

Desenvolvimento de um modelo preditivo relacionando a estrutura química com o índice de cetanos de diferentes tipos de biodiesel.

**DIOGO VIOLA DE NADAI**

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE – UENF

CAMPOS DOS GOYTACAZES, RJ  
OUTUBRO DE 2011

Desenvolvimento de um modelo preditivo relacionando a estrutura química com o índice de cetanos de diferentes tipos de biodiesel.

## **DIOGO VIOLA DE NADAI**

Tese apresentada ao Centro de Ciência e Tecnologia da Universidade Estadual do Norte Fluminense, como parte das exigências para obtenção do título de Doutor em Ciências Naturais.

Orientador: Prof. Dr. Carlos Eduardo Novo Gatts.  
Coorientador: Prof. Dr. Paulo Cesar Muniz de Lacerda Miranda.

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE – UENF**

CAMPOS DOS GOYTACAZES, RJ  
OUTUBRO DE 2011

Desenvolvimento de um modelo preditivo relacionando a estrutura química com o índice de cetanos de diferentes tipos de biodiesel.

## **DIOGO VIOLA DE NADAI**

Tese apresentada ao Centro de Ciência e Tecnologia da Universidade Estadual do Norte Fluminense, como parte das exigências para obtenção do título de Doutor em Ciências Naturais.

Comissão Examinadora:

---

**Orientador:**

**Prof. Dr. Carlos Eduardo Novo Gatts (UENF)**

---

**Coorientador:**

**Prof. Dr. Paulo César Muniz de Lacerda Miranda (UNICAMP)**

---

**1º Avaliador:**

**Prof. Dr. Helion Vargas (UENF)**

---

**2º Avaliador:**

**Prof. Dr. Victor Haber Pérez (UENF)**

---

**3º Avaliador:**

**Prof. Dr. Donato Alexandre Gomes Aranda (URFJ)**

## AGRADECIMENTOS

- ✓ Primeiramente a Deus pelas inúmeras bênçãos recebidas durante toda minha vida.
- ✓ Aos meus pais, Alvanir e Tânia, que sempre estiveram ao meu lado me apoiando em tudo, entendendo as dificuldades e as vezes que não pude ir pra casa, não há preço que pague a educação que vocês me deram. Sinto muita saudade e muito orgulho de ser filho de vocês.
- ✓ À minha esposa Érika, pelo amor, carinho e compreensão durante esta difícil jornada.
- ✓ À minha madrinha Luzia, que sempre me apoiou e acolheu como uma mãe.
- ✓ Aos meus tios, José e Mary, que me acolheram em sua casa, sempre me ajudam e me tratam como um verdadeiro filho.
- ✓ Ao meu primo Gustavo, pela grande amizade e pela ajuda em minha vida.
- ✓ Às minhas irmãs, Jamille e Letícia, que sempre me incentivaram.
- ✓ Aos meus falecidos avós Durval e Maria, pelo exemplo de dignidade, honestidade e família.
- ✓ À mãezinha e tia Celita pelas preces e palavras de incentivo.
- ✓ Ao meu orientador Carlos Gatts, pela amizade, paciência, confiança, incentivo, compreensão e disponibilidade durante todo este tempo que me orientou. Mais uma vez foi um grande prazer ser orientado por você.
- ✓ Ao meu coorientador Paulo Miranda, pela ajuda e incentivo nos momentos difíceis.
- ✓ Às professoras, Denise Ribeiro e Maria Priscila, e ao amigo Edson Terra, que em inúmeras oportunidades se dispuseram a me ajudar durante esta jornada.
- ✓ Ao amigo Pedro Castelo Branco, que sempre me ajudou.
- ✓ E a todos que torceram ou que de alguma maneira contribuíram para meu sucesso.

## RESUMO

O objetivo desta tese foi desenvolver metodologias alternativas para determinação do índice de cetanos de diferentes biocombustíveis. O trabalho foi dividido em três etapas:

A primeira na construção de um banco de dados a partir de espectros teóricos de RMN-<sup>1</sup>H, e posteriormente no estudo da aplicabilidade do mesmo a partir de quatro métodos multivariados: PCA e três métodos de clustering (hierárquico, não hierárquico e fuzzy).

A segunda na predição do índice de cetanos dos nove tipos de biodiesel presentes no banco utilizando três métodos multivariados em seqüência: PCA, fuzzy clustering e uma rede neural artificial. Nesta etapa os valores simulados obtiveram erros relativos variando entre 0,39% e 4,41% para esse conjunto de dados.

Na última etapa formatou-se uma nova base de dados utilizando descritores estruturais para traduzir a linguagem química relevante em linguagem matemática. A partir desta nova matriz de dados e do método de regressão não linear foi desenvolvido um modelo preditivo relacionando a estrutura química e o índice de cetanos de diferentes biocombustíveis. O coeficiente de correlação (0,93) apontou o aguçado poder de predição do modelo, e os resultados obtidos se mostraram muito positivos com 75% dos valores dentro de uma margem de erro inferior a 7% dos dados experimentais.

**Palavras-chave:** biodiesel, índice de cetanos, modelo preditivo.

## ABSTRACT

The aim of this work was to develop alternative methodologies for determining the cetane number of different biofuels. The work was divided into three steps:

The first was to build a database from theoretical  $^1\text{H-NMR}$  spectra, in order to study the applicability of statistical multivariate methods, such as Principal Component Analysis and clustering (hierarchical, non-hierarchical and fuzzy).

The second part was the prediction of the cetane number of nine types of biodiesel present in the database using three multivariate methods in sequence: PCA, fuzzy clustering and artificial neural network. At this stage the simulated values presented relative errors ranging from 0.39% to 4.41%.

The last step was to format a new database using a structural descriptor to translate the relevant chemical information into mathematical language. From this, a non-linear regression predictive model relating the chemical structure and cetane number of different biofuels was developed. The correlation coefficient (0.93) between simulated and reference data shows the sharp predictive power of the model.

**Key words:** biodiesel, cetane number, predictive model.

**LISTA DE FIGURAS**

Figura 1. Estrutura química de um triacilglicerol, genérico.	2
Figura 2. Esquema de uma reação de trans-esterificação.	3
Figura 3. Estruturas químicas dos ácidos mirístico, palmítico, esteárico, palmitoléico, oléico, erúcico, linoléico e linolênico.	5
Figura 4. Alguns tipos de matrizes vegetais com potencial para serem usadas como matérias primas para produção do biodiesel em cada região do país adaptada de <a href="http://www.perkinelmer.com.br">www.perkinelmer.com.br</a> .	7
Figura 5. Estrutura química do hidrocarboneto hexadec-1-eno.	12
Figura 6. Estrutura química do hidrocarboneto $\alpha$ -metil-naftaleno.	12
Figura 7. Estrutura química do hidrocarboneto hexadecano.	13
Figura 8. Estrutura química do hidrocarboneto 2,2,4,4,6,8,8-heptametilnonano.	13
Figura 9. Representação gráfica dos resultados do “clustering” hierárquico (dendrograma).	19
Figura 10. Partição de um conjunto de dados com 16 elementos em 3 “clusters” (GRUBESIC, 2006).	21
Figura 11. Classificação de um conjunto segundo a “fuzzy clustering” (GRUBESIC, 2006).	23
Figura 12. Esquema geral de uma rede neural.	24
Figura 13. Esquema do treinamento supervisionado.	25
Figura 14. Ilustração das sete fatias do espectro para o oleato de etila.	31
Figura 15. Distribuição do conjunto de dados no plano PC1 x PC2.	37
Figura 16. Dendrograma obtido do tratamento do banco de dados.	39
Figura 17. Resultados obtidos do “clustering” não hierárquico.	41
Figura 18. Projeção da nuvem de dados no plano PC1 x PC2 após a “fuzzy clustering”.	44
Figura 19. Arquitetura da rede utilizada neste trabalho.	51
Figura 20. Fórmula estrutural do estearato de etila.	59
Figura 21. Fórmula estrutural do palmitato de propila.	69
Figura 22. Fórmula estrutural do palmitato de isopropila.	70
Figura 23. Fórmula estrutural do palmitato de butila.	71
Figura 24. Fórmula estrutural do palmitato de isobutila.	71

Figura 25. Espectro de RMN- <sup>1</sup> H do erucicato de metila.	86
Figura 26. Espectro de RMN- <sup>1</sup> H do estearato de 2-etilexila.	87
Figura 27. Espectro de RMN- <sup>1</sup> H do estearato de butila.	88
Figura 28. Espectro de RMN- <sup>1</sup> H do estearato de etila.	89
Figura 29. Espectro de RMN- <sup>1</sup> H do estearato de isobutila.	90
Figura 30. Espectro de RMN- <sup>1</sup> H do estearato de isopropila.	91
Figura 31. Espectro de RMN- <sup>1</sup> H do estearato de metila.	92
Figura 32. Espectro de RMN- <sup>1</sup> H do estearato de propila.	93
Figura 33. Espectro de RMN- <sup>1</sup> H do linoleato de butila.	94
Figura 34. Espectro de RMN- <sup>1</sup> H do linoleato de etila.	95
Figura 35. Espectro de RMN- <sup>1</sup> H do linoleato de isopropila.	96
Figura 36. Espectro de RMN- <sup>1</sup> H do linoleato de metila.	97
Figura 37. Espectro de RMN- <sup>1</sup> H do linoleato de propila.	98
Figura 38. Espectro de RMN- <sup>1</sup> H do linolenato de butila.	99
Figura 39. Espectro de RMN- <sup>1</sup> H do linolenato de etila.	100
Figura 40. Espectro de RMN- <sup>1</sup> H do linolenato de isopropila.	101
Figura 41. Espectro de RMN- <sup>1</sup> H do linolenato de metila.	102
Figura 42. Espectro de RMN- <sup>1</sup> H do linolenato de propila.	103
Figura 43. Espectro de RMN- <sup>1</sup> H do miristato de butila.	104
Figura 44. Espectro de RMN- <sup>1</sup> H do miristato de etila.	105
Figura 45. Espectro de RMN- <sup>1</sup> H do miristato de metila.	106
Figura 46. Espectro de RMN- <sup>1</sup> H do miristato de propila.	107
Figura 47. Espectro de RMN- <sup>1</sup> H do oleato de 2-etilexila.	108
Figura 48. Espectro de RMN- <sup>1</sup> H do oleato de butila.	109
Figura 49. Espectro de RMN- <sup>1</sup> H do oleato de etila.	110
Figura 50. Espectro de RMN- <sup>1</sup> H do oleato de isobutila.	111
Figura 51. Espectro de RMN- <sup>1</sup> H do oleato de isopropila.	112
Figura 52. Espectro de RMN- <sup>1</sup> H do oleato de metila.	113
Figura 53. Espectro de RMN- <sup>1</sup> H do oleato de propila.	114
Figura 54. Espectro de RMN- <sup>1</sup> H do palmitato de 2-etilexila.	115
Figura 55. Espectro de RMN- <sup>1</sup> H do palmitato de butila.	116
Figura 56. Espectro de RMN- <sup>1</sup> H do palmitato de etila.	117
Figura 57. Espectro de RMN- <sup>1</sup> H do palmitato de isobutila.	118

Figura 58. Espectro de RMN- <sup>1</sup> H do palmitato de isopropila.	119
Figura 59. Espectro de RMN- <sup>1</sup> H do palmitato de metila.	120
Figura 60. Espectro de RMN- <sup>1</sup> H do palmitato de propila.	121
Figura 61. Espectro de RMN- <sup>1</sup> H do palmitoleato de metila.	122

## LISTA DE GRÁFICOS

Gráfico 1. Resultado das componentes principais resultantes da PCA.	36
Gráfico 2. Comparação entre o IC dos biocombustíveis obtidos a partir de dados da literatura e da simulação feita pela rede neural com as respectivas barras de erro.	53
Gráfico 3. Comparação entre as funções $-21x(\text{NHI})^{2/3}$ e $-16x(\text{NCA})^{1/3}$ no plano xy.	65
Gráfico 4. Comparação entre os dados da literatura com as respectivas barras de erro e os valores simulados pelo modelo preditivo.	74
Tabela 5. Composição química do biodiesel de Soja (etílico).	14
Gráfico 5. Frequências dos erros relativos observados na Tabela 16.	75
Gráfico 6. Frequências dos erros relativos observados para os ésteres do banco de dados.	76
Gráfico 7. Frequências dos erros relativos observados para os diferentes tipos de biodiesel do banco de dados.	76

## LISTA DE TABELAS

Tabela 1. Porcentagem em massa dos principais ácidos graxos que compõem óleos de relevância para o trabalho.	4
Tabela 2. Especificações do biodiesel B100. Versão adaptada de <a href="http://www.biodiesel.gov.br">www.biodiesel.gov.br</a> , os números sobrescritos referem-se às notas de rodapé.	9
Tabela 3. Ésteres usados com seus respectivos índices de cetanos.	30
Tabela 4. Tipos de biodiesel usados com seus respectivos índices de cetanos. Todos os valores foram obtidos de KNOTHE; DUNN; BAGBY (1997).	31
Tabela 5. Variação em ppm das sete fatias obtidas do espectro de RMN- <sup>1</sup> H e representação das características químicas correspondentes a cada região do espectro.	32
Tabela 6. Banco de dados com as proporções de hidrogênio em cada fatia do espectro.	34
Tabela 7. Funções de pertinência obtidas da “fuzzy clustering”.	49
Tabela 8. Comparação dos resultados simulados pela rede em relação aos dados da literatura com os respectivos erros relativos.	52
Tabela 9. Base de dados construída para formatação do modelo preditivo.	58
Tabela 10. Composição química do biodiesel de Soja (etílico).	61
Tabela 11. Variáveis referentes aos ésteres componentes do biodiesel de Soja (etílico) e sua composição química média.	62
Tabela 12. Ilustração dos cálculos feitos para a obtenção da linha do biodiesel de Soja (etílico).	62
Tabela 13. Ilustração das linhas referentes ao palmitato de etila e de propila e o IC calculado pela equação (4).	66
Tabela 14. Ilustração das linhas referentes ao palmitato de propila e de isopropila e o IC calculado pela equação (4).	70
Tabela 15. Ilustração das linhas referentes ao palmitato de butila e de isobutila e o IC calculado pela equação (4).	71
Tabela 16. Resultados obtidos pelo modelo, dados da literatura e os erros relativos.	72

## LISTA DE SIGLAS

Anfavea: Associação Nacional dos Fabricantes de Veículos Automotores

ANP: Agencia Nacional de Petróleo

ASTM D613: American Society for Testing and Materials - Standard Test Method for Cetane Number of Diesel Fuel Oil

CFR: Cooperative Fuel Research

BX: Mistura contendo X% de Biodiesel e (100-X)% de Diesel

B100: Biodiesel puro

B5: Mistura contendo 5% de Biodiesel e 95% de Diesel

B20: Mistura contendo 20% de Biodiesel e 80% de Diesel

CNPE: Conselho Nacional de Política Energética

FSF: Free Software Foundation

gNMR: Software gratuito de simulação de espectros de Ressonância Magnética Nuclear

GPL: General Public License

IC: Índice de Cetanos

MCTI: Ministério da Ciência, Tecnologia e Inovação

NCA: O Número de átomos de Carbono do Álcool utilizado na produção do respectivo biocombustível

NCR: O Número de átomos de Carbono presentes nas Ramificações

NCT: O Número Total de átomos de Carbono na molécula ou mistura

NHC: O Número de átomos de Hidrogênio ligados ao carbono Carbinólico

NHI: O Número de átomos de Hidrogênio ligados a átomos de carbono Insaturados

PCA: Principal Component Analysis

PC1: Componente Principal 1

PC2: Componente Principal 2

PC3: Componente Principal 3

PC's: Componentes Principais

ppm: partes por milhão

RMN-<sup>1</sup>H: Ressonância Magnética Nuclear de Hidrogênio

R<sup>2</sup>: Coeficiente de correlação

u.a.: unidades arbitrárias

## SUMÁRIO

1. REVISÃO BIBLIOGRÁFICA	1
1.1. O Biodiesel	1
1.1.1. A composição química do Biodiesel	2
1.1.2. O Biodiesel no Cenário Nacional e Internacional	6
1.1.3. Vantagens e desvantagens do biodiesel frente ao diesel	8
1.2. Os índices de qualidade do biodiesel	9
1.2.1. O Índice de Cetanos (IC)	12
1.3. Métodos de análise estatística multivariada	16
1.3.1. PCA	16
1.3.2. Análise de “Cluster”	18
1.3.2.1. “Clustering” Hierárquico	19
1.3.2.2. “Clustering” não hierárquico (k-médias)	20
1.3.2.3. “Fuzzy clustering”	21
1.3.3. Redes Neurais	23
1.3.4. Modelos de regressão	25
1.4. A Técnica de RMN- <sup>1</sup> H	27
2. OBJETIVO	28
3. OBJETIVOS ESPECÍFICOS	29
CAPÍTULO I – Análise exploratória	30
I.1. Métodos	30
I.1.1. Formatação do Banco de dados	30
I.1.2. O ambiente estatístico	35
I.2. Resultados e Discussão	35

1.2.1. Estudo da aplicabilidade do banco de dados	35
1.2.1.1. Tratamento dos Dados por PCA	36
1.2.1.2. Tratamento dos Dados por “Clustering” hierárquico	39
1.2.1.3. Tratamento dos Dados por “Clustering” não hierárquico (k-médias)	41
1.2.1.4. Classificação dos dados por “Fuzzy Clustering”	44
1.3. Conclusões	46
CAPÍTULO II – Modelo empírico	48
II.1. Metodologia	48
II.2. Resultados e Discussão	49
II.2.1. Fuzzy clustering	49
II.2.2. Predição do índice de cetanos via redes neurais	51
II.3. Conclusões	53
CAPÍTULO III – Modelo preditivo	55
III.1. Formatação da nova base de dados	58
III.2. O modelo	63
III.2.1. As variáveis com constantes negativas	64
III.2.2. As variáveis com constantes positivas	68
III.3. Conclusões	77
4. REFERÊNCIAS BIBLIOGRÁFICAS	79
5. ANEXOS	86

## 1. REVISÃO BIBLIOGRÁFICA

### 1.1. O Biodiesel

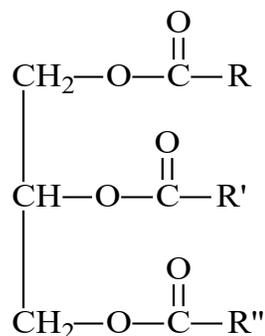
Durante as últimas décadas, a necessidade de se produzir energias alternativas às dos combustíveis fósseis resultou no desenvolvimento de novas tecnologias baseadas no uso de biomassa. A partir de resíduos vegetais e orgânicos é possível gerar diversos tipos de biocombustíveis, o biogás e o etanol são algumas destas opções (TAHERZADEH; KARIMI, 2008). Porém, no caso específico dos motores a compressão por ignição movidos a diesel, os esforços se concentraram no biodiesel (HAAS *et al*, 2006).

De acordo com a Lei nº 11.097, de 13 de janeiro de 2005, o biodiesel é definido como um “biocombustível derivado de biomassa renovável para uso em motores a combustão interna com ignição por compressão ou, conforme regulamentos para geração de outro tipo de energia, que possa substituir parcial ou totalmente combustíveis de origem fóssil” ([www.biodiesel.gov.br](http://www.biodiesel.gov.br)).

O biodiesel puro é corriqueiramente chamado de B100, entretanto este biocombustível é geralmente utilizado em misturas diesel-biodiesel (BX) que recebem denominações diferentes dependendo da porcentagem, em volume, de biodiesel. Por exemplo, uma mistura de 5% de biodiesel e 95% de diesel é denominada B5, já outra com 20% de biodiesel e 80% de diesel é denominada B20, e assim por diante.

### 1.1.1. A composição química do Biodiesel

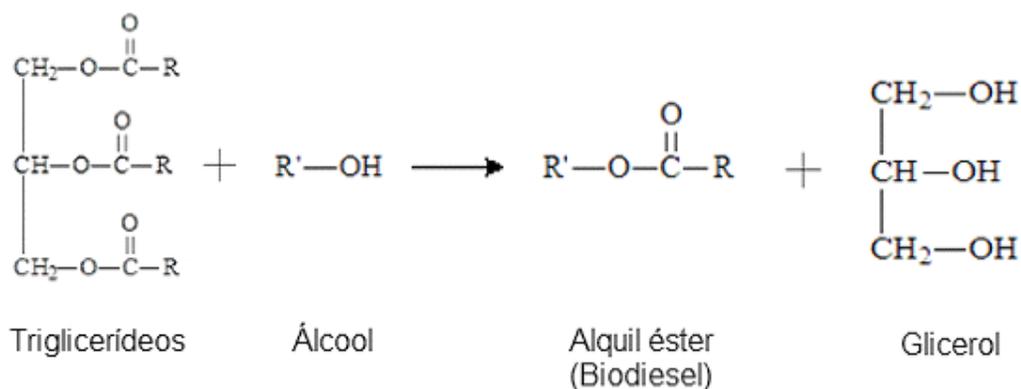
Os principais componentes das gorduras animais e dos óleos vegetais, potenciais matérias primas na produção do biodiesel, são os triacilgliceróis (ou triglicerídeos) (Figura 1), que são compostos por ésteres de ácidos graxos derivados do glicerol.



**Figura 1.** Estrutura química de um triacilglicerol, genérico.

Os triacilgliceróis normalmente contêm diferentes tipos de ácidos graxos em suas composições. Devido a essas variações na composição química destes compostos, os óleos e gorduras possuem propriedades físicas e químicas distintas, sendo essa diferença estrutural o principal parâmetro que influenciará em suas propriedades (KNOTHE; GERPEN; KRAHL, 2005).

O biodiesel é obtido através de reações de esterificação, trans-esterificação ou craqueamento desses ácidos graxos provenientes de gordura animal ou óleos vegetais ([www.biodiesel.gov.br](http://www.biodiesel.gov.br)). O processo mais usado atualmente é a trans-esterificação (Figura 2). Nesse processo, o óleo vegetal (ou a gordura animal) reage com um álcool para formar os alquil ésteres correspondentes (biodiesel), tendo como subproduto o glicerol.



**Figura 2.** Esquema de uma reação de trans-esterificação.

Vários álcoois podem ser utilizados na produção do biodiesel. A opção de um álcool ou outro fará com que o biodiesel formado tenha propriedades diferentes. Entretanto, por definição, qualquer álcool utilizado na reação de trans-esterificação formará um biodiesel (KNOTHE; GERPEN; KRAHL, 2005).

Como já mencionado, outro fator importante é a variação na composição dos óleos vegetais que são usados como matéria-prima para produção do biodiesel. Conforme a espécie da planta oleaginosa, diferenças na composição química do óleo vegetal são expressas por mudanças na relação molar entre os ácidos graxos presentes na estrutura (COSTA NETO; ROSSI, 2000; KNOTHE; DUNN; BAGBY, 1997). Essas variações podem ser visualizadas na Tabela 1, onde são mostradas as porcentagens em massa dos principais ácidos graxos que compõem alguns óleos vegetais utilizados neste trabalho.

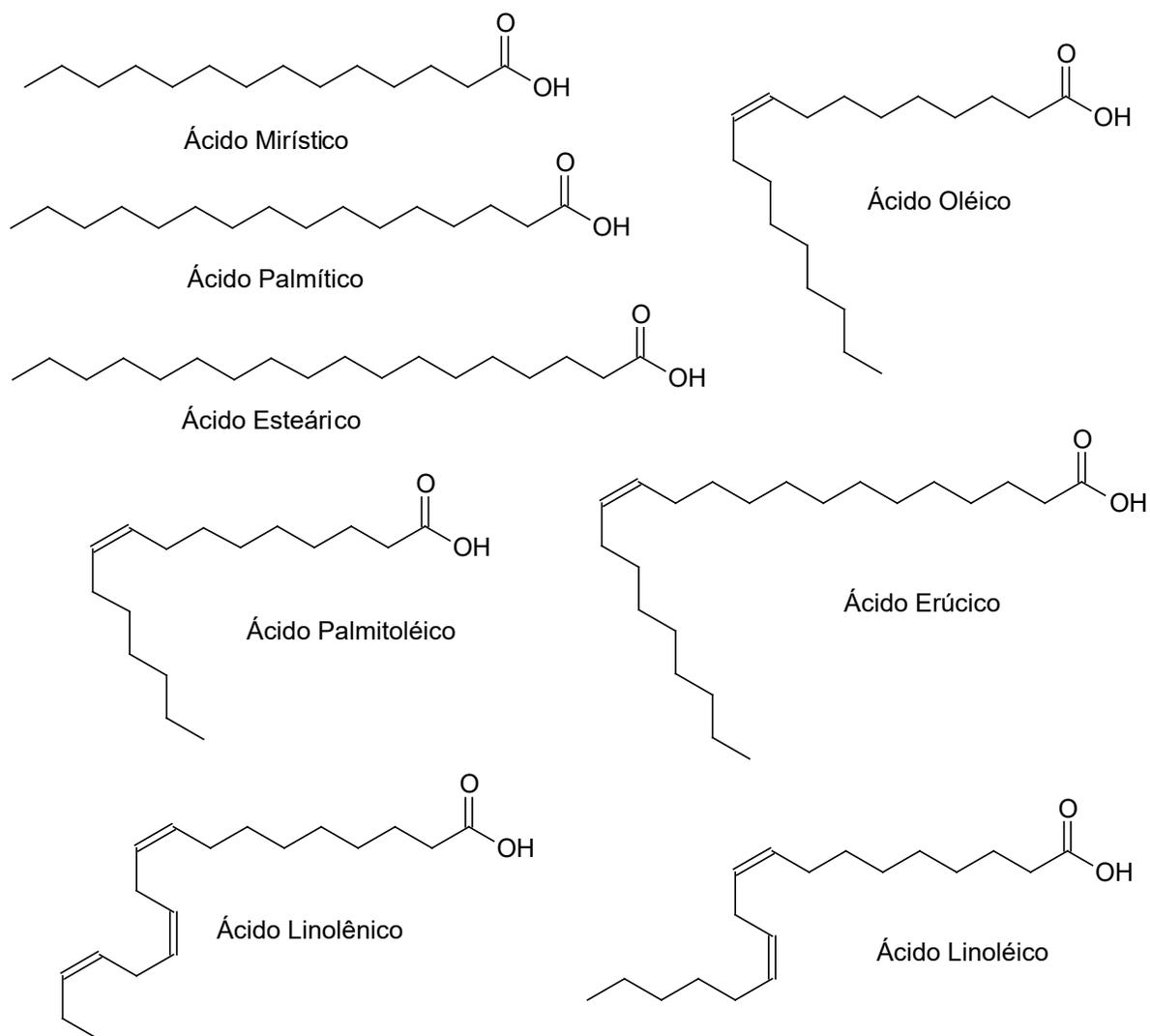
**Tabela 1.** Porcentagem em massa dos principais ácidos graxos que compõem óleos de relevância para o trabalho.

Óleo	Porcentagem (%) em massa dos principais ácidos graxos						
	Mirístico	Palmítico	Estearico	Oléico	Linoléico	Linolênico	Erúxico
Açafoa		6,4-7	2,4-29	9,7-13,8	75,3-80,5		
Algodão	0,8-1,5	22-24	2,6-5	19	50-52,5		
Colza	1,5	1-4,7	1-3,5	13-38	9,5-22	1-10	40-64
Girassol		3,5-6,5	1,3-5,6	14-43	44-68,7		
Palma	0,6-2,4	32-46,4	4-6,3	37-53	6-12		
Soja		2,3-11	2,4-6	22-30,8	49-53	2-10,5	

*Fonte:* (KNOTHE; DUNN; BAGBY, 1997)

Os ácidos graxos representados na Tabela 1 são os ácidos mirístico (14:0), palmítico (16:0), estearico (18:0), oléico (18:1), linoléico (18:2), linolênico (18:3) e erúxico (22:1), onde os índices entre parênteses (x,y) representam, respectivamente o número de carbonos e o número de ligações duplas presentes nas cadeias carbônicas.

As estruturas químicas destes ácidos bem como a estrutura do ácido palmitoléico (16:1) estão representadas na Figura 3 a seguir.



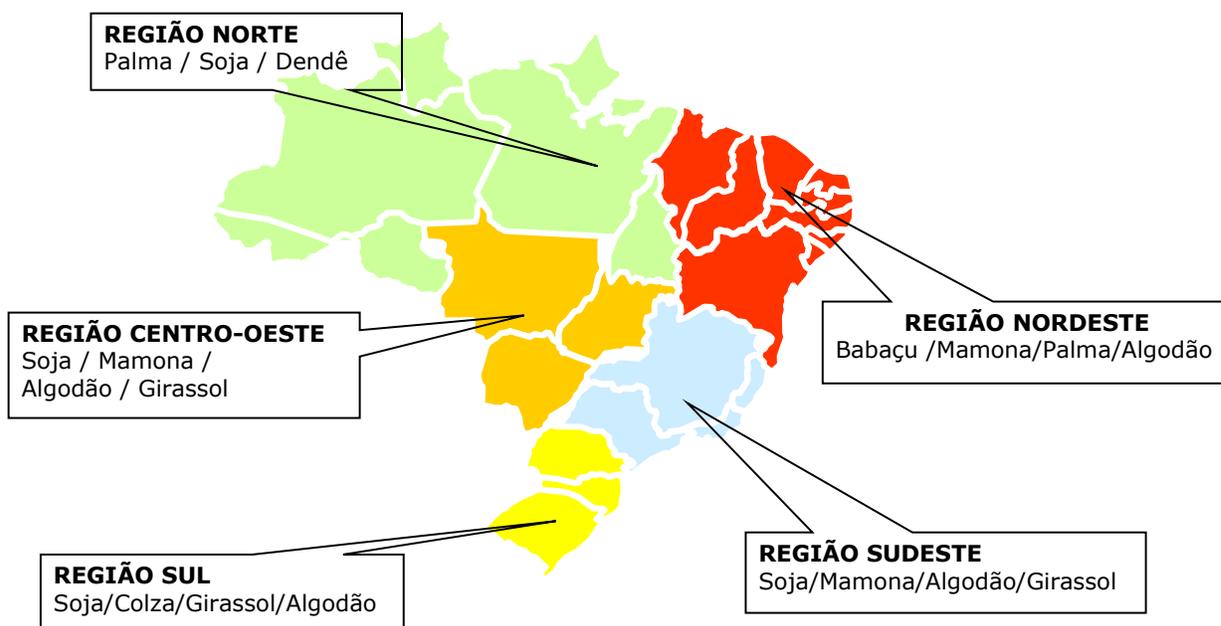
**Figura 3.** Estruturas químicas dos ácidos mirístico, palmítico, esteárico, palmitoléico, oléico, erúxico, linolênico e linoléico.

### **1.1.2. O Biodiesel no Cenário Nacional e Internacional**

Sabe-se que, nos dias de hoje, os combustíveis de origem fóssil são recursos finitos e que seus preços tendem a aumentar a cada dia. Uma possível alternativa em relação a esses combustíveis é a utilização do biodiesel (CONCEIÇÃO *et al*, 2005; ENCINAR; GONZÁLEZ; RODRÍGUEZ-REINARES, 2005). Vale a pena destacar que uma política nacional equilibrada de incentivo à produção de suas matérias primas, com o intuito de fomentar a produção desse biocombustível, poderia alavancar um desejado desenvolvimento sustentável para nosso país. Somente no ano de 2008, o uso do biodiesel evitou a importação de 1,1 bilhões de litros de diesel de petróleo resultando numa economia de cerca de US\$ 976 milhões (<http://www.anp.gov.br>).

Sua produção vem aumentando significativamente em muitas nações, como Estados Unidos da América, Itália, França, Alemanha, Áustria e Malásia, e está num patamar inicial em muitas outras (KNOTHE; MATHEAUS; RYAN III, 2003). Por ser um combustível de origem biológica, o biodiesel é recomendado pela União Europeia e classificado como um combustível com perspectivas para utilização no futuro (TOMASEVIC; SILER-MARINKOVIC, 2003).

Num país extenso como o nosso, onde são inúmeras as variedades climáticas, verifica-se que todas as regiões teriam capacidade de produzir plantas oleaginosas com potencial para se tornarem matérias primas na produção do biodiesel (Figura 4) ([www.perkinelmer.com.br](http://www.perkinelmer.com.br)).



**Figura 4.** Alguns tipos de matrizes vegetais com potencial para serem usadas como matérias primas para produção do biodiesel em cada região do país adaptada de [www.perkinelmer.com.br](http://www.perkinelmer.com.br).

No Brasil, a partir da Lei nº 11.097, publicada em 13 de janeiro de 2005, o biodiesel foi introduzido na matriz energética. Na época o percentual de biodiesel presente no óleo diesel era de 2%, esta proporção foi aumentando paulatinamente, até que a resolução nº 6/2009 do Conselho Nacional de Política Energética (CNPE) estabeleceu, a partir de 01 de janeiro de 2010, a especificação para a comercialização do biodiesel que vigora atualmente. Este deve ser adicionado ao óleo diesel na proporção de 5% em volume (B5), sendo obrigatória em todos os postos que revendem este combustível, estando sujeitos à fiscalização pela Agência Nacional de Petróleo (ANP) (<http://www.anp.gov.br>).

Relativo ao desempenho dos motores movidos a diesel, a ANP pronuncia-se da seguinte maneira quando questionada em relação à funcionalidade do B5:

“A adição de até 5% de biodiesel ao diesel de petróleo foi amplamente testada, dentro do programa de testes coordenado pelo Ministério da Ciência, Tecnologia e Inovação (MCTI), que contou com a participação da Associação Nacional dos Fabricantes de Veículos Automotores (Anfavea). Os resultados demonstraram, até o momento, não haver a necessidade de qualquer ajuste ou alteração nos motores e veículos que utilizem essa mistura” (<http://www.anp.gov.br>).

Ainda de acordo com a ANP, o Brasil está entre os maiores produtores e consumidores de biodiesel do mundo, com uma produção anual, em 2009, de 1,6 bilhões de litros e uma capacidade instalada, em janeiro de 2010, para cerca de 4,7 bilhões de litros (<http://www.anp.gov.br>).

### **1.1.3. Vantagens e desvantagens do biodiesel frente ao diesel**

O biodiesel é tecnicamente competitivo em relação ao diesel convencional (derivado do petróleo) e, em princípio, não necessita de nenhuma mudança na mecânica de distribuição nos atuais motores a diesel (KNOTHE; MATHEAUS; RYAN III, 2003). Outras vantagens desse biocombustível incluem a redução da quantidade de emissão da maioria dos gases (destaque para a redução de 100% na emissão dos óxidos de enxofre), lubricidade inerente, caráter não tóxico e biodegradabilidade duas vezes mais rápida do que o diesel puro (KNOTHE; MATHEAUS; RYAN III, 2003; FERRARI; OLIVEIRA; SCARBIO, 2005).

Como desvantagens do biodiesel pode-se citar o alto custo quando comparado ao óleo diesel, o aumento da emissão de óxidos de nitrogênio (NO<sub>x</sub>), a eventual competição com produtos alimentares no mercado das matérias-primas, a produção de grandes quantidades de glicerina prevista (subproduto do biodiesel) que poderá não ser

absorvida pelo mercado, a produção intensiva e não sustentável de matérias-primas para o biodiesel, que pode afetar a capacidade do solo ocasionando destruição da fauna e flora e conseqüentemente reduzindo a biodiversidade local (<http://www.energiasrenovaveis.com>).

## 1.2. Os índices de qualidade do biodiesel

Num país como o Brasil, em que a diversidade de matrizes vegetais com potencial para produção de biodiesel é muito grande, existe uma preocupação em relação à determinação dos índices de qualidade destes combustíveis, visto que as composições químicas destes óleos são distintas e resultam em propriedades físico-químicas com valores variados. As especificações para o B100 produzido aqui são descritas na resolução número 07 (Tabela 2) de 19/03/2008, da ANP.

**Tabela 2.** Especificações do biodiesel B100. Versão adaptada de [www.biodiesel.gov.br](http://www.biodiesel.gov.br), os números sobrescritos referem-se às notas de rodapé.

CARACTERÍSTICA	UNIDADE	LIMITE	MÉTODO		
			ABNT NBR	ASTM D	EN/ISO
Aspecto	-	LII (1)	-	-	-
Massa específica a 20° C	kg/m <sup>3</sup>	850-900	7148	1298	EN ISO 3675
Viscosidade Cinemática a 40°C	Mm <sup>2</sup> /s	3,0-6,0	10441	445	EN ISO 3104
Teor de Água, máx. (2)	mg/kg	500	-	6304	EN ISO 12937
Contaminação Total, máx.	mg/kg	24	-	-	EN ISO 12662
Ponto de fulgor, mín. (3)	°C	100,0	14598	93	EN ISO 3679
Teor de éster, mín	% massa	96,5	15764	-	EN 14103
Resíduo de carbono (4)	% massa	0,050	15586	4530	-
Cinzas sulfatadas, máx.	% massa	0,020	6294	874	EN ISO 3987
Enxofre total, máx.	mg/kg	50		5453	EN ISO 20846
Sódio + Potássio, máx.	mg/kg	5	15554	-	EN 14108

Cálcio + Magnésio, máx.	mg/kg	5	15553	-	EN 14538
Fósforo, máx.	mg/kg	10	15553	4951	EN 14107
Corrosividade ao cobre, 3h a 50 °C, máx.	-	1	14359	130	EN ISO 2160
Índice de Cetano (6)	-	Anotar	-	613	EN ISO 5165
Ponto de entupimento de filtro a frio, máx.	°C	19 (7)	14747	6371	EN 116
Índice de acidez, máx.	mg KOH/g	0,50	14448	664	EN 14104 (8)
Glicerol livre, máx.	% massa	0,02	15341	6584 (8)	EN 14105 (8)
Glicerol total, máx.	% massa	0,25	15344	6584 (8)	EN 14105 (8)
Mono, di, triacilglicerol (5)	% massa	Anotar	15342	6584 (8)	EN 14105 (8)
Metanol ou Etanol, máx.	% massa	0,20	15343	-	EN 14110
Índice de Iodo (5)	g/100g	Anotar	-	-	EN 14111
Estabilidade à oxidação a 110°C, mín.(2)	h	6	-	-	EN 14112 (8)

Notas: (1) Límpido e isento de impurezas com anotação da temperatura de ensaio. (2) O limite indicado deve ser atendido na certificação do biodiesel pelo produtor ou importador. (3) Quando a análise de ponto de fulgor resultar em valor superior a 130°C, fica dispensada a análise de teor de metanol ou etanol. (4) O resíduo deve ser avaliado em 100% da amostra. (5) Estas características devem ser analisadas em conjunto com as demais constantes da tabela de especificação a cada trimestre civil. Os resultados devem ser enviados pelo produtor de biodiesel à ANP, tomando uma amostra do biodiesel comercializado no trimestre e, em caso de neste período haver mudança de tipo de matéria-prima, o produtor deverá analisar número de amostras correspondente ao número de tipos de matérias-primas utilizadas. (6) Poderá ser utilizado como método alternativo o método ASTM D6890 para índice de cetano. (7) O limite máximo de 19°C é válido para as regiões Sul, Sudeste, Centro-Oeste e Bahia, devendo ser anotado para as demais regiões. O biodiesel poderá ser entregue com temperaturas superiores ao limite supramencionado, caso haja acordo entre as partes envolvidas. Os métodos de análise indicados não podem ser empregados para biodiesel oriundo apenas de mamona. (8) Os métodos referenciados demandam validação para as matérias-primas não previstas no método e rota de produção etílica.

