

Artigo

Metabolômica Baseada em RMN de ^1H e no Método dos Mínimos Quadrados Parciais Discriminante**Baddini, A. L. Q.; Netto, A. D. P.; Teixeira, V. L.****Rev. Virtual Quim.*, 2014, 6 (2), 396-415. Data de publicação na Web: 31 de janeiro de 2014<http://www.uff.br/rvq>**Metabolomics Based on ^1H NMR and Partial Least Squares-Discriminant Analysis**

Abstract: The metabolomic techniques are powerful tools to explain and predict phenotypes in biological systems. In this area, chemometric methods are employed in order to extract information from complex analytical data. Among the instrumental methods that allow obtaining analytical and chemometric methods, proton nuclear magnetic resonance (^1H NMR) and partial least squares discriminant analysis (PLS-DA) are the most used. This paper describes some practical issues in pretreatment ^1H NMR spectra and the use of PLS-DA with and application for differentiation of two species of brown algae, *Dictyota ciliolata* e *D. crenulata*.

Keywords: Metabolomics; PLS-DA; ^1H NMR; chemotaxonomy; brown algae; *Dictyota*.

Resumo

As técnicas de metabolômica são ferramentas poderosas para explicar e prever fenótipos em sistemas biológicos. Nesta área, os métodos quimiométricos são empregados para extrair informações de dados analíticos complexos. Dentre os métodos instrumentais para obtenção dos dados analíticos e os métodos quimiométricos, os mais utilizados são ressonância magnética nuclear de hidrogênio (RMN de ^1H) e o método dos mínimos quadrados parciais discriminante (*partial least squares-discriminant analysis*, PLS-DA). O presente trabalho descreve alguns aspectos práticos do processamento de espectros de RMN de ^1H e da utilização de PLS-DA com aplicação para diferenciação de duas espécies de algas pardas, *Dictyota ciliolata* e *D. crenulata*.

Palavras-chave: Metabolômica; PLS-DA; RMN de ^1H ; quimiotaxonomia; algas pardas; *Dictyota*.

* Universidade Federal Fluminense, Instituto de Biologia, Departamento de Biologia Marinha, ALGAMAR Laboratório de Produtos Naturais de Algas Marinhas, CEP 24020-141, Niterói-RJ, Brasil.

✉ valerialaneuville@gmail.com e valerialaneuville@id.uff.br

DOI: [10.5935/1984-6835.20140028](https://doi.org/10.5935/1984-6835.20140028)

Metabolômica Baseada em RMN de ^1H e no Método dos Mínimos Quadrados Parciais Discriminante

Ana Luísa Q. Baddini,^a Annibal D. Pereira Netto,^b Valéria L. Teixeira^{c,*}

^a Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro, Rua Senador Furtado 171, Maracanã, CEP 20270-021, Rio de Janeiro-RJ, Brasil.

^b Universidade Federal Fluminense, Instituto de Química, Departamento de Química Analítica, LAQFA Laboratório de Química Analítica Fundamental e Aplicada, CEP 24020-141, Niterói-RJ, Brasil.

^c Universidade Federal Fluminense, Instituto de Biologia, Departamento de Biologia Marinha, ALGAMAR Laboratório de Produtos Naturais de Algas Marinhas, CEP 24020-141, Niterói-RJ, Brasil.

* valerialaneuville@gmail.com e valerialaneuville@id.uff.br

Recebido em 17 de agosto de 2013. Aceito para publicação em 8 de janeiro de 2014

1. Introdução
2. Processamento dos espectros de RMN de ^1H e extração dos dados
3. Método dos mínimos quadrados parciais discriminante
4. Exemplo de aplicação: Seleção dos sinais representativos nos espectros de RMN de ^1H para diferenciação das espécies de algas *Dictyota crenulata* e *D. ciliolata*
 - 4.1. Material e métodos
 - 4.2. Resultados e discussão
5. Conclusões

1. Introdução

O termo metabolômica surgiu no final da década de noventa do século passado e refere-se ao estudo qualitativo e quantitativo de metabólitos em sistemas biológicos para relacionar variações dos metabólitos com mudanças genéticas ou perturbações nas condições ambientais.¹⁻⁴ Nesta área, são utilizados métodos instrumentais sofisticados para gerar os dados e métodos quimiométricos para extrair informações

destes.

Independente dos métodos instrumentais e quimiométricos, os trabalhos de metabolômica incluem as seguintes etapas: identificação da questão biológica a ser estudada, planejamento experimental, amostragem, preparação das amostras, aquisição, processamento e análise dos dados (Figura 1).

A identificação da questão biológica a ser estudada é o ponto de partida para um trabalho de metabolômica, seguida de um

planejamento experimental, que é uma etapa imprescindível em qualquer trabalho científico. A amostragem deve ser realizada para que as amostras representem significativamente o sistema biológico a ser estudado. O preparo da amostra é uma etapa que depende da natureza das amostras e tipo de análise que será feita.

Os métodos mais utilizados na aquisição dos dados de metabolômica são baseados em ressonância nuclear magnética (RMN) e cromatografia com espectrômetro de massas. Dependendo do tipo de dados gerados, as estratégias de processamento dos dados são diferentes. Neste trabalho serão abordados o processamento dos dados de ressonância magnética nuclear de hidrogênio unidimensional (1D), mostrando a importância e os cuidados que devem ser tomados nesta etapa.

As análises dos dados de metabolômica podem ser realizadas com métodos de reconhecimento de padrões não supervisionados e métodos supervisionados. O que distingue os métodos supervisionados dos não supervisionados é o conhecimento prévio de grupos. Neste artigo será apresentado o método dos mínimos quadrados parciais discriminante (*Partial Least Squares-Discriminant Analysis*, PLS-DA) que é o método de reconhecimento de padrões supervisionado mais utilizado. Também serão discutidos os aspectos práticos da aplicação do programa The Unscrambler® 9.5 com esta finalidade.

O trabalho apresenta, ainda, um exemplo de aplicação da metabolômica de algas pardas de duas espécies, sendo elas: *Dictyota ciliolata* e *Dictyota crenulata*. Estas duas espécies apresentam margens denticuladas e muitas vezes são de difícil distinção. Quatro são as características morfológicas que distinguem-as:

- Iridescência: *D. ciliolata* apresenta cor amarelo esverdeada e padrão de bandas transversais, enquanto *D. crenulata* apresenta ápices azul-esverdeados, bandas largas e pálidas transversais e manchas escuras quadradas ao longo de todo o talo;⁵

- Formato dos ápices: *D. ciliolata* apresenta ápice redondo a agudo, enquanto *D. crenulata* apresenta ápice redondo a obtuso;⁵

- Denticulação: *D. ciliolata* apresenta uma denticulação muito pequena a ciliada; *D. crenulata* apresenta denticulação triangular a crenulata;⁵

- Arranjo dos esporângios: *D. ciliolata* apresenta esporângios solitários; *D. crenulata* pode apresentar esporângios solitários ou agrupados.⁵

As Figuras 2 e 3 apresentam exemplos de espécimes de *Dictyota crenulata* e *D. ciliolata*, mostrando a variedade de formas e colorações que podem ser encontrados.

A distribuição da *D. crenulata* no litoral do Brasil, do Rio Grande do Norte ao Rio de Janeiro com ausência de registro de táxon nos litorais da Paraíba e Sergipe, proposta por Horta (2000), pode estar subestimada pela dificuldade taxonômica que este táxon apresenta. Crispino (2000) declara não reconhecer as duas espécies para o estado do Espírito Santo, apenas espécimes típicos de *D. crenulata*. Wysor & De Clerk (2003) reportaram que amostras do Panamá inicialmente identificadas como *D. ciliolata*, na verdade devem corresponder a uma forma de crescimento de *D. crenulata* com denticulações esparsas.

