Information Storage e Organization in the Brain", *Psychol. Rev.*, **1958**, 65, 386-408.
49. Widrow, B. e Hoff, M.E., "Adaptative Switching Circuits" *Institute of Radio Enginners. Western Electronic Show e Convention*, **1960**.
50. Minsky, M. e Papert, S., *Perceptrons: an Introduction to Computational Geometry*. MIT Press, Massachusetts, **1969**.
51. Sycara, K., "Using Case-Based Reasoning for plan Adaptation e Repair". *In Case Based Reasoning Workshop*, DARPA, Morgan Kaufmann, **1988**, 425-434.
52. Hopfield, J. J., "Neural Networks e Physical Systems with Emergent Collective Properties" . *Proc. Nat. Acad. Sci.*, **1982**, 79, 2554-2558.
53. Rumelhart, D.E. e McClelland, J.L., "*Parallel Distributed Processing*" , Vol 1:Foundations, MIT Press, **1986**.
54. Rumelhart, D. E.; . Hinton, G. E. e Williams, R. J., "Learning Representations by Back-propagation Error", *Nature*, **1986**, 323, 533-536.
55. Marr, D., *Vision*, New York : Freeman e Company, **1982**.
56. Levine, M., *Man e Machine Vision*, New York : Mc-Graaw-Hill, **1985**.
57. Churchland, P.S. e Sejnowski, T.J., *The Computational Brain*, Cambridge, MA : MIT Press, **1992**.
58. Suga, N., "Cortical computational maps for auditory imaging" , *Neural Networks*, vol.3, pp3-21, 1990.
59. Suga, N., "Computations of velocity e range in the bat auditory system for echo location", in *Computational Neuroscience*, E.L. Schwartz, ed., pp 213-231, Cambridge, MA : MIT Press, **1990**.
60. Braga, A.P.; Carvalho, A.L.F.; Ludermir, T.B.; "Redes Neurais Artificiais : Teoria e Aplicações. Ed. Livros Técnicos Científicos S.A., Rio de Janeiro , **2000**.
61. Mendel, J. M. e McLaren, R. W. "Adaptive, Learning, e Pattern Recognition Systems; Theory e Applications", chapter Reinforcement-learning control e pattern

recogniton systems, pp 287-318. New York: Academic Press, 1970.

62. Fukushima, K., "Cognitron: A Self-Organising Multilayered Neural Network", *Biological Cybernetics*, **1975**, 23, 121-134.
63. Kohonen, T., "Self-organized Formation of Topologically Correct Feature Maps", *Biological Cybernetics*, **1982**, 43, 59-69.
64. Grossberg, S., "Adaptive Pattern Recognition e Universal Recoding: Parallel Development e Coding of Neural Feature Detectors". *Biological Cybernetics*, **1976**, 24, 121-134.
65. Rumelhart, D.E. e Zipser, D., "Feature discovery by competitive learning", *Cognitive Science*, **1985**, 9, 75-112.
66. Minsky, M. L. "Steps Towards Artificial Intelligence". *In Proc. of the Institute of Radio Engineers*, **1961**, 49, 8-32.
67. Hampson, S. E. "*Connectionistic Problem Solving: Computational Aspects of Biological Learning*", Berlin: Birkhauser, **1990**.
68. Thorndike, E. L., "*Animal Intelligence*". Darien, **1911**.
69. Sutton, S.; Barto, A. G. e Willians, R. J. "Reinforcement Learning is Direct Adaptive Optimal Control". *In Proc. of the American Control Coference*, **1991**, 2143-2146.
70. Haykin, S., "*Neural Networks – A comprehensive Foundation*", Prentice Hall, New Jersey, **1994**.
71. G. Cybenko. "Approximation by Superpositons of a Sigmoid Function", *Mathematics of Control, Signals e Systems*, **1989**, 2, 303-314.
72. Welstead, S.T., "Neural Network e Fuzzy Logic Applications in C/C++", Wiley & Sons, New York, **1994**.
73. Wasserman, P.D., "Advanced Methods in Neural Computing, Van Nostre Reinhold, New York, **1993**.
74. Zupan, J.; Gasteiger, "J. *Neural Networks for Chemists : An Introduction*", VHC-Verlag, Weinheim, 1993.
75. Penza M, Cassano G "Recognition of adulteration of Italian wines by thin-film multisensor array e artificial neural networks", *Analytica Chimica Acta*, **2004**, 509, 159-

177.

76. Haddadnia J, Faez K, Ahmadi M.; "A fuzzy hybrid learning algorithm for radial basis function neural network with application in human face recognition" *Pattern Recognition*, **2003**, 36, 1187-1202.
77. Turkoglu, I.; Arslan, A.; Ilkay, E., "An intelligent system for diagnosis of the heart valve diseases with wavelet packet neural networks", *Computers in Biology e Medicine*, **2003**, 33, 319-331.
78. Aracena, J.; Goles, E., "Complexity of perceptron recognition for a class of geometric patterns", *Theoretical Computer Science*, **2003**, 299, 65-79.
79. Narendra, K.S. e Parthasarathy, K., " Identification e control of dynamic systems using neural networks. *IEEE Trans on Neural Networks*, **1992**, 1(1), 4-27.
80. Sciavicco, L. e Siciliano, B., "Modeling and control of robot manipulators", McGraw-Hill, **1996**.
81. Kodogiannis, V.; Lolis, A., "Forecasting financial time series using neural network e fuzzy system-based techniques" *Neural Computing Applications*, **2002**, 11, 90-102.
82. Rufino, A.R.; Santos, J.B.O.; Brant, A.J.C. e Emerenciano, V.P., "Simple Method for Identification of Skeletons of Aporphine Alkaloids from ^{13}C NMR Data Using Artificial Neural Networks", *Journal of Chemical Information e Computer Sciences*, **2004**, *In press*.
83. Haykin, S., "Redes neurais - Princípios e prática", ARTMED Editora S.A., **2001**, Pg.458.
84. Kohonen, T., " The neural phonetic typewriter", *Computer*, **1988**, 21(3), 11-22.
85. Hopfield, J.J. e Tank, D.W., " Neural computation of decision in optimization problems", *Biological Cybernetics*, **1985**, 52, 141-152.
86. Lacerda, E. e Carvalho, A., " Redes de base radial para avaliação de risco de crédito. *V Simpósio Brasileiro de Redes Neurais*, pp 259-264, Belo Horizonte – MG, Dezembro, **1998**.
87. Davis, G.E.; Lowell, M.D., *A Neural Network that predicts Psychiatric Length of Stay*. Ed. By D. e G.L. Davis, MD Computing, NY, Springer-Verlag, 1993.
88. Kallergi, M., "Computer-aided diagnosis of mammographic microcalcification clusters", *Medical Physics*, **2004**, 31(2), 314-326.