É fato que algumas das propriedades expostas são fáceis de serem medidas, como a massa específica por exemplo. Porém outras, como o índice de cetanos (IC), necessitam de aparelhagem pouco convencional e são, muitas vezes, de difícil acesso para boa parte dos pesquisadores da área.

A partir do método padrão americano ASTM D613, também empregado aqui, a medição do IC deve ser feita a partir de um motor específico chamado CFR (em inglês, Cooperative Fuel Research), com preço estimado em meio milhão de dólares, que possui alto custo operacional e de manutenção, além de necessitar de mão-de-obra altamente especializada (THE DEPARTMENT OF THE ENVIROMENT AND HERITAGE, 2004).

Diante das dificuldades apresentadas, a determinação deste índice através de métodos alternativos, e aqui se entendam como alternativos aqueles procedimentos que provenham o resultado a partir de aparelhagem de baixo (ou menor) custo ou de metodologias que possam substituir a utilização do referido motor, são de grande conveniência. Junta-se a isso o entendimento das relações estrutura x propriedades envolvidas, para uma melhor discussão da viabilidade de utilização de um dado biocombustível em escala comercial.

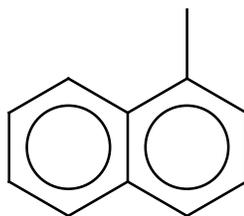
A ideia de utilizar métodos alternativos para a determinação de propriedades relativas a combustíveis não é novidade, LIMA e outros (2004) propõem uma alternativa para a detecção de aditivos indesejados em gasolinas. YANG e outros (2002) usam redes neurais para inferir o índice de cetanos e a densidade de diferentes tipos de óleo diesel a partir de dados de cromatografia líquida e gasosa. RAMIREZ-ARIZMENDI e HAMJE (2008) patentearam um método de determinação do índice de cetanos a partir de dados de infravermelho.

### 1.2.1. O Índice de Cetanos (IC)

As primeiras avaliações dos tipos de diesel se deram de maneira subjetiva. Era perceptível, por inspeções auditivas, que alguns combustíveis faziam com que o motor trabalhasse de maneira mais suave do que outros. Por isso, fazia-se necessário construir uma escala quantitativa para a classificação destes combustíveis. Na década de 30, Boerlage e Brooge desenvolveram, na Holanda, o primeiro procedimento para determinar a qualidade de ignição de um combustível diesel, o índice de cetanos. Neste método foram adotados dois padrões: o hexadec-1-eno (ceteno) (Figura 5), um composto de fácil oxidação, com valor atribuído do índice de cetanos igual a 100, e o  $\alpha$ -metil-naftaleno (Figura 6), de difícil oxidação e com valor atribuído do índice de cetanos igual a 0 (TAYLOR; MCCORMICK; MURPHY, 2004).



**Figura 5.** Estrutura química do hidrocarboneto hexadec-1-eno.



**Figura 6.** Estrutura química do hidrocarboneto  $\alpha$ -metil-naftaleno.

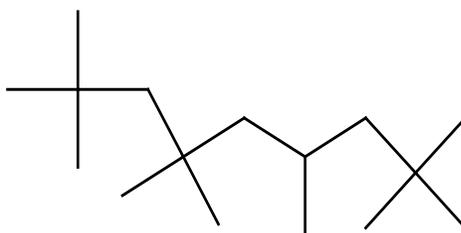
A medida do índice de cetanos era feita comparando-se os desempenhos: do diesel em teste e o de uma mistura destes padrões. O valor deste índice era dado a

partir de uma porcentagem em massa dos padrões que tivesse desempenho semelhante à do diesel em teste (TAYLOR; MCCORMICK; MURPHY, 2004).

Posteriormente, um teste semelhante foi feito nos Estados Unidos da América. Alegando dificuldades experimentais na preparação e purificação dos padrões utilizados por Boerlage e Brooge, os pesquisadores norte-americanos desenvolveram outra escala quantitativa para a classificação dos tipos de diesel, o índice de cetanos. Nesta escala (adotada atualmente), foram utilizados como padrões o hexadecano (em inglês conhecido como cetane) (Figura 7), com valor atribuído igual a 100 e o 2,2,4,4,6,8,8-heptametilnonano (Figura 8), com valor atribuído do índice de cetanos igual a 15 (KNOTHE; DUNN; BAGBY, 1997; TAYLOR; MCCORMICK; MURPHY, 2004).



**Figura 7.** Estrutura química do hidrocarboneto hexadecano.



**Figura 8.** Estrutura química do hidrocarboneto 2,2,4,4,6,8,8-heptametilnonano.

O diesel em teste tem seu desempenho comparado com uma mistura destes padrões, de maneira que o valor do índice de cetanos (IC) é obtido a partir da seguinte equação:

$$\text{IC} = (\% \text{ em vol. de hexadecano}) + 0,15(\% \text{ em vol. de heptametilnonano}) \quad (1)$$

Comparando-se, empiricamente, o método dos holandeses ao dos norte-americanos, chega-se à relação aproximada:

$$\text{Índice de cetanos} = 0,88 \times \text{Índice de cetenas} \quad (2)$$

Para que a combustão aconteça, é necessário que o combustível seja aquecido a certas temperaturas que possibilitem a quebra de ligações químicas formando os radicais livres. Uma vez que é atingida quantidade suficiente dos mesmos, é possível manter estável a reação de oxidação em cadeia, a ignição (TAYLOR; MCCORMICK; MURPHY, 2004). A velocidade deste processo radicalar está diretamente associada à estabilidade dos radicais formados durante a queima. Quanto mais instáveis forem os radicais produzidos, mais reativos eles serão e conseqüentemente mais rapidamente ocorrerá a combustão, e vice-versa (SOLOMONS, 2009).

O IC é o principal indicador da qualidade de um combustível em um motor a diesel. Da mesma forma que a octanagem está para a gasolina, o IC está para o diesel, e consiste em uma maneira relativa de classificar este combustível de acordo com o intervalo entre sua injeção e sua ignição.

O controle deste índice é fundamental, combustíveis com IC's muito altos ou muito baixos podem causar problemas operacionais nos motores. Se o IC for muito alto, a combustão ocorrerá antes do previsto, quando a mistura de ar e combustível no pistão ainda não é a ideal, isso resultará numa combustão incompleta e na emissão de fumaça. Já se o IC for muito baixo, a combustão demorará mais e podem ocorrer falhas

no motor como: trepidação, aquecimento lento ao ser acionado e fenômenos de combustão incompleta. Nos Estados Unidos da América, a maioria dos fabricantes de motores a diesel recomenda uma faixa de IC entre 40 e 50 (KNOTHE; GERPEN; KRAHL, 2005; KNOTHE, 2005).

Referente à estrutura química dos combustíveis, os principais fatores que interferem no IC são o número de ligações duplas carbono-carbono, a estrutura química do álcool que dá origem ao biocombustível e o tamanho da cadeia carbônica das moléculas. (KNOTHE; MATHEAUS; RYAN III, 2003; KNOTHE; BAGBY; RYAN III, 1997).

Para a medição deste indicador, é necessário que as amostras estejam puras, principalmente sem presença de peróxidos, pois estes são conhecidos por serem aditivos eficientes no aumento dos IC's, provocando interferências no valor medido. Os peróxidos podem ser formados a partir da oxidação dos hidrocarbonetos em estoque promovida pelo O<sub>2</sub> presente no ar (TAYLOR; MCCORMICK; MURPHY, 2004).

O IC é uma escala adimensional e o padrão ASTM D613 prevê os limites de reprodutibilidade dos resultados de 2,8 unidades para um IC de 40 (7%), 3,8 unidades para um IC de 48 (8%) e 4,8 unidades para um IC de 56 (8,6%). O padrão ASTM define reprodutibilidade como a diferença entre dois resultados para amostras idênticas, obtidas por operadores diferentes em laboratórios diferentes, que seria ultrapassado em apenas 1 de 20 casos. Em outras palavras, um combustível com IC igual a 48, obtido por operadores diferentes em laboratórios diferentes, teria de ter valores entre 44,2 e 51,8 para 95% das medições (TAYLOR; MCCORMICK; MURPHY, 2004). Diante disto, será utilizada para fins de comparação das simulações feitas com os resultados experimentais, uma margem de erro da ordem de 7% relativo ao valor medido no motor.

### **1.3. Métodos de análise estatística multivariada**

Em um experimento químico, o entendimento dos resultados envolve a análise de grande número de variáveis. Na maioria das vezes, apenas um pequeno número destas variáveis contém as informações mais relevantes para a elucidação de um dado problema, enquanto que a maioria das variáveis pouco, ou nada, adiciona à interpretação dos resultados em termos químicos (DASZYKOWSKI; WALCZAK; MASSART, 2003; NETO; MOITA, 1998).

Em função do elevado número de variáveis, técnicas estatísticas têm sido aplicadas para identificar e quantificar os fatores que causam influência significativa sobre os resultados dos experimentos. Esses métodos são usualmente referenciados como métodos de análise estatística multivariada. Os métodos multivariados consideram a correlação entre múltiplas variáveis analisadas simultaneamente (BRUNS; FAIGLE, 1985; BEEBE; KOWALSKI, 1987; TADINI; PARK; DING, 1997).

Neste trabalho foram utilizados seis métodos multivariados: PCA, análise de cluster (hierárquico, não hierárquico e fuzzy), redes neurais e regressão não linear. Com o objetivo de relacionar as estruturas químicas das moléculas que compõem diferentes tipos de biodiesel, com seus respectivos índices de cetanos.

#### **1.3.1. PCA**

No campo da modelagem empírica (via métodos estatísticos), pode-se destacar a PCA (em inglês, principal component analysis), uma vez que este tipo de análise multivariada atua buscando reduzir a dimensionalidade de uma dada matriz de dados.

Neste método, as  $n$ -variáveis originais geram, através de suas combinações lineares,  $m$ -componentes principais (onde  $m < n$ ), cuja principal característica é que estes componentes são capazes de reproduzir o conjunto de dados. Os componentes principais são obtidos em ordem decrescente de variância, ou seja, a componente principal 1 (PC1) detém mais informação estatística que a componente principal 2 (PC2), que possui mais informação estatística que a componente principal 3 (PC3), e assim sucessivamente.

Por ser um método exploratório, a PCA auxilia na elaboração de hipóteses gerais a partir dos dados coletados, contrastando com estudos direcionados nos quais hipóteses prévias são testadas. Este método permite a redução da dimensionalidade do espaço vetorial onde os pontos se encontram representados, sendo capaz de separar a informação importante da redundante e aleatória (NETO; MOITA, 1998; FERREIRA *et al*, 2002). Combinações lineares de menor dimensão são geralmente mais fáceis de interpretar, e servem como passo intermediário em análises de dados mais complexos (HARDLE; SIMAR, 2003).

Uma séria limitação da PCA reside no fato desta ser uma técnica que pressupõe relações lineares entre as variáveis. Como consequência, resulta uma espécie de proliferação de componentes principais associados a relações não lineares entre variáveis. Independente disso, a PCA pode funcionar muito bem como uma primeira aproximação, particularmente para sistemas que apresentam grande variabilidade. A interpretação dos resultados, no entanto, se torna mais simples quando se conjuga a PCA com técnicas de análise de aglomerados (“clusters”) (LAROSE, 2006; VENABLES; RIPLEY, 2002).

### **1.3.2. Análise de “Cluster”**

A análise de aglomerados visa à identificação de agrupamentos em um conjunto de dados. O número de componentes independentes  $k$  identificado pela PCA pode ser utilizado como restrição para o algoritmo de análise de aglomerados, no sentido de dividir o espaço vetorial definido pelo conjunto de dados em  $k$  regiões distintas, representando classes de similaridade em que os dados devem se distribuir.

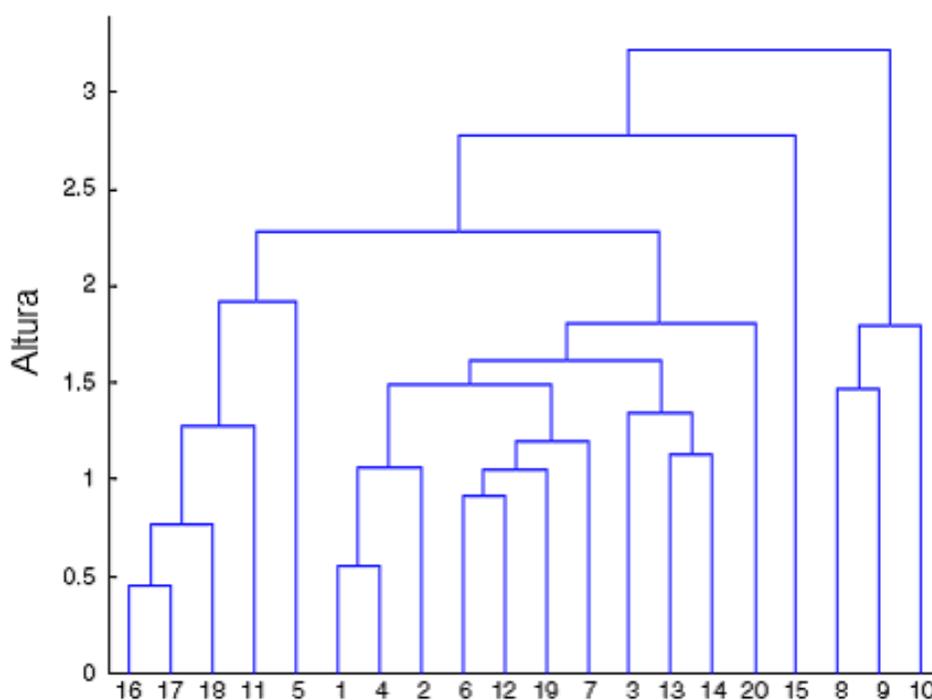
Uma característica importante deste tipo de análise é que muitas propriedades podem ser comparadas simultaneamente, sendo que cada propriedade pode ser representada em uma dimensão no espaço multidimensional (VENABLES; RIPLEY, 2002; MAQBOOL; BABRI, 2007; ERCANLI; BOYD, 2005).

Este tipo de análise tende a encontrar estruturas em conjuntos de dados. Esta ferramenta pode ser utilizada para uma análise exploratória dos dados, objetivando classificar objetos diferentes em um grupo, de maneira que os objetos de um dado “cluster” tenham alto grau de similaridade entre si e pouca com os pertencentes aos demais grupos. Este tipo de análise tem sido utilizado em muitos campos da ciência, como: engenharia, zoologia, medicina, antropologia, psicologia, marketing e geologia (CORTÉS; PALMA; WILSON, 2007; ALDENDERFER; BLASHFIELD, 1984; KAUFMAN; ROUSSEEUW, 1990; JONGMAN; TER BRAAK; VAN TONGEREN, 1995; CORTER, 1996; LEGENDRE; LEGENDRE, 1998; PILLAR, 1999; JERRAM; CHEADLE, 2000; EVERITT; LANDAU; LEESE, 2001).

Nos últimos trinta anos, mais de setenta algoritmos deste tipo foram relatados (XU; ZHANG; SHIH, 2006). Neste trabalho serão utilizados 3 tipos de “clustering”: hierárquico, não hierárquico e “fuzzy clustering”.

### 1.3.2.1. “Clustering” Hierárquico

Os métodos de “clustering” hierárquico são implementados usando um procedimento de maneira a aglomerar os objetos do conjunto de dados, dispondo-os de maneira hierárquica. Os resultados do tratamento são representados num diagrama em árvore chamado “dendrograma” (Figura 9) (ERCANLI; BOYD, 2005; XU; ZHANG; SHIH, 2006).



**Figura 9.** Representação gráfica dos resultados do “clustering” hierárquico (dendrograma).

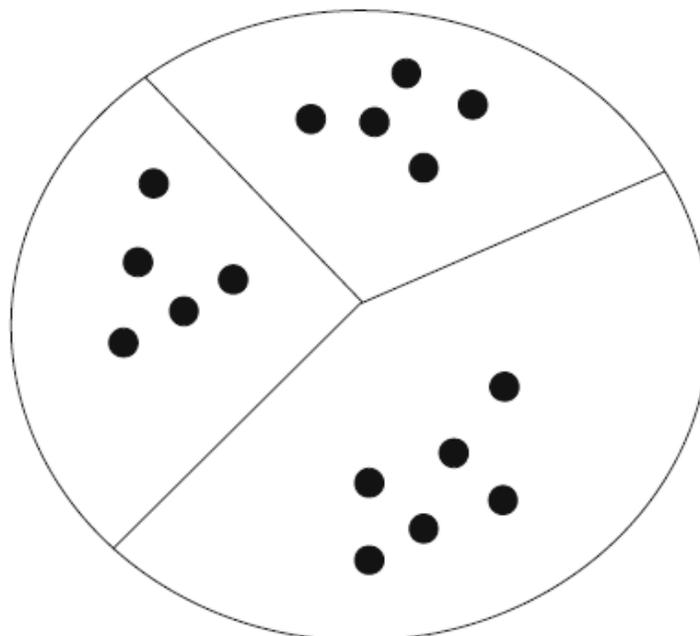
No início da fase de aglomeração (altura igual a zero), cada objeto compõe um “cluster”, de maneira que o número de “clusters” é igual ao número de objetos da matriz de dados. Utilizando a distância entre os pontos, os dois mais próximos são fundidos formando um novo, que substitui os antigos. Essa fusão ocorre até que, no final, reste apenas um “cluster” (XU; ZHANG; SHIH, 2006).

Neste método, inicialmente, é calculada a matriz (D) de dissimilaridades entre cada ponto do conjunto de dados. A aproximação dos dados (aglomeração) se dá utilizando a medida dos vizinhos mais próximos para a construção desta matriz D. Existe diversas maneiras de se calcular as distâncias entre as observações (de Mahalanobis, de Hausdorff, de Minkowski, de Hamming, de Levenshtein, de Kullback–Leibler), neste trabalho utilizou-se a distância euclidiana por ser a mais conhecida além de ser o padrão adotado pelo software de trabalho. (DE MAESSCHALCK; JOUAN-RIMBAUD; MASSART, 2000; HUTTENLOCHER; KLANDERMAN; RUCKLIDGE, 1993; GROENEN; MATHAR; HEISER, 1995; D'HULST; RODGERS, 1999; YUJIAN; BO, 2007; BIGI, 2003).

A avaliação do dendrograma se dá observando a altura das conexões entre as observações, quanto menor for esta altura, menor será a distância entre estas, ou seja, maior será o grau de similaridade. O contrário ocorre quando observamos as conexões de maior altura, neste caso observaríamos o grau de dissimilaridade entre os componentes do banco de dados.

### **1.3.2.2. “Clustering” não hierárquico (k-médias)**

Os métodos de “clustering” não hierárquico são implementados de maneira a distribuir os n-dados em k-“clusters” ( $n > k$ ), sendo que k deve ser previamente determinado. Em muitos casos, o valor de k não é conhecido, o que representa um problema adicional (XU; ZHANG; SHIH, 2006). Neste tipo de tratamento, um elemento pertence exclusivamente a um “cluster” (Figura 10).



**Figura 10.** Partição de um conjunto de dados com 16 elementos em 3 “clusters” (GRUBESIC, 2006).

Na prática, tais algoritmos são reiniciados várias vezes, de maneira que o ajuste das posições dos centros dos “clusters” é repetido até que um critério de convergência (erro satisfatório) seja atingido. Logo, esta se torna uma tarefa computacionalmente cara quando defrontamos grandes conjuntos de dados (DAVIDSON; RAVI, 2007).

### 1.3.2.3. “Fuzzy clustering”

O “fuzzy clustering” é uma técnica de análise que ajuda a melhor entender as ambigüidades presentes num dado conjunto de dados (NETO; MOITA, 1998; KASABOV, 1998).

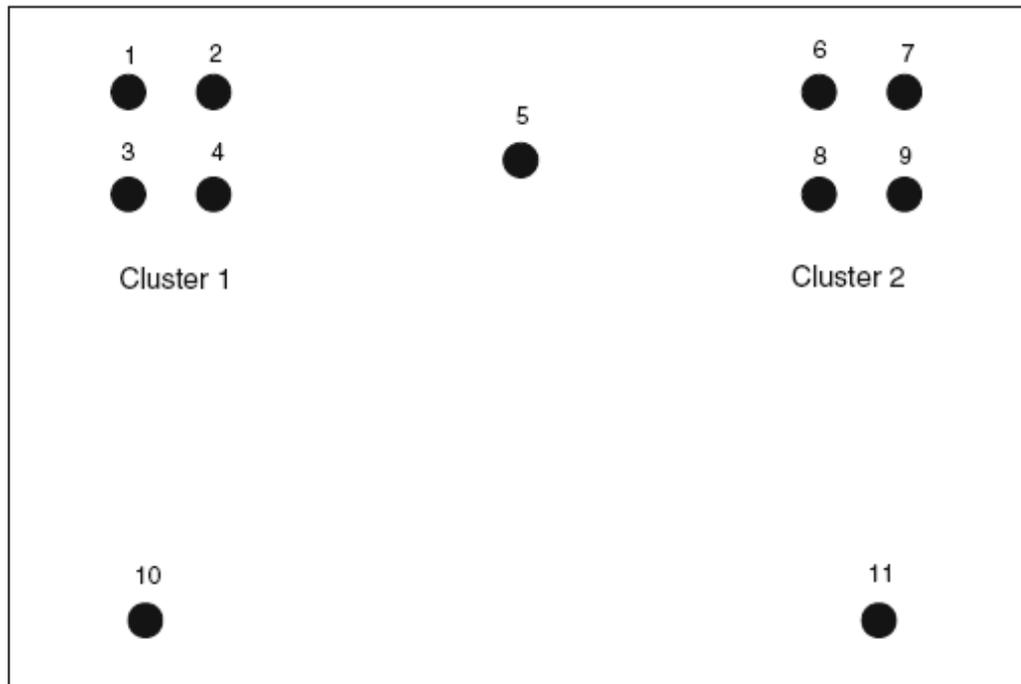
Diferentemente do método anterior, onde ou um elemento pertence a um “cluster”, ou pertence a outro, este tratamento se baseia em uma lógica contínua (SILER; BUCKLEY, 2005), isto é, que admite valores-verdade entre 0 e 1. A teoria de

conjuntos difusos leva a funções de pertinência contínuas no intervalo [0; 1]. Em outros termos, um ponto do conjunto de dados pode tanto pertencer exclusivamente a uma classe como ter sua função de pertinência distribuída entre outras, com a restrição de que a soma destas funções seja 1. Este método considera que não existe uma fronteira rígida e abrupta entre conjuntos, o que implica na possibilidade da sobreposição de “clusters” (VENABLES; RIPLEY, 2002; KASABOV, 1998).

O grau de pertinência de um ponto a um cluster é calculado em função da sua distância ao centro deste aglomerado. Desta forma, os pontos na fronteira de um cluster devem pertencer a este em menor grau quando comparados a outros que estejam mais próximos ao centro. Em outras palavras, qualquer variação na distância de um ponto ao centro de um cluster implicará na alteração da pertinência deste ponto ao conjunto em questão (KAYMAK; SETNES, 2002; YANG, 1993).

Este tipo de classificação flexível é particularmente importante para os tipos de biodiesel, pois estes são misturas de ésteres com estruturas químicas bastante distintas entre si. A possibilidade de pertinência a diferentes clusters em valores de 0 a 1 torna-se de grande utilidade na estimativa de propriedades para misturas (NADAI; MIRANDA; GATTS, 2009).

A Figura 11 a seguir exemplifica um resultado obtido deste tratamento. Pode-se observar que, o ponto 5 não pertence totalmente ao “cluster 1” nem totalmente ao “cluster 2”, e sim pertence aos dois em uma dada proporção.



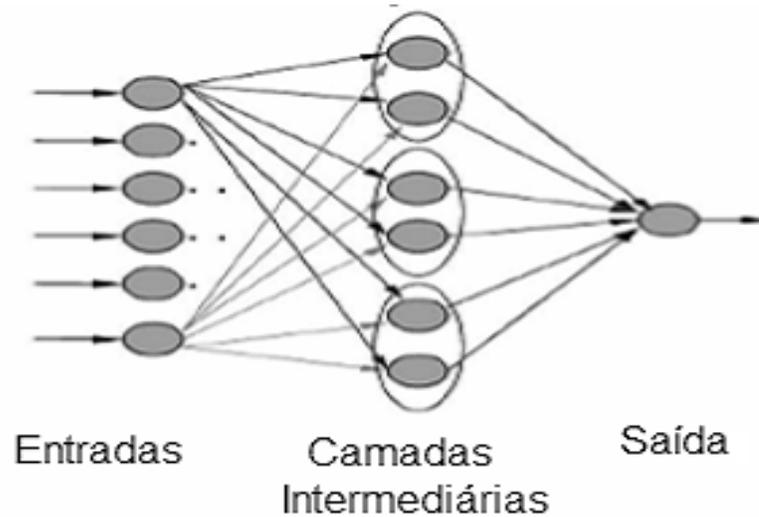
**Figura 11.** Classificação de um conjunto segundo a “fuzzy clustering” (GRUBESIC, 2006).

### 1.3.3. Redes Neurais

As redes neurais são algoritmos de computação não linear, que podem ilustrar uma interpretação física de características morfológicas através do reconhecimento de conexões nos dados de entrada ao sistema adaptável (AL-KHEDHER *et al*, 2007; CANAKCI; ERDIL; ARCAKLIOGLU, 2006).

As redes neurais têm seu funcionamento inspirado em sistemas nervosos biológicos. São compostas por elementos simples (neurônios) operando em paralelo, de maneira que o funcionamento do software é determinado pelas conexões entre seus elementos (HANSELMAN; LITTLEFIELD, 2003). Para este tratamento dos dados, é necessário que o banco de dados seja dividido em dois conjuntos: um conjunto para o treino da rede e o conjunto de validação dos resultados (YANG *et al*, 2002).

As redes neurais podem ser divididas em 3 partes: entradas (onde são inseridos os dados), camadas intermediárias (onde são processados os dados) e saída (onde são obtidos os resultados) (Figura 12).

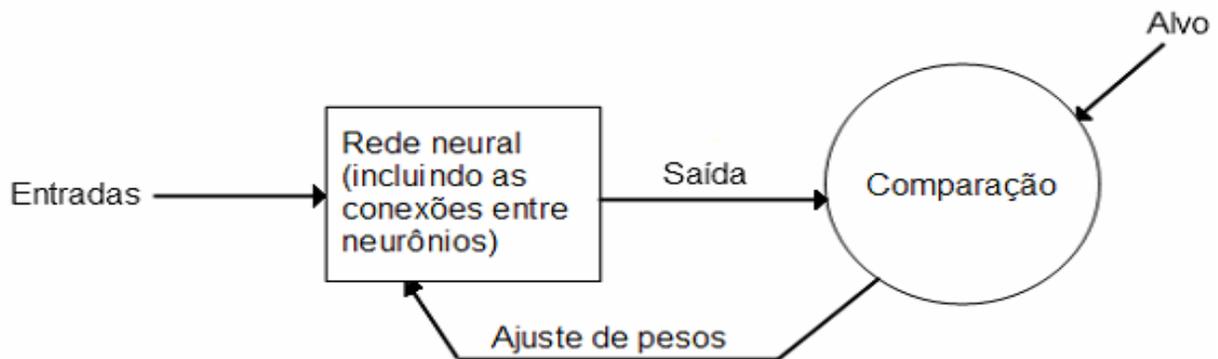


**Figura 12.** Esquema geral de uma rede neural.

A rede neural utilizada neste trabalho é a *feed-forward backpropagation*. Estas redes atuam como modelos de regressão não linear com flexibilidade suficiente para aproximar qualquer função suave (VENABLES; RIPLEY, 2002). Algoritmos deste tipo têm sido utilizados com sucesso em diferentes campos da ciência, como matemática, engenharia, medicina, economia, meteorologia, psicologia, robótica e neurologia (RAMADHAS *et al*, 2006; JOHNSON; BALKENIUS, 2007).

Este tipo de rede tem treinamento supervisionado (Figura 13), ou seja, pode ser treinada de maneira a ajustar os valores das conexões (pesos das conexões neurais) entre seus elementos, isso com o intuito de fazer com que as entradas levem a uma saída (resultado obtido) específica da rede, que por sua vez será comparada com o

alvo (resultado desejado). Este treino é processado até que a saída obtida seja compatível com o alvo desejado (HANSELMAN; LITTLEFIELD, 2003).



**Figura 13.** Esquema do treinamento supervisionado.

A rede neural é conhecida como um método de modelagem do tipo “caixa-preta”, uma vez que, a partir do conjunto de dados utilizado para treinamento da rede, foram determinados os pesos dados às conexões de cada neurônio em cada camada, não há como modificar os mesmos, impossibilitando o operador de interferir no modelo.

#### **1.3.4. Modelos de regressão**

Os métodos de regressão relacionam o comportamento de uma variável problema com pelo menos outra variável de entrada. Quando o modelo envolve duas variáveis este é chamado de simples, e caso o mesmo envolva mais de duas variáveis este será chamado de multivariado. Este relacionamento pode ser expresso por uma equação linear ou uma função não linear. A análise de regressão compreende, portanto, quatro tipos básicos de modelos:

- linear simples.
- não linear simples.
- linear multivariado.
- não linear multivariado.

Neste trabalho, utilizou-se um modelo de regressão não linear multivariado. Inicialmente são atribuídos valores aleatórios para os parâmetros ajustáveis (as constantes e os expoentes das variáveis em questão) do modelo, o processo de otimização consiste em ajustar os parâmetros fornecidos usando como critério a maximização do coeficiente de correlação entre os valores alvo e os simulados. O coeficiente de correlação obtido serve como medida do grau de relacionamento entre os conjuntos comparados, seu valor estará compreendido no intervalo  $[0;1]$ , e quanto maior ele for, mais explicativo será o modelo em questão.

O campo de utilização deste tipo de modelagem é bastante amplo, pode-se, por exemplo, utilizar um modelo de regressão para fazer previsões futuras de uma dada propriedade com base em dados de uma série temporal. Ou, pesquisadores interessados em simular os efeitos sobre uma variável A, em decorrência de alterações introduzidas nos valores de uma variável B, também podem optar por esta metodologia.

Os modelos de regressão estão entre os métodos estatísticos mais usados para investigar a relação entre variáveis e têm sido empregados em diversas áreas do conhecimento, tais como: computação, administração, engenharias, biologia, agronomia, saúde, sociologia, química, medicina, etc. (DICKMAN *et al*, 2004; BUNNELL, 2006).

#### 1.4. A Técnica de RMN-<sup>1</sup>H

Para o estudo das estruturas químicas dos ésteres que compõem o biodiesel, foi escolhida a espectroscopia de ressonância magnética nuclear de hidrogênio (RMN-<sup>1</sup>H), por ser uma técnica muito rápida, altamente versátil, sem restrições quanto ao ponto de ebulição da amostra e quantitativa. Essa técnica é reconhecida como um dos mais poderosos métodos analíticos para identificação de grupos estruturais numa molécula.

Nesta técnica, as amostras são submetidas a um campo magnético na região das ondas de rádio, dependendo da estrutura química da molécula ocorre absorção em frequências características. O espectro de RMN-<sup>1</sup>H é um registro gráfico das frequências dos picos de absorção dos átomos de hidrogênio ali presentes contra suas intensidades (SILVERSTEIN; WEBSTER, 2000).

A frequência de absorção dependerá da densidade eletrônica “ao redor” do átomo de hidrogênio em questão, fazendo uma aproximação, pode-se dizer que, a frequência em que um dado <sup>1</sup>H varia com o efeito indutivo dos grupos vizinhos a este átomo. Por exemplo, um átomo de hidrogênio ligado a um carbono saturado, terá deslocamento químico diferente de outro que esteja ligado a um carbono insaturado.

Sabendo que os principais centros universitários do país possuem algum modelo deste espectrômetro, e uma vez que a ideia do trabalho é partir destes dados para obter o IC, torna-se muito mais simples obter um espectro de RMN-<sup>1</sup>H de um determinado biodiesel e relacioná-lo ao IC do que realizar um teste num motor padrão, isso sem considerar o menor preço e a maior possibilidade de utilização de um espectrômetro frente ao motor CFR quando se pensa em futuras aquisições.

## 2. OBJETIVO

Tendo em vista as dificuldades experimentais expostas e o sucesso obtido anteriormente pelos pesquisadores mencionados para as tarefas a que se propuseram, o objetivo é desenvolver métodos alternativos inéditos utilizando algumas técnicas de análise estatística multivariada com base, inicialmente a partir de espectros de RMN-<sup>1</sup>H e posteriormente a partir de descritores estruturais das moléculas que compõem diferentes tipos de biodiesel, para a determinação de seus respectivos IC's.

### 3. OBJETIVOS ESPECÍFICOS

#### *Capítulo I – Análise exploratória*

A partir de espectros teóricos de RMN-<sup>1</sup>H e dados de IC disponíveis na literatura, construir um banco com dados contendo ésteres (substâncias simples) e biocombustíveis (misturas de ésteres).

Tratar este banco utilizando diferentes técnicas de análise multivariada, com a intenção de testar sua aplicabilidade como base de dados para a predição do IC de diferentes tipos de biodiesel.

#### *Capítulo II – Modelo empírico*

Utilizando o banco de dados previamente testado, fazer a predição do IC de diferentes tipos de biodiesel (misturas) a partir de dados de substâncias simples (ésteres) utilizando redes neurais.

#### *Capítulo III – Modelo preditivo*

No intuito de melhorar e simplificar o processo, construir uma nova matriz de dados utilizando descritores estruturais das moléculas componentes dos biocombustíveis.

Desenvolver um modelo preditivo a partir da nova matriz de dados e do método de regressão não linear, tendo como alvo determinar o IC destes biocombustíveis e estabelecer as relações estrutura x propriedade envolvidas.

## CAPÍTULO I – Análise exploratória

### I.1. Métodos

#### I.1.1. Formatação do Banco de dados

Inicialmente foram identificados os ésteres que constituem os mais variados tipos de biodiesel com dados de IC disponíveis na literatura, resultando num total de 35 ésteres (Tabela 3).

**Tabela 3.** Ésteres usados com seus respectivos índices de cetanos.

Éster	Ácido							
	Mirístico	Palmítico	Palmitoléico	Esteárico	Oléico	Linoléico	Linolênico	Erúico
Metílico	66,2 <sup>c</sup>	85,9 <sup>a</sup>	51,0 <sup>a</sup>	101,0 <sup>a</sup>	59,3 <sup>a</sup>	28,5 <sup>b</sup>	20,6 <sup>b</sup>	76,0 <sup>b</sup>
Etílico	66,9 <sup>c</sup>	93,1 <sup>a</sup>		97,7 <sup>a</sup>	67,8 <sup>a</sup>	39,6 <sup>a</sup>	22,7 <sup>c</sup>	
Propílico	71,0 <sup>d</sup>	85,0 <sup>a</sup>		90,9 <sup>a</sup>	58,8 <sup>a</sup>	44,0 <sup>a</sup>	26,7 <sup>c</sup>	
Isoprílico		82,6 <sup>a</sup>		96,5 <sup>a</sup>	86,6 <sup>a</sup>			
Butílico	73,0 <sup>d</sup>	91,9 <sup>a</sup>		92,5 <sup>a</sup>	61,6 <sup>a</sup>	53,5 <sup>a</sup>	26,8 <sup>c</sup>	
Isobutílico		83,6 <sup>a</sup>		99,3 <sup>a</sup>	59,6 <sup>a</sup>			
2-Etilexílico		98,2 <sup>a</sup>		115,5 <sup>a</sup>	88,2 <sup>a</sup>			

<sup>a</sup> (KNOTHE; MATHEAUS; RYAN III, 2003). <sup>b</sup> (KNOTHE; DUNN; BAGBY, 1997).

<sup>c</sup> (KNOTHE, 2005). <sup>d</sup> (SERDARI; LOIS; STOURNAS, 1999).