Tronholm *et al.* (2013) reiteraram que as diferenças morfológicas entre as duas espécies são sutis e fazem um estudo com sequenciamento de DNA para confirmar a distribuição geográfica destas espécies. No Brasil, as duas espécies são reconhecidas morfológica e quimicamente.^{6,10}

Muitos trabalhos sobre diversas atividades biológicas dos diterpenos produzidos por algas do gênero *Dictyota* já foram publicados.¹¹⁻¹⁹ Dentre esses estudos, podemos destacar a atividade antiviral contra o vírus da imunodeficiência humana adquirida tipo 1, o HIV-1, e o herpes simples humano (HSV-1) de diterpenos das espécies *D. menstrualis* e *D. pfaffi*.¹¹⁻¹⁹ Há grande potencial para descobertas de outras

atividades biológicas, sendo as macroalgas químicas de produtos naturais mais estudadas no Brasil em relação à

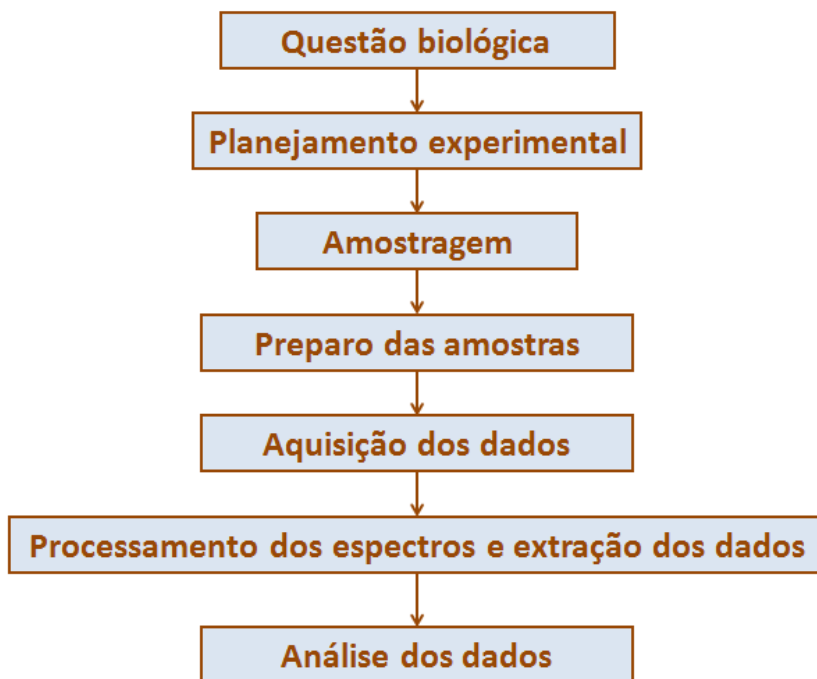


Figura 1. Esquematização das etapas gerais de um trabalho de metabolômica



Dictyota crenulata



Dictyota crenulata

Figura 2. Espécimes de *Dictyota crenulata*. Fotos obtidas em <http://www.botany.hawaii.edu/ReefAlgae/Algae%20in%20Hawaii%20pages/Dictyota%20in%20Hawaii,%20page.htm> e <http://biogeodb.stri.si.edu/bioinformatics/dfm/metas/view/13179>



Figura 3. Espécimes de *Dictyota ciliolata*. Fotos obtidas em http://www.seaweedafrica.org/search/images/view/?img_id=34473 e http://biogeodb.stri.si.edu/bocas_database/search/species/3533

2. Processamento dos espectros de RMN de ^1H e extração dos dados

A primeira etapa do processamento dos dados consiste em importar o arquivo de dados gerado pelo espectrômetro de RMN para o programa computacional. Neste estudo, seis programas de processamento de espectros foram avaliados, sendo eles: Top Spin[®] (Bruker Biospin, Alemanha), Vnmrj[®] (Agilent, EUA), MestReNova[®] (Mestrelab Reserch, Espanha), Automatics¹⁹, NMR Suite[®] (Chenomx, Canadá) e Metabolomic²⁰. Os dois primeiros programas são comercializados por fabricantes de duas marcas de espectrômetros de RMN, tendo, portanto, interface com os formatos de arquivos gerados por seus equipamentos. O MestReNova[®] e o NMR Suite[®] são programas comerciais que foram construídos para tratar dados gerados por qualquer espectrômetro de RMN. Os formatos de arquivo que podem ser abertos ou convertidos para serem

utilizados por estes programas estão listados na Tabela 1.

O Automatics¹⁹ e o Metabolomic²⁰ são programas acadêmicos construídos para atender e tratar dados gerados de várias marcas de espectrômetros, ambos processam arquivos no formato .fid da Bruker, sendo necessária a utilização de um programa conversor dos vários formatos para o formato .fid. O programa Automatics¹⁹ importa dados no formato XWIN-NMR e o Metabolomic²⁰ importa os dados espectrais em arquivo de texto.

A espectroscopia de ressonância magnética nuclear pulsada gera um sinal com decaimento livre de indução no domínio do tempo (*Free Induction Decay*, FID) – Figura 4A. Como este tipo de sinal não é interpretável pelo olho humano, é necessário transformá-lo para o domínio de frequência através da transformada de Fourier. A Figura 4B apresenta um espectro de RMN de ^1H gerado no domínio da frequência.

Todos os programas computacionais de

processamento de espectros avaliados convertem o FID automaticamente ou manualmente em espectros de domínio de frequência, além disso, fazem correção de fase (Figura 5A) e ajustam a linha base em zero (Figura 5B).

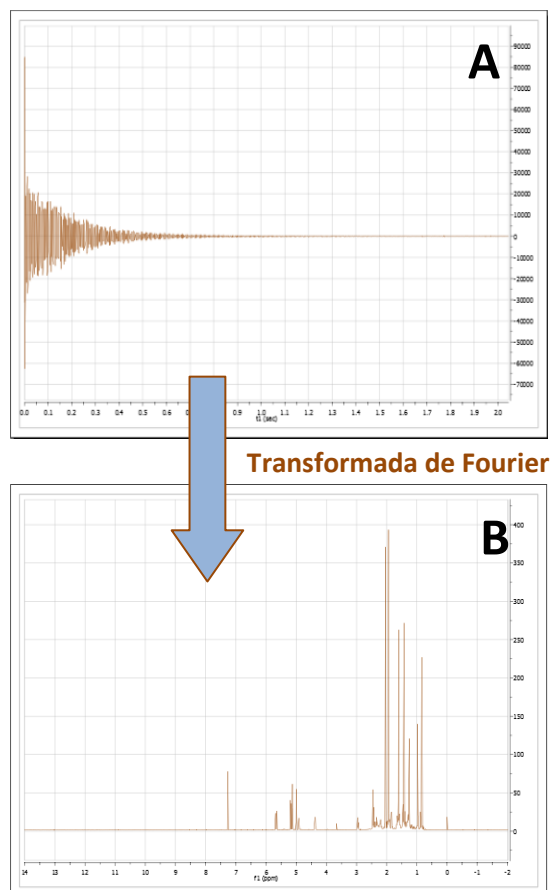


Figura 4. Exemplo de um espectro no domínio de tempo (A); Espectro no domínio de frequência após o cálculo da transformada de Fourier (B)

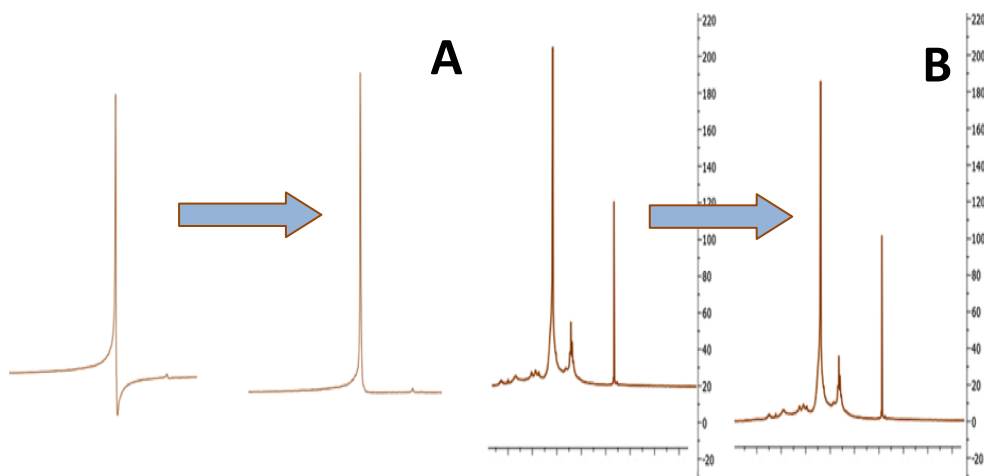


Figura 5. Espectros de RMN de ^1H : correção de fase (A); ajuste da linha base em zero (B)

Tabela 1. Programas usados para o processamento dos espectros de RMN de ^1H avaliados e tipos de arquivos para a importação

Programa	Tipo de arquivo para a importação
Top Spin 3.2 (Bruker Biospin, Alemanha)	.fid ou .ser
VnmrJ 3.2 (Agilent, EUA)	Arquivos da Agilent
MestRenova 6.0 e 8.0 (Mestrelab, Espanha)	.fid, .ser (Bruker) .fid (Varian) .jdf (JEOL) .jdx (JCAMP)
Automics ¹⁹	.fid (Bruker) XWIN-NMR Programa conversor para .fid (Bruker)
NMR Suite 7.6 (Chenomx, Canadá)	Programa conversor de vários formatos (Agilent, Bruker, Jeol, JCAMP-DX, Mestrelab e NMRpipe) em Chenomx
Metabolomic ²⁰	Arquivo de texto .fid (Bruker)