89. Crawford, E. D., "Use of algorithms as determinants for individual patient decision making: National comprehensive cancer network versus artificial neural networks", *Urology*, **2003**, 62(6A), 13-19.
90. Fraser, L.A. e Mulholle, D.A., *Fresen J. Anal. Chem.* **1999**, 365 (7), 631-634.
91. Elyashberg, M.E., "Expert systems for the determination of structures of organic molecules by spectral methods", *Uspekhi Khimii*, **1999**, 68 (7), 579-604.
92. Kvasnicka, V.; Sklenak, S. e Pospichal, J., "Application of recurrent neural network in chemistry: Prediction e classification of ^{13}C NMR chemical shifts in a series of monosubstituted benzenes", *Journal of Chemical Information e Computer Sciences*, **1992**, 32, 742-747.
93. Ivanciuc, O., "Artificial neural networks applications - Part 6. Use of nonbonded van der Waals e electrostatic intermolecular energies in the estimation of ^{13}C -NMR chemical shifts in saturated hydrocarbons", *Rev. Roum. Chim.*, **1995**, 40, 1093-1101.
94. Svozil, D.; Pospichal, J. e Kvasnicka, V., "Neural network prediction of carbon-13 NMR chemical shifts of alkanes", *Journal of Chemical Information e Computer Sciences*, **1995**, 35, 924-928 .
95. Thomas, S. e Kleinpeter, E., "Assignment of the ^{13}C NMR chemical shifts of substituted naphthalenes from charge density with an artificial neural network", *J. Prakt. Chem. Chem.-Ztg.*, **1995**, 337, 504-507.
96. Ivanciuc, O.; Rabine, J. P.; Cabrol-Bass, D.; Panaye, A. e Doucet, J.-P. ^{13}C NMR chemical shift prediction of sp^2 carbon atoms in acyclic alkenes using neural networks, *Journal of Chemical Information e Computer Sciences*, **1996**, 36, 644-653.
97. Li, Z.; Huang, Y.; Hu, F.; Sheng, Q. e Peng, S., "Neural networks in spectroscopy: Estimation e prediction of chemical shifts of ^{13}C NMR in alkanes by using subgraphs", *Bopuxue Zazhi*, **1997**, 14, 507-514.
98. Meiler, J.; Meusinger, R. e Will, M., "Neural network prediction of ^{13}C NMR chemical shifts of substituted benzenes", *Monats. Chem.*, **1999**, 130, 1089-1095.
99. Militão, J. S. L. T.; Emerenciano, V. P.; Ferreira, M. J. P., "Structure validation in computer-supported structure elucidation: C-13 NMR shift predictions for steroids", *Chemometrics e Intelligent Laboratory Systems*, **2003**, 67(1), 5-20.
100. Meiler, J. e Sanli, E. "Validation of structural proposals by substructure analysis e

- C-13 NMR chemical shift prediction" *Journal of Chemical Information e Computer Sciences*, **2002**, 42(2), 241-248.
- 101.** Meiler, J.; Will, M. e Meusinger, R., "Fast determination of ¹³C-NMR chemical shifts using artificial neural networks, *Journal of Chemical Information e Computer Sciences*, **2000**, 40, 1169-1176.
- 102.** Meiler, J.; Will, M. e Meier, W., "Using Neural Networks for ¹³C-NMR Chemical Shifts Prediction – Comparison with Traditional Methods", *J. Mag. Resson.*, **2002**, 157, 242-252.
- 103.** Clerc, J.T. e Sommerauer, H., " A minicomputer program based on additivity rules for the estimation of ¹³C NMR chemical shifts, *Anal. Chim. Acta*, **1977**, 95, 33-40.
- 104.** Meiler, J., "Software applications for chemistry, biochemistry e bioinformatics", available at www.jens-meiler.de ,1996-2002.
- 105.** Guinaudeau, H.; Leboeuf, M.; Cave, A., "Aporphinoid alkaloids 4", *Journal of Natural Products*, **1988**, 51, 389-474.
- 106.** Statistica Neural Networks, Inc. Tulsa, OK, USA, 2000.

Anexo I

Tabela A – Planilha com a codificação dos alcalóides aporfinicos, com seus respectivos deslocamentos químicos de RMN ¹³C.