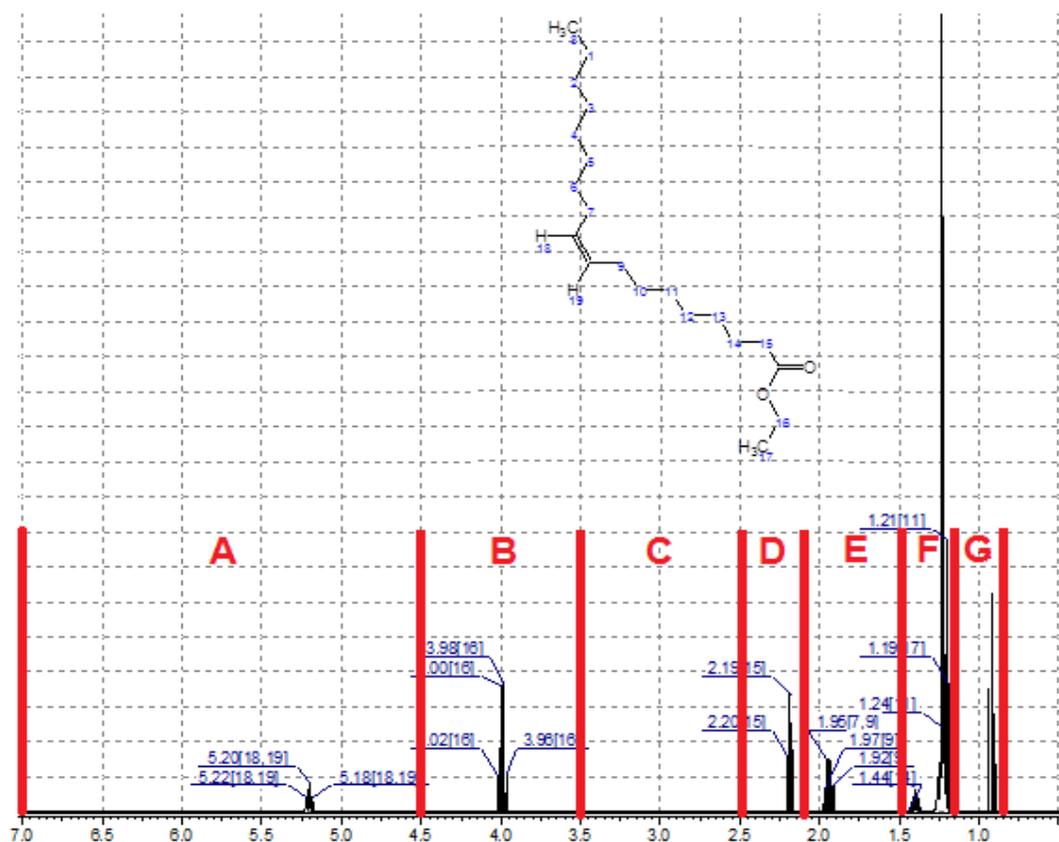
O passo seguinte foi simular os espectros de RMN-<sup>1</sup>H destes ésteres utilizando o programa gNMR (banco de espectros em anexo, Figuras 25 a 61) ([home.cc.umanitoba.ca/~budzelaa/gNMR/gNMR-license.html](http://home.cc.umanitoba.ca/~budzelaa/gNMR/gNMR-license.html)).

A pesquisa se estendeu para os tipos de biodiesel. O critério para sua escolha foi utilizar aqueles que tivessem índices de cetanos disponíveis na literatura. Assim, foram obtidos 9 tipos de biodiesel (Tabela 4), resultando num total de 44 elementos no banco de dados (35 ésteres e 9 tipos de biodiesel).

**Tabela 4.** Tipos de biodiesel usados com seus respectivos índices de cetanos. Todos os valores foram obtidos de KNOTHE; DUNN; BAGBY (1997).

Éster	Óleos					
	Açafrão	Colza	Palma	Soja	Girassol	Algodão
Metílico	49,8	54,4		46,2	46,6	51,2
Etílico			56,2	48,2		
Isopropílico				52,6		
Butílico				51,7		

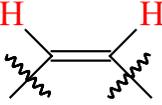
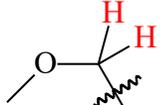
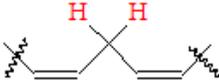
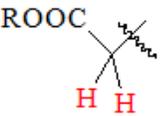
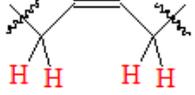
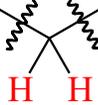
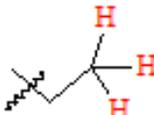
Com base nos espectros de RMN-<sup>1</sup>H simulados (exemplificado pelo oleato de etila na Figura 14), foi possível determinar 7 regiões específicas (nomeadas de A-G), referentes aos diferentes tipos de hidrogênios encontrados nestas moléculas. Na formatação do banco de dados foi utilizada a integração dos sinais obtidos em cada fatia para os espectros dos ésteres.



**Figura 14.** Ilustração das sete fatias do espectro para o oleato de etila.

Os valores dos intervalos de cada fatia do espectro, bem como os tipos de hidrogênio a que estas fatias correspondem, estão representados na Tabela 5.

**Tabela 5.** Variação em ppm das sete fatias obtidas do espectro de RMN-<sup>1</sup>H e representação das características químicas correspondentes a cada região do espectro.

Fatia do Espectro	Tipo de Hidrogênio correspondente	Estrutura química	Variação dos sinais obtidos	
			Máximo	Mínimo
A	Vinílico		7,00	4,50
B	Carbinólico		4,49	3,50
C	Bis-alfílico		3,49	2,50
D	$\alpha$ -carbonílico		2,49	2,10
E	Alílico		2,09	1,50
F	Parafínico		1,49	1,16
G	Metila terminal		1,15	0,85

Na construção do banco a partir de dados dos tipos de biodiesel, por não se conhecer a composição química das misturas queimadas que originaram os respectivos resultados experimentais, foi necessário, como forma de aproximação, calcular as porcentagens médias (obtidas a partir da Tabela 1) de cada éster que compunha o respectivo biodiesel. Vale menção o fato que os diferentes tipos de biodiesel são

constituídos por misturas de ésteres, sendo que a quantidade de cada componente varia com uma série de fatores, tais como, diferenças climáticas, de solo e fatores de estresse a que o vegetal pode ser submetido durante a colheita (MICHAEL; ROBERT, 1998; FANGRUI; MILFORD, 1999; AYHAN, 2003).

Tendo as porcentagens médias e os espectros de RMN-<sup>1</sup>H dos ésteres componentes de cada mistura, foi possível calcular o número de átomos de hidrogênio equivalentes em cada uma das sete fatias para os tipos de biodiesel. Para tanto, foi utilizado o mesmo procedimento (anteriormente aplicado às substâncias simples) em cada éster componente, e posteriormente foi multiplicando o valor obtido em cada uma das fatias pelas respectivas porcentagens (atribuídas a partir da Tabela 1). Por fim, foram somados os valores em cada fatia de cada componente resultando nos valores obtidos para as misturas.

Como parte do pré-tratamento do banco foi necessário normalizar os dados, isto é, a quantidade de H em cada fatia do espectro foi dividida pelo número total de átomos de hidrogênio da respectiva molécula. A partir de dados normalizados é possível fazer a comparação de moléculas com diferentes quantidades deste átomo em suas composições. O banco de dados obtido é mostrado na Tabela 6.

**Tabela 6.** Banco de dados com as proporções de hidrogênio em cada fatia do espectro.

Biocombustíveis	Fatias do espectro						
	A	B	C	D	E	F	G
Miristato de metila	0	0,1	0	0,0666	0	0,7333	0,1
Palmitato de metila	0	0,0882	0	0,0588	0	0,7647	0,0882
Palmitoleato de metila	0,0625	0,0937	0	0,0625	0,125	0,5625	0,0937
Estearato de metila	0	0,0789	0	0,0526	0	0,7894	0,0789
Oleato de metila	0,0555	0,0833	0	0,0555	0,1111	0,6111	0,0833
Linoleato de metila	0,1176	0,0882	0,0588	0,0588	0,1176	0,4705	0,0882
Linolenato de metila	0,1875	0,0937	0,125	0,0625	0,125	0,3125	0,0937
Miristato de etila	0	0,0625	0	0,0625	0	0,7812	0,0937
Palmitato de etila	0	0,0555	0	0,0555	0	0,8055	0,0833
Linoleato de etila	0,1111	0,0555	0,0555	0,0555	0,1111	0,5277	0,0833
Estearato de etila	0	0,05	0	0,05	0	0,825	0,075
Oleato de etila	0,0526	0,0526	0	0,0526	0,1052	0,6578	0,0789
Linolenato de etila	0,1764	0,0588	0,1176	0,0588	0,1176	0,3823	0,0882
Estearato de propila	0	0,0476	0	0,0476	0	0,7619	0,1428
Estearato de isopropila	0	0,0238	0	0,0476	0	0,7142	0,2142
Estearato de butila	0	0,0454	0	0,0454	0	0,7727	0,1363
Palmitato de isopropila	0	0,0263	0	0,0526	0	0,8421	0,0789
Palmitato de butila	0	0,05	0	0,05	0	0,75	0,15
Palmitato de isobutila	0	0,05	0	0,05	0	0,675	0,225
Palmitato de 2-etilexila	0	0,0416	0	0,0416	0	0,7291	0,1875
Miristato de propila	0	0,0588	0	0,0588	0	0,7058	0,1764
Miristato de butila	0	0,0555	0	0,0555	0	0,7222	0,1666
Oleato de propila	0,05	0,05	0	0,05	0,1	0,6	0,15
Oleato de isopropila	0,05	0,025	0	0,05	0,1	0,7	0,075
Oleato de butila	0,0476	0,0476	0	0,0476	0,0952	0,6190	0,1428
Oleato de isobutila	0,0476	0,0476	0	0,0476	0,0952	0,5476	0,2142
Oleato de 2-etilexila	0,04	0,04	0	0,04	0,08	0,62	0,18
Estearato de 2-etilexila	0	0,0384	0	0,0384	0	0,75	0,1730
Linoleato de propila	0,1052	0,0526	0,0526	0,0526	0,1052	0,4736	0,1578
Linoleato de butila	0,1	0,05	0,05	0,05	0,1	0,5	0,15
Linolenato de propila	0,1666	0,0555	0,1111	0,0555	0,1111	0,3333	0,1666
Linolenato de butila	0,1578	0,0526	0,1052	0,0526	0,1052	0,3684	0,1578
Erucicato de metila	0,0454	0,0681	0	0,0454	0,0909	0,6818	0,0681
Algodão (metílico)*	0,0717	0,0870	0,0305	0,0580	0,0822	0,5832	0,0870
Colza (metílico)*	0,0631	0,0773	0,0153	0,0515	0,0955	0,6196	0,0773
Açafroa (metílico)*	0,0953	0,0870	0,0447	0,0580	0,1012	0,5264	0,0870
Soja (metílico)*	0,0901	0,0867	0,0396	0,0578	0,1009	0,5377	0,0867
Girassol (metílico)*	0,0869	0,0862	0,0342	0,0575	0,1052	0,5434	0,0862
Palma (etílico)*	0,0337	0,0540	0,005	0,0540	0,0574	0,7146	0,0810
Soja (etílico)*	0,0851	0,0546	0,0374	0,0546	0,0954	0,5905	0,0820
Soja (isopropílico)*	0,0806	0,0259	0,0354	0,0518	0,0904	0,5366	0,1790
Soja (butílico)*	0,0766	0,0492	0,0336	0,0492	0,0859	0,5572	0,1478

\* Os nomes entre parênteses representam os álcoois utilizados na produção dos respectivos tipos de biodiesel (vide Figura 2).

### **1.1.2. O ambiente estatístico**

O tratamento dos dados foi realizado, em todas as etapas deste trabalho, com a utilização do ambiente estatístico R (<http://www.r-project.org>). O R foi escolhido por ser livremente distribuído sob condições estabelecidas na GPL (General Public License) da FSF (Free Software Foundation) (<http://www.fsf.org>), além de contar com extensa documentação e suporte de uma rede de usuários muito ativa.

## **1.2. Resultados e Discussão**

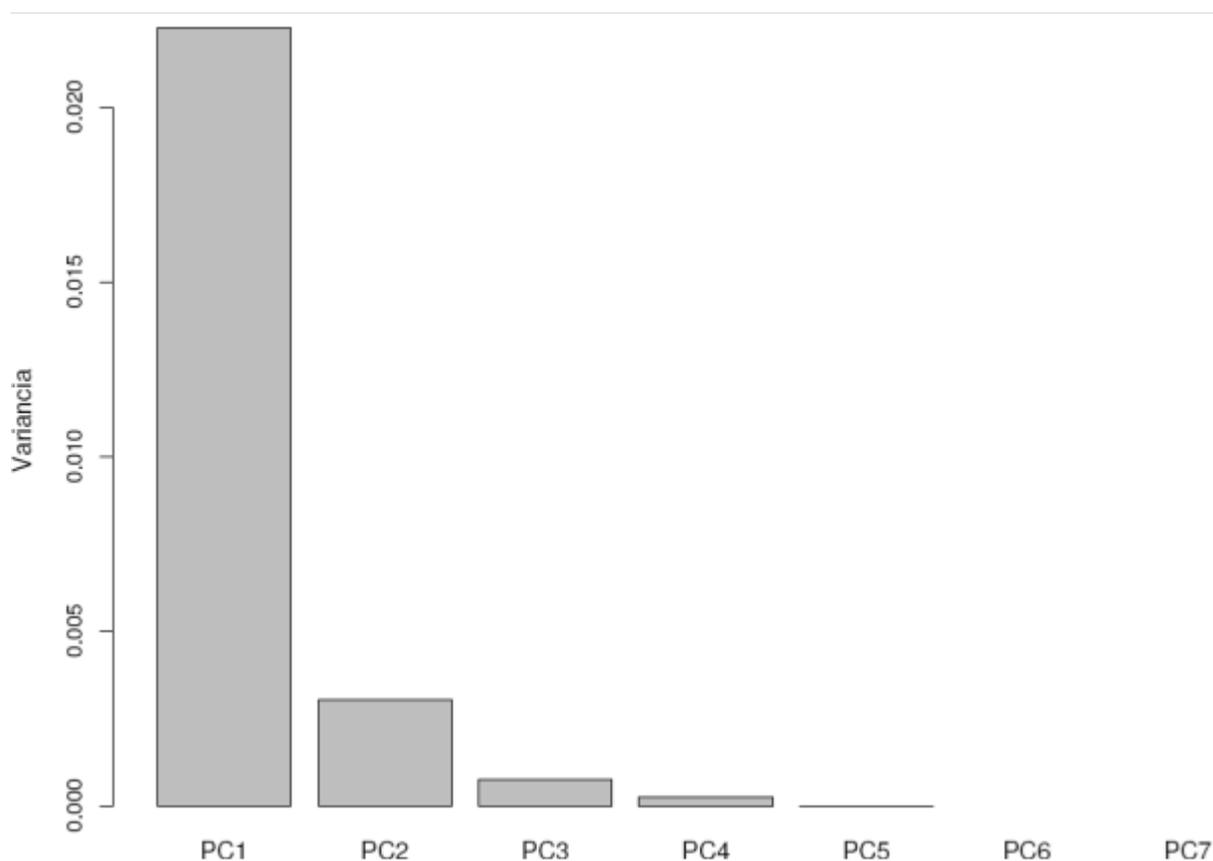
### **1.2.1. Estudo da aplicabilidade do banco de dados**

A construção da matriz de dados foi o estágio inicial do trabalho. Para que este banco possa ser considerado como “aplicável” devemos verificar se as propriedades químicas implícitas nele são discrimináveis estatisticamente e eficientes para fazer algum tipo de previsão relativa ao IC.

Nesta etapa do trabalho, foram utilizados quatro métodos de análise (PCA, “Clustering” hierárquico, “Clustering” não hierárquico e “Fuzzy Clustering”) sobre a matriz de dados, com o intuito de testar o desempenho de cada algoritmo. A proposta é verificar qual deles separa mais adequadamente as classes de similaridade convenientes na predição dos índices de cetanos, que seriam os principais fatores que interferem nesta propriedade.

### 1.2.1.1. Tratamento dos Dados por PCA

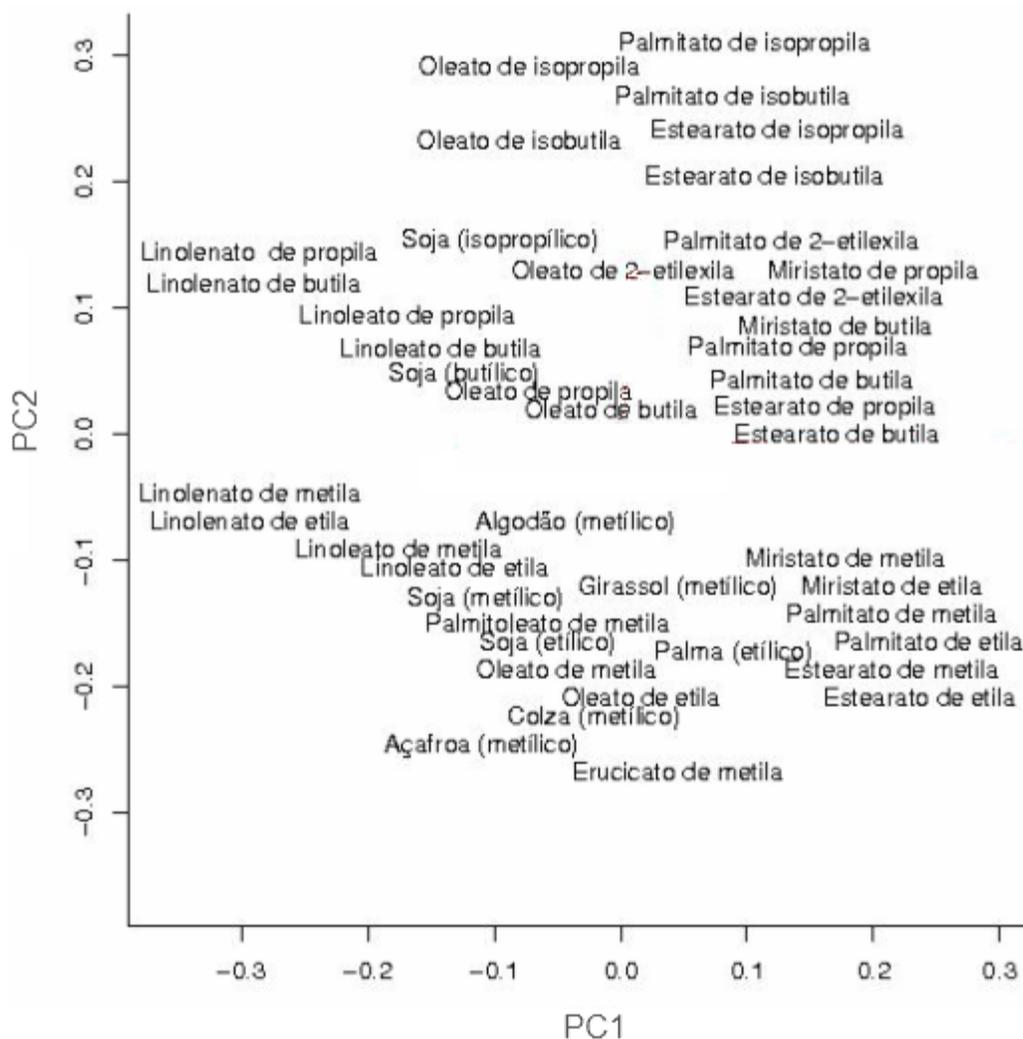
A partir das 7 variáveis de entrada (fatias A-G dos espectros de RMN-<sup>1</sup>H), a PCA indicou a existência de 5 componentes principais (Gráfico 1).



**Gráfico 1.** Resultado das componentes principais resultantes da PCA.

Enquanto os componentes principais (PC's) maiores são úteis para revelar estruturas nos dados, os de menor valor estão associados a eventos sutis, episódicos ou a ruído. Por isso, é importante ter cuidado na definição de um limite mínimo acima do qual os PC's serão considerados, pois informações relevantes podem ser desprezadas. Neste caso, por estarmos trabalhando com espectros simulados, não

existem ruídos presentes. Isso explica porque os autovalores associados aos componentes 6 e 7 são nulos. O PC1 somado ao PC2 equivale a 83,65% da variância do sistema.



**Figura 15.** Distribuição do conjunto de dados no plano PC1 x PC2.

A partir da Figura 15, que representa a projeção da nuvem de dados no plano definido pelo primeiro e segundo componentes principais, pode-se visualizar uma boa discriminação dos dados feita por este método. Ésteres com estruturas químicas

similares estão próximos entre si. Por exemplo, o palmitoleato de metila e o oleato de metila, que diferem em suas estruturas químicas apenas por duas unidades de  $\text{CH}_2$  (vide Figura 3), da mesma forma o miristato de etila e palmitato de etila, assim como vários outros.

Os biocombustíveis estão próximos dos ésteres que são seus principais componentes (vide Tabela 1). Por exemplo, o biodiesel de soja (metílico) ao linoleato de metila, da mesma forma o biodiesel de palma (etílico) ao palmitato de etila e ao estearato de etila.

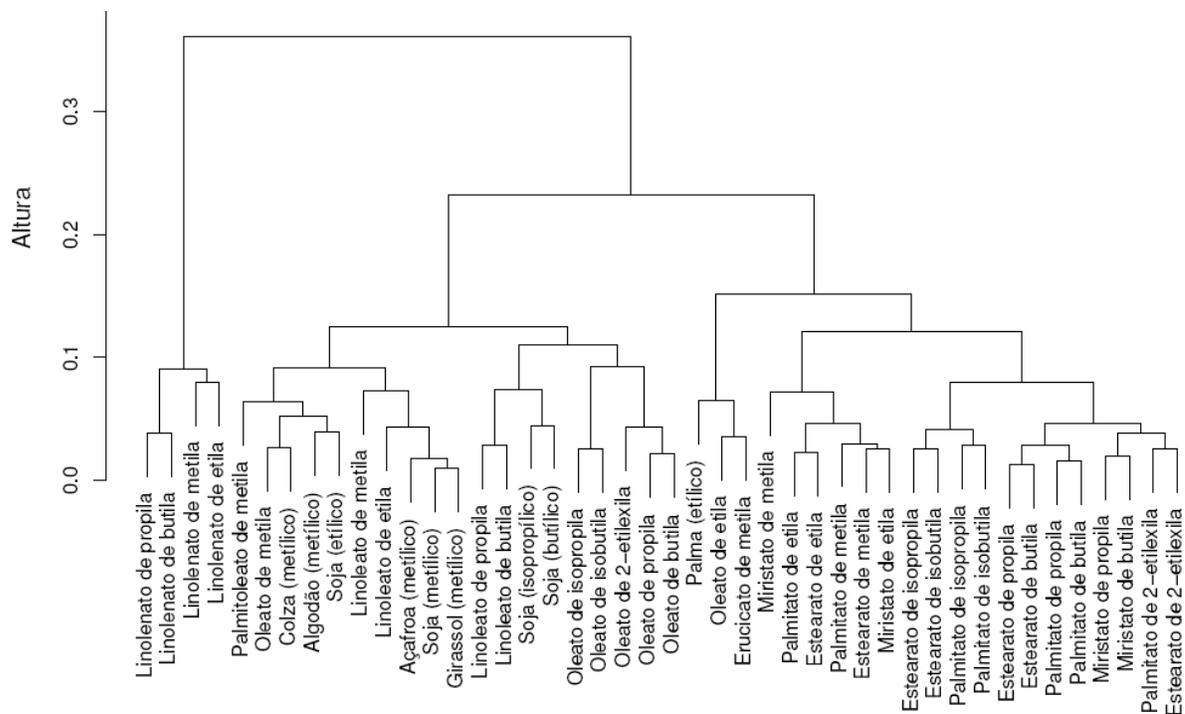
Avaliando o eixo das abscissas da Figura 15, observa-se que a PC1 é governada primordialmente pelo número de insaturações  $\text{C}=\text{C}$  presentes nas cadeias carbônicas dos ésteres (à medida que cresce o valor neste eixo verifica-se uma diminuição do número destas ligações).

Fazendo agora uma avaliação no eixo das ordenadas ainda sobre a Figura 15, tem-se que a PC2 é governada pela estrutura química do álcool utilizado na produção dos biocombustíveis, sendo esta componente principal uma combinação entre o tamanho da cadeia carbônica com o número de ramificações presentes neste álcool (à medida que cresce o valor neste eixo verifica-se um aumento do tamanho da cadeia carbônica do álcool, bem como um aumento no número de ramificações presentes no mesmo).

Apesar de considerar apenas combinações lineares no processamento dos dados, a PCA discrimina de maneira bastante interessante este conjunto de observações. Isso porque tanto o número de duplas ligações nas cadeias carbônicas (PC1) como a estrutura química do álcool utilizado na produção do biodiesel (PC2), que de acordo com este método equivalem a 83,65% da variância do sistema, são

parâmetros apontados por KNOTHE; BAGBY; RYAN III (1997) como determinantes no IC destes biocombustíveis.

### 1.2.1.2. Tratamento dos Dados por “Clustering” hierárquico



**Figura 16.** Dendrograma obtido do tratamento do banco de dados.

Avaliando o dendrograma (Figura 16) em função do grau de dissimilaridade, ou seja, de “cima para baixo”, vê-se que, num estágio inicial, este tratamento separa os ésteres linolênicos dos demais elementos do banco de dados. Posteriormente, o método separa basicamente os elementos saturados dos insaturados, salvas duas exceções (o oleato de etila e o erucicato de metila que são elementos com uma insaturação e foram posicionados ao lado de elementos saturados).

Considerando os índices de cetanos dos componentes do banco (Tabelas 3 e 4), os elementos que possuem os menores valores desta propriedade são justamente os ésteres linolênicos, e os ésteres com cadeia carbônica saturada têm, em linhas gerais, índices de cetanos maiores do que os com cadeia insaturada. Ou seja, este tratamento sinaliza que o banco está discriminando os dados da maneira adequada relativamente à variação do IC.

Este método propõe os ésteres linolênicos como sendo “muito diferentes” (estão em ramos distintos) dos outros elementos do banco de dados. Ele “enxerga” também que os elementos saturados são pouco semelhantes em relação aos insaturados. Como já mencionado anteriormente, os índices de cetanos dos elementos presentes neste banco de dados também obedecem a esta hierarquia.

No que diz respeito à composição química dos diferentes tipos de biodiesel este tratamento comete dois “equivocos”: o biodiesel de colza (metílico) está afastado do erucicato de metila (principal componente da mistura) no dendrograma, e o biodiesel de palma (etílico), que é composto majoritariamente por ésteres saturados, está classificado como um elemento muito semelhante a dois ésteres insaturados (oleato de etila e erucicato de metila). Em outras palavras, este método calcula que a distância entre o erucicato de metila é grande em relação ao biodiesel de colza (metílico), e que a distância entre o biodiesel de palma (etílico) e dois ésteres insaturados (oleato de etila e erucicato de metila) é pequena, o que do ponto de vista químico não é conveniente.

Quimicamente falando, ainda que este método cometa alguns equivocos, o principal critério para a discriminação dos dados foi o número de ligações duplas C=C presentes nas cadeias carbônicas, e este, de acordo com KNOTHE; BAGBY; RYAN III

(1997) é o principal fator que influencia no IC dos elementos presentes neste banco de dados, o que é um bom indicativo.

### 1.2.1.3. Tratamento dos Dados por “Clustering” não hierárquico (k-médias)

Para a realização deste método é necessário que se pré-estabeleça o número de classes de similaridade que o software deverá utilizar para distribuir os elementos do banco de dados, e este número foi determinado a partir dos resultados obtidos da análise de componentes principais. A PCA indicou a existência de 5 PC's (Gráfico 1), o que acarretou na opção de 5 “clusters” para esta etapa.

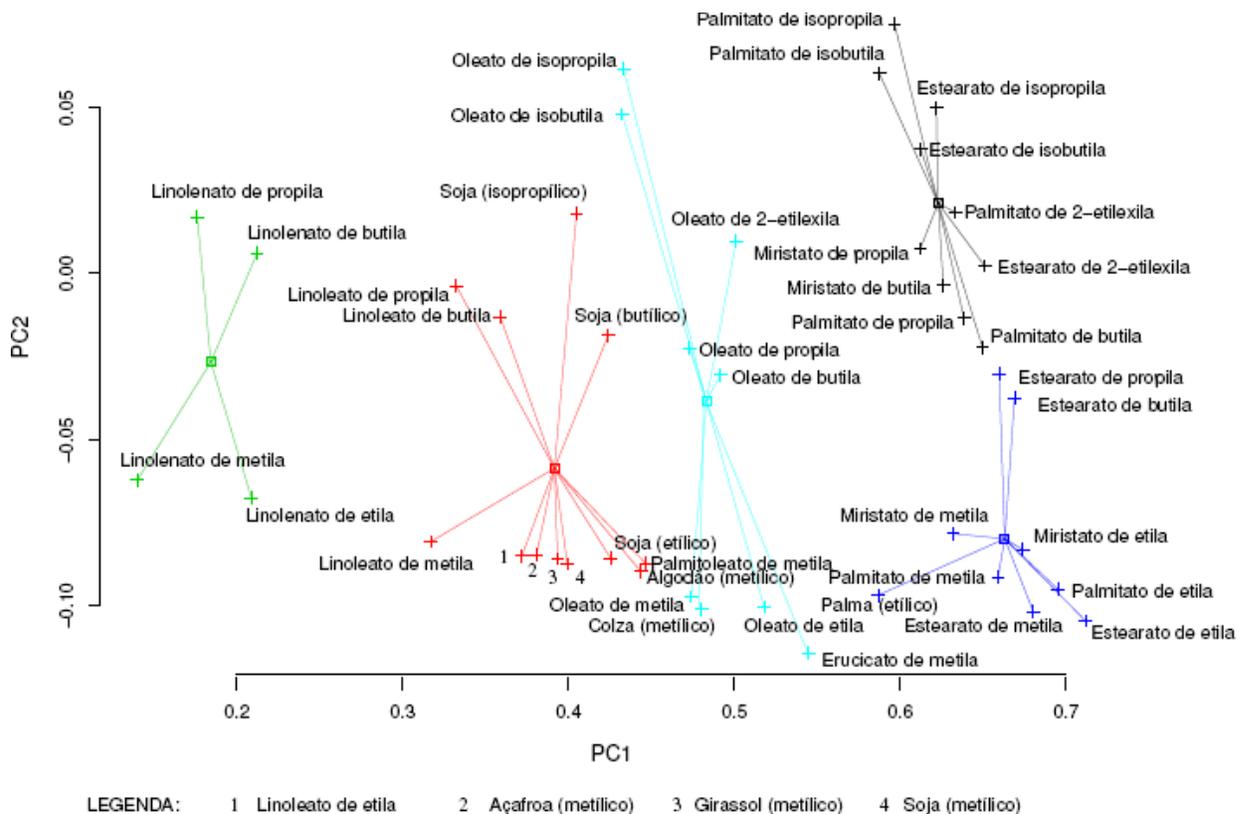


Figura 17. Resultados obtidos do “clustering” não hierárquico.

A partir da Figura 17 observamos um “cluster” contendo ésteres com 3 duplas ligações entre carbonos (ésteres linolênicos), outro que possui elementos com 2 insaturações entre carbonos (ésteres linoléicos e biocombustíveis com constituintes majoritários dos mesmos), um terceiro grupo com elementos com 1 dupla ligação entre carbonos (ésteres oléicos e biocombustíveis com constituintes majoritários destes), e dois outros conjuntos que possuem elementos com cadeia carbônica saturada (um com ésteres derivados de álcoois ramificados, e de álcoois com 3 ou 4 átomos de carbonos; e outro com ésteres derivados de álcoois com 1 ou 2 átomos de carbono e um biodiesel com constituinte majoritário de ésteres com cadeia carbônica saturada).

Houve três exceções em relação à descrição anterior: o palmitoleato de metila, éster com uma insaturação que foi classificado junto com elementos contendo duas insaturações e outros dois ésteres saturados, o estearato de propila e o estearato de butila, estes derivados de álcoois contendo 3 e 4 átomos de carbono respectivamente, que foram classificados junto com ésteres saturados derivados de álcoois com 1 ou 2 átomos de carbono.

Este tratamento mostrou que como critério primário a divisão dos “clusters” se deu primordialmente pelo número de insaturações C=C presentes nas cadeias carbônicas (mesmo critério da PC1 na PCA). Como temos 4 tipos de elementos quando faz-se referência a este critério (com três, duas, uma e sem ligações deste tipo) presentes no banco de dados e 5 conjuntos de aglomerados, um dos grupos teria que ser dividido em duas partes.

A partir da Figura 17, observa-se que o conjunto de elementos saturados foi o escolhido. A divisão deste conjunto ocorreu com base no tamanho e no número de ramificações da cadeia do álcool utilizado na produção do éster (critério secundário

aqui, e o mesmo da PC2 na PCA). Em outras palavras, de acordo com esta discriminação dos dados, a influência do tamanho e do número de ramificações presentes na cadeia dos álcoois utilizados na produção dos ésteres é mais pronunciada em elementos saturados do que em elementos insaturados.

A partir dos valores de IC obtidos de dados da literatura (Tabela 3) e avaliando apenas os ésteres palmíticos, esteáricos e oléicos (famílias de ésteres com mais dados de IC disponíveis para uma devida comparação). Observa-se que o conjunto de elementos mais afetado pela estrutura química do álcool precursor é o dos ésteres oléicos, porque existe uma maior variação nos IC destes elementos quando comparados aos outros em questão. Como mencionado anteriormente, o critério secundário utilizado aqui é o mesmo em relação à PCA, porém ao avaliar os índices de cetanos das substâncias puras, o conjunto (dos elementos saturados) a que este critério foi aplicado não é, “em tese”, o que deveria (dos elementos com uma insaturação).

O aparente ponto negativo mencionado não invalida esta discriminação dos dados, pois os critérios utilizados aqui são os principais fatores que interferem no IC (KNOTHE; MATHEAUS; RYAN III, 2003; KNOTHE; BAGBY; RYAN III, 1997). É interessante ressaltar ainda, que os tipos de biodiesel (misturas de ésteres) foram posicionados nos “clusters” junto com seus respectivos constituintes majoritários (ésteres que se encontram em maior percentagem nas misturas). A soma destes fatores dá indícios de que este banco de dados é viável para ser utilizado como base de dados na predição do IC.



com duas instaurações C=C, agora foram classificados no “cluster” com apenas uma insaturação deste tipo.

Porém, diferentemente do método anterior, onde ou um elemento pertence a um “cluster”, ou pertence a outro, no “fuzzy clustering” as observações são classificadas com porcentagens relativas a cada conjunto (as funções de pertinência). Isso significa dizer, que para cada ponto da matriz de dados é atribuída uma porcentagem referente a cada “cluster”.

Para o conjunto de dados utilizado neste trabalho, necessita-se justamente de uma ferramenta que nos proporcione um “meio-termo” entre os conjuntos, de maneira que um elemento seja classificado como correspondente a um “cluster”, porém seja capaz de pertencer a outro em uma dada proporção, já que se trabalha com misturas de ésteres (tipos de biodiesel) quimicamente muito diferentes entre si.

Levando-se em conta as descrições anteriores, tanto o fuzzy clustering quanto o clustering não hierárquico (k-médias), separaram adequadamente as classes de similaridade convenientes na predição dos índices de cetanos, sem que houvesse diferenças significativas entre estes métodos. Porém, devido à referida ferramenta extra que o agrupamento fuzzy nos proporciona, este método foi escolhido para ser utilizado na etapa seguinte do trabalho, a predição dos índices de cetanos de diferentes tipos de biodiesel presentes neste banco de dados.

### 1.3. Conclusões

O objetivo desta fase inicial do trabalho foi realizar uma análise exploratória dos dados verificando a aplicabilidade da matriz construída na predição dos índices de cetanos de diferentes tipos de biodiesel. A partir dos resultados obtidos pode-se considerar que o banco construído acenou positivamente neste sentido, pois os métodos de análise estatística implementados evidenciaram justamente as propriedades químicas relevantes na determinação da referida propriedade.

Avaliando os resultados método a método, temos:

- A PCA identificou a existência de cinco componentes principais. O PC de maior variância refere-se ao número de insaturações C=C presentes nas cadeias carbônicas e o PC2 à estrutura química do álcool utilizado na produção do respectivo biocombustível. Sendo que a soma entre estes dois componentes representa 83,65% da variância do sistema.

- O dendrograma obtido dispôs os dados utilizando como critério principal para discriminação o número de insaturações C=C, porém mesmo esta sendo a principal variável que provoca a diminuição do IC tratamentos posteriores envolveram mais critérios na separação dos dados, e conseqüentemente produziram resultados mais consistentes.

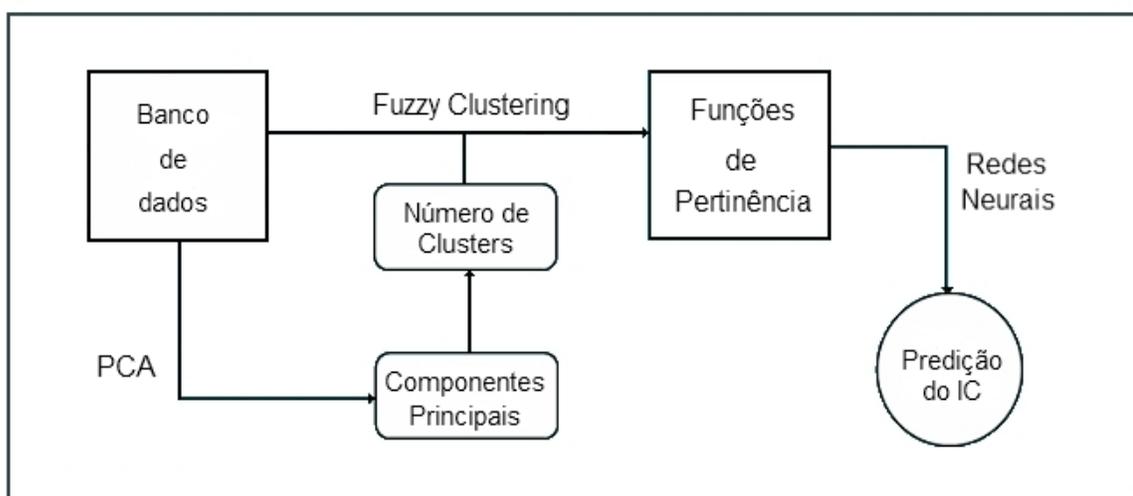
- Tendo como base as cinco componentes principais obtidas da PCA, o “clustering” não-hierárquico (k-médias) discriminou o conjunto de dados em cinco classes de similaridade. Esta se deu em função de dois fatores: o número de insaturações C=C presentes nas cadeias carbônicas e a estrutura química dos álcoois precursores na produção dos biocombustíveis presentes na matriz de dados.