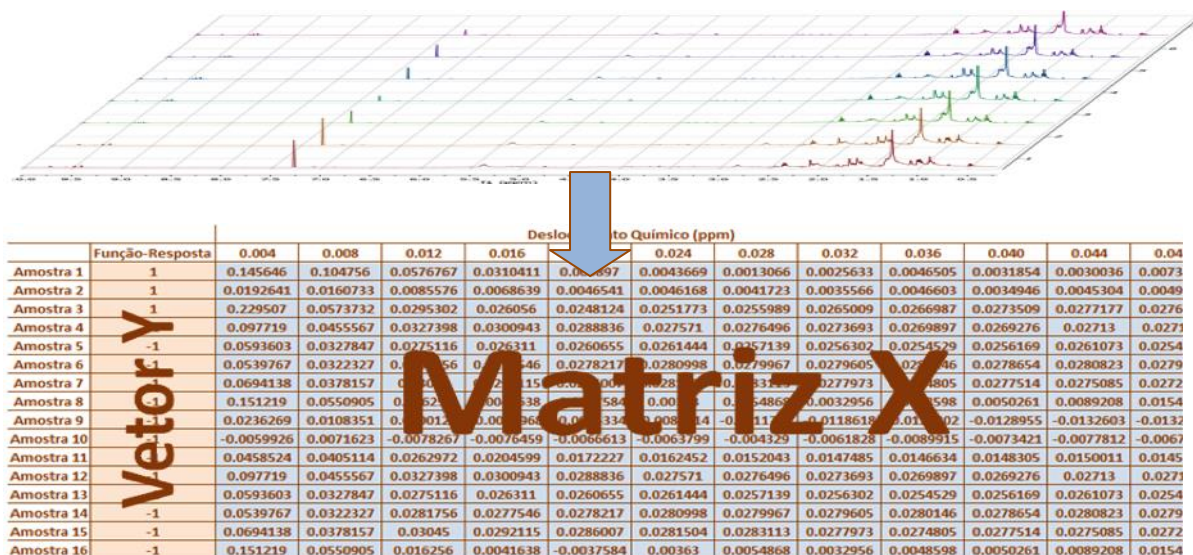


Figura 6. Organização dos dados espectrais em uma matriz X de dados

Utilizando os espectros no domínio de frequência, com correção de fase e linha base em zero, é desejável organizar os espectros em uma matriz de dados de acordo com a Figura 6 para realizar os cálculos multivariados. Contudo, pode ocorrer uma

série de problemas que inviabilizam ou atrapalham os cálculos multivariados de reconhecimentos de padrões:

- Os espectros de RMN de ^1H podem apresentar sinais de ruído, sinais indesejados ou sinais de solventes;

- A intensidade dos sinais entre duas amostras pode diferir devido a diferenças de concentração;

- Pequenas variações nos deslocamentos químicos devido a diferenças na temperatura de aquisição do espectro, pH das soluções e força iônica.

A identificação de sinais indesejáveis e ruído para posterior retirada pode ser realizada por deconvolução. Na deconvolução de espectros de RMN ocorre a

identificação dos sinais pelo formato do espectro, isto é realizado pelo reconhecimento de curvas provenientes de equações Lorentzianas ou Gaussianas.²¹ Na Figura 7, a curva em vermelho representa o espectro de RMN de ¹H original, a cor azul representa os sinais deconvoluídos e o sinal na cor preta representa o ruído. A soma dos sinais deconvoluídos e do ruído compõe o espectro original. Os programas o Top Spin® e o MestReNova® apresentam a ferramenta de deconvolução.

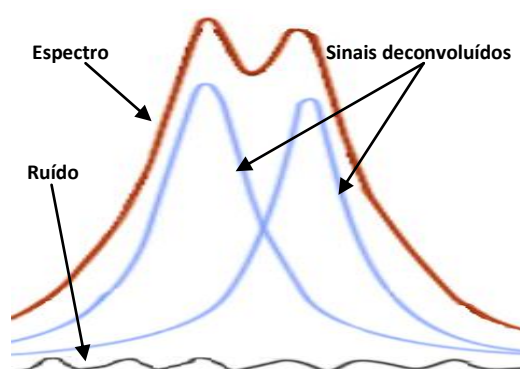


Figura 7. Exemplo de espectro de RMN deconvoluido

Para resolver o problema de diferentes alturas de sinais nos espectros obtidos devido às diferenças de concentrações entre as amostras, basta dividir todos os dados por uma área ou altura de sinal do espectro; este processo é denominado normalização. A Figura 8A apresenta três espectros de amostras com concentrações diferentes e a Figura 8B apresenta o resultado após normalização dos dados pela área do sinal principal. Se o sinal principal fosse muito

variável em relação aos sinais dos outros metabólitos, poderia ser utilizada a área total do espectro ou a área de uma faixa do espectro. Esta possibilidade de tratamento é comum nos programas avaliados, e mesmo que não houvesse este recurso, a normalização poderia ser realizada em programa de tratamento de planilhas eletrônicas, como por exemplo, Excel® ou Origin®.

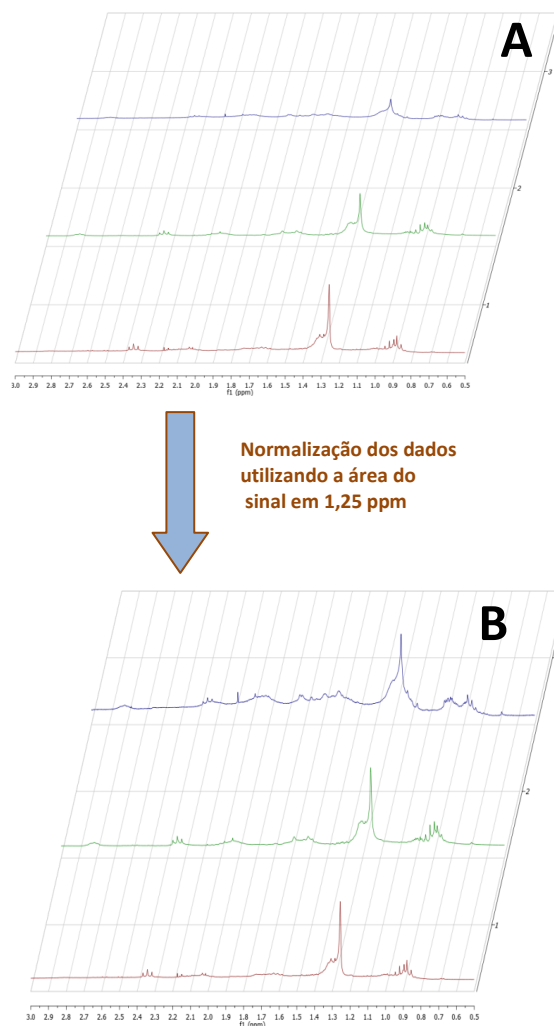


Figura 8. Espectros não normalizados (A); Espectros normalizados pela área do sinal principal (B)

O alinhamento dos espectros é uma etapa importante para posterior organização da matriz de dados. Alguns programas disponibilizam o alinhamento do espectro utilizando um sinal do próprio espectro, geralmente um sinal de solvente. Este procedimento não é eficaz para os espectros de RMN de ^1H porque as variáveis responsáveis pela variação nos deslocamentos químicos não afetam os sinais da mesma maneira.²² O alinhamento ideal é realizado com um espectro considerado como padrão em relação aos espectros das outras amostras do estudo de metabolômica.¹⁴ O espectro padrão deve apresentar todos os sinais para estes servirem como referência no alinhamento dos demais espectros. Os programas

MestReNova (a partir da versão 7.0), VnmrJ 3.2, Automics¹⁹ e Metabolomic²⁰ apresentam esta ferramenta de alinhamento. Há ainda opções de algoritmos de alinhamento que não estão nos pacotes avaliados, como por exemplo, o *Correlation Optimized Warping* (COW). No COW, o espectro é dividido em segmentos de tamanhos iguais e é escolhido um número de “pontos de folga”. O número de “pontos de folga” corresponde ao máximo de pontos que o espectro a ser alinhado irá percorrer no eixo horizontal para encontrar a melhor correlação com o espectro padrão. O algoritmo então testa a melhor correlação entre os segmentos do espectro padrão e do espectro que está sendo alinhado.^{23,24}

O *binning* ou *bucketing* é um tratamento

que reduz o tamanho dos dados espectrais.²⁵ Há vários tipos de *binings*, o tipo mais comum divide o espectro em *bins* equidistantes (geralmente utiliza-se a faixa de 0,04 ppm), calcula a área de cada *bin* e substitui a faixa do espectro por um ponto (Figura 9).²² Este recurso está disponível no MestReNova, Automics¹⁹, NMR Suite 7.6 e Metabonomic.²⁰ A vantagem do uso deste recurso é que na redução dos dados, o problema do desalinhamento dos sinais pode desaparecer e a desvantagem é uma possível perda de informação. O programa Automics¹⁹ ainda possui outros dois tipos de *binings*: no primeiro tipo o algoritmo identifica os sinais e verifica a largura dos sinais para ser utilizada como faixa do *bin* equidistante; o segundo tipo foi denominado de *binning* inteligente,

onde o algoritmo identifica os sinais e utiliza faixas variáveis de *bin*.²⁵

O importante na aplicação de qualquer tratamento nos espectros de RMN de ¹H é observar o resultado após a aplicação de cada procedimento e se possível testar vários parâmetros diferentes, como por exemplo, tamanhos de faixas diferentes no *binning*.