	ESQ01	ESQ02	ESQ03	ESQ04	ESQ05	ESQ06	ESQ07	ESQ08	ESQ09	ESQ10	ESQ11	ESQ12	C1	C2	C3	C4	C5	C6	C7	C8
APF001	1	0	0	0	0	0	0	0	0	0	0	0	43.7	28.8	115.5	150.9	144.4	127.6	54.8	130
APF002	1	0	0	0	0	0	0	0	0	0	0	0	43.5	29.2	116.4	149.6	147	126.1	55.3	130.3
APF003	1	0	0	0	0	0	0	0	0	0	0	0	42.8	29.2	108.3	146.9	142.6	112.3	54	129.4
APF004	1	0	0	0	0	0	0	0	0	0	0	0	42.9	28.9	107.9	146.7	142.2	111.9	53.8	129.6
APF005	1	0	0	0	0	0	0	0	0	0	0	0	42.7	24.2	140.9	134.8	144	108.1	54.1	119.1
APF006	1	0	0	0	0	0	0	0	0	0	0	0	157.1	178.3	118.4	152.8	154.1	125.3	131.8	126.4
APF007	0	0	0	1	0	0	0	0	0	0	0	0	60.9	32.5	114.4	150.4	145.2	121.2	120.9	129.5
APF008	1	0	0	0	0	0	0	0	0	0	0	0	43.1	29.1	107.1	146.6	141.6	116.2	53.5	126.5
APF009	1	0	0	0	0	0	0	0	0	0	0	0	42.8	28.7	113.8	151.6	148.2	124.6	54.1	130
APF010	0	0	0	1	0	0	0	0	0	0	0	0	61	31.9	110.5	145	142.1	117.1	122.7	130.8
APF011	0	0	0	0	0	0	0	1	0	0	0	0	49.3	29.6	118.5	148.3	143.2	124.4	120.7	130.2
APF012	0	0	0	0	1	0	0	0	0	0	0	0	45.4	19.5	148.5	143.8	143.5	117.6	170.6	120.3
APF013	0	1	0	0	0	0	0	0	0	0	0	0	156.5	176	112.9	153.2	154.9	124.7	132.7	124.8
APF014	0	1	0	0	0	0	0	0	0	0	0	0	156.7	175	108.6	148.2	151.4	125.7	132.4	123
APF015	1	0	0	0	0	0	0	0	0	0	0	0	156.9	178	113.2	153.4	154.9	125.8	131.5	124.8
APF016	0	0	0	0	0	0	0	0	0	0	1	0	139.6	119	152.7	146.2	147.2	109.7	162.7	135.2
APF017	0	0	0	0	0	0	0	0	1	0	0	0	148.1	123.2	118.9	147.1	?	150.7	147.5	138.3
APF018	1	0	0	0	0	0	0	0	0	0	0	0	43.8	29.5	117.4	151.7	146.8	127.8	54.8	130.5
APF019	1	0	0	0	0	0	0	0	0	0	0	0	44	29.5	111.8	145.9	141.5	116	55	129.6
APF020	1	0	0	0	0	0	0	0	0	0	0	0	41.3	21	148.8	145.9	148.8	109.2	53.2	118.5
APF021	1	0	0	0	0	0	0	0	0	0	0	0	42.5	23.8	150.8	144.7	147	120.9	54.1	124.5
APF022	1	0	0	0	0	0	0	0	0	0	0	0	42.5	23.4	144.5	135.2	140.3	108.1	53.7	118.2
APF023	1	0	0	0	0	0	0	0	0	0	0	0	42.7	23.5	140.8	135.2	144.1	106.7	54	118.7
APF024	0	1	0	0	0	0	0	0	0	0	0	0	52.8	23.4	110.7	149	149.3	130.5	62.8	123.5
APF025	0	0	0	0	1	0	0	0	0	0	0	0	45.6	19.7	143.6	140.4	138.6	115.3	170.6	124.2
APF026	1	0	0	0	0	0	0	0	0	0	0	0	157.2	179	117	159.3	117.1	133.2	132.1	131.2
APF027	0	1	0	0	0	0	0	0	0	0	0	0	156.7	176.8	114.6	159.6	115.7	133.1	133.1	130.2
APF028	0	1	0	0	0	0	0	0	0	0	0	0	38.5	30.5	112.9	151.8	146	125.7	128.4	128.2
APF029	1	0	0	0	0	0	0	0	0	0	0	0	41.2	30.7	111.9	142.4	141.1	117.8	145.2	135.5
APF030	0	1	0	0	0	0	0	0	0	0	0	0	50	24.7	147.8	150.7	146.6	120.3	141.7	121.9
APF032	0	0	0	1	0	0	0	0	0	0	0	0	60.2	31.2	106.6	146.6	155.7	124.3	120.8	131.5
APF033	0	0	0	1	0	0	0	0	0	0	0	0	61.2	31.7	111.4	145.5	143.1	117.9	119.3	130.6
APF035	0	0	0	0	0	0	0	1	0	0	0	0	148	118.2	149.9	126.8	?	143.2	147.2	131.8
APF036	0	0	0	0	0	0	0	0	1	0	0	0	128.7	152.9	114.1	146.5	?	150.2	140.5	132.6
APF037	0	0	0	0	0	0	0	0	0	1	0	0	157.6	176.5	119.2	148.5	?	117.9	122.5	131.8
APF038	0	1	0	0	0	0	0	0	0	0	0	0	52.4	28.4	110.5	148.4	141.4	119.1	62	123
APF039	0	1	0	0	0	0	0	0	0	0	0	0	53.2	23.4	147.3	144.3	138.3	118.5	62.8	115.7
APF040	1	0	0	0	0	0	0	0	0	0	0	0	43.2	28.9	110.1	145.9	141.5	119.1	53.8	123.9
APF041	0	1	0	0	0	0	0	0	0	0	0	0	53.3	28.8	109.5	145.7	141.4	119.3	58.6	123.8
APF042	0	1	0	0	0	0	0	0	0	0	0	0	53.1	28.9	111.3	152	145.1	126.8	62.5	128.4

Tabela A - Continuação

C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	N-ME 1	N-ME 2	N-Et(CH2)	N-Et(CH3)
APF001	126	36.7	130.3	124.8	112.8	147.8	147.2	?	?	?	?	?	?
APF002	129.8	38.7	129.1	125.1	149.9	144.9	116	122.9	?	?	?	?	?
APF003	126.4	37.4	129.1	122.8	148.5	145.6	114.5	123.4	?	?	?	?	?
APF004	126.2	37.1	128.8	113.3	147	143.8	107.1	120.1	?	?	?	?	?
APF005	130.8	37.6	128.9	122.5	148.3	144.5	113.6	123.4	?	?	?	?	?
APF006	118.4	112.3	133.6	127.2	128.1	126.9	128.2	128.9	?	?	?	?	?
APF007	124.7	125.2	125.8	133.3	106.4	147.7	147	105.3	?	45.4	45.4	?	?
APF008	127	36.4	127.9	123.5	110.5	147.5	148.1	111	?	?	?	?	?
APF009	127.9	37.8	129.3	123.3	148.2	145.5	113.8	122.4	?	?	?	?	?
APF010	126	125.1	138	128.8	127.3	126.2	126.7	127.6	?	45.3	45.3	?	?
APF011	124.9	124.4	128.7	123.2	109.1	148.2	146.8	112	?	?	?	?	?
APF012	119.4	73.1	144.7	119.4	128.7	114.7	156.4	111.4	?	?	?	?	?
APF013	119.9	114.3	133.3	127.2	128.1	127.7	128.4	129.6	?	30.2	?	?	?
APF014	121.3	114.3	133.1	128.3	128.5	127	127.4	129	?	30.1	?	?	?
APF015	119.5	113.2	133.5	127.2	128.3	127.2	128.3	129	?	?	?	?	?
APF016	121	129	135	180	116	145.9	115.2	127.9	?	?	?	?	?
APF017	119.3	181.5	132	135	125.1	134.4	131.1	128.1	?	?	?	?	?
APF018	131.4	37.8	137.4	133.2	129.2	127.9	128.9	128.9	?	?	?	?	?
APF019	127.4	37.1	137.1	123.9	129	115.8	159.2	112.8	?	?	?	?	?
APF020	134.5	34.7	139.6	123.2	128.2	114.4	155.5	114.1	?	?	?	?	?
APF021	133.2	38.3	137.7	119.7	153.7	118.6	128.9	120.1	?	?	?	?	?
APF022	129.5	37.5	137.4	119.7	156	110.6	128	120.8	?	?	?	?	?
APF023	129.6	36.9	128.9	113.5	147.2	143.7	106.9	120.4	?	?	?	?	?
APF024	129.9	33.7	123	122.8	114.2	144.7	145.2	145.5	?	40.2	?	?	?
APF025	115.3	73.1	143.2	119.1	128.3	114.7	156.5	111.9	?	?	?	?	?
APF026	116.4	110.6	133.5	127.2	123	126.3	128.4	128.7	?	?	?	?	?
APF027	117.8	112.4	133.3	127.2	123.6	127.1	128.7	129.4	?	30.1	?	?	?
APF028	118.9	112.1	132.4	133	127.9	126.2	127.3	127.8	?	160.8	?	?	?
APF029	119.5	101.9	127.6	127.6	128.5	105.8	158.5	106.6	?	?	?	?	?
APF030	121.3	102.9	129.3	118.5	108.4	146.2	148.8	106.7	?	40.7	?	?	?
APF032	125.6	121.5	126.6	131.7	115.7	119.6	126.9	151.3	?	44.6	44.6	?	?
APF033	126.8	119.7	122.7	130.4	121.5	127.2	111.2	153.6	?	45.2	45.2	?	?
APF035	119.7	182	131.5	125.7	124.6	134.6	130.2	128.5	?	?	?	?	?
APF036	119.8	180.9	130.2	135.5	125.2	134.1	131.1	128.1	?	?	?	?	?
APF037	145.3	144.6	130.4	130.1	124.9	128.3	130.1	122.7	?	35.6	?	?	?
APF038	127	35.6	139.2	121.5	153.3	111.2	127.8	121.8	?	43.4	?	?	?
APF039	125.2	34.3	130.5	124.6	110.9	147.1	147.1	111.6	?	44	?	?	?
APF040	129.1	36.1	128.7	133.3	114.1	159.2	112.4	128.1	?	?	?	?	?
APF041	128.2	34	127.9	133.1	114	158.2	112.4	128.2	?	43.8	?	?	?
APF042	127.5	34.8	127.8	132.8	113.7	158.5	113.1	128.4	?	43.6	?	?	?