- Os resultados do “fuzzy clustering” foram basicamente os mesmos do “clustering” não-hierárquico, com a particularidade de que neste tratamento existe a possibilidade de sobreposição dos “clusters”. Esse diferencial é muito interessante do ponto de vista dos tipos de biodiesel, pois estes são misturas de ésteres que na maioria das vezes são quimicamente muito diferentes entre si, o que leva a situações de interface entre os conjuntos de aglomerados. Por este motivo o “fuzzy clustering” foi o tratamento escolhido para ser utilizado na próxima etapa do trabalho, o desenvolvimento de um modelo empírico.

## CAPÍTULO II – Modelo empírico

### II.1. Metodologia

Verificada a consistência da matriz de dados inicial, o passo seguinte foi utilizar este banco como base de dados para predição do IC de diferentes tipos de biodiesel. A metodologia utilizada nesta etapa do trabalho está exposta no Esquema 1.



**Esquema 1.** Metodologia utilizada para predição dos índices de cetanos.

A metodologia empregada pode ser dividida em três etapas: pré-tratamento (PCA), classificação dos dados (fuzzy clustering) e predição dos índices de cetanos (redes neurais).

Em resumo, sobre o banco inicial (dados de RMN-<sup>1</sup>H, Tabela 6) foi feito um pré-tratamento (PCA) com intuito de determinar o número de aglomerados a ser utilizado na etapa seguinte, posteriormente na classificação (reformatação) dos dados um “fuzzy clustering” foi empregado ainda sobre o banco inicial, resultando numa nova matriz de

dados (as funções de pertinência de cada observação em cada cluster). E finalmente, os resultados das funções de pertinência obtidos deste tratamento foram utilizados como dados de entrada numa rede neural artificial visando à predição dos índices de cetanos de diferentes tipos de biodiesel.

## //.2. Resultados e Discussão

### //.2.1. Fuzzy Clustering

Como descrito anteriormente, observou-se que o fuzzy clustering foi o método que proporcionou os melhores resultados, e conseqüentemente foi escolhido para ser utilizado na predição do IC via redes neurais. Lembrando que para a implementação deste método foi necessário pré-determinar o número de grupos que o software deveria distribuir as observações da base de dados. Este número (cinco) foi obtido a partir do número de componentes principais obtidos da PCA.

Ao submetermos um conjunto de dados a um fuzzy clustering é gerada uma matriz com a porcentagem de cada observação em cada cluster, chamada de funções de pertinência. Como a matriz utilizada possui 44 dados de entrada e foi determinado a utilização de 5 “clusters”, as funções de pertinência constituem uma matriz de 44x5 (Tabela 7).

**Tabela 7.** Funções de pertinência obtidas da “fuzzy clustering”

Biocombustíveis	“Cluster” 1	“Cluster” 2	“Cluster” 3	“Cluster” 4	“Cluster” 5
Miristato de metila	0,845228	0,045002	0,060943	0,027558	0,021268
Palmitato de metila	0,032316	0,883209	0,016977	0,05779	0,009708
Palmitoleato de metila	0,049923	0,018883	0,889673	0,014024	0,027497
Estearato de metila	0,052373	0,130788	0,035191	0,758378	0,02327
Oleato de metila	0,45215	0,086129	0,333954	0,058934	0,068832

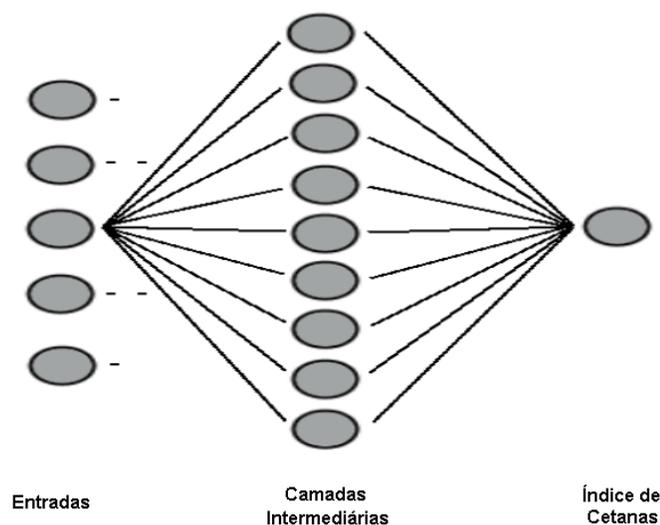
Biocombustíveis	“Cluster” 1	“Cluster” 2	“Cluster” 3	“Cluster” 4	“Cluster” 5
Linoleato de metila	0,037736	0,023669	0,064078	0,019629	0,854889
Linolenato de metila	0,063413	0,042731	0,095326	0,036212	0,762318
Miristato de etila	0,86184	0,042047	0,051994	0,025362	0,018756
Palmitato de etila	0,067649	0,314521	0,041154	0,551175	0,025501
Estearato de etila	0,019018	0,056791	0,012325	0,903915	0,00795
Oleato de etila	0,84078	0,051266	0,056482	0,030282	0,021191
Linoleato de etila	0,153041	0,081663	0,386992	0,06493	0,313374
Linolenato de etila	0,042462	0,028136	0,065546	0,023719	0,840137
Estearato de propila	0,076109	0,502509	0,044612	0,349727	0,027043
Estearato de isopropila	0,029307	0,095375	0,018714	0,844651	0,011953
Estearato de butila	0,07148	0,362733	0,043067	0,496189	0,026532
Estearato de isobutila	0,029489	0,080061	0,019464	0,858274	0,012712
Palmitato de propila	0,040738	0,863233	0,020875	0,06336	0,011794
Palmitato de isopropila	0,10736	0,697067	0,050912	0,11688	0,027781
Palmitato de butila	0,074386	0,414333	0,044373	0,439733	0,027175
Palmitato de isobutila	0,074219	0,775593	0,036429	0,09358	0,020179
Palmitato de 2-etilexila	0,019272	0,05575	0,012564	0,904279	0,008136
Miristato de propila	0,632511	0,142514	0,104195	0,076826	0,043955
Miristato de butila	0,745924	0,08999	0,080833	0,050968	0,032285
Oleato de propila	0,420412	0,085446	0,364507	0,058828	0,070806
Oleato de isopropila	0,035781	0,862465	0,019142	0,071568	0,011045
Oleato de butila	0,587023	0,083616	0,217438	0,055475	0,056449
Oleato de isobutila	0,470361	0,086417	0,316741	0,058908	0,067573
Oleato de 2-etilexila	0,056506	0,744983	0,031386	0,14866	0,018466
Estearato de 2-etilexila	0,143484	0,246781	0,107176	0,425612	0,076947
Linoleato de propila	0,145874	0,070454	0,550285	0,05478	0,178606
Linoleato de butila	0,116362	0,038414	0,770673	0,027944	0,046606
Linolenato de propila	0,01596	0,010198	0,026198	0,008503	0,939141
Linolenato de butila	0,016198	0,01034	0,026636	0,008619	0,938206
Erucicato de metila	0,358019	0,327359	0,117986	0,139711	0,056924
Algodão (metílico)	0,049585	0,018576	0,89141	0,013774	0,026654
Colza (metílico)	0,161185	0,050129	0,695522	0,036169	0,056995
Açafoa (metílico)	0,065423	0,026103	0,847338	0,019561	0,041575
Soja (metílico)	0,122454	0,055557	0,662159	0,042645	0,117184
Girassol (metílico)	0,116854	0,052358	0,68377	0,040091	0,106928
Palma (etílico)	0,261911	0,070683	0,546682	0,050105	0,070619
Soja (etílico)	0,092572	0,0393	0,767615	0,029782	0,070731
Soja (isopropílico)	0,080197	0,027929	0,835229	0,020476	0,036169
Soja (butílico)	0,055746	0,020373	0,880657	0,015047	0,028176

Esta reformatação dos dados pode ser encarada como uma etapa classificatória do trabalho, o objetivo na reorganização dos dados foi permitir uma melhor classificação das misturas, uma vez que estas seriam o objeto de desejo na etapa de predição.

A idéia agora é utilizar esta nova matriz (com os valores de pertinência dos pontos aos aglomerados) como base de dados na etapa de predição. Para tanto, o método escolhido para relacionar as funções de pertinência com os respectivos índices de cetanos dos biocombustíveis foi uma rede neural artificial do tipo *feed forward backpropagation*.

### **//.2.2. Predição do índice de cetanos via redes neurais**

A arquitetura da rede utilizada neste trabalho foi do tipo 5-9-1 (Figura 19). Isto significa, 5 neurônios na camada de entrada, correspondentes às proporções de cada observação em cada um dos cinco “clusters” (funções de pertinência do tratamento anterior), 9 neurônios na camada intermediária, onde foram processados os dados, e 1 neurônio na última camada, correspondente ao índice de cetanos.



**Figura 19.** Arquitetura da rede utilizada neste trabalho.

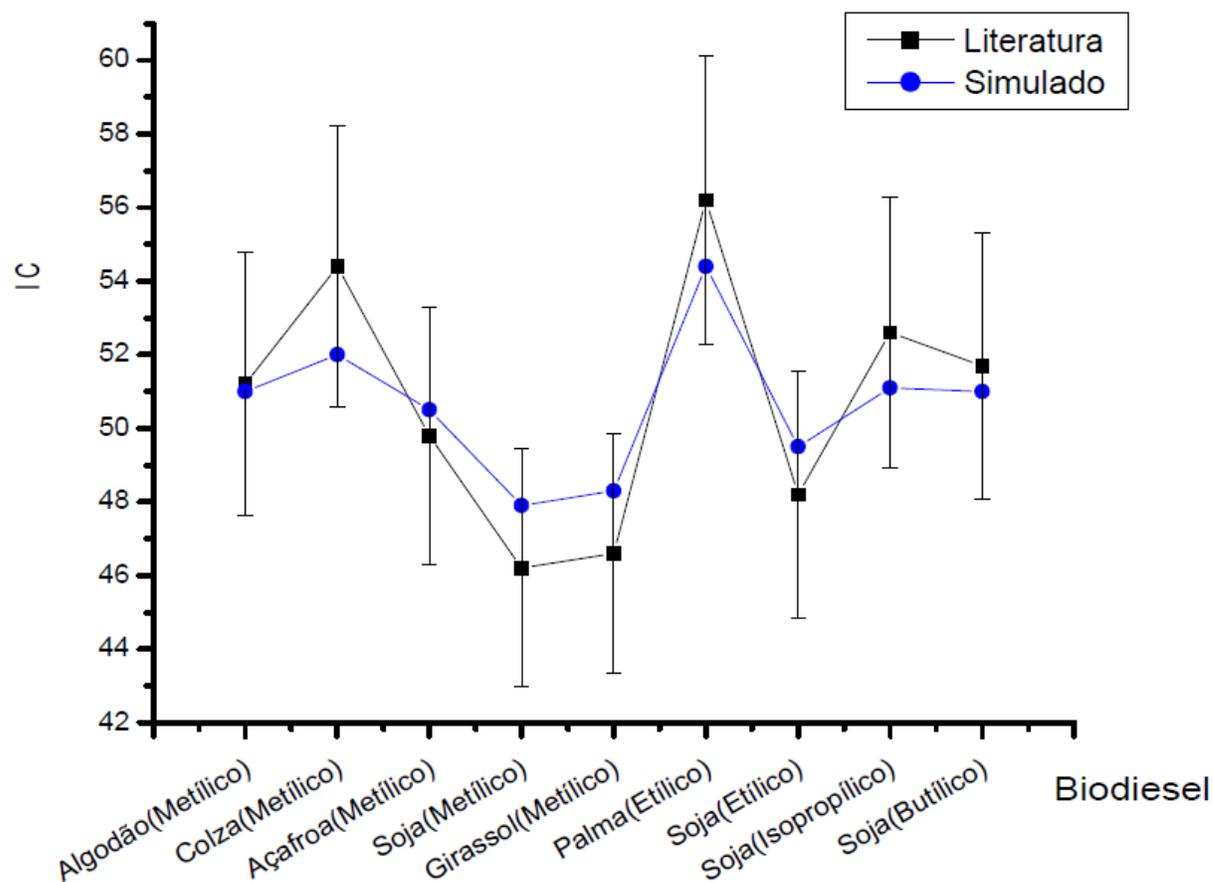
Os dados de entrada foram divididos em duas partes, uma matriz 35x5, para treinamento da rede (contendo os ésteres – substâncias puras), e o restante, uma matriz 9x5 (os tipos de biodiesel – misturas de ésteres) foi usada para validação dos resultados. Através desta divisão foi possível fazer a predição do IC de misturas (tipos de biodiesel) a partir de dados de substâncias puras (ésteres).

A comparação entre os resultados simulados e os dados obtidos da literatura com os respectivos erros relativos são mostrados na Tabela 8:

**Tabela 8.** Comparação dos resultados simulados pela rede em relação aos dados da literatura com os respectivos erros relativos.

Biodiesel	IC – Literatura	IC – Rede Neural	Erro Relativo (%)
Algodão (Metílico)	51,2	51,0	0,39
Colza (Metílico)	54,4	52,0	4,41
Açafroa (Metílico)	49,8	50,5	1,40
Soja (Metílico)	46,2	47,9	3,68
Girassol (Metílico)	46,6	48,3	3,65
Palma (Etílico)	56,2	54,4	3,20
Soja (Etílico)	48,2	49,5	2,70
Soja (Isopropílico)	52,6	51,1	2,85
Soja (Butílico)	51,7	51,0	1,36

Estes resultados também podem ser visualizados através do Gráfico 2 a seguir, em que estão representadas as respectivas barras de erro (respeitando os 7% explanados anteriormente) relativas às medias experimentais:



**Gráfico 2.** Comparação entre o IC dos biocombustíveis obtidos a partir de dados da literatura e da simulação feita pela rede neural com as respectivas barras de erro.

### //3. Conclusões

De antemão destaca-se ineditismo do trabalho quando é levada em consideração a ideia de se obter informações a partir de espectros teóricos de RMN-<sup>1</sup>H de biocombustíveis, e fazer a predição dos índices de cetanos dos mesmos. A metodologia empregada para tanto (PCA para pré-tratamento dos dados, Fuzzy Clustering para classificação e uma rede neural para simulação dos resultados) também não fora identificada em outros trabalhos desta natureza até a data de envio desta tese.

A utilização da PCA no pré-tratamento dos dados foi determinante no sucesso do trabalho, pois permitiu a resolução de um problema chave na etapa de classificação: a determinação do número de classes de similaridade presentes no banco de dados inicial.

A ideia de utilizar os valores das funções de pertinência como dados de entrada para uma rede neural (reformatação dos dados) se deu no intuito de permitir uma melhor classificação dos tipos de biodiesel, uma vez que o fuzzy clustering é reconhecido por ser um poderoso instrumento na classificação de misturas.

A modelagem empírica via redes neurais é conhecida como do tipo “caixa preta”, uma vez que são determinados os pesos das conexões entre os neurônios, o operador não pode modificar estes valores, impossibilitando qualquer tentativa de interferência nos resultados alcançados. Todavia os resultados obtidos aqui demonstram o forte poder de predição destes tipos de algoritmos, legitimando a diversidade de campos científicos a que este método foi empregado.

Apesar de funcionar apenas como uma aproximação, o uso de porcentagens médias para os ésteres componentes dos diferentes tipos de biodiesel não “atrapalhou” na determinação dos índices de cetanos. Uma vez que todos os valores obtidos foram excelentes, com resultados (variações entre 0,39% e 4,41%) dentro dos 7% relativos à margem de erro das medidas experimentais. E isto vem ratificar todos os pontos positivos expostos nesta conclusão.

### **CAPÍTULO III – Modelo preditivo**

O modelo empírico desenvolvido no capítulo anterior produziu excelentes resultados, entretanto existem algumas situações inconvenientes relativamente à reprodutibilidade dos resultados, caso desejássemos agregar outros biocombustíveis àquele conjunto de dados. Tendo como base a metodologia utilizada, a reprodutibilidade dos resultados estaria condicionada a satisfazer três estágios:

1º estágio – A manutenção do número de componentes principais na PCA: O acréscimo de uma ou mais observações no conjunto de dados de RMN poderia resultar na variação do número componentes principais apontados inicialmente. O efeito então seria em cadeia, mudar-se-ia o número de classes de similaridade, modificando então o número de clusters na etapa subsequente e conseqüentemente a matriz de pertinência, o que certamente acarretaria em valores diferentes para os índices de cetanos obtidos.

Caso a adição de uma ou mais observações no conjunto de dados de RMN não variasse o número componentes principais apontados inicialmente, o que também é possível, o problema poderia ocorrer na etapa subsequente.

2º estágio – A manutenção da matriz de pertinência no Fuzzy Clustering: Com o acréscimo de novas observações no banco, os centros dos clusters poderiam ser mantidos (sem considerar a influência destes novos dados) ou ser recalculados (considerando também os novos pontos do conjunto de dados).

Para que se obtivesse reprodutibilidade, a matriz de pertinência atual não poderia ser modificada. Sendo assim, a opção mais sensata seria de fixar a atual posição dos centros dos clusters, desta forma os novos pontos seriam apenas classificados de

acordo com o sistema atual, enquanto os atuais pontos da matriz de dados não teriam rigorosamente nenhuma alteração na matriz de pertinência.

Numa segunda hipótese, todos os dados (os atuais somados aos eventuais) participariam da definição dos centros dos aglomerados. Com um diferente número de pontos na matriz de dados o fuzzy clustering muito provavelmente calcularia diferentes posições para os centros dos clusters, interferindo diretamente na pertinência de cada observação em cada conjunto. Com a variação da matriz de pertinência os resultados dos índices de cetanos alcançados certamente seriam outros. Outra situação, muito improvável, mas possível, seria que o acréscimo de uma ou mais observações não variasse os centros dos clusters na etapa de classificação. O resultado disto seria que os atuais pontos da matriz de dados não teriam rigorosamente nenhuma alteração na matriz de pertinência e os novos pontos seriam adicionados a esta matriz com seus respectivos valores.

Diante das situações expostas e pensando na reprodutibilidade dos atuais resultados, a escolha mais prudente seria manter a posição atual dos centros dos clusters, o que garantiria a manutenção da matriz de pertinência atual.

3º estágio – A manutenção do conjunto de treino utilizado na rede neural: Nesta etapa teríamos duas situações possíveis. A primeira havendo alteração no conjunto utilizado para o treinamento da rede. Nesta situação os valores dos IC obtidos certamente mudariam, uma vez que um diferente número de observações no conjunto de treinamento da rede resultaria em alterações nas conexões (pesos) entre os neurônios. A segunda não havendo alteração no conjunto utilizado para o treinamento da rede. Esta seria a única possibilidade em que obteríamos reprodutibilidade para todos os resultados, pois o conjunto de treino da rede seria rigorosamente o mesmo e

as conexões entre os neurônios na rede também. Em outras palavras, para manutenção do atual conjunto de treinamento da rede, os novos dados deveriam constar, obrigatoriamente, apenas no conjunto de validação dos resultados. O que do ponto de vista prático deixa a desejar.

Por algumas vezes foi dito que os resultados obtidos provavelmente seriam diferentes dos atuais. Isso não necessariamente significa que os mesmos estivessem fora dos 7% relativos às barras de erro das medidas experimentais, mas seria leviano fazer qualquer previsão quanto a isso, uma vez que não há como prever tais valores.

Em resumo, caso desejássemos determinar o IC de um biocombustível em estudo pela metodologia descrita no capítulo anterior, com resultados reproduzíveis para os dados atuais, seria necessário preservar a matriz de dados utilizada para treinamento da rede. Para tanto, a adição de uma ou mais observações não poderia alterar nem o número de componentes principais obtidos anteriormente nem a posição dos centros dos clusters na etapa de classificação, além disso, o conjunto utilizado no treinamento da rede não poderia ser modificado, isto é, os novos dados em estudo somente poderiam ser utilizados no conjunto de validação da rede.

Tais discussões não desqualificam o modelo empírico desenvolvido, pois este funciona muito bem para a função a qual lhe foi dada, de prever o IC para todos os tipos de biodiesel presentes neste banco de dados. Porém, ao se ampliar o leque de trabalho, provavelmente, o referido modelo passaria a ser insuficiente para tal função.

Todas as restrições mencionadas dificultam a utilização do método desenvolvido para futuras simulações. Neste cenário, surgiu a necessidade de desenvolver uma nova metodologia mais simples, prática e principalmente reproduzível sob quaisquer aspectos. Neste sentido foi proposta uma nova metodologia, que relaciona as estruturas

químicas dos diferentes biocombustíveis com seus respectivos IC's através de cinco descritores estruturais e um modelo de regressão não linear.

### III.1. Formatação da nova base de dados

O passo inicial para a criação do modelo preditivo foi desenvolver um novo banco de dados. Com base nos estudos realizados nos capítulos anteriores, onde foi possível identificar quais as principais variáveis que influem no IC dos biocombustíveis utilizados, formatou-se uma nova base de dados (Tabela 9). Nesta, foram utilizados cinco descritores estruturais para traduzir a linguagem química relevante em linguagem matemática:

- 1) O Número de átomos de Hidrogênio ligados a átomos de carbono Insaturados (NHI)
- 2) O Número de átomos de Carbono presentes nas Ramificações (NCR)
- 3) O Número de átomos de Carbono do Álcool utilizado na produção do respectivo biocombustível (NCA)
- 4) O Número Total de átomos de Carbono na molécula ou mistura (NCT)
- 5) O Número de átomos de Hidrogênio ligados ao carbono Carbinólico (NHC)

**Tabela 9.** Base de dados construída para formatação do modelo preditivo.

Biocombustíveis	NHI	NCR	NCA	NCT	NHC
<b>Miristato de metila</b>	0	0	1	15	3
<b>Palmitato de metila</b>	0	0	1	17	3
<b>Palmitoleato de metila</b>	2	0	1	17	3
<b>Estearato de metila</b>	0	0	1	19	3
<b>Oleato de metila</b>	2	0	1	19	3
<b>Linoleato de metila</b>	4	0	1	19	3
<b>Linolenato de metila</b>	6	0	1	19	3
<b>Miristato de etila</b>	0	0	2	16	2
<b>Palmitato de etila</b>	0	0	2	18	2
<b>Estearato de etila</b>	0	0	2	20	2
<b>Oleato de etila</b>	2	0	2	20	2

Biocombustíveis	NHI	NCR	NCA	NCT	NHC
Linoleato de etila	4	0	2	20	2
Linolenato de etila	6	0	2	20	2
Estearato de propila	0	0	3	21	2
Estearato de isopropila	0	1	3	21	1
Estearato de butila	0	0	4	22	2
Estearato de isobutila	0	1	4	22	2
Palmitato de propila	0	0	3	19	2
Palmitato de isopropila	0	1	3	19	1
Palmitato de butila	0	0	4	20	2
Palmitato de isobutila	0	1	4	20	2
Palmitato de 2-etilexila	0	2	8	22	2
Miristato de propila	0	0	3	17	2
Miristato de butila	0	0	4	18	2
Oleato de propila	2	0	3	21	2
Oleato de isopropila	2	1	3	21	1
Oleato de butila	2	0	4	22	2
Oleato de isobutila	2	1	4	22	2
Oleato de 2-etilexila	2	2	8	24	2
Estearato de 2-etilexila	0	2	8	24	2
Linoleato de propila	4	0	3	21	2
Linoleato de butila	4	0	4	22	2
Linolenato de propila	6	0	3	21	2
Linolenato de butila	6	0	4	22	2
Erucicato de metila	2	0	1	23	3
Algodão (metílico)	2,54	0	1	18,82	3
Colza (metílico)	2,38	0	1	20,86	3
Açafroa (metílico)	3,24	0	1	18,86	3
Soja (metílico)	3,11	0	1	18,86	3
Girassol (metílico)	3,02	0	1	18,90	3
Palma (etílico)	1,26	0	2	19,16	2
Soja (etílico)	3,11	0	2	19,86	2
Soja (isopropílico)	3,11	1	3	20,86	1
Soja (butílico)	3,11	0	4	21,86	2

Para um melhor entendimento de como foi moldada a matriz de dados anterior, tomemos dois componentes (um éster e um biodiesel) como exemplos:

Exemplo 1: O estearato de etila (Figura 20):



**Figura 20.** Fórmula estrutural do estearato de etila.

Esta molécula não tem ligações duplas C=C, logo não possui átomos de hidrogênio ligados a estes tipos de carbono (NHI = 0); Não há ramificações na cadeia carbônica (NCR = 0); Este é um éster de etila, ou seja, o álcool utilizado para a produção deste biocombustível foi o etílico (álcool precursor com dois átomos de carbono, NCA = 2); O carbono carbinólico (átomo de carbono na posição  $\alpha$  ao oxigênio) possui dois átomos de hidrogênio ligados a ele (NHC=2); Sua fórmula geral é  $C_{20}H_{40}O_2$ , isto é, um total de vinte átomos de carbono em sua cadeia (NCT = 20).

A partir do raciocínio desenvolvido, podemos escrever a linha referente ao estearato de etila da seguinte forma:

Linha da Tabela 9 referente ao estearato de etila.

Biocombustível	NHI	NCR	NCA	NCT	NHC
Estearato de etila	0	0	2	20	2

Este procedimento foi aplicado a todos os ésteres (substâncias puras) presentes no banco de dados.

Tendo em vista que os tipos de biodiesel são misturas de ésteres com estruturas químicas bem distintas, o procedimento descrito a seguir busca um valor proporcional para cada biodiesel em cada uma das cinco variáveis.

Exemplo 2: O biodiesel de Soja (etílico):

Fazendo uma analogia com a Tabela 1, é razoável supor que tenhamos a seguinte composição química para o biodiesel de soja (Tabela 10):

**Tabela 10.** Composição química do biodiesel de Soja (etílico).

Biodiesel	Porcentagem (%) em massa dos principais ésteres				
	Palmitato de etila	Estearato de etila	Oleato de etila	Linoleato de etila	Linolenato de etila
Soja (etílico)	2,3 – 11,0	2,4 – 6,0	22,0 – 30,8	47,0 – 55,0	2,0 – 10,5

Como as composições químicas dos diferentes tipos de biodiesel são variáveis utilizamos como aproximação o valor médio destas:

Palmitato de etila 6,65%

Estearato de etila 4,2%

Oleato de etila 26,4%

Linoleato de etila 51%

Linolenato de etila 6,25%

Neste caso a soma das porcentagens foi igual a 94,5%, assim, para que a soma fosse igual a 100%, foi aplicado um fator de correção:

Palmitato de etila  $6,65\% / 0,945 = 7,0\%$

Estearato de etila  $4,2\% / 0,945 = 4,4\%$

Oleato de etila  $26,4\% / 0,945 = 28,0\%$

Linoleato de etila  $51\% / 0,945 = 54,0\%$

Linolenato de etila  $6,25\% / 0,945 = 6,6\%$

Determinadas as porcentagens médias relativas a cada éster, o próximo passo foi determinar os valores das variáveis (Tabela 11), assim como no exemplo anterior, para cada um dos componentes do biodiesel em questão:

**Tabela 11.** Variáveis referentes aos ésteres componentes do biodiesel de Soja (etílico) e sua composição química média.

Ésteres	NHI	NCR	NCA	NCT	NHC	Porcentagem (%) em massa
Palmitato de etila	0	0	2	18	2	7,0
Estearato de etila	0	0	2	20	2	4,4
Oleato de etila	2	0	2	20	2	28,0
Linoleato de etila	4	0	2	20	2	54,0
Linolenato de etila	6	0	2	20	2	6,6

Feito isso, cada valor foi multiplicado pela porcentagem do respectivo éster e os valores foram somados em cada coluna (Tabela 12), resultando na linha referente ao biodiesel de soja (etílico):

**Tabela 12.** Ilustração dos cálculos feitos para a obtenção da linha do biodiesel de Soja (etílico).

Ésteres e <b>Biodiesel</b>	NHI	NCR	NCA	NCT	NHC
Palmitato de etila	0 (x0,07)	0 (x0,07)	2 (x0,07)	18 (x0,07)	2 (x0,07)
Estearato de etila	0 (x0,044)	0 (x0,044)	2 (x0,044)	20 (x0,044)	2 (x0,044)
Oleato de etila	2 (x0,28)	0 (x0,28)	2 (x0,28)	20 (x0,28)	2 (x0,28)
Linoleato de etila	4 (x0,54)	0 (x0,54)	2 (x0,54)	20 (x0,54)	2 (x0,54)
Linolenato de etila	6 (x0,066)	0 (x0,066)	2 (x0,066)	20 (x0,066)	2 (x0,066)
<b>Soja (etílico)</b>	<b>3,11</b>	<b>0</b>	<b>2</b>	<b>19,86</b>	<b>2</b>

Este procedimento foi aplicado a todas as misturas constantes no banco de dados.

### III.2. O modelo

Sobre o novo banco foi empregado o método de regressão não linear multivariado, criou-se então um modelo preditivo correlacionando às estruturas químicas das moléculas que compõem os biocombustíveis da base de dados com seus respectivos índices de cetanos.

Inicialmente foram escolhidos aleatoriamente os seguintes valores:  $a = 10$ ,  $b = 4$ ,  $c = -15$ ,  $d = 2$ ,  $e = 8$ ,  $f = 1,5$ ,  $g = 7$ , como parâmetros para a equação (3) a seguir:

$$\text{IC} = a \cdot \text{NCT} + b \cdot \text{NCR} + c \cdot (\text{NHI})^d + e \cdot (\text{NCA})^f + g \cdot \text{NHC} \quad (3)$$

Nesta equação foram utilizadas cinco variáveis de entrada (NHI, NCR, NCA, NCT, NHC) para a obtenção de uma variável resposta (IC). O processo de otimização dos valores iniciais consistiu em aperfeiçoar os parâmetros fornecidos, usando como critério a maximização do coeficiente de correlação ( $R^2$ ) entre os valores simulados e os dados da literatura. O algoritmo utilizado para tanto foi um Gauss-Newton e o coeficiente de correlação ( $R^2$ ) ajustado foi de 0,93. O modelo obtido após a otimização dos parâmetros é expresso pela equação (4) a seguir:

$$\text{IC} = 5,5 \cdot \text{NCT} + 3,5 \cdot \text{NCR} - 21 \cdot (\text{NHI})^{2/3} - 16 \cdot (\text{NCA})^{1/3} + 0,6 \cdot \text{NHC} \quad (4)$$

Antes de avaliar individualmente os parâmetros da equação anterior, devemos lembrar que existem algumas variáveis correlacionadas. Tanto o NCA como o NCR são condicionados ao NCT, pois uma variação na estrutura química relacionada a estas

partes da molécula modificam também o NCT. Por exemplo, ao comparar um biocombustível qualquer derivado do álcool etílico a esse mesmo biocombustível, porém derivado do álcool propílico observaríamos que tanto o NCA como o NCT seriam modificados. A princípio, a ideia da comparação era modificar “apenas” o NCA e observar o efeito no IC, e posteriormente tirar as conclusões relativas àquela variável, porém como também alteramos outra variável, este raciocínio não é direto. Em decorrência disto, a avaliação de apenas uma variável por vez, para variáveis correlacionadas, não é trivial.

Apesar de se referirem a mesma parte da molécula (relativo ao álcool precursor na produção do biocombustível), o NHC é apenas parcialmente dependente do NCA, quando se altera o tamanho da cadeia do álcool que dá origem ao biocombustível ocorrerá necessariamente uma variação no NCA, entretanto o NHC pode variar ou não. Por exemplo, comparando-se o miristato de metila (NCA=1) com o miristato de etila (NCA=2) o NHC varia de 3 para 2, porém quando confrontamos o miristato de etila (NCA=2) com o miristato de propila (NCA=3) o valor da variável em questão permanece constante. Em resumo, nem sempre uma variação no tamanho da cadeia do álcool mudará o valor do NHC, por isso é dito que estas variáveis são parcialmente dependentes.

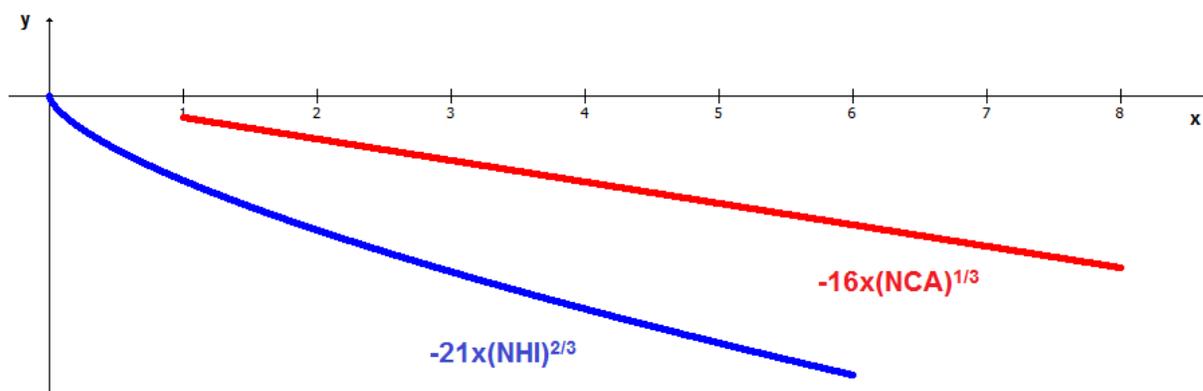
Na verdade, a única variável totalmente independente das outras é o NHI.

### **III.2.1. As variáveis com constantes negativas**

Avaliando agora a equação (4) observamos que duas variáveis (NHI e NCA) possuem constantes negativas.

Apesar de o NCA possuir valor máximo maior do que o NHI (8 contra 6 – Tabela 9), a segunda variável infere mais intensamente no resultado do IC do que a primeira, isto advém de tanto o seu expoente ( $2/3$  contra  $1/3$ ) como o módulo de sua constante (21 contra 16) serem maiores, ocasionando em valores máximos de  $-69,3$  para o NHI e  $-32,0$  para o NCA.

Um comparativo entre as funções  $-21x(\text{NHI})^{2/3}$  e  $-16x(\text{NCA})^{1/3}$  é mostrado no Gráfico 3 a seguir.



**Gráfico 3.** Comparação entre as funções  $-21x(\text{NHI})^{2/3}$  e  $-16x(\text{NCA})^{1/3}$  no plano  $xy$ .

O Gráfico 3 foi construído nos intervalos entre  $[1;8]$  para o NCA e  $[0;6]$  para o NHI, pois estes são os valores existentes para as variáveis em questão nesta matriz. Para esta base de dados, sendo o NHI não nulo, este sempre terá um maior poder de interferência no IC maior do que o NCA.

O maior poder de diminuição do IC concedido ao NHI é um excelente indicativo, pois é sabido por [8,9] que o número de ligações C=C (neste caso entendido como o NHI) é o principal fator causador da diminuição do IC. Quimicamente falando, quando os combustíveis possuem estes tipos de ligações, durante a queima formam-se radicais alila (quimicamente mais estáveis por efeito de ressonância) que freiam o processo radicalar, aumentando o tempo de ignição e conseqüentemente diminuindo o valor do

IC. Desta forma, não só o valor negativo de sua constante, mas também o maior poder de interferir negativamente no IC dado ao NHI fazem todo o sentido.

O valor negativo da constante atribuída ao NCA nos levaria a crer que um aumento no valor desta variável refletiria num decréscimo do IC, no entanto o cenário é exatamente o inverso, na verdade, esta variável possui um efeito aditivo no valor do IC. Como descrito anteriormente, o NCA é relacionado ao NCT, e a constante negativa atribuída a sua função atua como uma espécie de suavização no resultado da equação (4), pois quando o NCA aumenta o NCT (variável com constante positiva) também cresce e o saldo obtido entre estes dois parâmetros é positivo em relação ao valor do IC.

Para melhor entendimento, observemos a Tabela 13 com as variáveis relativas ao palmitato de etila e ao palmitato de propila.

**Tabela 13.** Ilustração das linhas referentes ao palmitato de etila e de propila e o IC calculado pela equação (4).

Biocombustíveis	NHI	NCR	NCA	NCT	NHC	IC simulado
Palmitato de etila	0	0	2	18	2	80,1
Palmitato de propila	0	0	3	19	2	82,7

A diferença entre as duas estruturas químicas é, apenas, a adição de mais um carbono no álcool que dá origem ao biocombustível. Relativamente ao NCA, este valor passa de 2 para 3 (-20,2 a -23,1 = -2,9 no IC), no entanto o NCT também aumenta uma unidade (+5,5 no IC). A combinação dos dois fatores promove uma variação de  $-2,9 + 5,5 = +2,6$  no IC, sendo o valor calculado para o palmitato de etila de 80,1, enquanto para o palmitato de propila o resultado foi 2,6 unidades maior, 82,7.

Ao contrário do que se imaginaria a primeira vista, em razão do valor negativo de sua constante, a situação exposta no exemplo anterior demonstra o efeito aditivo promovido pelo aumento do NCA.

O efeito da parte alcoólica das moléculas sobre o IC é variável. Caso esta fosse à parte determinante no valor do IC, devido ao aumento do número de radicais instáveis metileno ( $\text{CH}_2$ ) e conseqüentemente do NCT, seria esperado que o IC crescesse na seguinte ordem: metílicos < etílicos < propílicos < butílicos. Entretanto, ao se observar os valores experimentais obtidos isto não é constatado, a exceção são ésteres mirísticos, linoleicos e linolênicos que obedecem a referida ordem. Isto nos dá a ideia, de que, em alguns casos, esta região da molécula é determinante, ou pelo menos, relevante no valor do IC (mirísticos, linoléicos e linolênicos). Já em outros casos, diferentes partes da molécula sobrepõem o efeito da parte alcoólica (palmíticos, esteáricos e oléicos).

Na verdade o raciocínio passa por uma avaliação conjunta das três principais variáveis do modelo envolvidas (NCT, NCA e NHI).

Especificamente para os mirísticos (ésteres saturados,  $\text{NHI} = 0$ , com as menores cadeias desta base de dados), o valor do NCT é relativamente pequeno, tendo o NCA então, uma inferência maior no valor do IC, desta forma a referida ordem de crescimento dos índices é “obedecida”.

À medida que o NCT vai aumentando e não há nenhum termo “negativo” com valor suficientemente grande (situação relativa aos ésteres com cadeias saturadas maiores,  $\text{NHI} = 0$ , e os ésteres oleicos,  $\text{NHI} = 2$ ), o papel do NCA fica abreviado no resultado do IC, em conseqüência disso à ordem de crescimento não é mais seguida.