Todos os programas avaliados fazem a extração dos dados espectrais, permitindo, desta forma, a exportação dos dados de RMN de ¹H em arquivos de formatos diversos. Com estes dados, é montada uma matriz de acordo com a Figura 6 para posterior importação por programas estatísticos de cálculos multivariados.

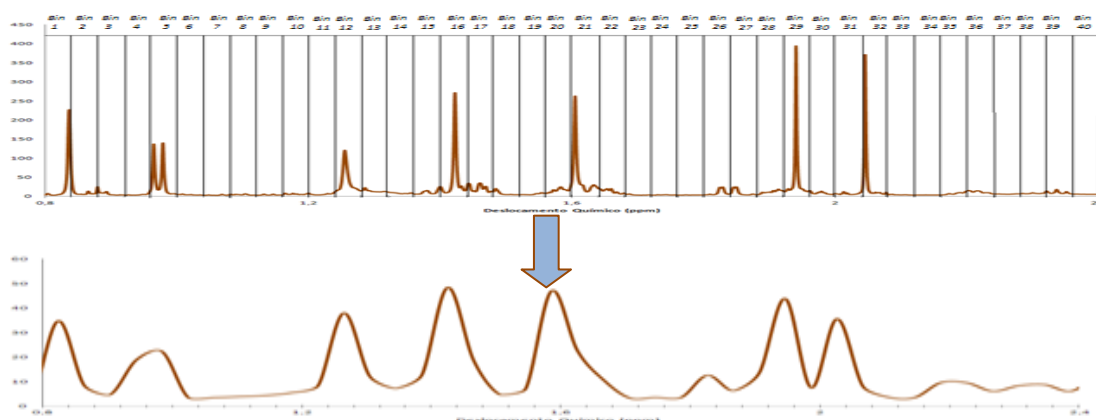


Figura 9. Espectro de RMN ¹H sem tratamento e *bins* equidistantes (A), resultado após a aplicação do *binning* (B)

3. Método dos mínimos quadrados parciais discriminante

O método dos mínimos quadrados parciais (*Partial Least Squares*-PLS) é um método de regressão em que o vetor da função-resposta *Y* e a matriz *X* são decompostos em componentes principais, também denominadas variáveis latentes:^{26,27}

$$X = T P^T + E \quad (\text{Eq. 1})$$

$$Y = U C^T + F \quad (\text{Eq. 2})$$

Onde:

P e *C* são os pesos de *X* e *Y*, respectivamente; *T* e *U* são os escores de *X* e *Y*, respectivamente; *E* e *F* são os resíduos de *X* e *Y*, respectivamente.

No PLS, os valores dos escores nas matrizes *T* e *U* são modificados até que o melhor modelo linear seja determinado entre eles e assim, a maior covariância entre *X* e *Y* seja obtida. Isso é feito em uma série de etapas e há diversos algoritmos propostos, dentre eles: NIPALS (*Non-linear Iterative*

Partial Least Squares)²⁷, SIMPLS²⁸ (Statistically Inspired Modification of Partial Least Squares), Kernel,^{29,30} Wide Kernel,³¹ Ding & Gentleman³² etc. No presente trabalho será apresentada uma aplicação do programa The Unscrambler® 9.5 que utiliza o algoritmo NIPALS para o cálculo PLS.

Nos cálculos PLS utilizados em reconhecimento de padrões, o método passa a ser chamado de método dos mínimos quadrados parciais discriminante (*Partial Least Squares-Discriminant Analysis*, PLS-DA) e se a função-resposta não for numérica, como na distinção das espécies de algas do exemplo a seguir, o valor +1 é atribuído na função-resposta para um grupo e para outro grupo é atribuído zero (ou outro valor, como -1).

No PLS-DA, a interpretação dos gráficos de escores e pesos passa a ter importância maior que nos trabalhos que utilizam os modelos PLS para previsão, como no caso da calibração multivariada. No gráfico de escores é desejável haver segregação entre os grupos em que se deseja discriminação. O gráfico de pesos apresenta o resultado mais importante do PLS-DA, pois as variáveis que mais contribuem para o modelo irão apresentar os maiores valores em módulo. Para os trabalhos de metabolômica com RMN, o gráfico de pesos mostra que sinais do espectro são os responsáveis pela discriminação dos grupos.

Além dos gráficos de escores e pesos, outras características dos modelos devem ser observadas:

- Número de variáveis latentes para a construção do modelo;
- Variância explicada em cada variável latente;
- Gráfico de calibração que apresenta os valores previstos pelo modelo *versus* os valores reais ou experimentais. Neste gráfico, o coeficiente de determinação (R^2) e o erro médio quadrático de calibração (RMSEC) devem ser avaliados;
- Gráfico da validação cruzada que apresenta os valores previstos pelo modelo

(com a retirada de uma ou mais amostras) *versus* os valores reais ou experimentais. O coeficiente de determinação (Q^2) e o erro médio quadrático da validação cruzada (RMSECV) também devem ser avaliados neste gráfico.

O RMSEC e RMSECV são calculados pelas seguintes equações:

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}} \quad (\text{Eq. 3})$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}'_i)^2}{n-1}} \quad (\text{Eq. 4})$$

Onde:

\hat{y}_i representa o valor previsto pelo modelo;

\hat{y}'_i representa o valor previsto pelo modelo na validação cruzada;

y_i representa o valor real;

n corresponde ao número de amostras.

Para ter um bom modelo PLS é desejável que R^2 e Q^2 sejam próximos a 1, os valores de RMSEC e RMSECV tenham os valores mais baixos possíveis e que se tenha alta variância explicada (próximo a 100 %). Contudo, muitas vezes o modelo não apresenta resultados satisfatórios devido ao excesso de informação dos espectros de RMN. Nestes casos é recomendável realizar seleção de variáveis.^{33,34}

Há diversos métodos de seleção de variáveis em modelos PLS que podem ser classificados em três categorias:

- Métodos de Filtro: utilizam os resultados do cálculo PLS sem seleção para identificar um subconjunto de variáveis importantes. Os gráficos de pesos ou coeficientes de regressão podem ser utilizados para identificar as variáveis importantes a serem selecionadas;³⁴

- Métodos *Wrapper*: são baseados em algum tipo de aprendizagem supervisionada e utilizam duplo procedimento iterativo. O método *wrapper* mais utilizado é o algoritmo genético. Este algoritmo é inspirado na teoria da evolução biológica e seleção natural de Darwin. Os modelos matemáticos são os indivíduos e as variáveis os genes. Os modelos com as melhores combinações de variáveis são selecionados para a próxima geração. O “cruzamento” entre os modelos e a mutação ocorre diversas vezes e ao fim do processo obtém-se um modelo mais ajustado a função-resposta.³⁴

- Métodos *Embedded*: são baseados em um único processo iterativo, como exemplo o iPLS. No iPLS o espectro é subdividido em faixas equidistantes e em seguida são obtidos modelos PLS de cada faixa. O modelo final é construído com as melhores faixas.³⁴

Um método de filtro, selecionando as variáveis importantes pelo gráfico de coeficientes de regressão no programa The Unscrambler® 9.5, foi utilizado neste trabalho.

4. Exemplo de aplicação: Seleção dos sinais representativos nos espectros de RMN de ¹H para diferenciação das espécies de algas *Dictyota crenulata* e *D. ciliolata*

4.1. Material e métodos

Os extratos brutos de cinco amostras de *D. crenulata* e oito amostras de *D. ciliolata* foram obtidos por maceração estática em acetona e secagem à pressão reduzida a 40 °C.⁹

Para as análises por RMN de ¹H cerca de 10 mg de extrato bruto foram dissolvidos em 500 µL de clorofórmio deuterado. As análises foram feitas em um espectrômetro Varian

Unity Plus 300 e o tetrametilsilano (TMS) foi usado como padrão interno para a referência de deslocamento químico (0 ppm).