Tabela A - Continuação

	ESQ01	ESQ02	ESQ03	ESQ04	ESQ05	ESQ06	ESQ07	ESQ08	ESQ09	ESQ10	ESQ11	ESQ12	C1	C2	C3	C4	C5	C6	C7	C8
APF043	1	0	0	0	0	0	0	0	0	0	0	0	43.1	29.2	110.7	151.9	144.1	126.6	53.8	128.2
APF044	0	0	0	1	0	0	0	0	0	0	0	0	60.7	32.2	114.8	162.5	150.8	125.2	128.1	130.1
APF045	0	0	0	0	0	0	0	1	0	0	0	0	59.4	32.5	116	150.5	145.5	124.2	127.3	129.3
APF046	1	0	0	0	0	0	0	0	0	0	0	0	43	29	111.8	152	145	126	53.7	128.6
APF047	0	1	0	0	0	0	0	0	0	0	0	0	53.6	29.1	108.9	145.9	140.6	119.8	62.7	127.7
APF048	1	0	0	0	0	0	0	0	0	0	0	0	42.8	23.2	148.4	144.5	138.7	118.2	54	115.5
APF049	0	1	0	0	0	0	0	0	0	0	0	0	52.9	23.3	145.5	138.5	148	118.8	62.8	116.3
APF050	1	0	0	0	0	0	0	0	0	0	0	0	42.7	23.7	149.2	149.9	145.1	125.3	53.9	128.1
APF051	0	1	0	0	0	0	0	0	0	0	0	0	52.9	23.6	149.5	149.8	145.1	122.6	62.7	125.6
APF052	1	0	0	0	0	0	0	0	0	0	0	0	41.6	31.1	111.8	151.4	145.9	118.6	141.6	125
APF053	1	0	0	0	0	0	0	0	0	0	0	0	41.1	30.7	107.5	145.2	141.9	117.3	141.8	118.2
APF054	1	0	0	0	0	0	0	0	0	0	0	0	41.3	31.3	110.7	150.8	144.5	118.5	140.4	125.8
APF055	0	0	1	0	0	0	0	0	0	0	0	0	55.3	36	116.6	145.9	144.2	126.3	139.7	132.7
APF056	0	0	1	0	0	0	0	0	0	0	0	0	49.6	24.9	111.8	149.8	144.6	124.1	140.7	129.7
APF057	0	0	1	0	0	0	0	0	0	0	0	0	44.8	30.3	112.2	151.3	145.9	117.9	130.8	124.9
APF058	0	0	0	0	1	0	0	0	0	0	0	0	46.3	26.4	106.2	150.5	141.8	115.4	169.6	132.8
APF059	0	0	0	0	1	0	0	0	0	0	0	0	42.5	29.6	107	146.8	141.9	114.5	61.4	128
APF060	1	0	0	0	0	0	0	0	0	0	0	0	43.2	29.2	111.8	152.2	145.2	126.6	53.5	129
APF061	1	0	0	0	0	0	0	0	0	0	0	0	43.6	29.6	108	146.8	142.5	116.3	53.6	128.7
APF062	0	1	0	0	0	0	0	0	0	0	0	0	53.5	28.9	109.6	146.6	142.6	123.6	61.7	126.3
APF063	0	1	0	0	0	0	0	0	0	0	0	0	53	23.8	150.1	145.2	150.2	122.6	62.6	122.9
APF064	1	0	0	0	0	0	0	0	0	0	0	0	42.8	29	106.7	146.1	141	115.9	53.2	127
APF065	1	0	0	0	0	0	0	0	0	0	0	0	41.5	27.4	105.3	144.9	140	114.4	51.4	125.5
APF066	0	1	0	0	0	0	0	0	0	0	0	0	52.9	28.7	106.8	146.7	142.8	113.6	62.4	125.4
APF067	1	0	0	0	0	0	0	0	0	0	0	0	42.4	22.7	149.5	138.9	145.5	115.5	53.5	117.9
APF068	1	0	0	0	0	0	0	0	0	0	0	0	42.4	23.1	150.6	145.5	150.2	122.2	53.7	122.4
APF069	1	0	0	0	0	0	0	0	0	0	0	0	43.3	24.1	140.6	135.2	144.1	110.8	53.7	119.4
APF070	1	0	0	0	0	0	0	0	0	0	0	0	42.9	23.5	140.5	135.2	143.9	110.6	53.5	119.1
APF071	1	0	0	0	0	0	0	0	0	0	0	0	145	123.4	106.4	152	145.3	119.6	156.7	135.3
APF072	1	0	0	0	0	0	0	0	0	0	0	0	140	129	105.6	155.3	163.2	124.5	161.4	146.2
APF073	1	0	0	0	0	0	0	0	0	0	0	0	144.5	119	148.5	147.4	145.6	115.6	156.5	131
APF074	1	0	0	0	0	0	0	0	0	0	0	0	145	123.2	106.4	151	145.5	119.8	156.7	135.3
APF075	1	0	0	0	0	0	0	0	0	0	0	0	142.2	128.2	105.7	166.1	163	124.3	161.3	145.5
APF076	1	0	0	0	0	0	0	0	0	0	0	0	144	123.3	102.1	151.9	147.3	104.4	144.3	135.2
APF077	0	1	0	0	0	0	0	0	0	0	0	0	53.5	29.1	107.9	147.3	142	115.5	67.2	128.8
APF078	0	1	0	0	0	0	0	0	0	0	0	0	50.3	31.1	111.4	151.1	145.8	125.8	143.2	129.3
APF079	0	1	0	0	0	0	0	0	0	0	0	0	50.4	31.2	110.6	150.7	144.6	125.8	142.5	129.7
APF080	0	1	0	0	0	0	0	0	0	0	0	0	50.7	31	106.6	144.6	141.2	117.3	142.8	127.5
APF081	0	1	0	0	0	0	0	0	0	0	0	0	34.7	30.3	107.7	147.1	143.2	117.3	53.3	128.1
APF082	1	0	0	0	0	0	0	0	0	0	0	0	42.9	23.7	146.6	138.7	148.8	118.6	54.1	116.3