A explicação para os ésteres linoleicos (NHI = 4) e linolênicos (NHI = 6), que possuem valores de NCT grandes, é o aumento do termo negativo NHI que puxa o IC para baixo, tendo o NCA neste caso apenas um papel de “ajuste fino” no valor do IC, assim a referida ordem de crescimento dos índices volta a “funcionar”.

### **III.2.2. As variáveis com constantes positivas**

Avaliaremos agora os outros três parâmetros (NCT, NCR e NHC), que possuem constantes positivas.

Devido à dificuldade descrita anteriormente em associar intuitivamente o NCA ao IC, e com o objetivo de elucidar melhor a interferência do álcool precursor no valor da propriedade, utilizou-se outra variável relativa a esta parte da molécula, o NHC. Esta variável descreve o número de átomos de hidrogênio presentes no carbono  $\alpha$  ao oxigênio e possui valores pequenos (+0,6; +1,2 ou +1,8 no valor do IC), se encaixando como uma espécie de refinamento do modelo.

Entre as duas outras variáveis, o NCT tem um peso maior do que o NCR no resultado do IC, pois os valores deste parâmetro variam entre 15 e 24, enquanto os valores obtidos para o NCR são apenas 0, 1 ou 2 para esta matriz de dados (Tabela 9). Além disso, a constante que acompanha o NCT no modelo é maior do que a constante do NCR (5,5 versus 3,5).

O maior poder de interferência no aumento do IC concedido ao NCT é um indicativo positivo, pois de acordo com [8,9] o aumento da cadeia carbônica do combustível é o principal fator responsável pelo crescimento da propriedade em questão. Quimicamente falando, isso pode ser explicado devido ao aumento do número de radicais instáveis metileno ( $\text{CH}_2$ ), que tendem a acelerar o processo de combustão,

diminuindo o tempo de ignição e conseqüentemente elevando o valor do IC. Desta forma, não só o valor positivo de sua constante, mas também o maior poder de interferir positivamente no IC conferido ao NCT fazem todo o sentido.

Especificamente para este banco de dados, as ramificações nas cadeias carbônicas, quando presentes, ocorrem apenas na parte alcoólica das moléculas, parte esta que possui duas outras variáveis (NCA e NHC) que a descrevem. Esta particularidade não permite aqui uma discussão mais aprofundada da interferência de ramificações no valor do IC.

De qualquer forma, a análise da interferência do NCR no IC deve ser feita em função da posição da ramificação na cadeia carbônica. Nesta base de dados a ramificação ocorre em duas posições: no carbono carbinólico ou no carbono  $\alpha$  ao carbono carbinólico.

Quando a ramificação ocorre no carbono carbinólico a avaliação não é direta, pois este carbono se refere diretamente a outra variável, o NHC. A presença de uma ramificação nesta posição altera o número de átomos de hidrogênio ligados ao carbono em questão, sendo o NHC decrescido de uma unidade. Para melhor entendimento, observemos as fórmulas estruturais do palmitato de propila e do palmitato de isopropila (Figuras 21 e 22 respectivamente) e as variáveis relativas a estas moléculas (Tabela 14).



**Figura 21.** Fórmula estrutural do palmitato de propila.



**Figura 22.** Fórmula estrutural do palmitato de isopropila.

**Tabela 14.** Ilustração das linhas referentes ao palmitato de propila e de isopropila e o IC calculado pela equação (4).

Biocombustíveis	NHI	NCR	NCA	NCT	NHC	IC simulado
<b>Palmitato de propila</b>	0	0	3	19	2	82,7
<b>Palmitato de isopropila</b>	0	1	3	19	1	85,6

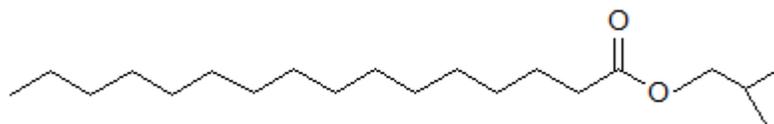
As duas estruturas químicas diferem apenas em relação ao álcool precursor utilizado, os dois álcoois possuem 3 carbonos (mesmo NCA), todavia quando utiliza-se o álcool isopropílico forma-se uma molécula ramificada no carbono carbinólico (Figura 22), por outro lado o álcool propílico forma um biocombustível de cadeia carbônica reta, isto é, sem ramificações (Figura 21). O resultado disso é que os referidos biocombustíveis diferem em apenas duas variáveis, o NCR e o NHI.

Ao se comparar o cálculo do IC dos mesmos através da equação (4), o palmitato de isopropila (IC = 85,6) terá um adicional de +3,5 (NCR = 1) – 0,6 (NHC decrescendo de 2 a 1) = +2,9 unidades em relação ao palmitato de propila (IC = 82,7).

Por outro lado, quando a ramificação ocorre no carbono  $\alpha$  ao carbono carbinólico, a avaliação passa a ser direta, pois os outros parâmetros do modelo permanecem inalterados. Para melhor entendimento, observemos as fórmulas estruturais do palmitato de butila e do palmitato de isobutila (Figuras 23 e 24 respectivamente) e as variáveis relativas a estas moléculas (Tabela 15).



**Figura 23.** Fórmula estrutural do palmitato de butila.



**Figura 24.** Fórmula estrutural do palmitato de isobutila.

**Tabela 15.** Ilustração das linhas referentes ao palmitato de butila e de isobutila e o IC calculado pela equação (4).

Biocombustíveis	NHI	NCR	NCA	NCT	NHC	IC simulado
<b>Palmitato de butila</b>	0	0	4	20	2	85,9
<b>Palmitato de isobutila</b>	0	1	4	20	2	89,4

Novamente as duas estruturas químicas em questão diferem apenas em relação ao álcool precursor utilizado, os dois álcoois possuem 4 carbonos (mesmo NCA), porém, desta vez, a ramificação formada a partir da utilização do álcool isobutílico ocorre no carbono  $\alpha$  ao carbinólico (Figura 24), sendo assim esta não altera os outros descritores do banco de dados. Em outras palavras, quando as variáveis dos dois biocombustíveis são comparadas, a partir da Tabela 9 têm-se que a única diferença presente está no NCR.

Ao se comparar o cálculo do IC dos referidos biocombustíveis através da equação (4), o palmitato de isobutila (IC = 85,9) terá um adicional de +3,5 unidades (NCR = 1) em relação ao palmitato de butila (IC = 89,4).

Poderia parecer um contrassenso o valor positivo no modelo para o aumento de ramificações (NCR) de uma cadeia, visto que teríamos a presença de carbonos mais ramificados nas moléculas e, conseqüentemente, durante a combustão, ocorreria à formação de radicais mais estáveis, o que tenderia a frear processo radicalar. Pois bem, ao mesmo tempo em que este raciocínio é válido, o aumento do NCR provoca também um aumento do NCT, e KNOTHE; MATHEAUS; RYAN III (2003) ao fazer estudos com ésteres derivados de álcoois ramificados verificou que este aumento (da cadeia carbônica do biocombustível, representado aqui pelo NCT) aparentemente compensa o efeito negativo da formação destes radicais e chega à conclusão que, de forma geral, os biocombustíveis derivados de álcoois ramificados têm valores de IC um pouco mais elevados do que os derivados de álcoois metílicos e etílicos.

Os resultados obtidos através do modelo preditivo assim como os dados da literatura com os respectivos erros relativos podem ser visualizados na Tabela 16.

**Tabela 16.** Resultados obtidos pelo modelo, dados da literatura e os erros relativos.

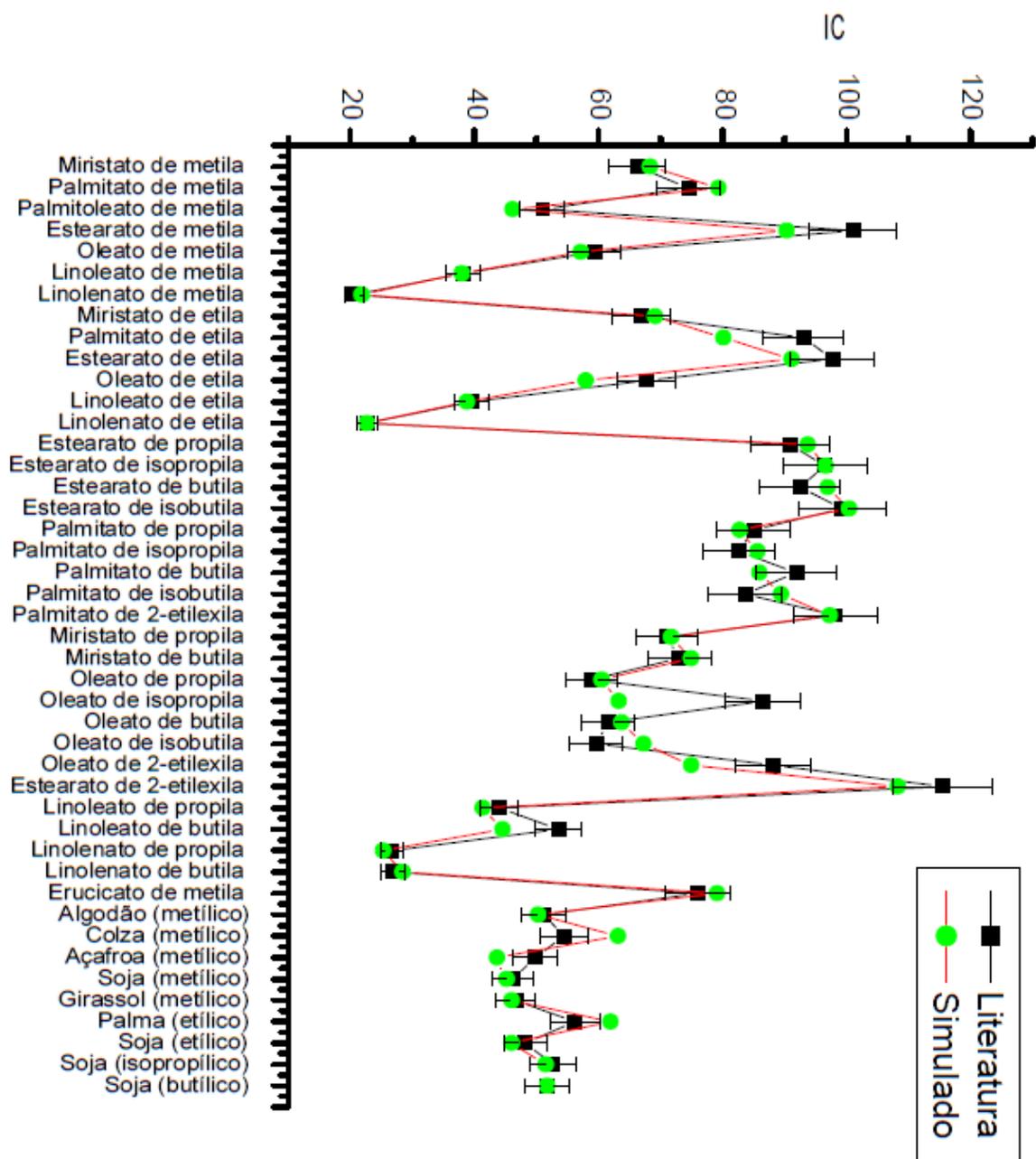
Biocombustíveis	IC simulado	IC literatura	Erro relativo
Miristato de metila	68,3	66,2 <sup>c</sup>	3,2
Palmitato de metila	79,3	74,5 <sup>c</sup>	6,4
Palmitoleato de metila	46,1	51,0 <sup>a</sup>	9,8
Estearato de metila	90,3	101,0 <sup>a</sup>	10,6
Oleato de metila	57,1	59,3 <sup>a</sup>	3,7
Linoleato de metila	37,9	38,2 <sup>a</sup>	0,8
Linolenato de metila	21,8	20,6 <sup>b</sup>	5,8
Miristato de etila	69,1	66,9 <sup>c</sup>	3,3
Palmitato de etila	80,1	93,1 <sup>a</sup>	14,0
Estearato de etila	91,1	97,7 <sup>a</sup>	6,7
Oleato de etila	57,9	67,8 <sup>a</sup>	13,4
Linoleato de etila	38,7	39,6 <sup>a</sup>	2,3
Linolenato de etila	22,6	22,7 <sup>c</sup>	0,4

Biocombustíveis	IC simulado	IC literatura	Erro relativo
Estearato de propila	93,7	90,9 <sup>a</sup>	3,1
Estearato de isopropila	96,6	96,5 <sup>a</sup>	0,1
Estearato de butila	96,9	92,5 <sup>a</sup>	4,7
Estearato de isobutila	100,4	99,3 <sup>a</sup>	1,1
Palmitato de propila	82,7	85,0 <sup>a</sup>	2,7
Palmitato de isopropila	85,6	82,6 <sup>a</sup>	3,6
Palmitato de butila	85,9	91,9 <sup>a</sup>	6,5
Palmitato de isobutila	89,4	83,6 <sup>a</sup>	6,9
Palmitato de 2-etilexila	97,2	98,2 <sup>a</sup>	1,0
Miristato de propila	71,7	71,0 <sup>d</sup>	1,0
Miristato de butila	74,9	73,0 <sup>d</sup>	2,6
Oleato de propila	60,5	58,8 <sup>a</sup>	2,9
Oleato de isopropila	63,2	86,6 <sup>a</sup>	27,0
Oleato de butila	63,7	61,6 <sup>a</sup>	3,4
Oleato de isobutila	67,2	59,6 <sup>a</sup>	12,7
Oleato de 2-etilexila	74,9	88,2 <sup>a</sup>	15,1
Estearato de 2-etilexila	108,2	115,5 <sup>a</sup>	6,3
Linoleato de propila	41,3	44,0 <sup>a</sup>	6,1
Linoleato de butila	44,5	53,5 <sup>a</sup>	16,8
Linolenato de propila	25,2	26,7 <sup>c</sup>	4,5
Linolenato de butila	28,4	26,8 <sup>c</sup>	6,0
Erucicato de metila	79,1	76,0 <sup>b</sup>	4,1
Algodão (metílico)	50,2	51,2 <sup>b</sup>	1,9
Colza (metílico)	63,1	54,4 <sup>b</sup>	16,0
Açafoa (metílico)	43,6	49,8 <sup>b</sup>	12,4
Soja (metílico)	45,1	46,2 <sup>b</sup>	2,4
Girassol (metílico)	45,9	46,6 <sup>b</sup>	1,5
Palma (etílico)	61,9	56,2 <sup>b</sup>	10,1
Soja (etílico)	45,9	48,2 <sup>b</sup>	4,8
Soja (isopropílico)	51,4	52,6 <sup>b</sup>	2,3
Soja (butílico)	51,7	51,7 <sup>b</sup>	0,0

<sup>a</sup> (KNOTHE; MATHEAUS; RYAN III, 2003). <sup>b</sup> (KNOTHE; DUNN; BAGBY, 1997).

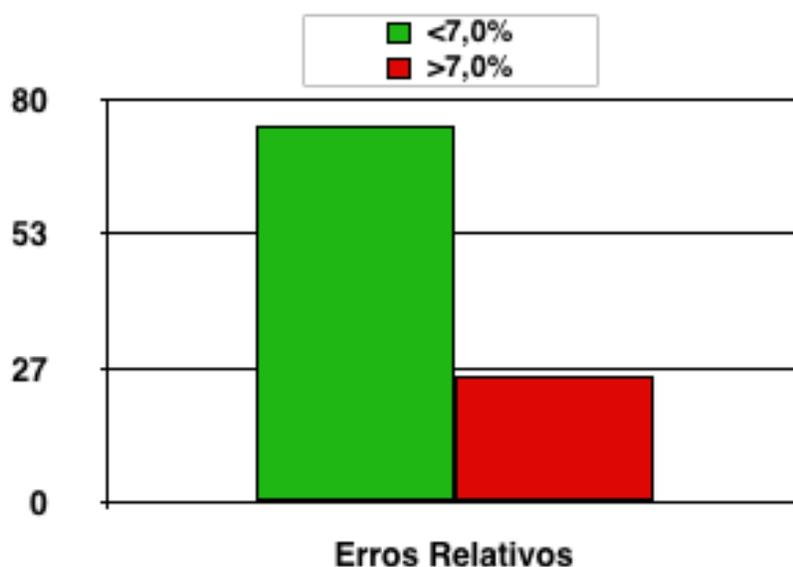
<sup>c</sup> (KNOTHE, 2005). <sup>d</sup> (SERDARI; LOIS; STOURNAS, 1999).

Os resultados obtidos a partir da equação (4) que podem ser observados na Tabela 16 estão representados no Gráfico 4 a seguir:



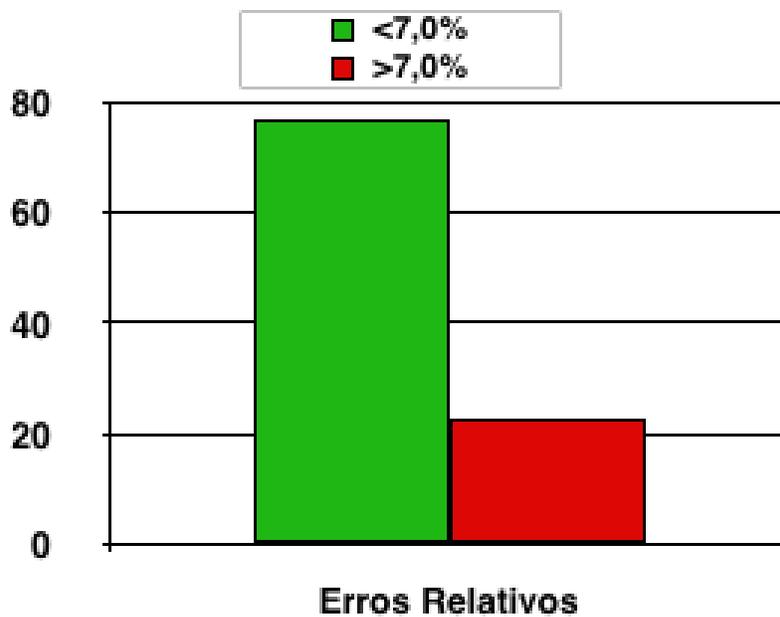
**Gráfico 4.** Comparação entre os dados da literatura com as respectivas barras de erro e os valores simulados pelo modelo preditivo.

A partir da Tabela 16 observa-se que 75% (33 entre 44 elementos) das observações, com erros relativos variando entre 0,0% e 6,9%, obtiveram diferenças inferiores a 7,0%, isto é, dentro da margem de erro admitida. Enquanto os outros 25% (11 entre 44 elementos), com valores variando entre 9,8% e 27,0%, obtiveram erro relativo fora da margem admitida. O gráfico 5 mostra os resultados obtidos sob esta ótica:



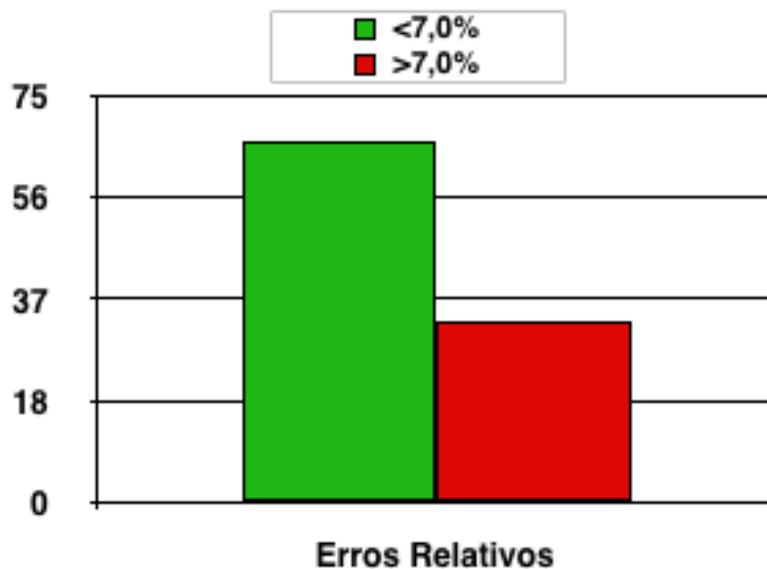
**Gráfico 5.** Frequências dos erros relativos observados na Tabela 16.

Quando separamos os dados em substâncias puras e misturas, os resultados obtidos não diferem muito da proporção exposta no gráfico anterior. Para as substâncias puras (Gráfico 6) observa-se que 27 (77,1%) das 35 constantes na base de dados obtiveram diferenças inferiores a 7,0%.



**Gráfico 6.** Frequências dos erros relativos observados para os ésteres do banco de dados.

Enquanto para os diferentes tipos de biodiesel (Gráfico 7) observa-se que 6 (66,7%) dos 9 constantes na base de dados obtiveram diferenças inferiores a 7,0%.



**Gráfico 7.** Frequências dos erros relativos observados para os diferentes tipos de biodiesel do banco de dados.

### **III.3. Conclusões**

O modelo empírico desenvolvido no capítulo anterior funcionou muito bem para a tarefa que lhe foi dada, de prever o IC para os todos os tipos de biodiesel presentes na matriz de dados. Porém, as situações inconvenientes, relativamente à reprodutibilidade dos resultados, quando desejássemos agregar outros biocombustíveis àquele conjunto de dados, direcionaram os esforços no caminho uma nova metodologia, que fosse mais simples, prática e principalmente reprodutível sob quaisquer aspectos.

É muito importante mencionar que o sucesso obtido neste capítulo do trabalho, advém, intimamente, do entendimento das variáveis envolvidas na predição dos IC's ocorrida nos capítulos 1 e 2 desta tese. As principais variáveis anteriormente identificadas foram transcritas em linguagem matemática e utilizadas nesta etapa. O número de classes de similaridade (cinco) propostas no modelo empírico é igual ao número variáveis utilizadas no modelo preditivo. Todas estas considerações promovem o modelo preditivo como uma evolução do modelo empírico.

O coeficiente de correlação obtido (0,93) apontou o aguçado poder de predição do modelo. Os resultados se mostraram muito positivos, com a esmagadora maioria (três quartos do total) obtendo valores dentro de uma margem de erro inferior a 7% dos dados experimentais. No caso específico dos tipos de biodiesel, o uso de porcentagens médias é uma potencial fonte de erro, o ideal seria que soubéssemos a composição química das amostras que foram queimadas, e, portanto, determinássemos o valor de cada variável para cada biodiesel. Ainda assim, 6 dos 9 tipos de biodiesel apresentaram resultados dentro da margem de erro experimental, o que comprova o forte poder de predição do modelo desenvolvido.

Do ponto de vista químico, o modelo destaca duas variáveis, o NCT com a maior contribuição para o aumento do IC e o NHI com o efeito inverso. Por motivos já colocados anteriormente, esta linha de raciocínio está de acordo com os conceitos químicos envolvidos e conseqüentemente com os relatos da literatura citados. Essa correlação (estrutura química x propriedade) é muito importante, pois mostra que os resultados obtidos aqui não foram fruto de uma simples coincidência matemática.

Estes tipos de estudos, que visam o esclarecimento das relações estrutura química x propriedade, podem auxiliar de sobremaneira no desenvolvimento de novos combustíveis. Uma vez que, quando estiverem quantificados estes relacionamentos, pesquisadores da área poderão pré-estabelecer as propriedades físico-químicas específicas em função da variação nas estruturas químicas dos combustíveis em questão.

Diante da dificuldade experimental envolvida na obtenção deste índice, as principais características do modelo são a simplicidade e a praticidade. Um bom aluno de segundo grau, com noções básicas de química orgânica e que saiba operar uma calculadora científica tem condições de calcular o IC utilizando este método.

Quando nos deparamos com o alto custo envolvido na obtenção desta propriedade, todos os resultados experimentais dos IC's foram obtidos a partir de dados da literatura e o ambiente estatístico R utilizado é gratuito, desta forma a "despesa" do trabalho para a instituição foi praticamente zero, sendo necessário apenas um computador para o seu desenvolvimento.

#### 4. REFERÊNCIAS BIBLIOGRÁFICAS

- ALDENDERFER, M., BLASHFIELD, R. **Cluster Analysis**. Thousand Oaks, USA: Sage Publications, 1984.
- AL-KHEDHER, M. A., PEZESHKI, C., MCHALE, J. L., KNORR, F. J. Quality classification via Raman identification and SEM analysis of carbon nanotube bundles using artificial neural networks. **Nanotechnology**. v.18, artigo número 355703, 2007.
- AYHAN, D. Biodiesel fuels from vegetable oils via catalytic and non-catalytic supercritical alcohol transesterification and other methods: a survey. **Energy Conv Manage**. v.44, p.2093-2109, 2003.
- BEEBE, K. R., KOWALSKI, B. R. An introduction to multivariate calibration and analysis. **Analytical Chemistry**. v.59, p.1007A-1010A, 1987.
- BIGI, B. Using Kullback-Leibler Distance for Text Categorization. Springer-Verlag Berlin Heidelberg, 2003.
- BRUNS, R. E., FAIGLE, J. F. G. Quimiometria. **Química Nova**. v.8, n.2, p.84-99, 1985.
- BUNNELL, R., EKWARU, J. P., SOLBERG, P., WAMAI, N., BIKAAKO-KAJURA, W., WERE, W., COUTINHO, A., LIECHTY, C., MADRAA, E., RUTHERFORD, G., MERMIN, J. Changes in sexual behavior and risk of HIV transmission after antiretroviral therapy and prevention interventions in rural Uganda. **AIDS**. v.20, n.1, p.85-92, 2006.
- CANAKCI, M., ERDIL, A., ARCAKLIOGLU, E. Performance and exhaust missions of a biodiesel engine. **Applied Energy**. v.83, p.594–605, 2006.
- CONCEIÇÃO, M. M., CANDEIA, R. A., DANTAS, H. J., SOLEDADE, L. E. B., FERNANDES JR, V. J., SOUZA, A. G. Rheological Behavior of Castor Oil Biodiesel. **Energy & Fuels**. v.19, p.2185-2188, 2005.

- CORTER, J. E. **Tree Models of Similarity and Association**. Thousand Oaks, CA: Sage Publications, 1996.
- CORTÉS, J. A., PALMA, J. L., WILSON, M. Deciphering magma mixing: The application of cluster analysis to the mineral chemistry of crystal populations. **Journal of Volcanology and Geothermal Research**. v.165, p.163-188, 2007.
- COSTA NETO, P. R., ROSSI, L. F. S. The utilization of used frying oil for the production of biodiesel. **Química Nova**, v.23, n.4, p.531-537, 2000.
- DASZYKOWSKI, M., WALCZAK, B., MASSART, D. L. Projection methods in chemistry. **Chemometrics and Intelligent Laboratory Systems**. v.65, p.97-112, 2003.
- DAVIDSON, I., RAVI, S. S. The complexity of non-hierarchical clustering with instance and cluster level constraints. **Data Min Knowl Disc**. v.14, p.25-61, 2007.
- DE MAESSCHALCK, R., JOUAN-RIMBAUD, D., MASSART, D. L., The Mahalanobis distance. **Chemometrics and Intelligent Laboratory Systems**. V. 50 (1), p. 1-18, 2000.
- D'HULST, R., RODGERS, G. J. The Hamming distance in the minority game. **Physica A: Statistical Mechanics and its Applications**. V. 270 (3-4), p. 514-525, 1999.
- DICKMAN, P. W., SLOGGETT, A., HILLS, M., HAKULINEN, T. Regression models for relative survival. **Statistics in Medicine**. v.23, p.51-64, 2004.
- ENCINAR, J. M., GONZÁLEZ, J. F., RODRÍGUEZ-REINARES, A. Biodiesel from Used Frying Oil: Variables Affecting the Yields and Characteristics of the Biodiesel. **Ind. Eng. Chem. Res**. v.44, p.5491-5499, 2005.
- ERCANLI, T., BOYD, D. B. Evaluation of Computational Chemistry Methods: Crystallographic and Cheminformatics Analysis of Aminothiazole Methoximes. **J. Chem. Inf. Model**. v.45, p.591-601, 2005.

- EVERITT, B., LANDAU, S., LEESE, M., **Cluster Analysis**. Oxford: Oxford University Press, 2001.
- FANGRUI, M. A., MILFORD, A. H. Biodiesel production: A review. **Bioresource Technol.** v.70, p.1-15, 1999.
- FERREIRA, E. C., RODRIGUES, S. H. B. G., FERREIRA, M. M. C., NÓBREGA, J. A., NOGUEIRA, A. R. A. Análise exploratória dos teores de constituintes inorgânicos em sucos e refrigerantes de uva. **Eclética Química**, v.27 n.especial, 2002.
- FERRARI, R. A., OLIVEIRA, V. DA S., SCARBIO, A. Biodiesel de soja - Taxa de Conversão em Ésteres Etílicos, Caracterização Físico-Química e Consumo em Gerador de Energia. **Química Nova**. v.28, p.19-23, 2005.
- GROENEN, P. J. F., MATHAR R., HEISER W. J. The majorization approach to multidimensional scaling for Minkowski distances. **Journal of classification**. V. 12 (1), p. 3-19, 1995.
- GRUBESIC, T. H. On The Application of Fuzzy Clustering for Crime Hot Spot Detection. **Journal of Quantitative Criminology**. v.22, p.77-105, 2006.
- HAAS, M. J., MCALOON, A. J., YEE, W. C., FOGLIA, T. A. A process model to estimate biodiesel production costs. **Bioresource Technology**. v.97, p.671-678, 2006.
- HANSELMAN, D., LITTLEFIELD, B. **MATLAB 6: Curso Completo**, São Paulo: Editora Parentice Hall, 2003.
- HARDLE, W., SIMAR, L. **Applied Multivariate Statistical Analysis**. Berlin: Springer, 2003.

- [home.cc.umanitoba.ca/~budzelaa/gNMR/gNMR-license.html](http://home.cc.umanitoba.ca/~budzelaa/gNMR/gNMR-license.html), acessado em 07/05/2006.
- HUTTENLOCHER, D. P., KLANDERMAN, G. A., RUCKLIDGE, W. A. Comparing Images Using the Hausdorff Distance. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v. 15 (9), p. 850-863, 1993.
- JERRAM, D. A., CHEADLE, M. J. On the cluster analysis of grains and crystals in rocks. **American Mineralogist**. v.85, p.47-67, 2000.
- JOHNSON, M., BALKENIUS, C. Neural network models of haptic shape perception. **Robotics and Autonomous Systems**. v.55, p.720-727, 2007.
- JONGMAN, R. H. G., TER BRAAK, C. J. F., VAN TONGEREN, O. F. R. **Data Analysis in Community and Landscape Ecology**. Cambridge University Press, 1995.
- KASABOV, N. K. **Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering**. 2<sup>nd</sup> edition: Massachusetts, USA: The MIT Press, 1998.
- KAUFMAN, L., ROUSSEEUW, P. **Finding Groups in Data: An Introduction to Cluster Analysis**. John Wiley & Sons, 1990.
- KAYMAK, U.; SETNES, M. Fuzzy Clustering With Volume Prototypes and Adaptive Cluster Merging. **IEEE TRANSACTIONS ON FUZZY SYSTEMS**, V. 10 (6), 2002.
- KNOTHE G. Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. **Fuel Processing Technology**. v.86, p.1059-1070, 2005.
- KNOTHE, G., BAGBY, M. O., RYAN III, T. W. Cetane numbers of fatty compounds: influence of compound structure and various potential cetane improvers. **Society of Automotive Engineers**. artigo número 971681, 1997.

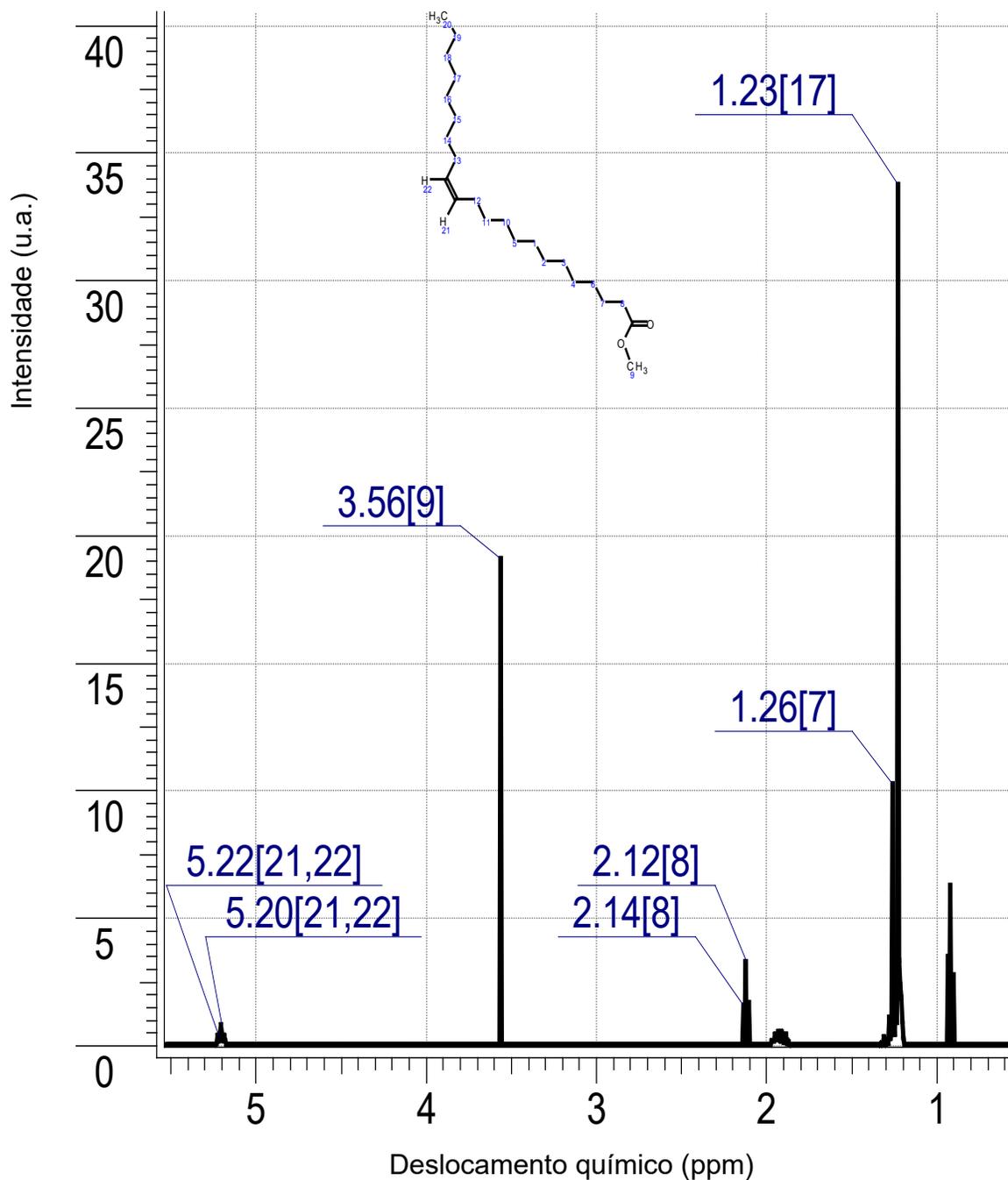
- KNOTHE, G., DUNN, R. O., BAGBY, M. O. Biodiesel: The Use of Vegetable Oils and Their Derivatives as Alternative Diesel Fuels. **Fuels and Chemicals from Biomass**, Washington, D.C.: American Chemical Society, 1997.
- KNOTHE, G., GERPEN, J. V., KRAHL, J. **The Biodiesel Handbook**. 1<sup>a</sup> ed. Champaign, Illinois: AOCS Press, 2005.
- KNOTHE, G., MATHEAUS, A. C., RYAN III, T. W. Cetane Numbers of Branched and Straight-Chain Fatty Esters Determined in an Ignition Quality Tester. **Fuel**. v.82, p.971-975, 2003.
- LAROSE, D. T. **Data Mining Methods and Models**. Hoboken, NJ: Wiley, 2006.
- LEGENDRE, P., LEGENDRE, L. **Numerical Ecology**. Amsterdam: Elsevier, 1998.
- LIMA, J. A. P., MASSUNAGA, M. S. O., VARGAS, H., MIRANDA, L. C. M. Photothermal Detection of Adulterants in Automotive Fuels. **Analytical chemistry**. v. 76 (1), p. 114-119, 2004.
- MAQBOOL, O., BABRI, H. A. Hierarchical Clustering for Software Architecture Recovery. **IEEE Transactions on Software Engineering**. v.33, n.11, p.759-780, 2007.
- MICHAEL, S. G., ROBERT, L. M. Combustion of fat and vegetable oil derived fuels in diesel engines. **Progr Combust Sci**. v.24, p.125-64, 1998.
- NADAI, D. V., MIRANDA, P. C. M. L., GATTS, C. E. N. Study of the correlation between the chemical structure of biodiesel and their respective cetane number. **International Review of Chemical Engineering**. v. 1, p. 598-602, 2009.
- NETO, J. M. M., MOITA, G. C. Uma Introdução À Análise Exploratória De Dados Multivariados. **Química Nova**. v.21, n.4, p.467-469, 1998.
- PILLAR, V. How sharp are classifications. **Ecology**. v.80, p.2508-2516, 1999.

- RAMADHAS, A. S., JAYARAJ, S., MURALEEDHARAN, C., PADMAKUMARI, K. Artificial neural networks used for the prediction of the cetane number of biodiesel. **Renewable Energy**. Technical note. v.31, p.2524-2533, 2006.
- RAMIREZ-ARIZMENDI, L., HAMJE, H. D. Fourier transform infrared (FTIR) chemometric method to determine cetane number of diesel fuels containing fatty acid alkyl ester additives. **United States Patent 7420170**. 01 set. 2007, 09 fev. 2008.
- SERDARI, A., LOIS, E., STOURNAS S. Impact of Esters of Mono- and Dicarboxylic Acids on Diesel Fuel Quality, **Ind. Eng. Chem. Res.** v.38, p.3543-3548, 1999.
- SILER, W., BUCKLEY, J. J. **Fuzzy Expert Systems and Fuzzy Reasoning**. Hoboken, USA: Wiley, 2005.
- SILVERSTEIN, R. M., WEBSTER, F. X. **Identificação Espectrométrica de Compostos Orgânicos**. 6ª ed. Rio de Janeiro: LTC, 2000.
- SOLOMONS, T. W. G. **Química Orgânica**. v.1, 9ª ed. Rio de Janeiro: LTC, 2009.
- TADINI, C. C., PARK, S. W., DING, H. Análise multivariada dos processos químicos: Parte I: Interpretação de dados via PCA (análise dos componentes principais). In: V SIMPÓSIO DE INICIAÇÃO CIENTÍFICA DA UNIVERSIDADE DE SÃO PAULO. 5., 1997, São Paulo. **Anais Do V Sicsup. São Paulo**. São Paulo: 1997. v.2, p.11.31.
- TAHERZADEH, M. J., KARIMI K., Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review. **International Journal of Molecular Sciences**. v.9, p. 1621–1651, 2008.
- TAYLOR, J. D., MCCORMICK, R. L., MURPHY, M. J. Compendium of Experimental Cetane Number Data. **National Renewable Energy Laboratory**, Relatório técnico de outubro. Golden, Colorado: EUA, 2004.