O processamento dos espectros foi realizado usando o programa MestReNova® 8.0 que foi escolhido por apresentar todas as ferramentas de interesse, dentre elas, a deconvolução dos sinais e o alinhamento pelo espectro padrão.

A obtenção dos modelos PLS-DA foi feito no programa The Unscrambler® 9.5.

4.2. Resultados e discussão

4.2.1. Avaliação dos espectros de RMN ¹H e tratamento de dados

Os extratos brutos das algas das espécies *D. ciliolata* e *D. crenulata* obtidos empregando solventes apolares, contêm ácidos graxos, diterpenos e esteróis.^{9,10,35-37} Um espectro de RMN de ¹H típico para estes extratos está apresentado na Figura 10A. São observados muitos sinais na região de 0,5 a 3,0 ppm devido aos hidrogênios das metilas dos diterpenos e apresenta o sinal intenso em 1,25 ppm que corresponde aos hidrogênios dos ácidos graxos.

Através da comparação dos resultados obtidos do *binning* com o espectro original (Figura 10A) pôde-se escolher a faixa do *binning* utilizada. A faixa usualmente escolhida é a de 0,04 ppm, mas foi observada perda de informação (Figura 10B). Já com a aplicação do *binning* com a faixa de 0,004 ppm (Figura 10C) não foi observada mudança significativa em relação ao espectro original. Não há problema em utilizar uma faixa de *bin* tão estreita porque foi realizado alinhamento por um espectro padrão e seleção de variáveis na obtenção do modelo PLS-DA.

Os ácidos graxos não permitem a discriminação das espécies de algas do gênero *Dictyota*.^{9,10,35-37} Deste modo, o sinal correspondente a estas substâncias foi alinhado em 1,250 ppm usando o sinal do

TMS como referência. Tanto o sinal dos ácidos graxos obtido após alinhamento, quanto aqueles considerados como ruído, foram retirados após deconvolução. A deconvolução foi realizada empregando o modo sintético do MestReNova® 8.0. A Figura

10D apresenta o resultado referente à retirada dos sinais após deconvolução e *binning* com faixa de 0,004 ppm. A Figura 11 resume as etapas utilizadas no processamento de dados neste trabalho.

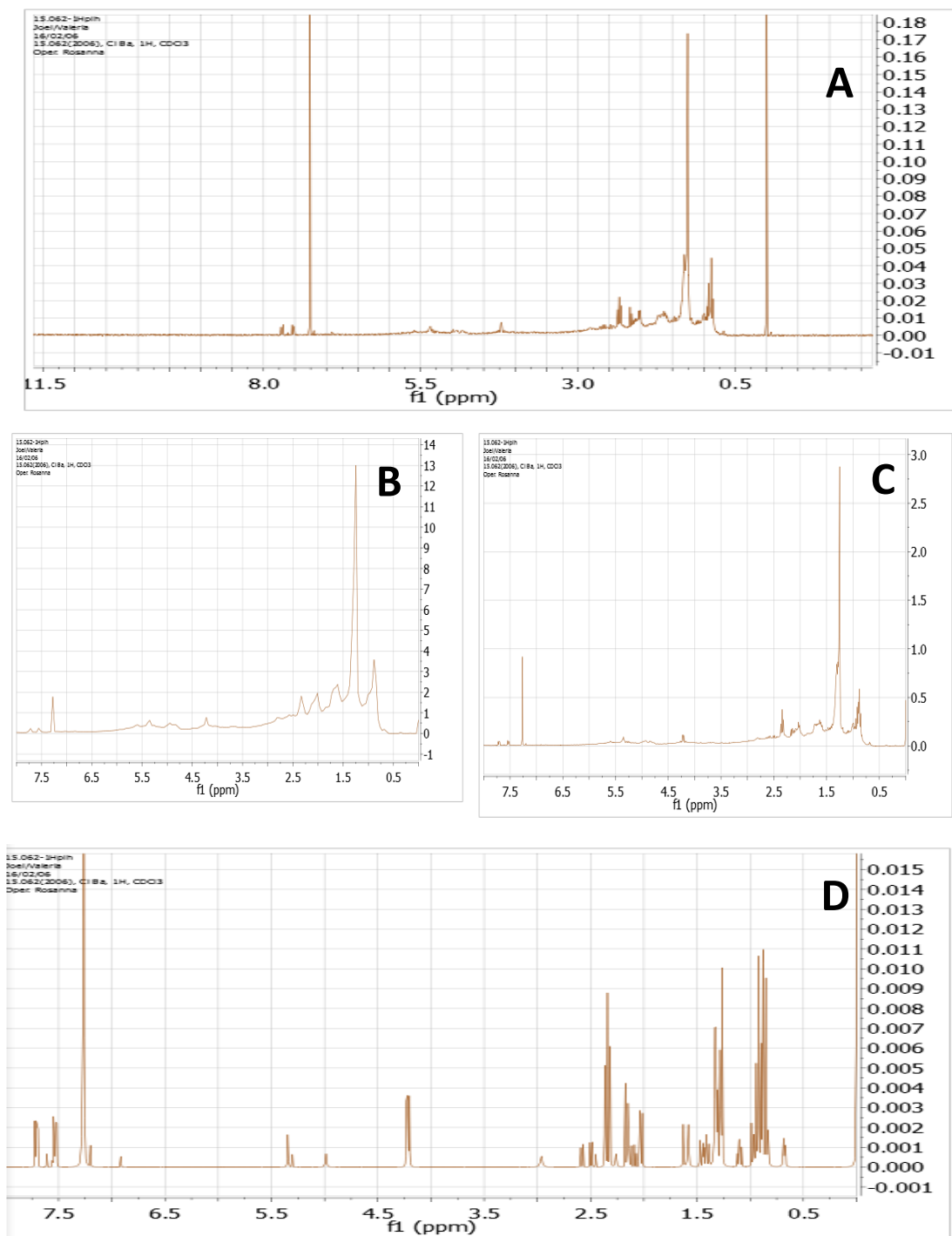


Figura 10. Espectro de RMN de ¹H de uma das amostras de *Dictyota ciliolata*(A); Espectro com aplicação de *binning* utilizando a faixa de 0,04 ppm (B); Espectro com aplicação de *binning* utilizando a faixa de 0,004 ppm (C); Espectro com aplicação de *binning* (faixa de 0,004 ppm) e retirada de ruído e sinal em 1,250 ppm após deconvolução (D)

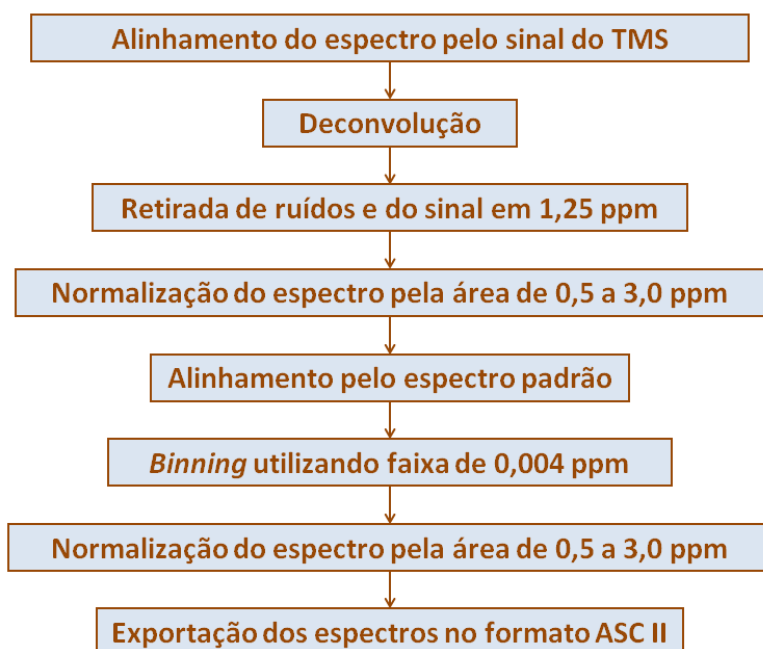


Figura 11. Etapas de tratamento dos espectros

Embora o alinhamento pelo sinal de referência do TMS tenha sido eficaz para identificar o sinal referente a ácidos graxos, este tipo de alinhamento não foi eficaz para os demais sinais do espectro, os quais foram corretamente alinhados através do espectro padrão.