Tabela A - Continuação

	C10	C11	C12	C13	C14	C15	C16	C17	C18	N-ME 1	N-ME 2	N-Er(CH2)	N-Er(CH3)
APF043	127.9	129.7	123.8	114	144.9	145.2	111.3	?	?	?	?	?	?
APF044	126.1	132.7	132.7	122.3	126.5	126.5	125.7	?	?	45.2	45.2	?	?
APF045	125.4	129.3	132.4	121.6	126.7	126.7	125.7	?	?	49.1	?	?	?
APF046	127.6	128.2	132.9	113.5	158.4	113.2	128.2	?	?	?	?	?	?
APF047	124.3	129.9	124.1	112.1	145.1	144.7	114	?	?	43.9	?	?	?
APF048	127.6	130.8	124.6	110.8	147	147.2	111.7	?	?	?	?	?	?
APF049	131.1	128.4	124.7	111.1	147.5	147.4	111	?	?	44	?	?	?
APF050	129.5	131.1	122.6	108.3	146.2	145.8	107.9	?	?	?	?	?	?
APF051	130.1	130.8	122.6	108.4	146.5	146	108	?	?	43.6	?	?	?
APF052	129.7	134.8	125.9	126.9	122.7	128	125.7	?	?	?	?	?	?
APF053	127.5	133.9	123.5	126.9	122.1	127	125	?	?	?	?	?	?
APF054	129.9	130.1	117.8	109.1	145.7	149.2	105.7	?	?	?	?	?	?
APF055	127.3	133.1	126.3	126.9	127	127.1	131.9	47.8	?	152.1	?	?	?
APF056	121.7	120.8	129.6	121.1	147.1	148.8	104.6	14.1	?	42.5	?	?	?
APF057	128.1	106.7	137.2	128.4	127.2	119.9	123.4	66.9	?	81.1	?	?	?
APF058	116.3	72.7	144.9	119.6	113.7	160.1	109.5	34.2	?	?	?	?	?
APF059	124.4	74.7	145.9	121.4	112.9	159.7	108.7	22.2	?	?	?	?	?
APF060	129.1	37.5	136.6	132.3	127	127.8	128.4	?	?	?	?	?	?
APF061	129.1	37.4	135.4	131.4	127	127.5	128.1	?	?	?	?	?	?
APF062	126.3	25.8	116.5	132	127	107.4	156.1	?	?	43.7	?	?	?
APF063	132	34.8	135.9	131.5	127	127.7	127.8	?	?	44	?	?	?
APF064	127.7	36.7	137.3	121.9	115.1	157	113.7	?	?	?	?	?	?
APF065	124.8	35.6	135.2	122.2	126.6	111.9	110.5	?	?	?	?	?	?
APF066	128	35.3	138.4	120	156.2	120	128.4	?	?	43.6	?	?	?
APF067	131.1	36.4	134.7	132.1	126.5	127.6	127.6	?	?	?	?	?	?
APF068	131.1	36.5	135.1	131.9	127.7	127.7	126.9	?	?	?	?	?	?
APF069	129.7	37.3	134.8	131.5	126.5	126.8	126.9	?	?	?	?	?	?
APF070	128.1	37.1	136.1	121	127.7	113.5	158.6	?	?	?	?	?	?
APF071	122	182.5	132	134.7	128.3	134.2	128.7	?	?	?	?	?	?
APF072	139.6	179.9	129.8	135.2	130.4	135.4	131.2	?	?	?	?	?	?
APF073	122.8	182.5	131.5	134.6	128.9	134.3	127.7	?	?	?	?	?	?
APF074	121.7	181.2	126.9	129.1	110.3	153.8	109.9	?	?	?	?	?	?
APF075	136.8	177.9	118	137.8	123.6	135.1	115.4	?	?	?	?	?	?
APF076	121.6	182	133.2	123.1	156	116.1	129.1	?	?	?	?	?	?
APF077	122.4	66.2	134.8	130.5	127.9	130	129.5	?	?	43.5	?	?	?
APF078	119.2	110.7	135	124.5	122.6	126.4	127.7	?	?	40.3	?	?	?
APF079	118.2	101.4	130.6	118.6	109.4	145.9	106.7	?	?	40.2	?	?	?
APF080	117.1	100.7	129.5	118.5	108.7	146	106.4	?	?	40.3	?	?	?
APF081	125.4	38.9	135.4	130.8	126.9	127	127.7	?	?	157.4	?	?	?
APF082	133.3	37.5	135.5	132.5	126.9	127.4	127.8	?	?	?	?	?	?