- THE DEPARTMENT OF THE ENVIRONMENT AND HERITAGE - AUSTRALIAN GOVERNMENT. Measuring Cetane Number: options for diesel and alternative diesel fuels. Discussion Paper, NATIONAL FUEL QUALITY STANDARDS, 2004.
- TOMASEVIC, A. V., SILER-MARINKOVIC, S. S. Methanolysis of used frying oil. **Fuel Process. Technol.** v.81, p.1, 2003.
- VENABLES, W. N., RIPLEY, B. D. **Modern Applied Statistics with S.** 4<sup>a</sup> ed.: Springer, 2002.
- [www.anp.gov.br](http://www.anp.gov.br) acessado em 14/02/11.
- [www.biodiesel.gov.br](http://www.biodiesel.gov.br) acessado em 10/08/06.
- [www.energiasrenovaveis.com](http://www.energiasrenovaveis.com) acessado em 14/02/2011.
- [www.fsf.org](http://www.fsf.org) acessado em 23/11/2007.
- [www.perkinelmer.com.br/downloads/biodiesel/Dra%20Ana%20Rita%20Fraga20Drummond.PDF](http://www.perkinelmer.com.br/downloads/biodiesel/Dra%20Ana%20Rita%20Fraga20Drummond.PDF). acessado em 29/06/07.
- [www.r-project.org](http://www.r-project.org) acessado em 23/11/2007.
- XU, J., ZHANG, Q., SHIH, C. V-cluster algorithm: A new algorithm for clustering molecules based upon numeric data. **Molecular Diversity.** v.10, p.463-478, 2006.
- YANG, H., RING, Z., BRIKER, Y., MCLEAN, N., FRIESEN, W., FAIRBRIDGE, C. Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by LC and GC–MS. **Fuel.** v. 81, p. 65-77, 2002.
- YANG, M. S. A survey of fuzzy clustering. **Mathematical and Computer Modelling.** V.18 (11), p.1-16, 1993.
- YUJIAN, L., BO, L., A Normalized Levenshtein Distance Metric. . **IEEE Transactions on Pattern Analysis and Machine Intelligence.** v. 29 (6), p. 1091-1095, 2007.

## 5. ANEXOS

Espectros teóricos de RMN-<sup>1</sup>H simulados pelo programa gNMR. No eixo das abscissas o deslocamento químico em ppm, e no eixo das ordenadas a intensidade dos picos em unidades arbitrárias.



**Figura 25.** Espectro de RMN-<sup>1</sup>H do erucicato de metila.

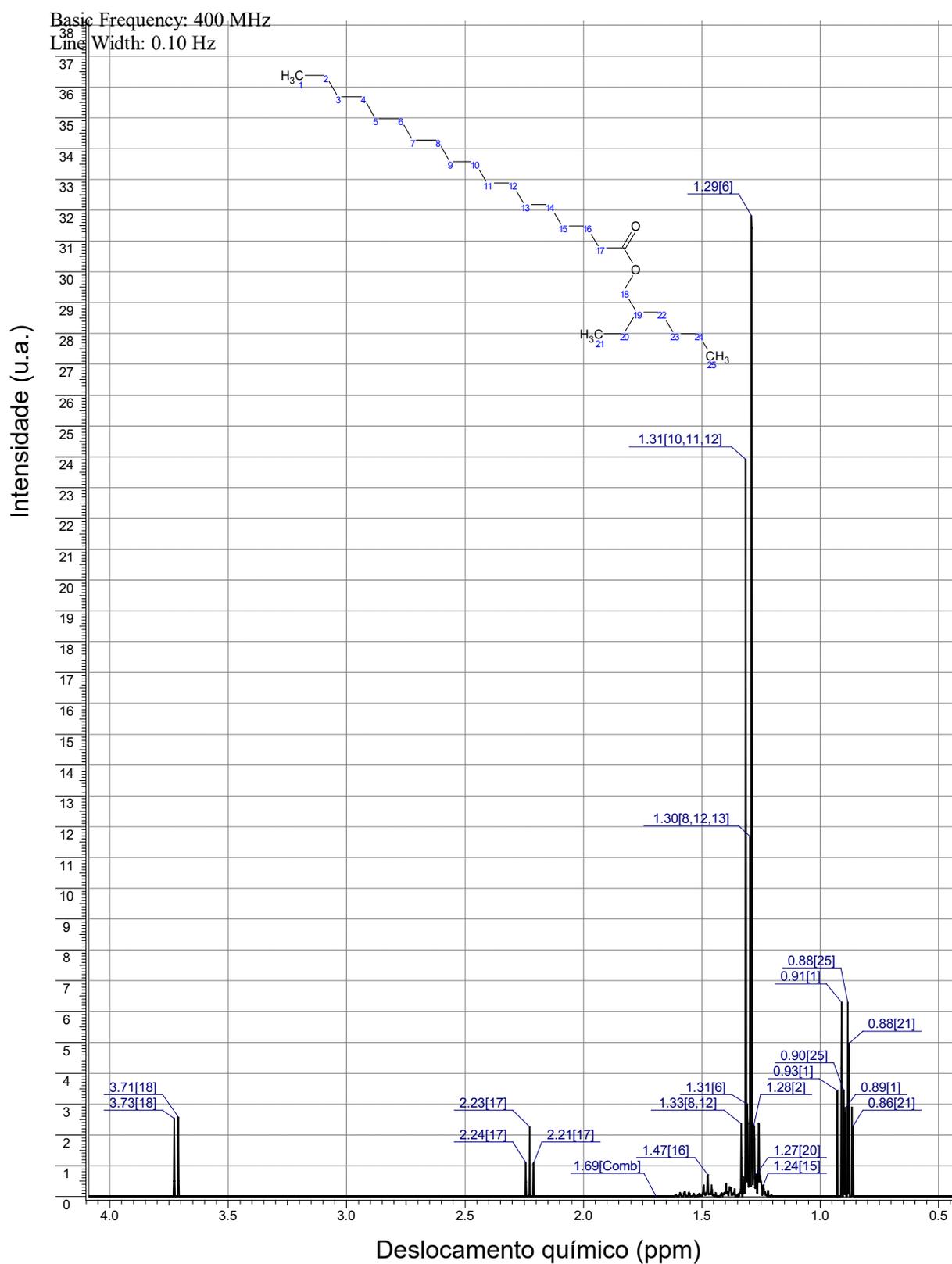


Figura 26. Espectro de RMN-<sup>1</sup>H do estearato de 2-etilexila.

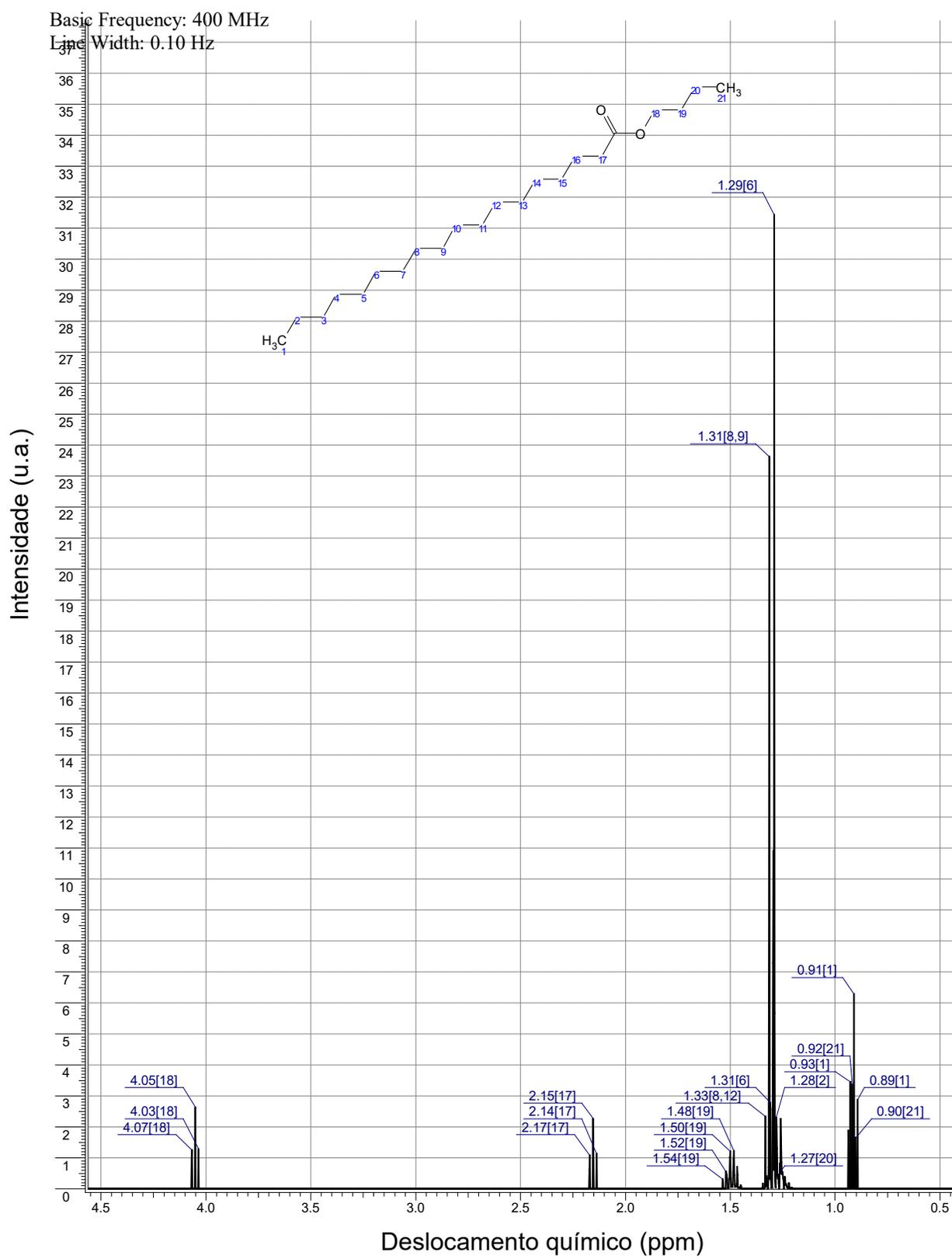


Figura 27. Espectro de RMN-<sup>1</sup>H do estearato de butila.

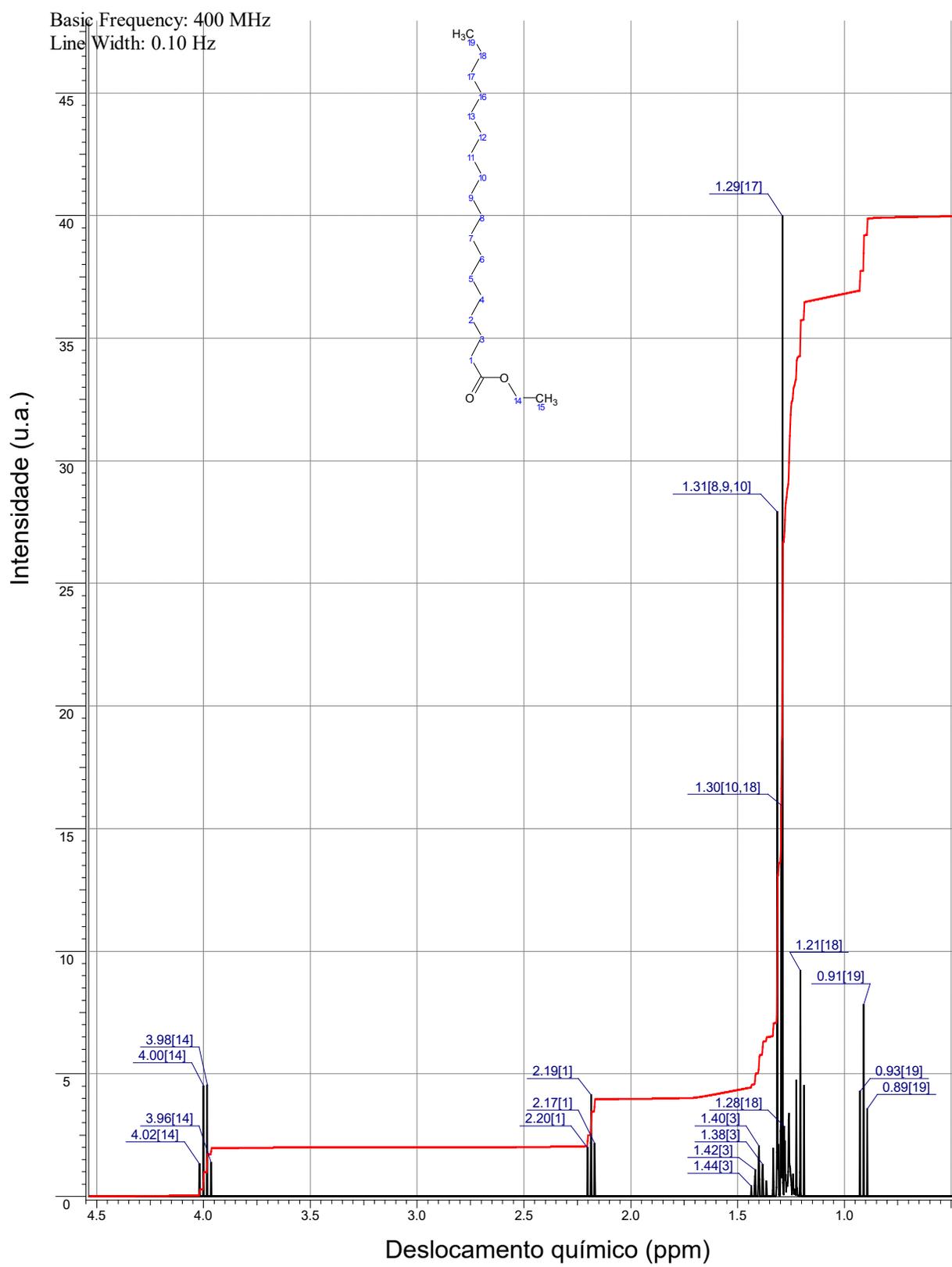


Figura 28. Espectro de RMN-<sup>1</sup>H do estearato de etila.

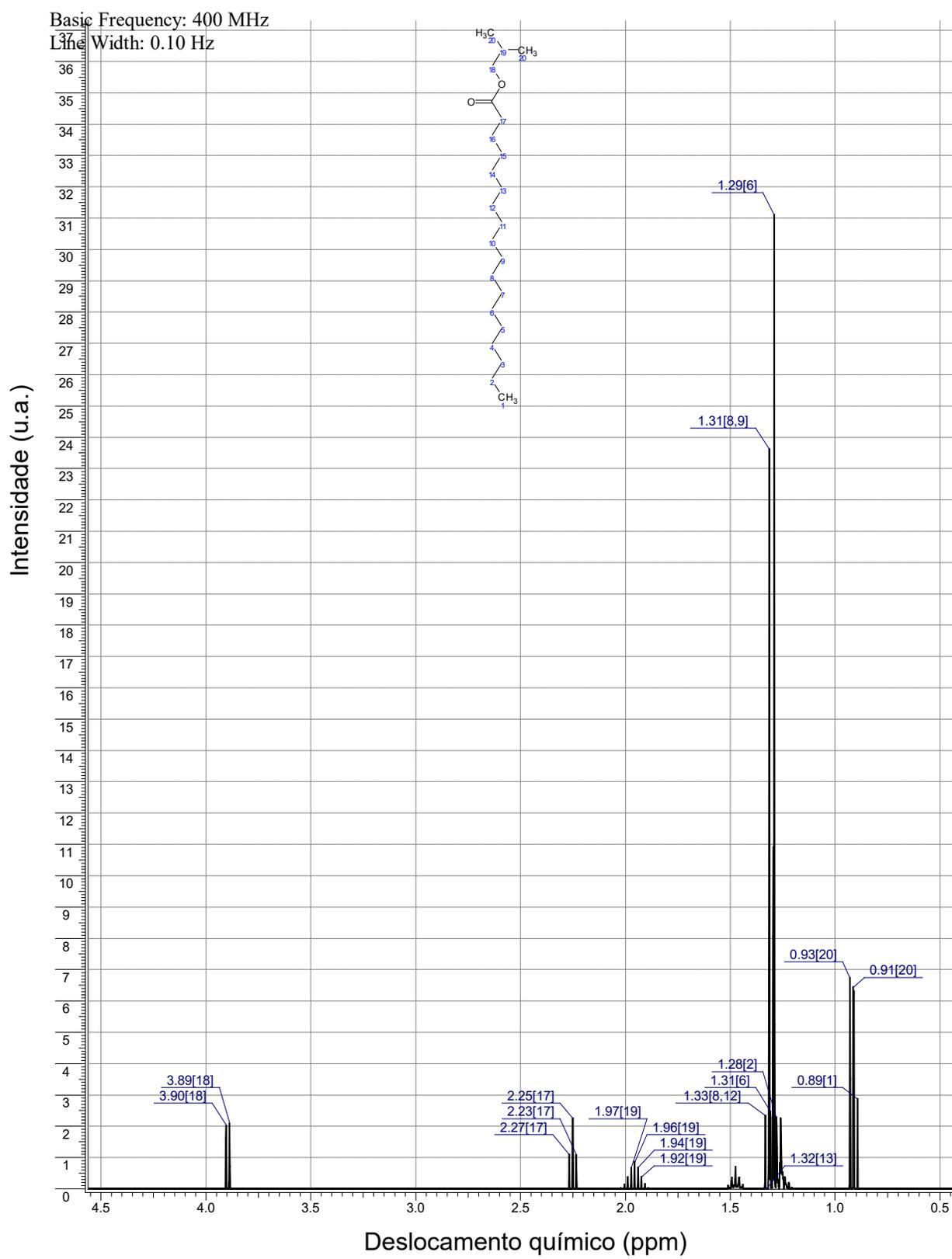


Figura 29. Espectro de RMN-<sup>1</sup>H do estearato de isobutila.

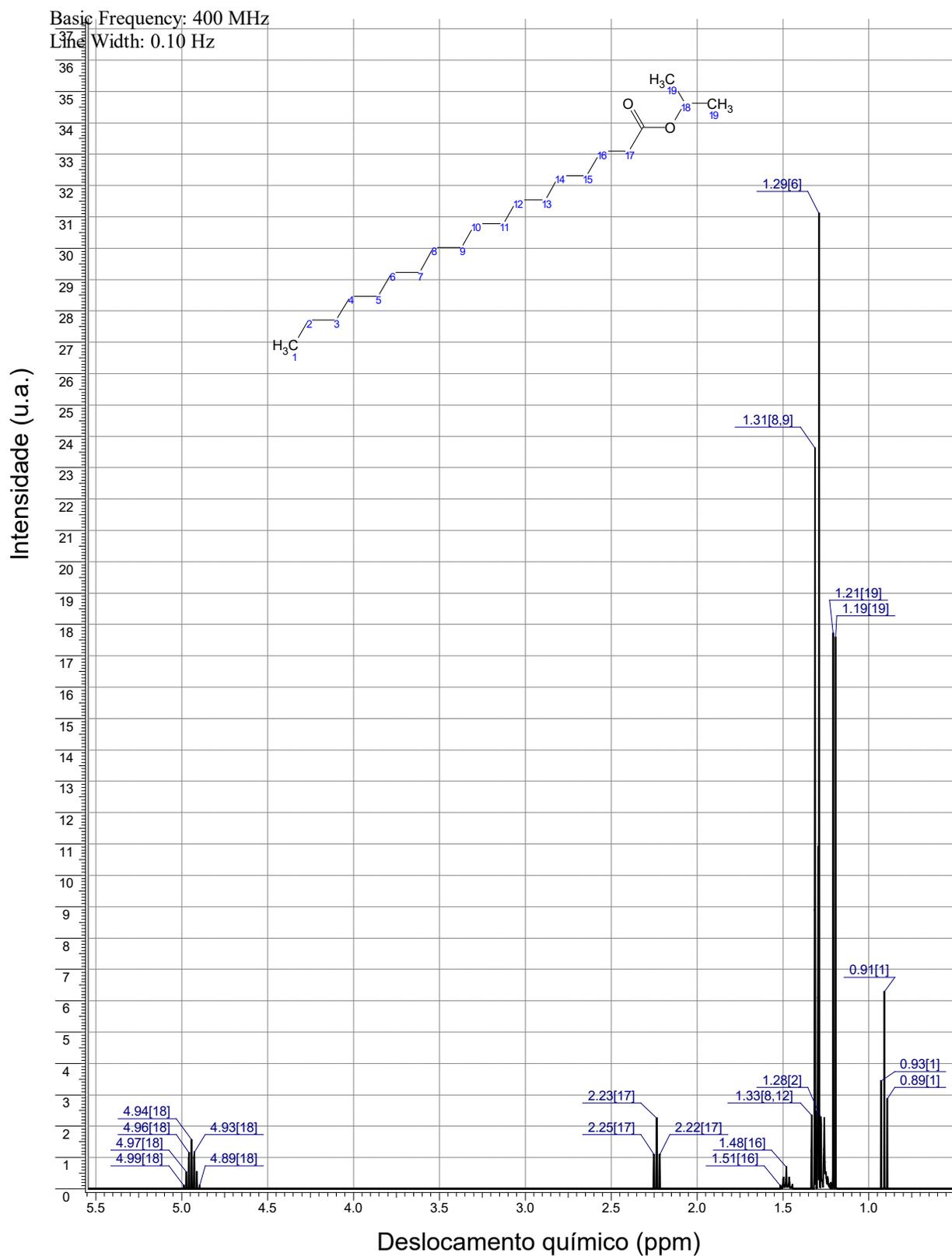


Figura 30. Espectro de RMN-<sup>1</sup>H do estearato de isopropila.

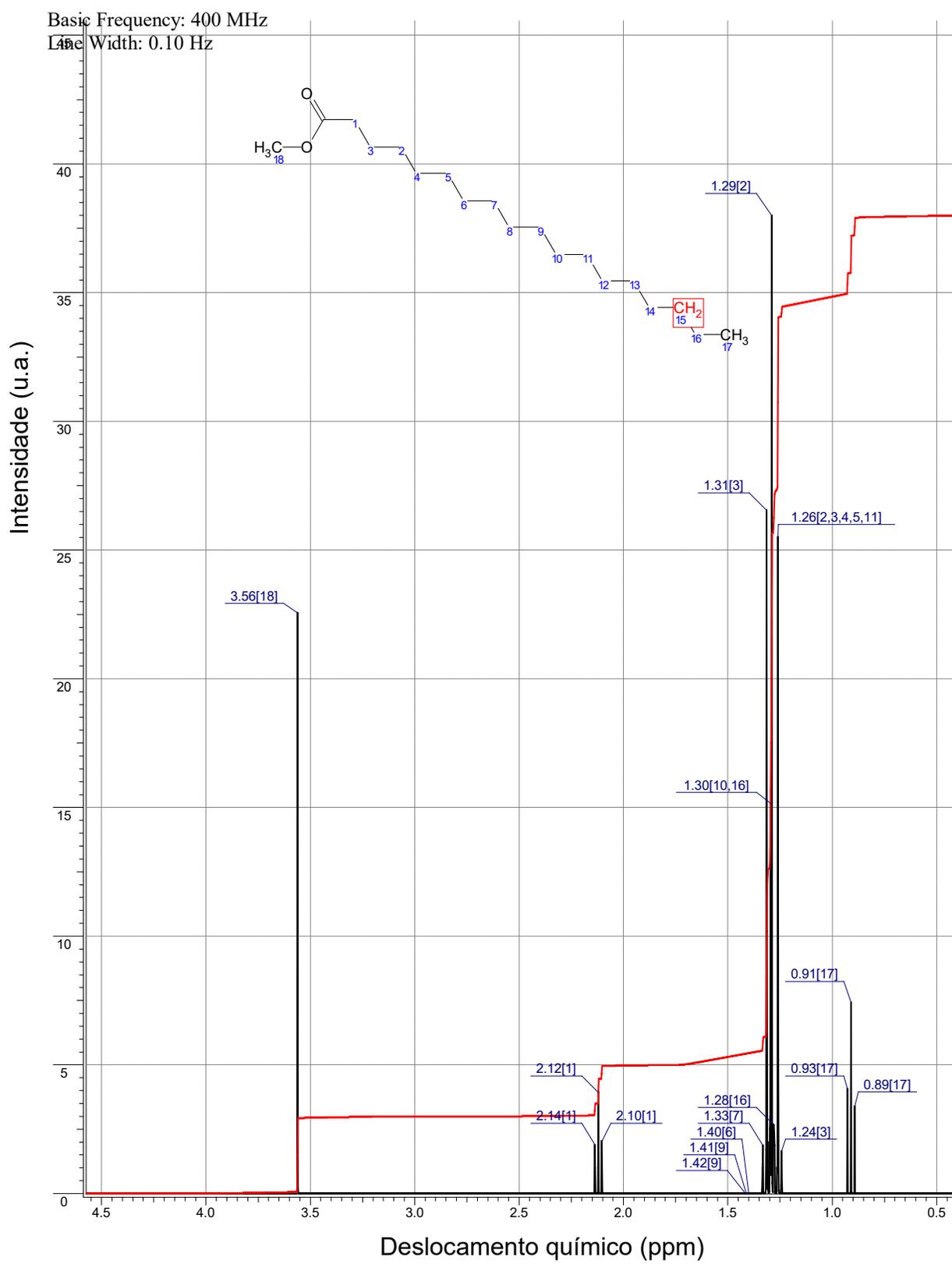


Figura 31. Espectro de RMN-<sup>1</sup>H do estearato de metila.

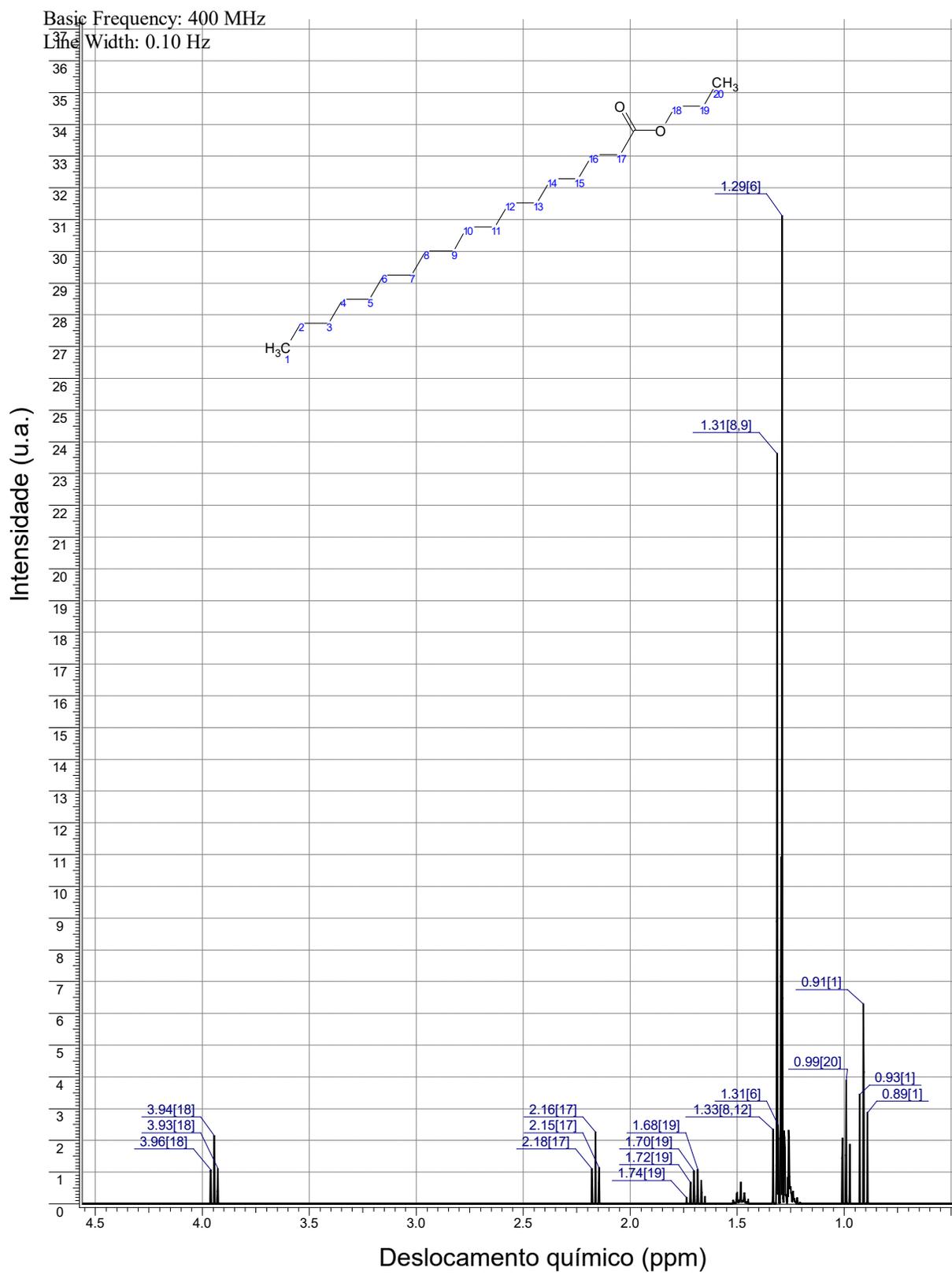


Figura 32. Espectro de RMN- $^1H$  do estearato de propila.

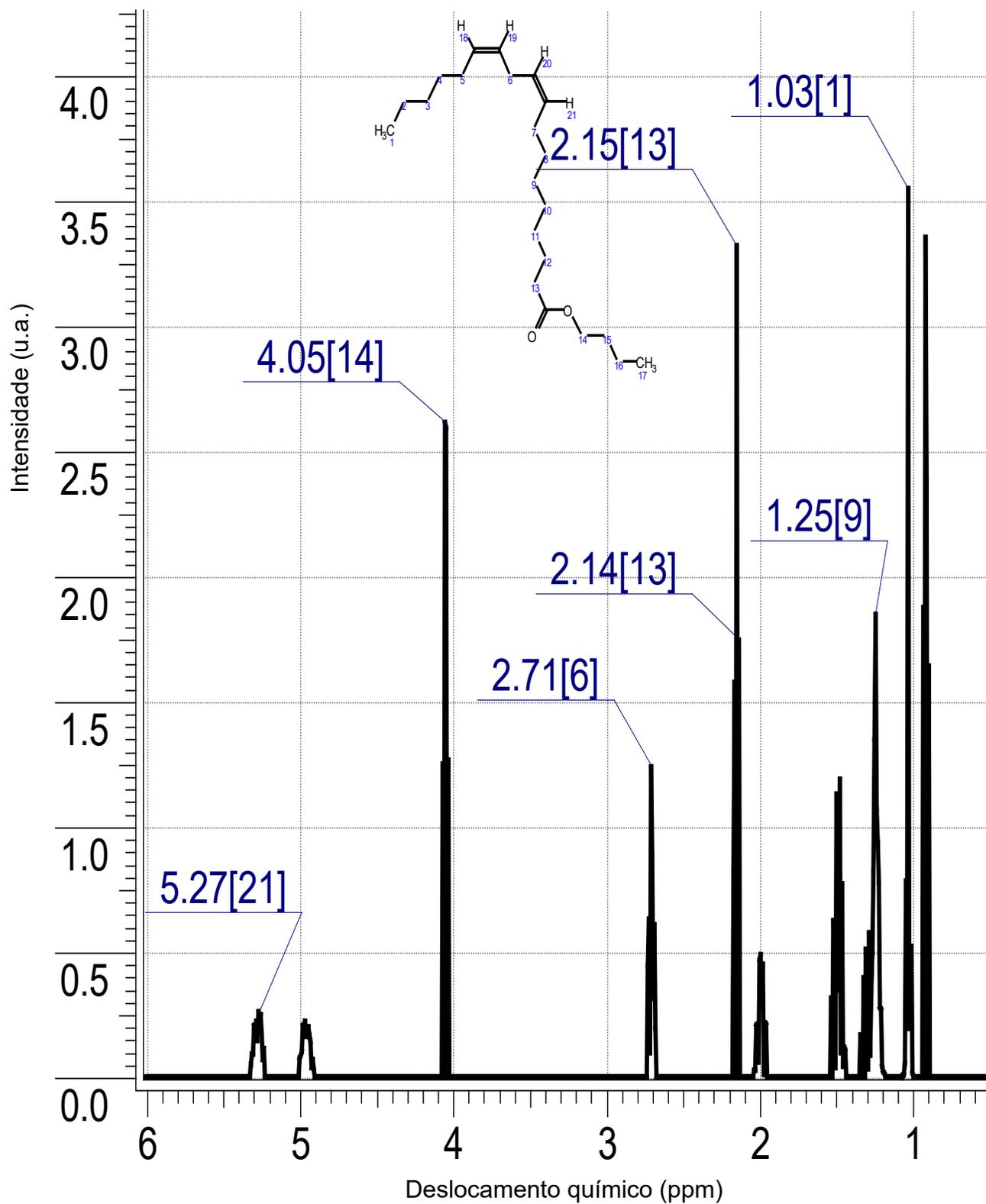


Figura 33. Espectro de RMN- $^1\text{H}$  do linoleato de butila.

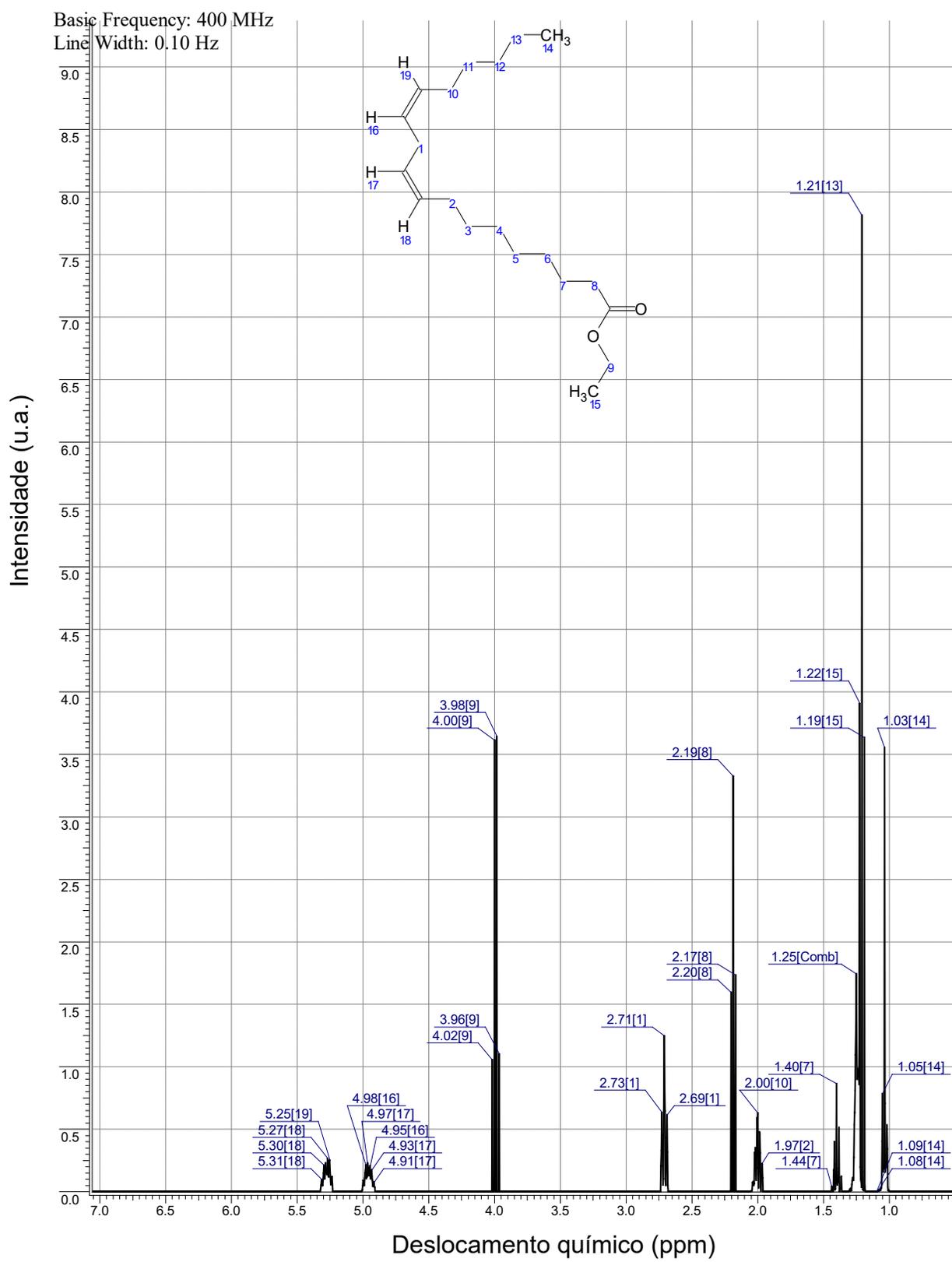


Figura 34. Espectro de RMN-<sup>1</sup>H do linoleato de etila.