A escolha do espectro padrão de cada espécie foi realizada pela presença dos sinais dos produtos majoritários. Entretanto, como estes espectros foram obtidos de extratos brutos, eles apresentam sinais de outras substâncias. Os produtos majoritários da *D. ciliolata* e *D. crenulata* são respectivamente o acetato de dictyol B e o 4 β -acetoxydictyodial A,⁹ cujas estruturas moleculares são

representadas na Figura 12.

As Figuras 13A e 13B apresentam os espectros padrão da *D. ciliolata* e *D. crenulata* respectivamente, onde alguns deslocamentos químicos dos produtos majoritários foram destacados.³⁸⁻³⁹

A Figura 13C apresenta todos os espectros utilizados no trabalho após os tratamentos mencionados acima. Grande semelhança entre os dois grupos de espectros é observada na Figura, mostrando a necessidade da aplicação de um método de reconhecimento de padrões para destacar quais os deslocamentos químicos são relevantes nos espectros de RMN de ¹H de cada espécie.

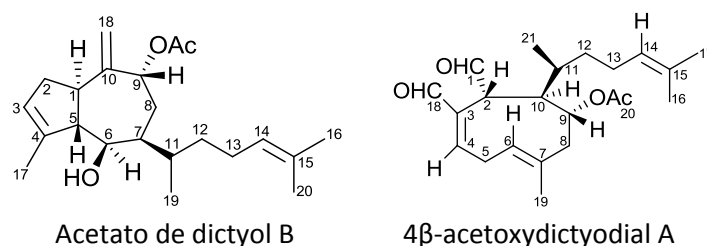


Figura 12. Produtos majoritários das espécies *D. ciliolata* e *D. Crenulata*

A normalização dos espectros foi realizada usando o valor da área da região espectral de 0,5 a 3,0 ppm por apresentar maior importância em termos de quantidade de sinais. Esta etapa foi realizada antes do

alinhamento pelo espectro padrão para que os espectros tivessem sinais com valores semelhantes para serem alinhados e foi realizada após o *binning*, pois nesta etapa ocorre alteração na intensidade dos sinais.

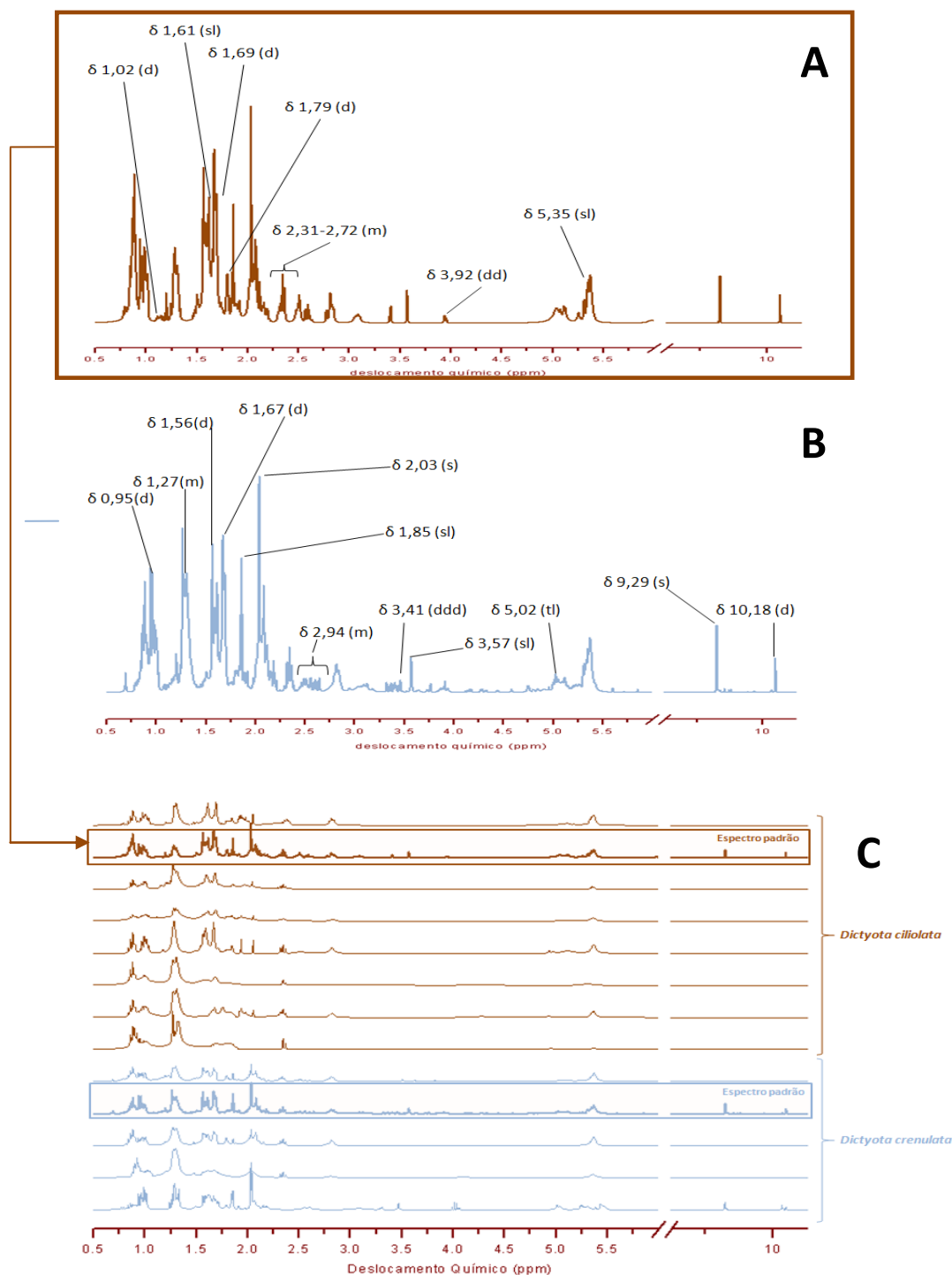


Figura 13. Espectro padrão da espécie *D. ciliolata* com alguns sinais destacados referentes ao acetato de dictyol B³⁸ (A); Espectro padrão da espécie *D. crenulata* com alguns sinais destacados referentes ao 4 β -acetoxydictyodial A³⁹ (B); Espectros de *D. ciliolata* e *D. crenulata* após tratamentos dos espectros de RMN de ^1H (C)

4.2.2. Obtenção e avaliação do modelo PLS-DA

Devido à natureza do problema estudado, a função-resposta atribuída à espécie *Dicyota ciliolata* foi +1 e à espécie *D. crenulata* foi -1. As regiões de -0,1 a 0,1 ppm e 7,24 a 7,26 ppm foram excluídas dos cálculos porque correspondem aos sinais do TMS e clorofórmio, respectivamente.

O melhor modelo de PLS-DA para diferenciação das espécies *D. ciliolata* e *D. crenulata* foi obtido com os dados centrados na média e seleção de 110 variáveis originais pelo gráfico de coeficiente de regressão. Este modelo com três variáveis latentes explica

86,3 % da variância de X.

Os gráficos de calibração com a previsão do modelo sem a retirada de amostras e com a validação cruzada do tipo “leave one out” mostram boa correlação apresentando respectivamente R^2 igual a 0,964 e Q^2 igual a 0,850 (Figuras 14A e 14B). As Figuras 14A e 14B apresentam respectivamente o RMSEC e RMSECV obtidos. O gráfico 2D de escores (Figura 14C) apresenta segregação entre as espécies. O gráfico de pesos da VL1 da Figura 14D mostra que os picos mais importantes para *D. ciliolata* ocorrem em δ 0,880, δ 1,300, δ 1,936 e δ 2,344 e para *D. crenulata* os sinais importantes ocorrem em δ 2,032; δ 0,936 e δ 1,616.