Tabela A - Continuação

	ESQ01	ESQ02	ESQ03	ESQ04	ESQ05	ESQ06	ESQ07	ESQ08	ESQ09	ESQ10	ESQ11	ESQ12	C1	C2	C3	C4	C5	C6	C7	C8
APF083	0	1	0	0	0	0	0	0	0	0	0	0	41.9	30.5	107.2	147	143.5	113.6	49.1	125.3
APF084	1	0	0	0	0	0	0	0	0	0	0	0	41.9	22.1	160.4	144.7	154.7	122.4	53	116.9
APF085	1	0	0	0	0	0	0	0	0	0	0	0	42.4	28.9	106.9	145.3	138.6	113.9	53	127.4
APF086	1	0	0	0	0	0	0	0	0	0	0	0	42.7	28.6	104.9	146.9	142.1	113.7	53.6	125.4
APF087	1	0	0	0	0	0	0	0	0	0	0	0	42.3	23.9	142.8	134.8	139.1	107.6	53.7	119
APF088	0	1	0	0	0	0	0	0	0	0	0	0	41.4	23.5	141.3	134.5	138.7	108.9	48.9	119.3
APF089	0	0	0	0	1	0	0	0	0	0	0	0	44.6	26	106	151.8	143.4	114.3	172.6	133.4
APF090	0	0	0	0	1	0	0	0	0	0	0	0	42.5	29.6	107	146.8	141.9	114.5	61.4	128
APF092	0	0	0	0	1	0	0	0	0	0	0	0	44.9	19.1	146.8	142.8	143.1	116.3	171.4	121.8
APF093	0	0	0	0	1	0	0	0	0	0	0	0	45.4	19.7	147.1	143.4	145.2	116.4	172	122.7
APF094	0	0	0	0	1	0	0	0	0	0	0	0	45.6	26.4	105.4	151.5	142.3	116.5	170.3	133.1
APF095	0	0	0	0	1	0	0	0	0	0	0	0	45.8	26.4	105.6	151.2	142.4	115.6	169.7	132.8
APF096	0	0	0	0	1	0	0	0	0	0	0	0	47	26.3	106	141.4	141.4	116.6	170.3	132.9
APF097	0	0	0	0	0	0	1	0	0	0	0	0	140.2	113.9	146.6	142.5	142.6	111.5	163	127.3
APF098	0	1	0	0	0	0	0	0	0	0	0	0	50.4	31.1	110.8	150.6	145	125.3	142.4	130
APF099	0	0	1	0	0	0	0	0	0	0	0	0	44.8	29.9	107.8	145.4	142.1	117.5	130.6	123.5
APF100	0	0	1	0	0	0	0	0	0	0	0	0	44.9	30	105.7	145.5	142.4	112.6	132.3	125.9
APF101	0	1	0	0	0	0	0	0	0	0	0	0	53.1	29.1	110.1	151.5	143.9	126.5	62.3	128.1
APF102	0	1	0	0	0	0	0	0	0	0	0	0	52.4	29.1	110.8	150.8	141.7	125.4	62.6	129.8
APF103	0	1	0	0	0	0	0	0	0	0	0	0	53.1	29.2	110.8	150.7	142.1	125.4	62.7	129.9
APF104	0	1	0	0	0	0	0	0	0	0	0	0	53.3	28.8	113.2	147.9	141.9	126.6	62.5	129.7
APF105	0	0	0	0	0	1	0	0	0	0	0	0	61.4	24.3	110.2	153.9	145.6	127.7	70.3	124.1
APF106	0	0	0	0	0	1	0	0	0	0	0	0	61.5	24	109.8	153.6	145.9	127.6	69.9	124.4
APF107	0	0	0	0	0	1	0	0	0	0	0	0	61.5	23.5	108	148	142.2	118.2	69.7	119.8
APF108	0	0	0	0	0	1	0	0	0	0	0	0	61.3	24.2	110.1	153.9	145.6	127.7	70.4	123.9
APF109	0	0	0	0	0	1	0	0	0	0	0	0	61.4	24.2	110	153.9	144.5	128	70.3	124
APF110	0	0	0	0	0	1	0	0	0	0	0	0	61.8	24	106.4	148.5	143.1	116.7	69.8	121.8
APF111	0	0	0	0	0	1	0	0	0	0	0	0	61.4	23.4	113.6	151.1	143.9	127.1	69.7	124.1
APF112	0	0	0	0	0	1	0	0	0	0	0	0	61.5	19.3	138.6	135.3	144.7	110.7	69.5	114.3
APF113	0	0	0	0	0	1	0	0	0	0	0	0	60.3	23.8	110.6	152.9	143	126	69.1	125.2
APF114	0	0	0	0	0	1	0	0	0	0	0	0	60.8	23.4	111	153	146.9	124.5	69.3	122.9
APF115	0	0	0	0	0	1	0	0	0	0	0	0	61	23.9	110.6	150.9	143.6	119.2	69.8	119.6
APF116	0	0	0	0	0	1	0	0	0	0	0	0	61.5	23.9	109.6	148.8	140.2	118.9	69.7	120.3
APF117	0	1	0	0	0	0	0	0	0	0	0	0	53.2	23.6	139.1	134.8	143.2	110.4	62.3	119.1
APF118	1	0	0	0	0	0	0	0	0	0	0	0	44.1	25.1	140.1	134.5	144.5	109.1	54.9	118.4
APF119	1	0	0	0	0	0	0	0	0	0	0	0	145	119.7	136.2	138.2	150.7	124.4	146.3	131.5
APF120	1	0	0	0	0	0	0	0	0	0	0	0	42.6	27.3	109.7	145.8	141.8	119.7	53.9	122.9
APF121	0	1	0	0	0	0	0	0	0	0	0	0	53.9	29.6	107.4	147.1	142.7	116.9	62.5	127.3
APF122	0	1	0	0	0	0	0	0	0	0	0	0	53.5	29.6	116.1	151.2	146.1	127.2	62.9	129.6
APF123	0	1	0	0	0	0	0	0	0	0	0	0	53.9	29.7	106.9	147.1	142.4	117.6	62.6	127.3
APF124	0	1	0	0	0	0	0	0	0	0	0	0	53.7	29.5	116.6	150.9	145	126.9	63	130.1
APF125	1	0	0	0	0	0	0	0	0	0	0	0	169	140.9	111.5	135.5	130.2	115.3	162	111.7