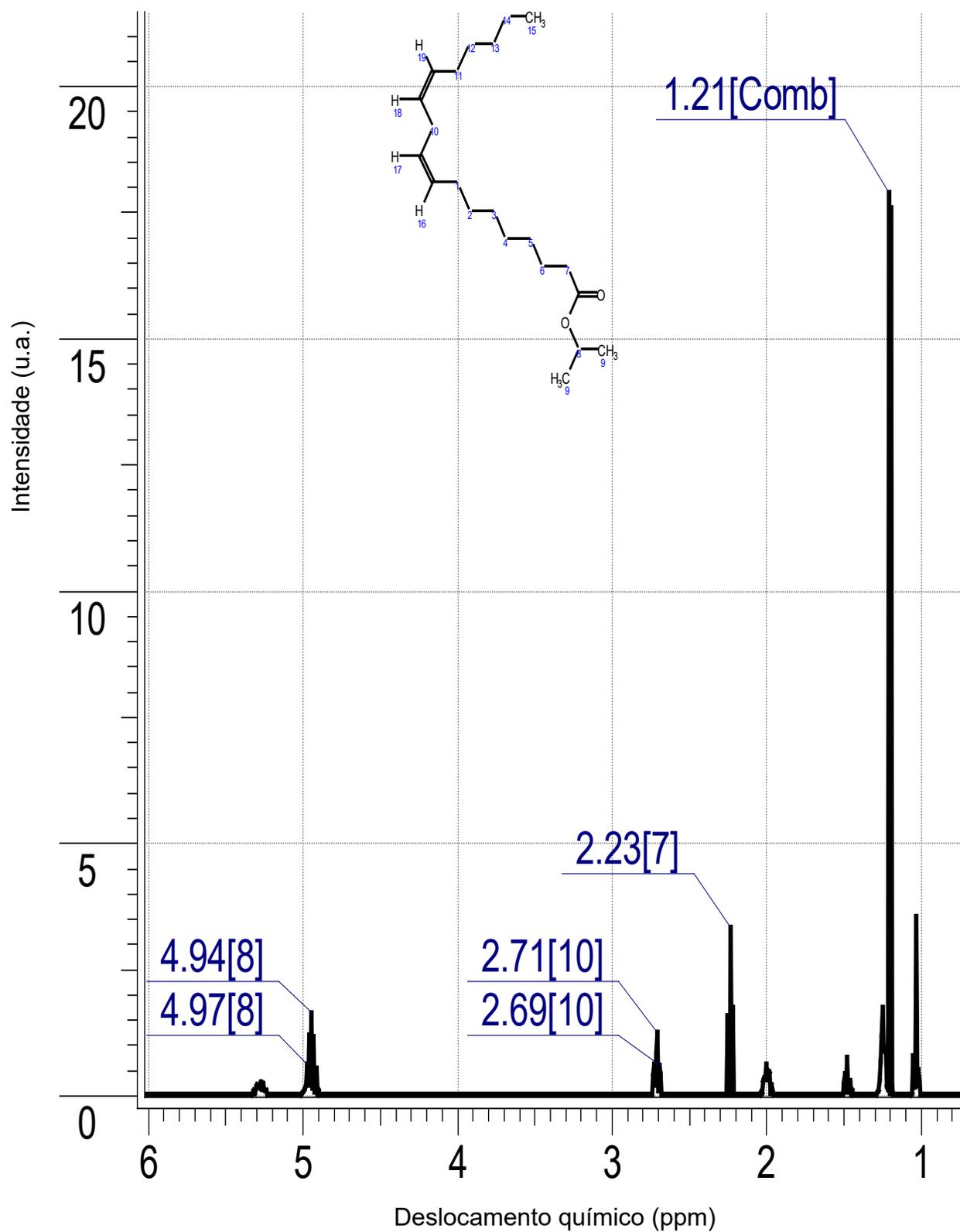


Figura 35. Espectro de RMN- $^1\text{H}$  do linoleato de isopropila.

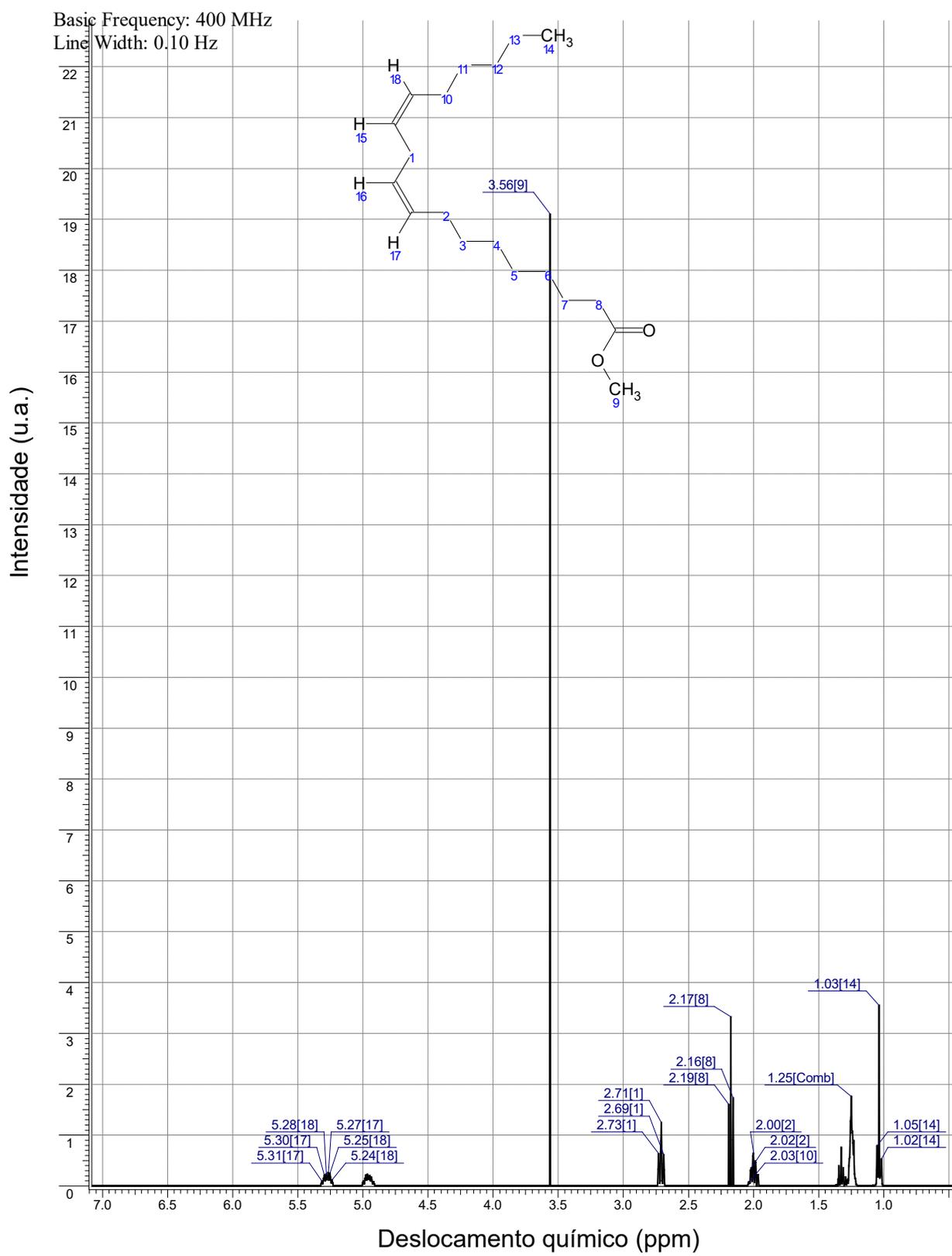


Figura 36. Espectro de RMN-<sup>1</sup>H do linoleato de metila.

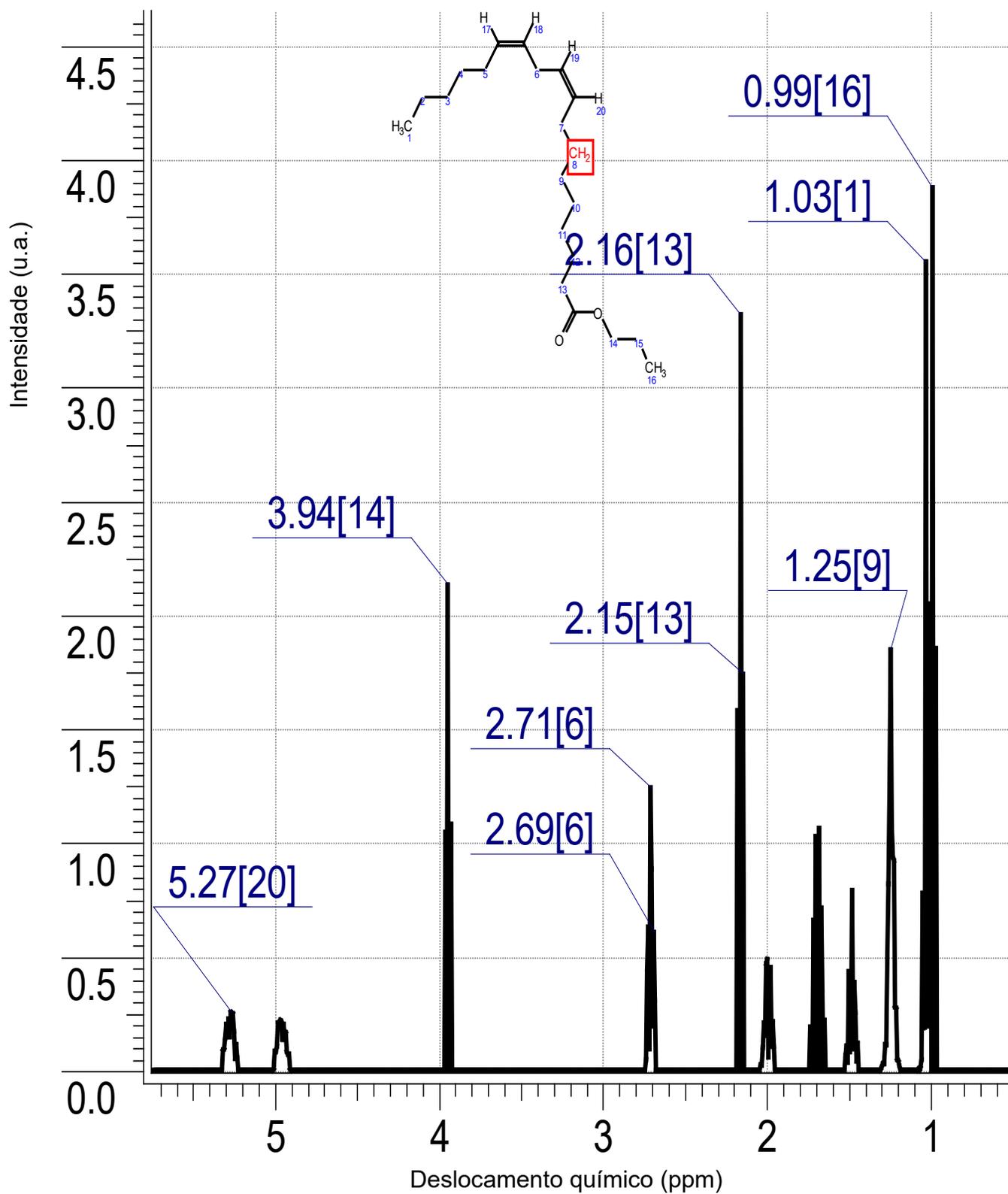


Figura 37. Espectro de RMN-<sup>1</sup>H do linoleato de propila.

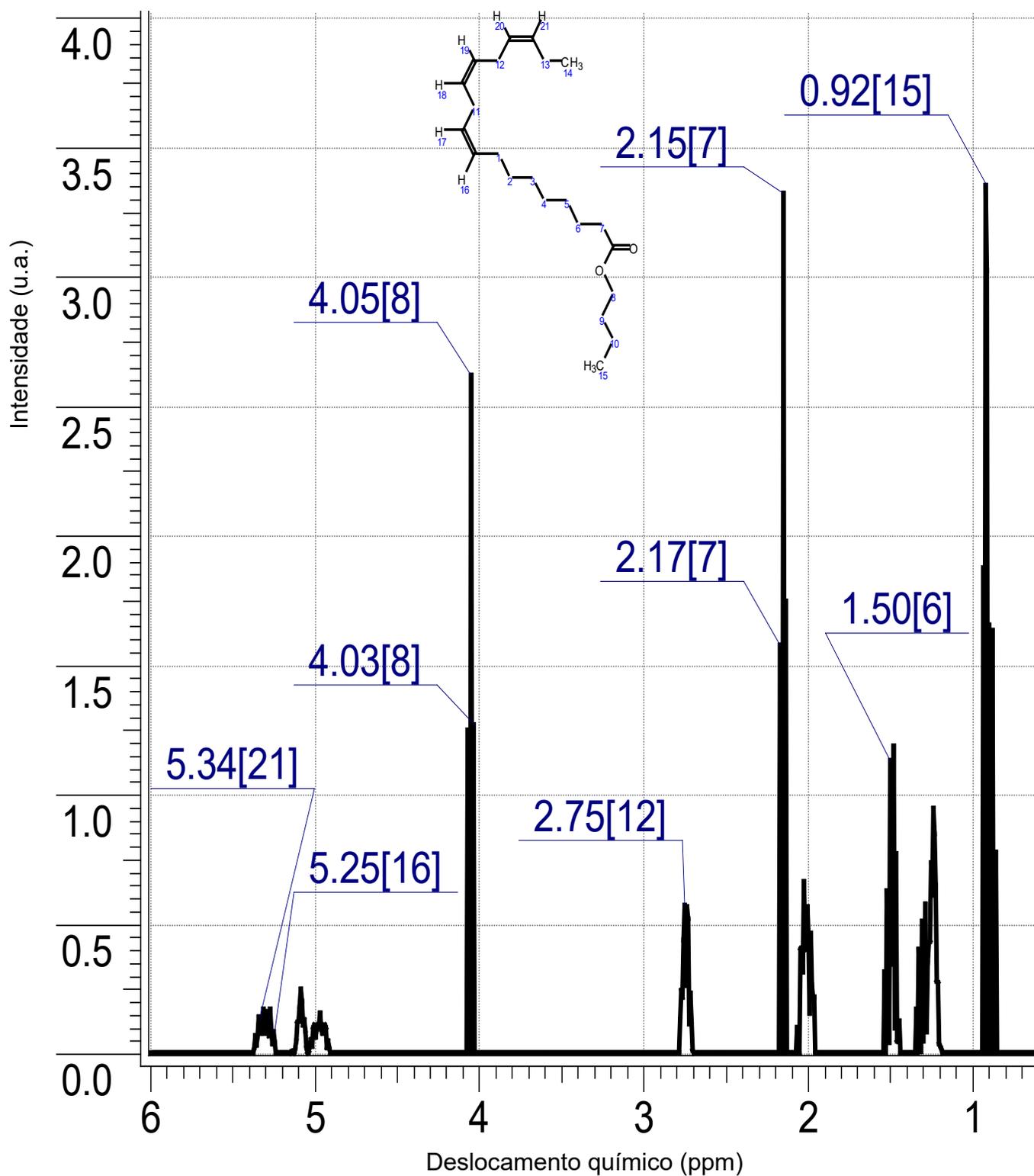


Figura 38. Espectro de RMN-<sup>1</sup>H do linolenato de butila.

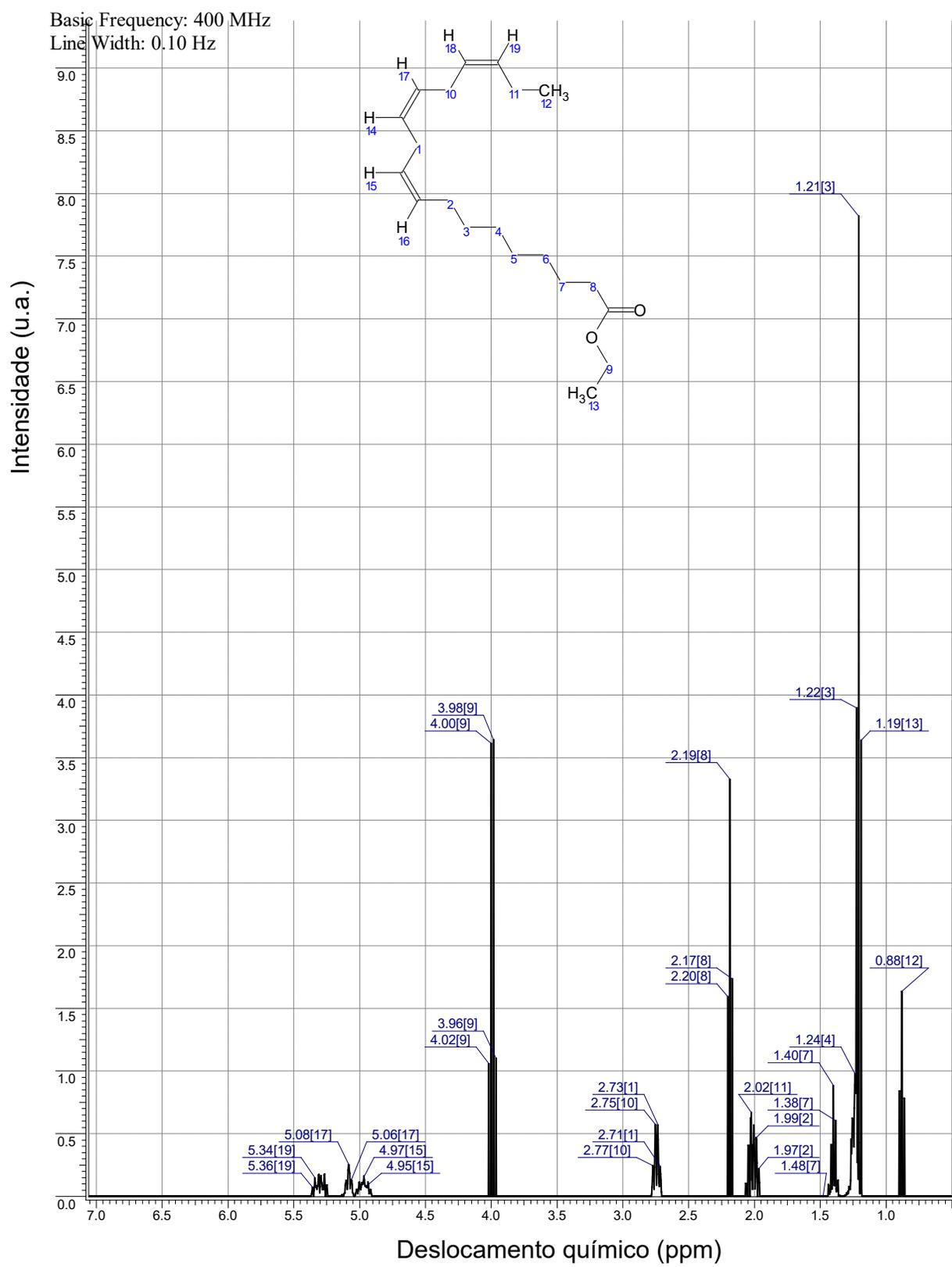


Figura 39. Espectro de RMN-<sup>1</sup>H do linolenato de etila.

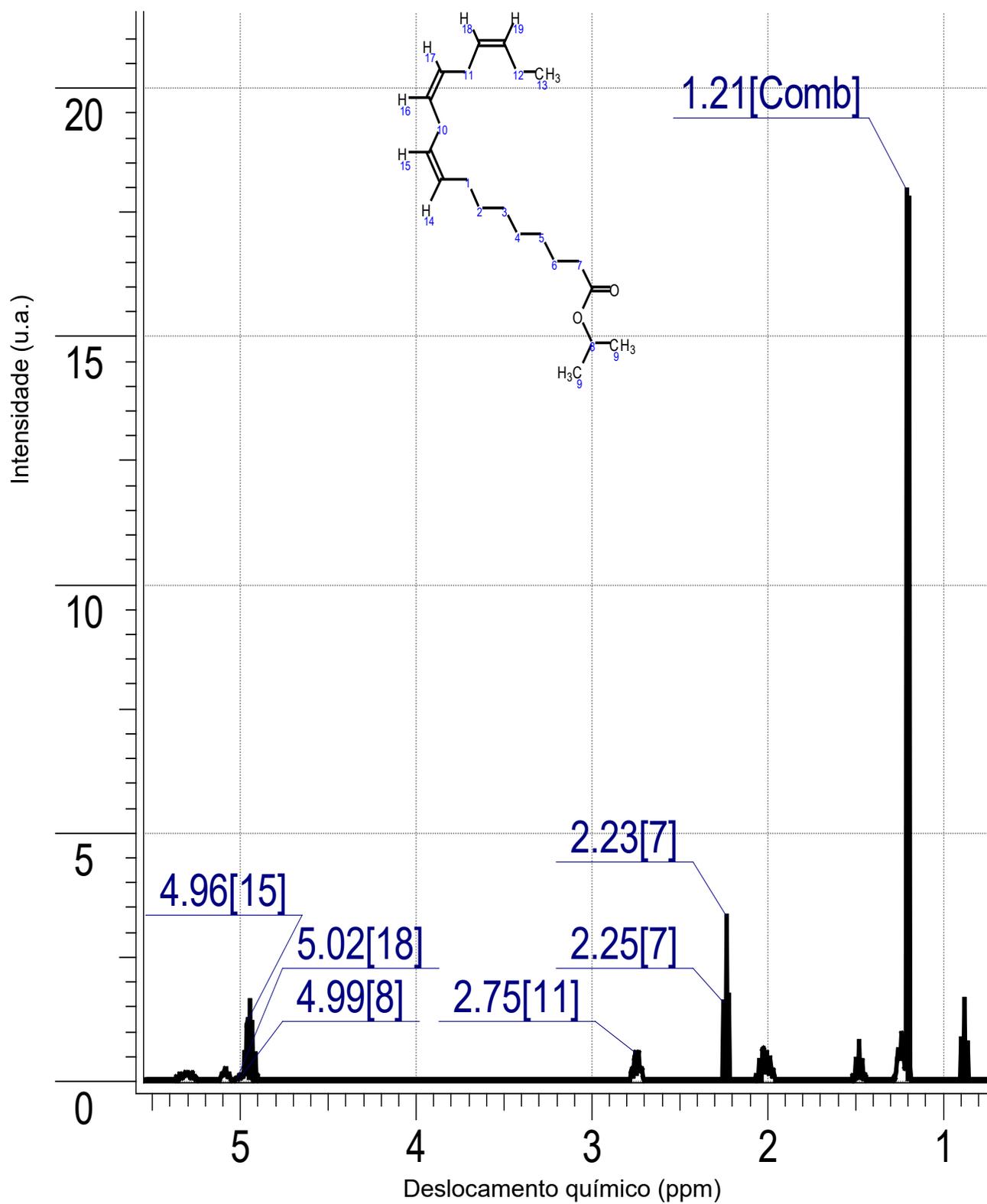


Figura 40. Espectro de RMN- $^1\text{H}$  do linolenato de isopropila.

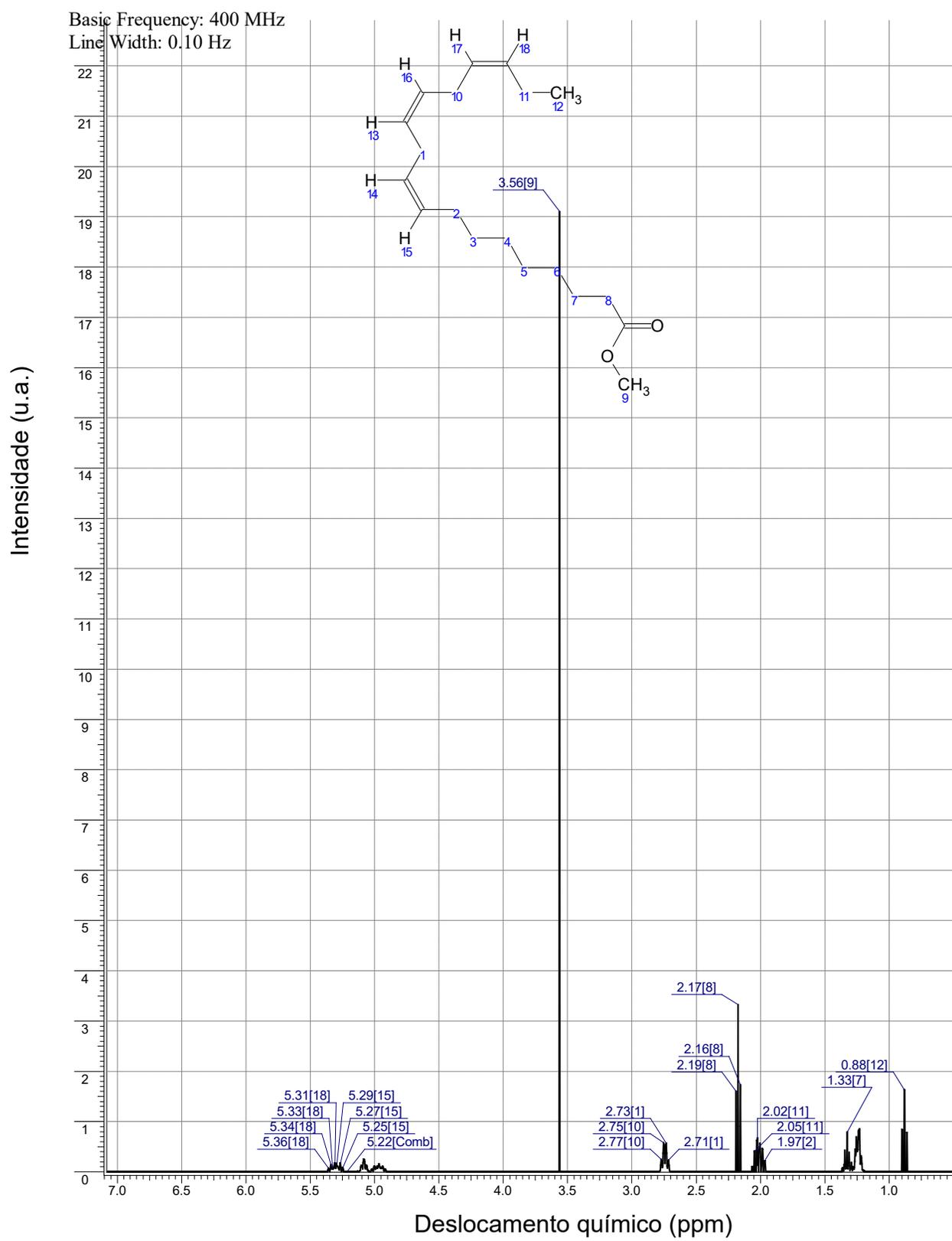


Figura 41. Espectro de RMN-<sup>1</sup>H do linolenato de metila.

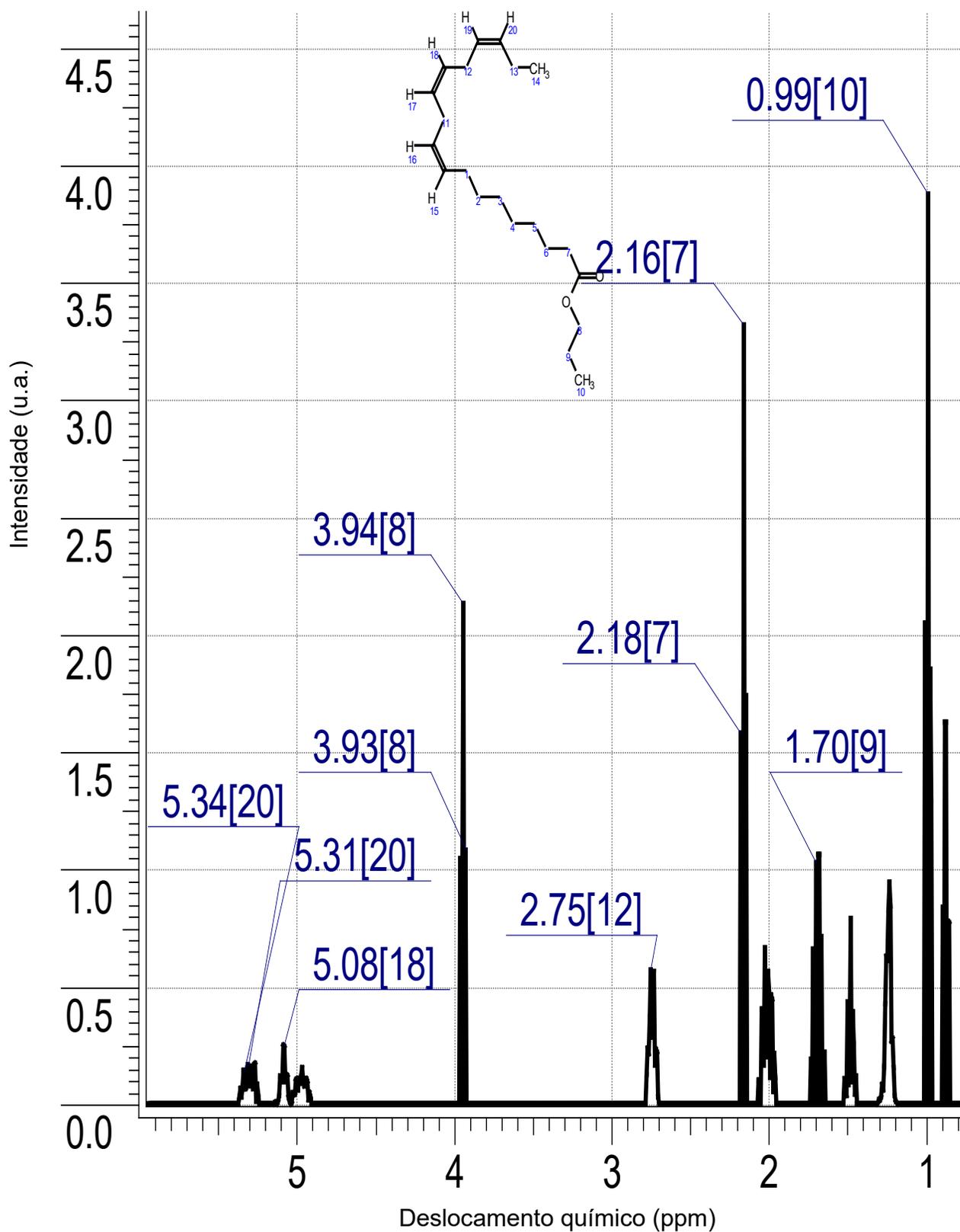


Figura 42. Espectro de RMN- $^1\text{H}$  do linolenato de propila.

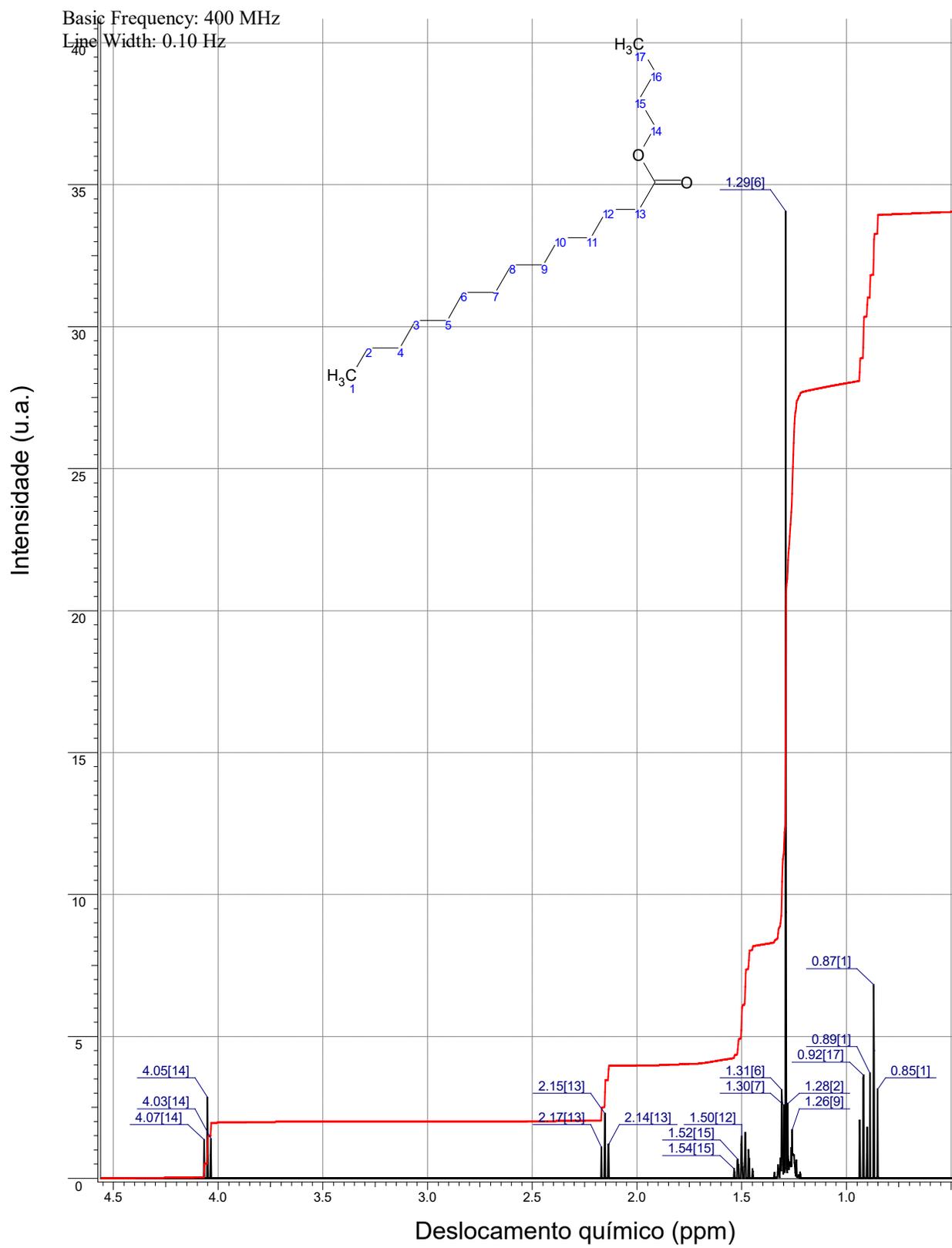


Figura 43. Espectro de RMN-<sup>1</sup>H do miristato de butila.

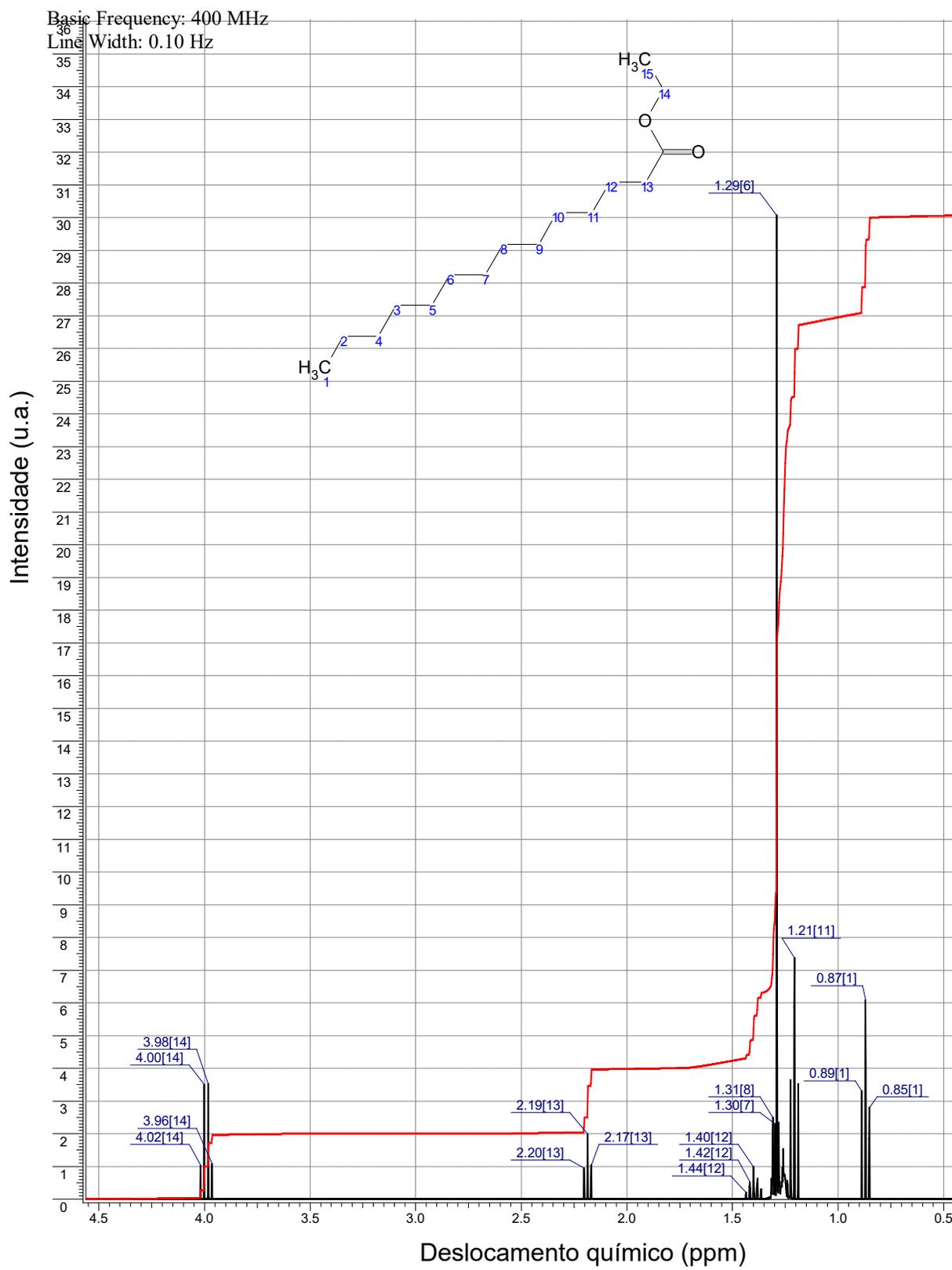


Figura 44. Espectro de RMN-<sup>1</sup>H do miristato de etila.