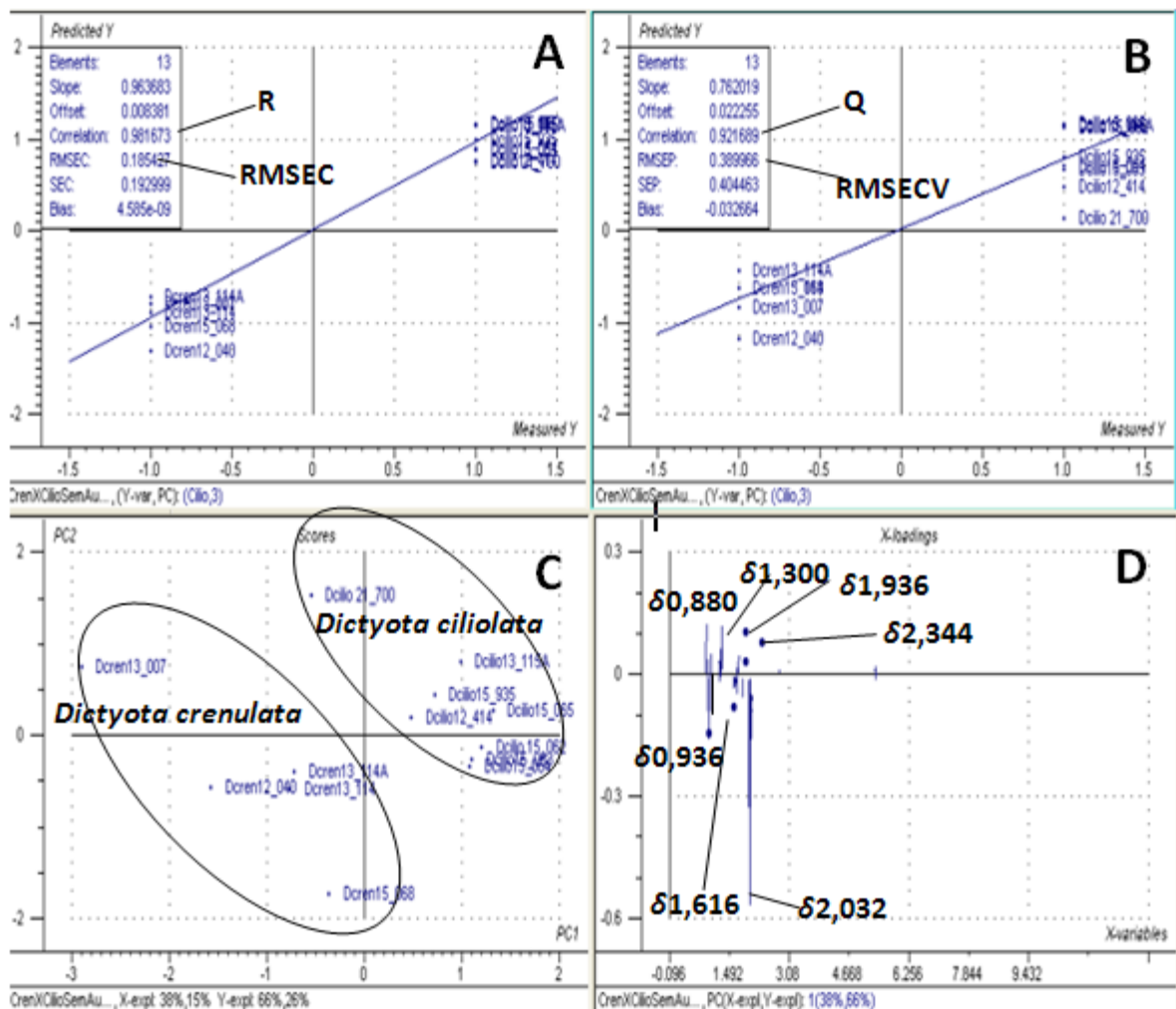


Figura 14. Gráfico de calibração (A); gráfico de validação cruzada do tipo “leave one out” (B); gráfico de escores de VL2 X VL1(C); gráfico de pesos em VL1 (D)

Os resultados para *D. crenulata* confirmam que o 4 β -acetoxydictyodial A é a substância importante para diferenciação desta espécie em relação a *D. ciliolata*. Os sinais δ 2,03, δ 0,936 e 1,616 correspondem respectivamente aos hidrogênios dos carbonos 20, 21 e 5.³⁹

Por outro lado, os resultados para *D. ciliolata* mostraram que a informação relevante para diferenciação desta espécie está em sinais dos hidrogênios do esqueleto guaiano prenilado, esqueleto do diterpeno majoritário acetato de dictyol B e de outros diterpenos presentes na amostra. Os sinais δ

3,444 e δ 1,300 correspondem respectivamente aos hidrogênios do esqueleto guaiano prenilado 1, 19 (Figura 15).³⁸ Outro sinal relevante ocorre em δ 1,936, que corresponde aos hidrogênios de metila do grupamento acetato.³⁸ Há ainda outro sinal importante para diferenciação de *D. ciliolata* que é o sinal em δ 0,880, este corresponde a hidrogênios de esteróis. Os esteróis abundantes nestas algas são o fucosterol e o isofucosterol cujas estruturas moleculares estão apresentadas na Figura 15. O sinal em δ 0,880 corresponde aos hidrogênios do carbono 27 destes esteróis.⁴⁰

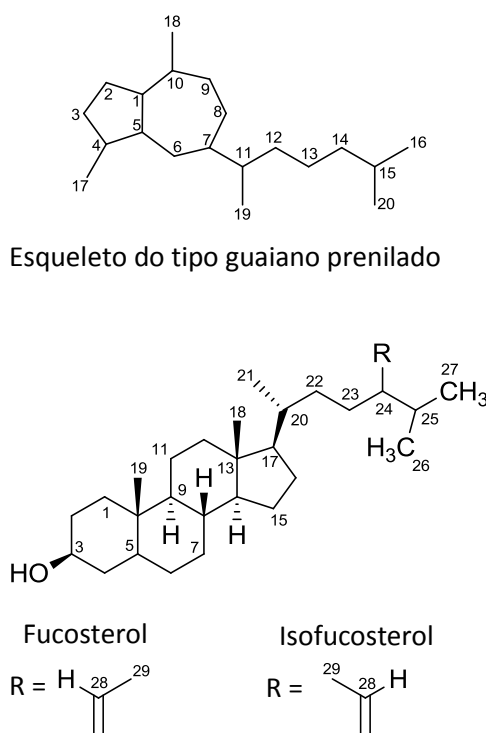


Figura 15. Estruturas moleculares do esqueleto guaiano prenilado, fucosterol e isofucosterol

5. Conclusões

Os trabalhos de metabolômica de RMN ¹H compreendem diversas etapas, dentre as quais os tratamentos dos espectros e os cuidados na utilização do método quimiométrico são etapas imprescindíveis para a obtenção de resultados confiáveis.

No trabalho foi apresentada uma aplicação de metabolômica baseada em RMN ¹H e PLS-DA para diferenciação entre *Dictyota ciliolata* e a *D. crenulata*. O modelo obtido mostrou que na *D. ciliolata* os sinais mais importantes são os referentes aos esteróis (δ 0,880) e aos guaianos prenilados (δ 1,300; δ 1,936; δ 2,344) por apresentarem maior abundância nos espectros. Na *D. crenulata* a diferenciação pode ser feita pelos sinais do

diterpeno acetilado (δ 2,032; δ 0,936; δ 1,260), que é o seu produto majoritário.