Tabela A - Continuação

	C10	C11	C12	C13	C14	C15	C16	C17	C18	N-ME 1	N-ME 2	N-E(CH2)	N-E(CH3)
APF083	125.7	34.5	138.2	120.1	156.4	110.6	128.9	120.6	?	161.9	?	?	?
APF084	130	36.1	136	123.7	128.7	113.6	160.8	112.5	?	?	?	?	?
APF085	128.8	38.2	138.7	110.7	154.5	101.9	160.1	106.9	?	?	?	?	?
APF086	128.1	38.1	138.8	112.9	157.6	97.7	160.3	106.7	?	?	?	?	?
APF087	130.4	38.2	139.1	111	155	100.7	159.3	104.9	?	?	?	?	?
APF088	125.7	34.9	138.1	110.3	153.9	102.2	159.8	107.2	?	162.1	?	?	?
APF089	117.1	73.1	141.7	117.9	129	114.7	157.9	111.3	?	?	?	?	?
APF090	124.4	74.7	145.9	121.4	128.5	112.9	159.7	108.8	?	?	?	?	?
APF092	116.7	72.9	145.2	119.9	129.3	113.9	156.3	110.6	?	?	?	?	?
APF093	117.4	73.1	145.2	121.1	129.5	113.6	159.5	109.5	?	?	?	?	?
APF094	114.8	73.2	144.6	108.6	158.6	99	158	104.2	?	?	?	?	?
APF095	113.6	72.5	145.2	106.6	157.6	98.4	161.3	101.1	?	?	?	?	?
APF096	117	42.3	146.5	121.2	128.8	114.4	159.6	110.9	?	?	?	?	?
APF097	119.4	42.1	144.6	121.3	129.6	113.5	156	112.5	?	?	?	?	?
APF098	118.8	101.6	129.4	123.6	112.3	142	147.2	106	?	40.3	?	?	?
APF099	130	106	137.3	129.8	127.1	119.5	127.4	122.9	?	80.8	?	?	?
APF100	132.3	105.7	137.5	119.6	157.7	111.5	128	107.6	?	80.6	?	?	?
APF101	127	34.4	129.1	124.2	111.4	147.1	147.7	110.6	?	43.4	?	?	?
APF102	128.8	35.6	129.6	119.8	143.6	149	110.7	118.6	?	43.6	?	?	?
APF103	128.8	35.7	130.4	125.1	108.4	145.9	146	107.8	?	43.5	?	?	?
APF104	125.8	34.1	130.1	123.5	110.1	145.4	144.9	114.1	?	42.9	?	?	?
APF105	118.6	29.5	124	123	111.4	148.2	148.8	111.4	?	44	54.8	?	?
APF106	118.4	28.9	123.9	122	111.4	146.5	145.9	114.5	?	43.6	54.3	?	?
APF107	118.3	28.6	123.7	122.5	112.1	146.2	145.5	114.3	?	42.6	53.4	?	?
APF108	118.5	29.6	123.9	122.8	111.6	148.2	148.4	112.3	?	43.8	54.7	?	?
APF109	118.6	29.5	123.8	123.1	111.8	148.1	148.1	112.3	?	43.5	54.7	?	?
APF110	117.9	28.7	123.3	122	110.2	148.2	148.8	111.6	?	43.6	54.3	?	?
APF111	117.1	28.7	123.8	121.8	110.9	146.6	145.8	114.5	?	43.2	53.7	?	?
APF112	118.7	28.6	122.6	121.9	109.5	148	148	111.6	?	43.6	54.4	?	?
APF113	118.3	30.6	124.3	120.2	143.5	149.7	111.5	119.6	?	42.9	53.5	?	?
APF114	121	30.4	125.2	123.8	147.1	152.4	112.3	122.6	?	43.1	53.9	?	?
APF115	119.2	30.7	124.7	124.6	143.6	152.5	111.8	125.3	?	44	54.8	?	?
APF116	117.7	30.3	123.8	119.2	140.2	147.6	110.9	120.8	?	43.4	54.2	?	?
APF117	127.4	34.1	127.1	123.5	110	147.5	147.5	111.1	?	43.5	?	?	?
APF118	130.6	38.1	135.6	133.1	127.1	127.7	127	128.9	?	?	?	?	?
APF119	123.3	182.5	125.7	133.6	129.4	122.6	160.2	111.4	?	?	?	?	?
APF120	125.1	35.1	128.4	124.3	113.5	147.5	146.1	114.8	?	?	?	?	?
APF121	127.3	27.5	130.3	126.3	123.8	115.4	151.2	147.1	?	44	?	?	?
APF122	129.9	35.4	137.3	132.9	128.8	127.3	127.8	128.6	?	44	?	?	?
APF123	127.6	27.8	130.4	124.4	124.4	115.8	151.3	145.8	?	44.1	?	?	?
APF124	127.5	35.6	137.3	133.2	128.3	127.4	127.7	128.7	?	44.1	?	?	?
APF125	131	180.5	134.3	134.8	128.3	134.6	128.8	127.4	?	?	?	?	?

Tabela A - Continuação

	ESQ01	ESQ02	ESQ03	ESQ04	ESQ05	ESQ06	ESQ07	ESQ08	ESQ09	ESQ10	ESQ11	ESQ12	C1	C2	C3	C4	C5	C6	C7	C8
APF124	0	1	0	0	0	0	0	0	0	0	0	0	53.7	29.5	116.6	150.9	145	126.9	63	130.1
APF125	1	0	0	0	0	0	0	0	0	0	0	0	169	140.9	111.5	135.5	130.2	115.3	162	111.7
APF126	1	0	0	0	0	0	0	0	0	0	0	0	164.8	136.8	110.2	135.3	130.8	100.2	149.4	119.3
APF127	0	1	0	0	0	0	0	0	0	0	0	0	46.6	31.1	135.5	149.4	151.3	120.1	57	130.1
APF128	0	1	0	0	0	0	0	0	0	0	0	0	50.8	34.4	150.8	148.2	152.3	125.1	57.5	133.2
APF129	0	0	0	0	0	0	1	0	0	0	0	0	46.1	26.3	105.4	150.1	141.7	114.1	171.2	132.9
APF130	0	0	0	0	0	0	0	0	0	0	0	1	41.5	23.6	143.6	135.1	137.9	108.4	49.9	118.4
APF131	0	1	0	0	0	0	0	0	0	0	0	0	60	66.5	111.8	152.1	143.6	125.8	63.1	132
APF132	0	1	0	0	0	0	0	0	0	0	0	0	53.5	29	108.7	145.8	140.7	119.5	62.7	123.9
APF133	0	1	0	0	0	0	0	0	0	0	0	0	53.3	28.7	113.5	148.2	142.3	126.3	62.5	129.6
APF134	0	1	0	0	0	0	0	0	0	0	0	0	52.9	28.4	109.2	146.5	140.6	119.7	62.4	126.7
APF135	0	1	0	0	0	0	0	0	0	0	0	0	53.3	28.8	109.7	145.8	140.7	119.5	62.5	123.6
APF136	0	1	0	0	0	0	0	0	0	0	0	0	53.6	29.2	106.1	146.6	141.7	116.6	62.4	126.6
APF137	0	1	0	0	0	0	0	0	0	0	0	0	49.8	23.2	106.3	146.5	141.6	116.3	64.3	126.9
APF138	0	1	0	0	0	0	0	0	0	0	0	0	52	25.3	106.5	146.4	141.4	116	?	127.4
APF139	0	1	0	0	0	0	0	0	0	0	0	0	52.5	28.6	111.8	151.4	144.3	125.7	61.9	127.5
APF140	0	1	0	0	0	0	0	0	0	0	0	0	48.9	22.8	107.4	146.8	142.4	116.5	64.6	127.6
APF141	1	0	0	0	0	0	0	0	0	0	0	0	42.7	29.1	107.9	141.8	141.8	114.8	60.4	127.2
APF142	0	1	0	0	0	0	0	0	0	0	0	0	49.3	17.2	139.5	134.9	143.9	110.7	64.2	119.3
APF143	0	1	0	0	0	0	0	0	0	0	0	0	53	29.2	127.9	126.5	126.2	132.6	62.5	131.5
APF144	0	1	0	0	0	0	0	0	0	0	0	0	52.2	27.8	110.3	150.6	134.9	127.4	61.8	130.6
APF145	0	1	0	0	0	0	0	0	0	0	0	0	52.9	28.8	122	143.1	147.1	129.1	62.4	129.8
APF146	1	0	0	0	0	0	0	0	0	0	0	0	42.7	28.4	110.9	146.5	141.6	119.7	53.2	127.8
APF147	1	0	0	0	0	0	0	0	0	0	0	0	42.6	28.5	115.7	148.9	143.2	125.1	53.2	126.8
APF148	0	1	0	0	0	0	0	0	0	0	0	0	52.9	28.4	110.3	146.5	141.6	119.2	62.1	127.4
APF149	0	0	0	0	0	1	0	0	0	0	0	0	61.5	23.5	108	148	142.2	118.2	69.7	119.3
APF150	0	1	0	0	0	0	0	0	0	0	0	0	52.8	28.4	109.2	146.2	140.8	119.4	62.4	126.9
APF151	1	0	0	0	0	0	0	0	0	0	0	0	40.8	24.8	111.2	152.8	144.3	125.7	51.9	126.4
APF152	0	0	0	0	0	1	0	0	0	0	0	0	61.4	24.3	110.2	153.9	145.6	127.7	70.3	124.1
APF153	0	1	0	0	0	0	0	0	0	0	0	0	60.6	56.2	114.4	149	143.8	125.8	62.8	131.8
APF154	0	1	0	0	0	0	0	0	0	0	0	0	52.9	28.7	106.8	146	141.6	116	61.7	126.7
APF155	0	1	0	0	0	0	0	0	0	0	0	0	36.1	29.6	111	152.5	145.2	127.4	53.6	129.4
APF156	0	0	0	0	0	0	0	0	0	0	0	1	41.9	30.7	111.2	152	145.8	127.3	50.3	126.6
APF157	0	0	0	0	0	0	0	0	0	0	0	1	42	29	107.5	147	143.2	118	50.5	128.8
APF158	0	1	0	0	0	0	0	0	0	0	0	0	41.9	30.8	107.3	146.9	143.1	117.1	49.2	126.4
APF159	0	1	0	0	0	0	0	0	0	0	0	0	42	31	111.5	152.4	146.1	127.9	49.4	125.3
APF160	0	1	0	0	0	0	0	0	0	0	0	0	67.2	27.3	107.2	148.9	143.9	116.2	76.4	123
APF161	0	1	0	0	0	0	0	0	0	0	0	0	42.1	30.9	110.8	152.3	145.3	127.8	49.5	128.4
APF162	0	0	0	0	0	0	0	1	0	0	0	0	59.4	32.5	116	150.5	145.5	124.2	127.3	129.3
APF163	0	1	0	0	0	0	0	0	0	0	0	0	53	29.2	107.3	146.8	143.1	113.9	62.9	125.7
APF164	0	1	0	0	0	0	0	0	0	0	0	0	53.9	29.5	108.1	148.5	144.3	115.3	64.4	126.9
APF165	0	1	0	0	0	0	0	0	0	0	0	0	54	29	115.7	150.1	145.5	125.6	64.6	128.8