Referências Bibliográficas

- ¹ Nicholson, J.K.; Lindon, J. C.; Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999**, *29*, 1181. [[CrossRef](#)] [[PubMed](#)]
- ² Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. Systematic functional analysis of the yeast genome. *Trends in Biotechnology* **1998**, *16*, 373. [[CrossRef](#)]
- ³ Dunn, W.B.; Goodacre, R.; Neyses, L.; Mamas, M. Integration of metabolomics in heart disease and diabetes research: current achievements and future outlook. *Bioanalysis* **2011**, *3*, 2205. [[CrossRef](#)] [[PubMed](#)]
- ⁴ Smolinskaa, A.; Blancheta, L.; Buydensa, L.M.C.; Wijmengaa, S. S. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Analytica Chimica Acta* **2012**, *750*, 82. [[CrossRef](#)] [[PubMed](#)]
- ⁵ Tronholm, A.; Afonso-Carrillo, J.; Sanson, M.; Leliaert, F.; Fernández-García, C.; De Clerck, O. Taxonomy of the Dictyota ciliolata-crenulata complex (Dictyotales, Phaeophyceae) *Phycologia* **2013**, *52*, 171. [[CrossRef](#)]
- ⁶ De Paula, J. C.; Tese de Doutorado, Universidade Federal do Rio de Janeiro, Brasil, 2007.
- ⁷ Horta, P.A.; Tese de Doutorado, Universidade de São Paulo, Brasil, 2000.
- ⁸ Crispino, L. M. B.; Tese de Mestrado, Universidade Estadual Paulista, Brasil, 2000.
- ⁹ Wysor, B.; De Clerck, O. An Updated and Annotated List of Marine Brown Algae (Phaeophyceae) of the Caribbean Coast of the Republic of Panama. *Botanica Marina* **2003**, *46*, 151. [[CrossRef](#)]
- ¹⁰ De-Paula, J. C.; Bueno, L.B.; Cavalcanti, D.N.; Yoneshigue-Valentin, Y.; Teixeira, V. L. Diterpenes from the Brown Alga Dictyota crenulata. *Molecules* **2008**, *13*, 1253. [[CrossRef](#)] [[PubMed](#)]
- ¹¹ Pinto, A. M. V.; Leite, J. P. G.; Ferreira, W. J.; Cavalcanti, D. N.; Villaça, R. C.; Giongo, V.; Teixeira, V. L.; Paixão, I. C. N. P. Marine natural seaweed products as potential antiviral drugs against Bovine viral diarrhoea virus. *Revista Brasileira de Farmacognosia* **2012**, *22*, 813. [[CrossRef](#)]
- ¹² Pereira, H. S.; Leão-Ferreira, L. R.; Moussatché, N.; Teixeira, V. L.; Cavalcanti, D. N.; da Costa, L. J.; Diaz, R.; Frugulhetti, I. C. P. P. Antiviral activity of diterpenes isolated from the Brazilian marine alga Dictyota menstrualis against human immunodeficiency virus type 1 (HIV-1). *Antiviral Research* **2004**, *64*, 69. [[CrossRef](#)] [[PubMed](#)]
- ¹³ Pereira, H. S.; Leão-Ferreira, L. R.; Moussatché, N.; Teixeira, V. L.; Cavalcanti, D. N.; Costa, L. J.; Diaz, R.; Frugulhetti, I. C. P. P. Effects of Diterpenes Isolated from the Brazilian Marine Alga Dictyota menstrualis on HIV-1 Reverse Transcriptase. *Planta Medica* **2005**, *71*, 1019. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴ Abrantes, J. L.; Barbosa, J.; Cavalcanti, D. N.; Pereira, R. C.; Fontes, C. L. F.; Teixeira, V. L.; Souza, T. L. M.; Paixão, I. C. P. The Effects of the Diterpenes Isolated from the Brazilian Brown Algae Dictyota pfaffii and Dictyota menstrualis against the Herpes Simplex Type-1 Replicative Cycle. *Planta Medica* **2010**, *76*, 339. [[CrossRef](#)] [[PubMed](#)]
- ¹⁵ Barbosa, J. P.; Pereira, R. C.; Abrantes, J. L.; Cirne-Santos, C. C.; Rebello, M. A.; Frugulhetti, I. C. P. P.; Teixeira, V. L. In vitro Antiviral Diterpenes from the Brazilian Brown Alga Dictyota pfaffii. *Planta Medica* **2004**, *70*, 856. [[CrossRef](#)] [[PubMed](#)]
- ¹⁶ Cirne-Santos, C. C.; Teixeira, V. L.; Castello-Branco, L. R.; Frugulhetti, I. C. P. P.; Bou-Habib, D. C. Inhibition of HIV-1 Replication in Human Primary Cells by a Dolabellane Diterpene Isolated from the Marine Algae Dictyota pfaffii. *Planta Medica* **2006**, *72*, 295. [[CrossRef](#)] [[PubMed](#)]

- ¹⁷ Cirne-Santos, C. C.; Souza, T. M. L.; Teixeira, V. L.; Fontes, C. F. L.; Rebello, M. A.; Castello-Branco, L. R. R.; Abreu, C. M.; Tanuri, A.; Frugulhetti, I. C. P. P.; Bou-Habib, D. C. The dolabellane diterpene Dolabelladienetriol is a typical noncompetitive inhibitor of HIV-1 reverse transcriptase enzyme. *Antiviral Research* **2008**, *77*, 64. [CrossRef] [PubMed]
- ¹⁸ Garrido, V.; Teixeira, G. A. P. B.; Teixeira, V. L.; Ocampo, P.; Ferreira, W. J.; Cavalcanti, D. N.; Campos, S. M. N.; Pedruzzi, M. M. B.; Olaya, P.; Cirne-Santos, C. C.; Giongo, V.; Paixão, I. C. P. Evaluation of the acute toxicity of dolabelladienetriol, a potential antiviral from the brown alga *Dictyota paffii*, in BALB/c mice. *Revista Brasileira de Farmacognosia* **2011**, *21*, 209. [CrossRef]
- ¹⁹ Wang, T.; Shao, K.; Chu, Q.; Ren, Y.; Mu, Y.; Qu, L.; He, J.; Jin, C.; Xia, B. Automics: an integrated platform for NMR-based metabolomics spectral processing and data analysis. *BMC Bioinformatics* **2009**, *10*, 83. [CrossRef] [PubMed]
- ²⁰ Izquierdo-García, J. L.; Rodríguez, I.; Kyriazis, A.; Villa, P.; Barreiro, P.; Desco, M.; Ruiz-Cabello, J. A novel R-package graphic user interface for the analysis of metabolomic profiles. *BMC Bioinformatics* **2009**, *10*, 363. [CrossRef] [PubMed]
- ²¹ Siteio do Mestrelab Research. Disponível em: <<http://mestrelab.com/>> Acesso em 23 de Julho de 2013.
- ²² Izquierdo-García, J. L.; Villa, P.; Kyriazis, A.; Del Puerto-Nevado, L.; Pérez-Rial, S.; Rodríguez, I.; Hernandez, N.; Ruiz-Cabello, J. Descriptive review of current NMR-based metabolomic data analysis packages. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2011**, *59*, 263. [CrossRef] [PubMed]
- ²³ Nielsen, N. P. V.; Carstensen, J. M.; Smedsgaard, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A* **1998**, *805*, 17. [CrossRef]
- ²⁴ Tomasi, G.; Van Den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* **2004**, *18*, 231. [CrossRef]
- ²⁵ De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Tsiorkova, E.; Rietzschel, E. R.; De Buyzere, M. L.; Gillebert, T. C.; Bekaert, S.; Martins, J. C.; Van Criekinge, W. NMR-Based Characterization of Metabolic Alterations in Hypertension Using an Adaptive, Intelligent Binning Algorithm. *Analytical Chemistry* **2008**, *80*, 3783. [CrossRef] [PubMed]
- ²⁶ Worley, B.; Power, R. Multivariate Analysis in Metabolomics. *Current Metabolomics* **2013**, *1*, 92. [CrossRef]
- ²⁷ Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Analytica Chimica Acta* **1986**, *185*, 1. [CrossRef]
- ²⁸ De Jong, S. SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **1993**, *18*, 251. [CrossRef]
- ²⁹ De Jong, S.; Ter Braak, J. F. C. Comments on the PLS kernel algorithm. *Journal of Chemometrics* **1994**, *8*, 169. [CrossRef]
- ³⁰ Dayal, B. S. Improved PLS algorithms. *Journal of Chemometrics*. **1994**, *11*, 73. [CrossRef]
- ³¹ Rännar, S.; Geladi, P.; Wold, S. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. *Journal of Chemometrics* **1994**, *8*, 111. [CrossRef]
- ³² Ding, B.; Gentleman, R. Classification Using Generalized Partial Least Squares. *Journal of Computational and Graphical Statistics* **2005**, *14*, 280. [CrossRef]
- ³³ Saeys, Y.; Inza, I.; Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507. [CrossRef] [PubMed]
- ³⁴ Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems* **2012**, *118*, 62. [CrossRef]
- ³⁵ Cronin, G.; Lindquist, N.; Hay, M.E.; Fenical, W. Effects of storage and extraction procedures on yields of lipophilic metabolites from the brown seaweeds *Dictyota ciliolata* and *D. menstrualis*. *Marine Ecology Progress Series* **1995**, *119*, 265. [CrossRef]

- ³⁶ Kirkup, M. P.; Moore, R. E. Identity of sanadaol with β -crenulat, a diterpene from the brown alga *Dictyota crenulata*. *Phytochemistry* **1983**, *22*, 2527. [[CrossRef](#)]
- ³⁷ Sun, H. H.; McEnroe, F. J.; Fenical, W. Acetoxycrenulide, a new bicyclic cyclopropane-containing diterpenoid from the brown seaweed *Dictyota crenulata*. *Journal of Organic Chemistry* **1983**, *48*, 1903. [[CrossRef](#)]
- ³⁸ Freitas, O. S. P. *Tese de Doutorado*, Universidade Federal Fluminense, Brasil, 2007.
- ³⁹ Siamopoulou, P.; Bimplakis, A.; Iliopoulou, D.; Vagia, C.; Cos, P.; Berghe, D. V.; Roussis, V. Diterpenes from the brown algae *Dictyota dichotoma* and *Dictyota linearis*. *Phytochemistry* **2004**, *65*, 2025. [[CrossRef](#)] [[PubMed](#)]
- ⁴⁰ Alarif, W. M.; Ayyad, S.N.; Al-lihaibi, S. S. Acyclic Diterpenoid From The Red Alga *Gracilaria Foliifera*. *Revista Latinoamericana de Química* **2010** *38*, 52. [[Link](#)]