C09	C10	C11	C12	C13	C14	C15	C16	C17	C18	N-ME 1	N-ME 2	N-EK(CH2)	N-EK(CH3)
APF126	132.3	179	131.8	136.6	128.3	133.9	125.9	123.8	?	?	?	?	?
APF127	124.7	38.2	130.2	129.5	113.7	142	144.1	111	?	153.4	?	?	?
APF128	131.7	39.2	131.1	130.9	113.5	144.2	145.4	112.3	?	154.7	?	?	?
APF129	118.7	43.1	146.6	111	157.9	96.7	161	102.7	26.9	?	?	?	?
APF130	128.2	34.9	139.5	115.5	147.6	107.4	159.1	111.9	?	169	?	22.4	?
APF131	129.4	35.3	128.8	119.8	143.9	149.6	111.1	119.3	?	43.5	?	?	?
APF132	127.2	34.5	128.9	124.8	112	147.1	147.6	110.9	?	44	?	?	?
APF133	125.9	34.2	129.2	124.1	110	147.6	148.1	110.7	?	43.8	?	?	?
APF134	123.5	33.7	129.1	123	113.6	145.3	145.4	114.9	?	43.6	?	?	?
APF135	127.2	34.9	130.2	125.8	108.7	145.8	145.8	108.2	?	43.9	?	?	?
APF136	126.4	34.3	128.3	123.4	111.2	147.6	148.2	110.5	?	44	?	?	?
APF137	122.5	70	141.3	121.4	127.8	112.5	159.1	109	?	39.5	?	?	?
APF138	123.5	81.5	139.4	122.2	128	112.2	159	109.3	?	40.7	?	?	?
APF139	128.6	34.3	136.2	131.5	126.6	127.1	127.5	127.9	?	43.6	?	?	?
APF140	123.4	69.8	138.8	128.6	127	127.8	127.6	123.8	?	40.6	?	?	?
APF141	124.7	83.2	136.4	129.6	126.7	127.4	127.4	123.1	?	?	?	?	?
APF142	124.1	69.7	138.7	128.7	125.7	126.9	126.9	123.6	?	39	?	?	?
APF143	134.9	34.7	129.8	127.8	147	152	111.4	123.3	?	44	?	?	?
APF144	127.4	33.5	129.6	123.4	111.3	147.7	148.7	110.9	?	42.3	?	?	?
APF145	133.9	33.5	129.8	130	112.4	149.8	138.8	122	?	43.9	?	?	?
APF146	123.5	36.8	135.7	132.4	125.9	126.2	128.1	128.1	?	?	?	?	?
APF147	129.3	36.9	136.3	132.1	126.5	127.2	127.5	127.7	?	?	?	?	?
APF148	123.5	34.4	135.7	132.4	126	126.2	127.5	128.1	?	43.6	?	?	?
APF149	118.3	28.6	123.7	122.5	112.1	146.2	145.5	114.3	?	42.6	53.4	?	?
APF150	122.7	33.9	126.9	124.7	111.3	143.8	145.7	116	?	43.6	?	?	?
APF151	120.8	32.3	126.2	123.1	111.6	147.3	148.3	111.6	?	?	?	?	?
APF152	118.6	29.5	124	123	111.4	148.2	148.8	111.4	?	44	54.8	?	?
APF153	125.8	33.7	128.7	123.9	111.4	148.1	148.7	110.9	?	43.5	?	?	?
APF154	126.4	25.8	115.6	125.8	102.4	150.8	135.9	146.8	?	43.5	?	?	?
APF155	125.1	38	129.8	124	109.1	146.9	146.8	108.4	?	162	?	?	?
APF156	128.8	34	136.7	131.5	128.3	126.9	127.7	128.6	?	169.7	?	22.6	?
APF157	126	33.5	136	130.5	127	127.5	128	128.5	?	167	?	22.5	?
APF158	124.4	33.4	134.9	130.2	126.9	126.9	127.7	128.6	?	161.8	?	?	?
APF159	128.6	34.1	136.1	131.5	128.4	127.1	127.8	128.6	?	162.1	?	?	?
APF160	119	68.8	137.2	127.3	126.3	128.8	127.6	124.7	?	48.9	?	?	?
APF161	124.9	34.1	130.6	124.5	108.8	146.7	146.6	108.8	?	162.2	?	?	?
APF162	125.4	128.2	129.3	132.4	121.6	126.7	126.7	125.7	?	49.1	?	?	?
APF163	128.7	35.3	130.2	124.1	146.8	152	111.5	122.2	?	44	?	?	?
APF164	129.2	35.8	129.6	124.6	146.6	150.5	116.5	123.9	?	43.9	?	?	?
APF165	128.4	36.1	125.9	125.6	147	150.2	116	123.5	?	44	?	?	?