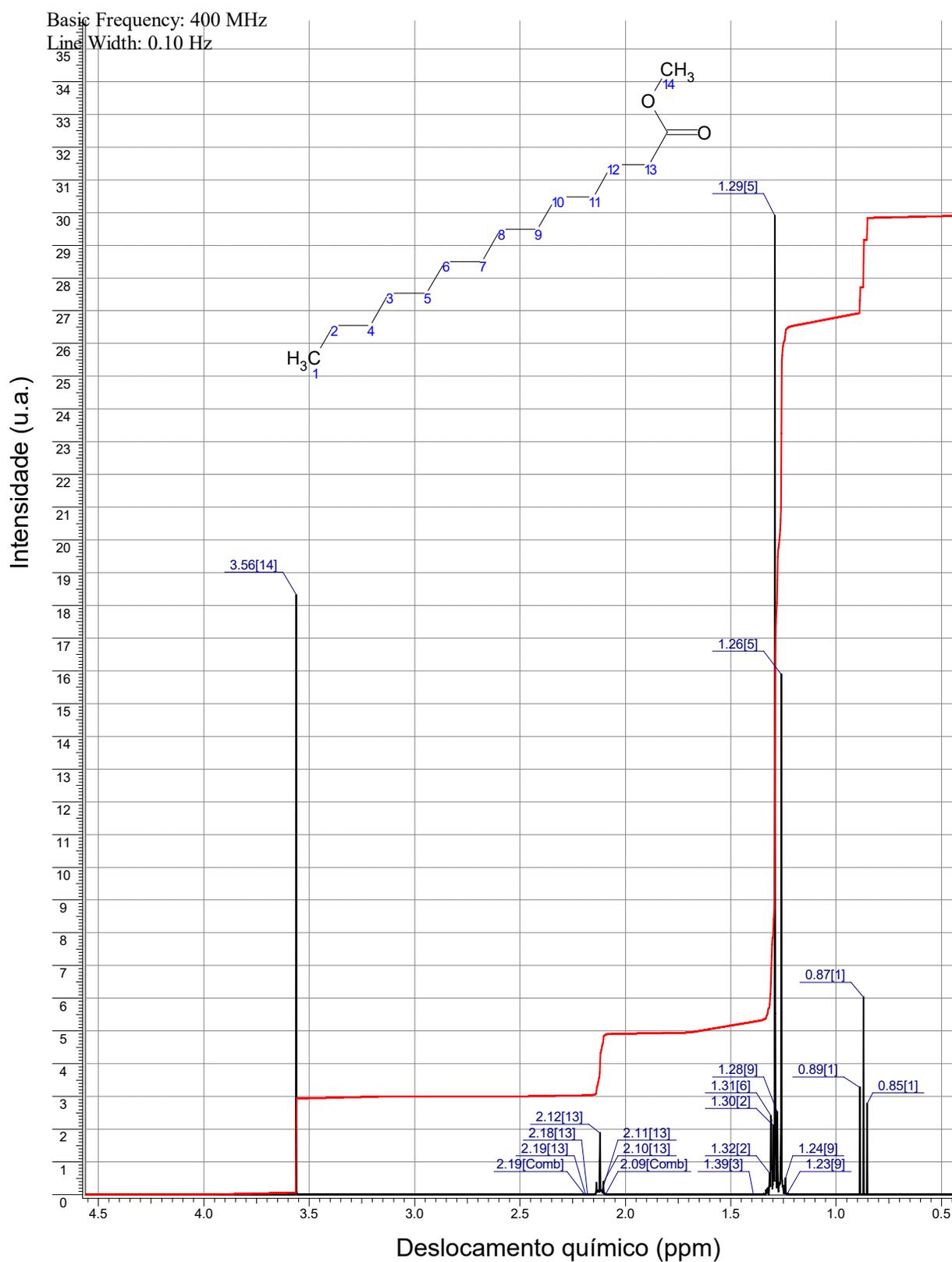


Figura 45. Espectro de RMN- $^1H$  do miristato de metila.

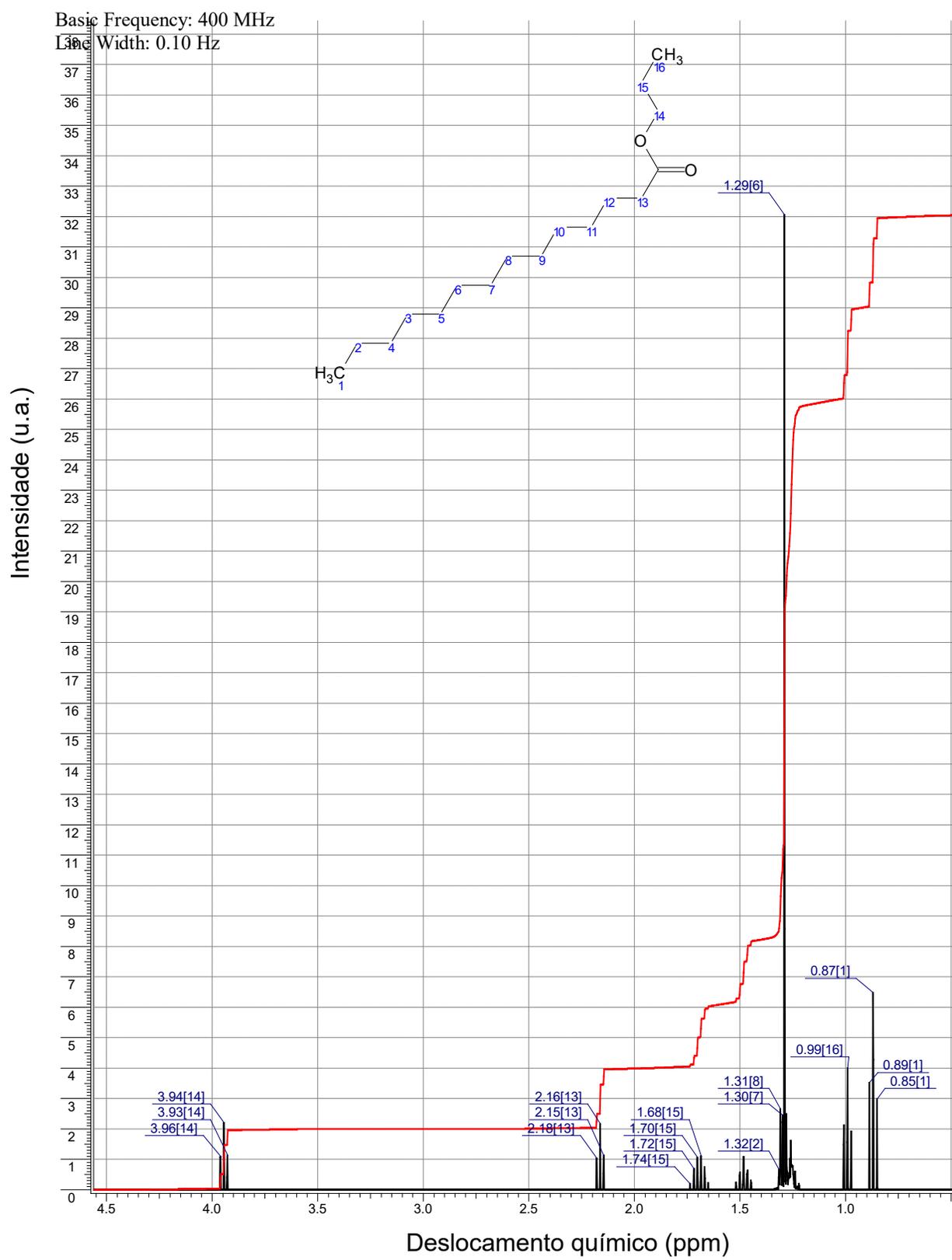


Figura 46. Espectro de RMN- $^1\text{H}$  do miristato de propila.

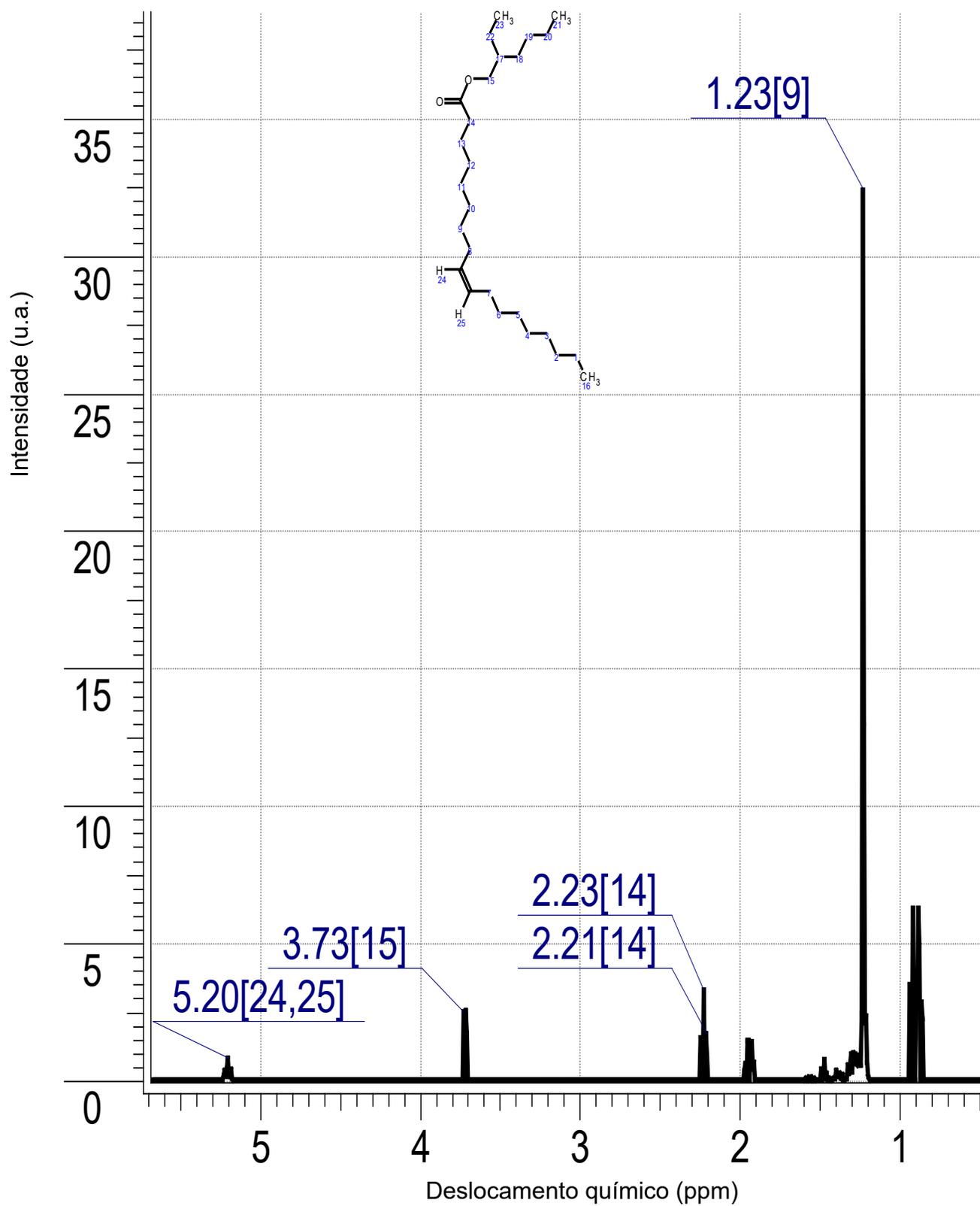


Figura 47. Espectro de RMN- $^1\text{H}$  do oleato de 2-etilexila.

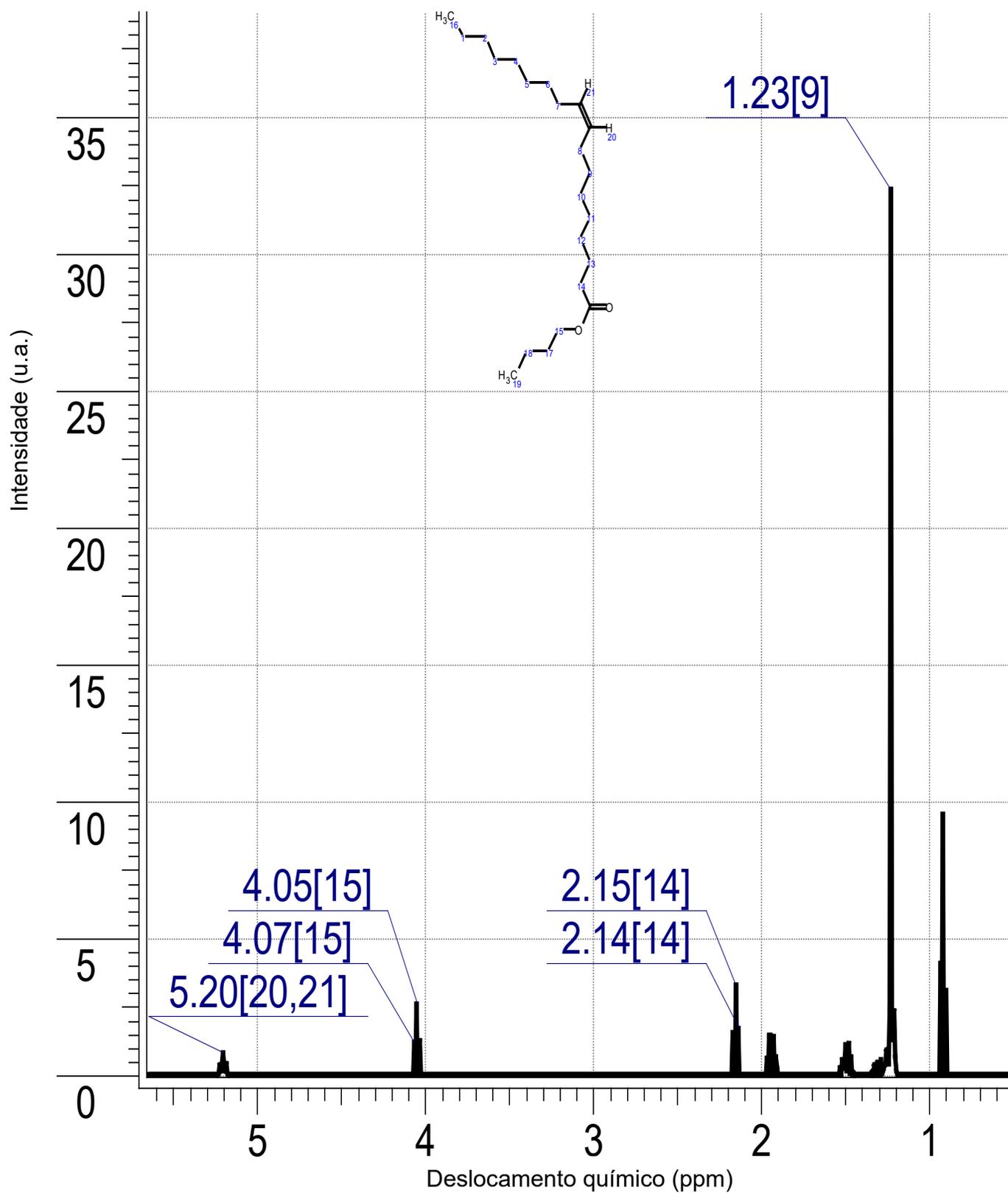


Figura 48. Espectro de RMN- $^1\text{H}$  do oleato de butila.

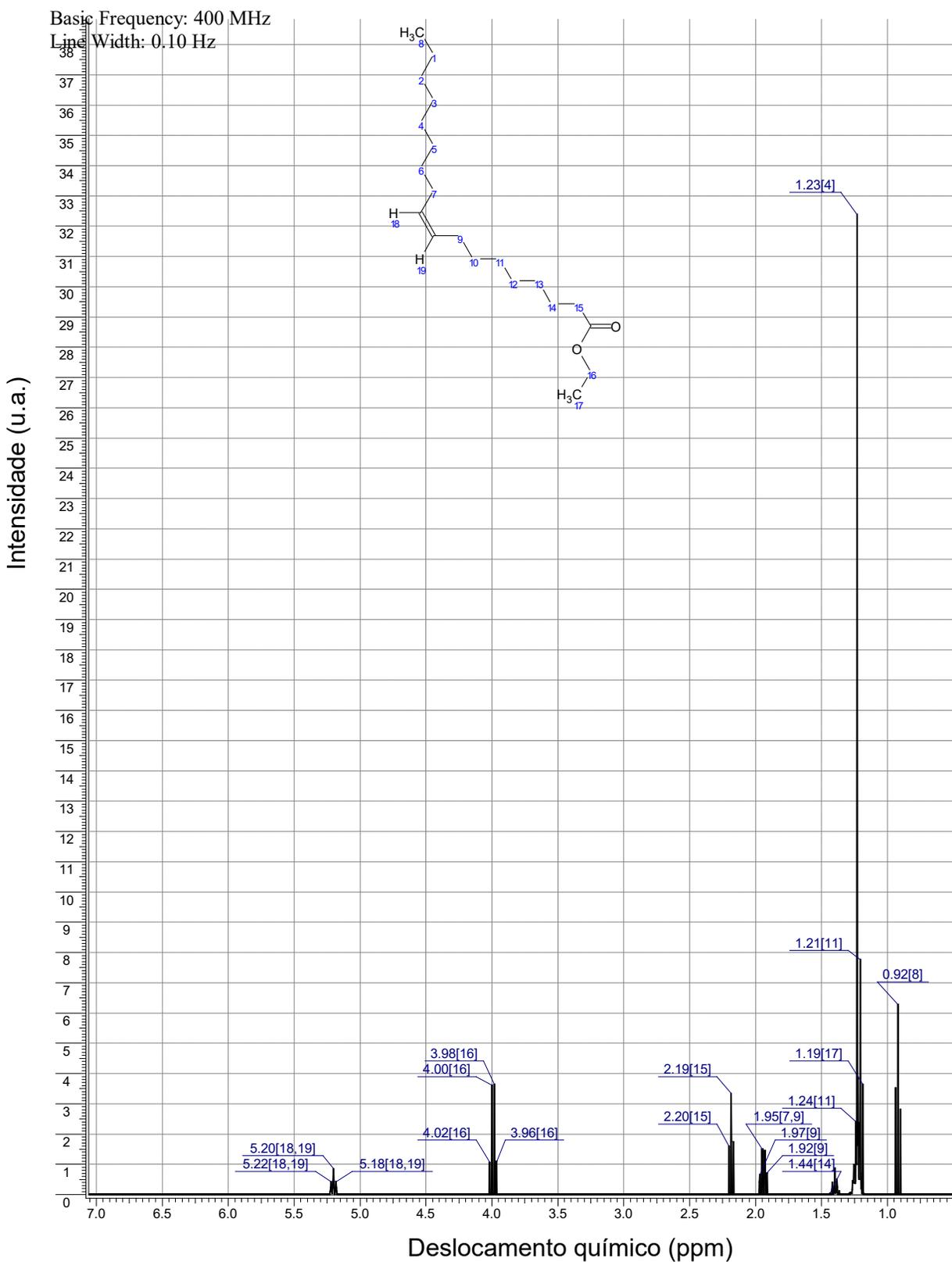


Figura 49. Espectro de RMN-<sup>1</sup>H do oleato de etila.

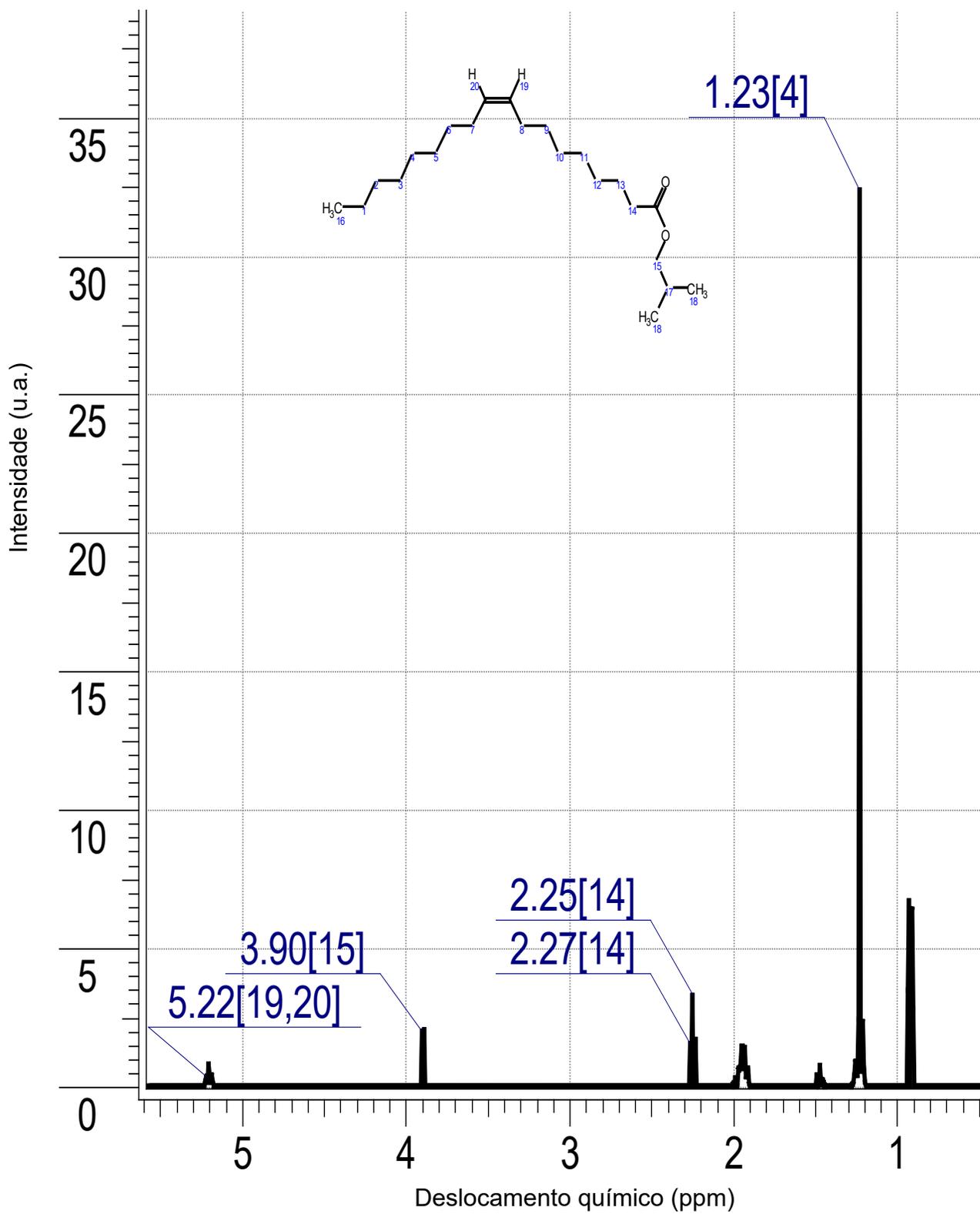


Figura 50. Espectro de RMN- $^1\text{H}$  do oleato de isobutila.

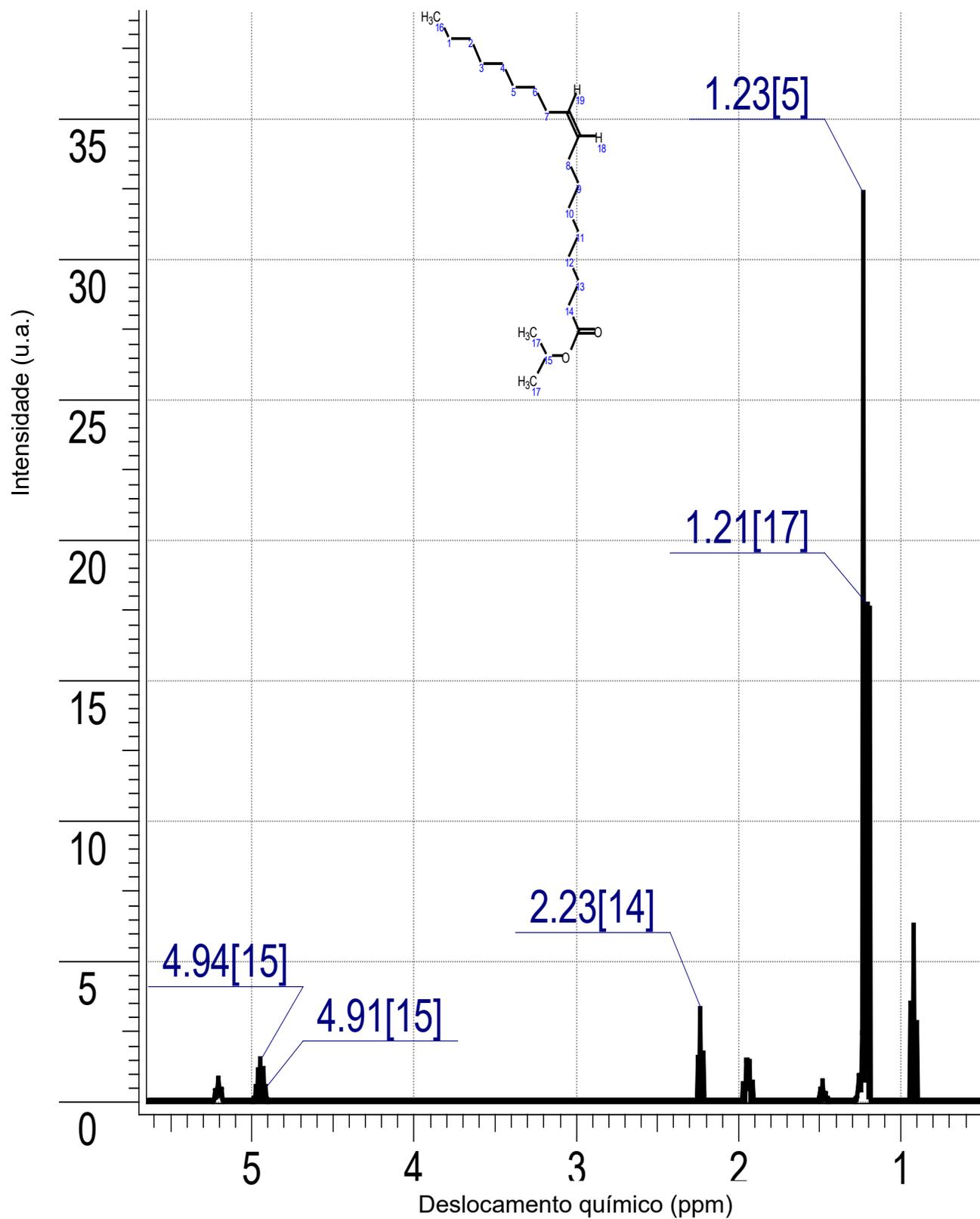


Figura 51. Espectro de RMN- $^1\text{H}$  do oleato de isopropila.

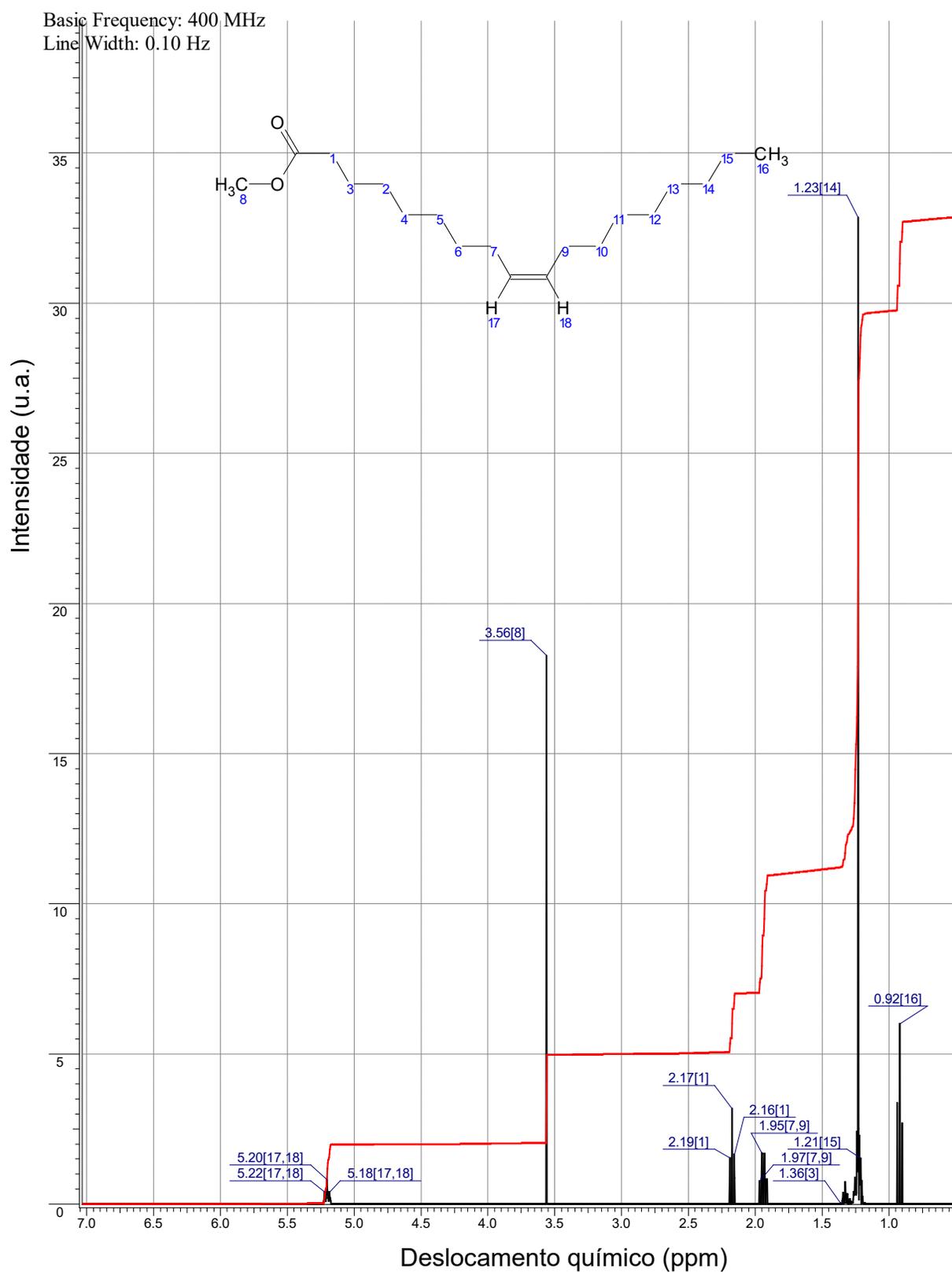


Figura 52. Espectro de RMN-<sup>1</sup>H do oleato de metila.

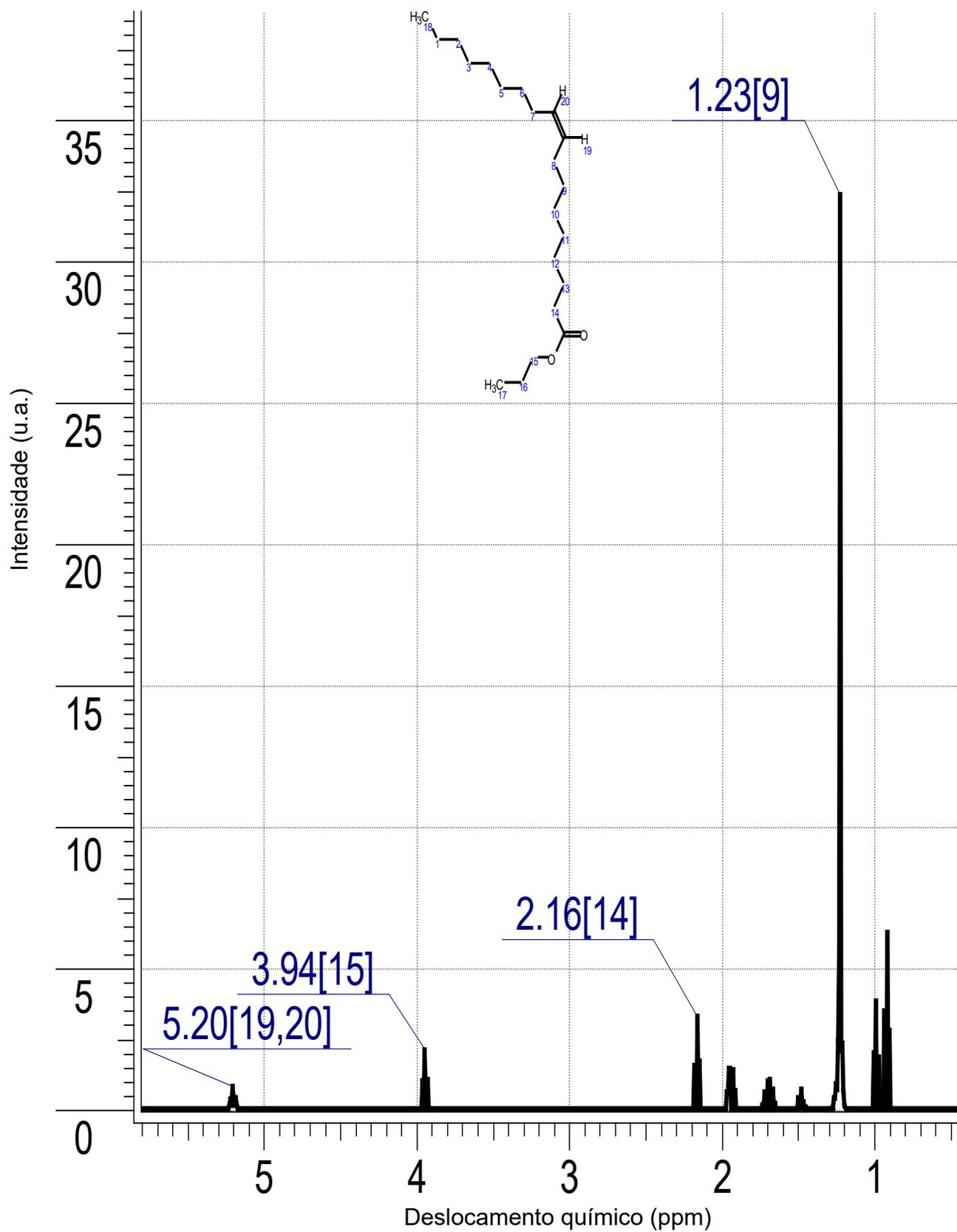


Figura 53. Espectro de RMN- $^1\text{H}$  do oleato de propila.

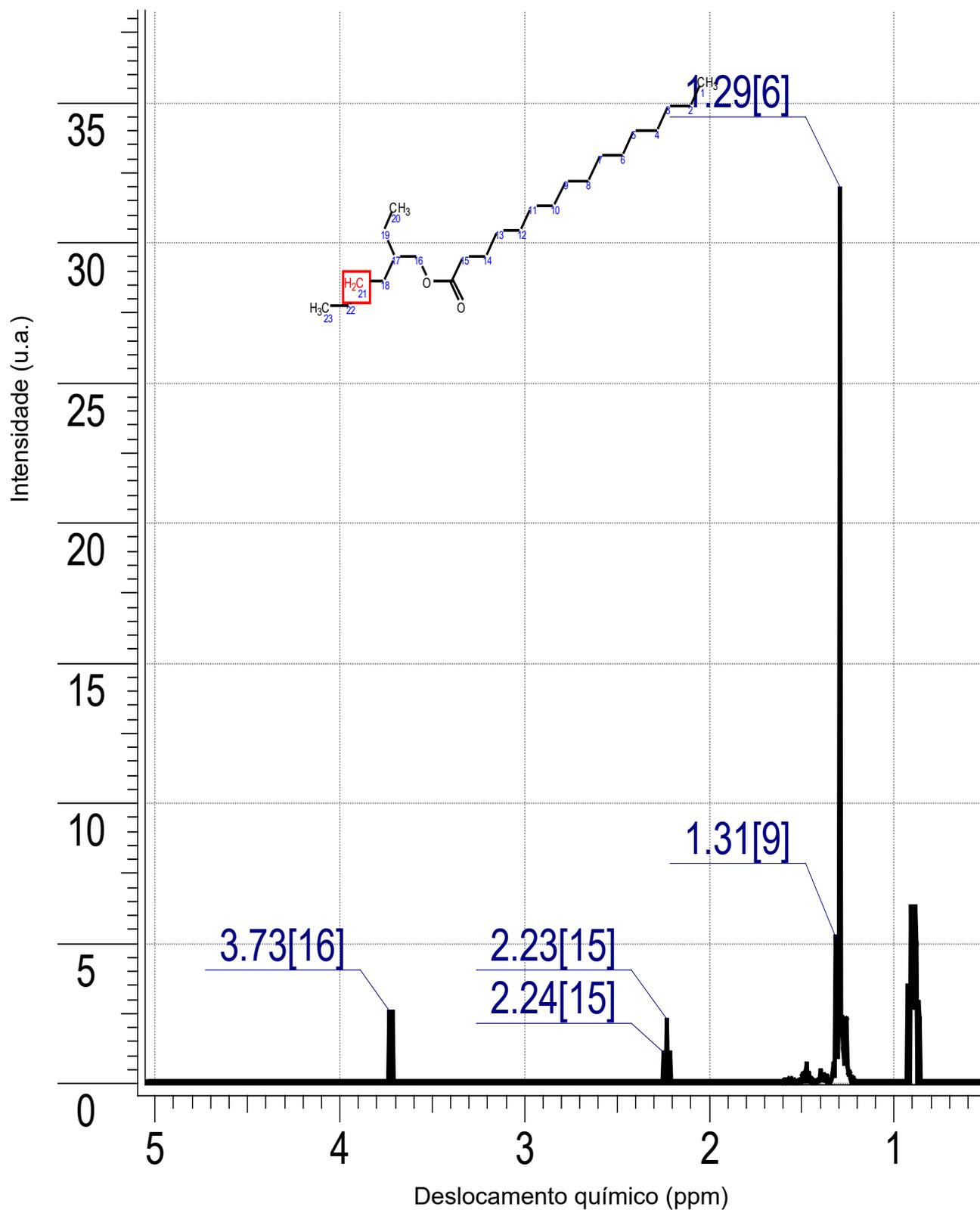


Figura 54. Espectro de RMN-<sup>1</sup>H do palmitato de 2-etilexila.

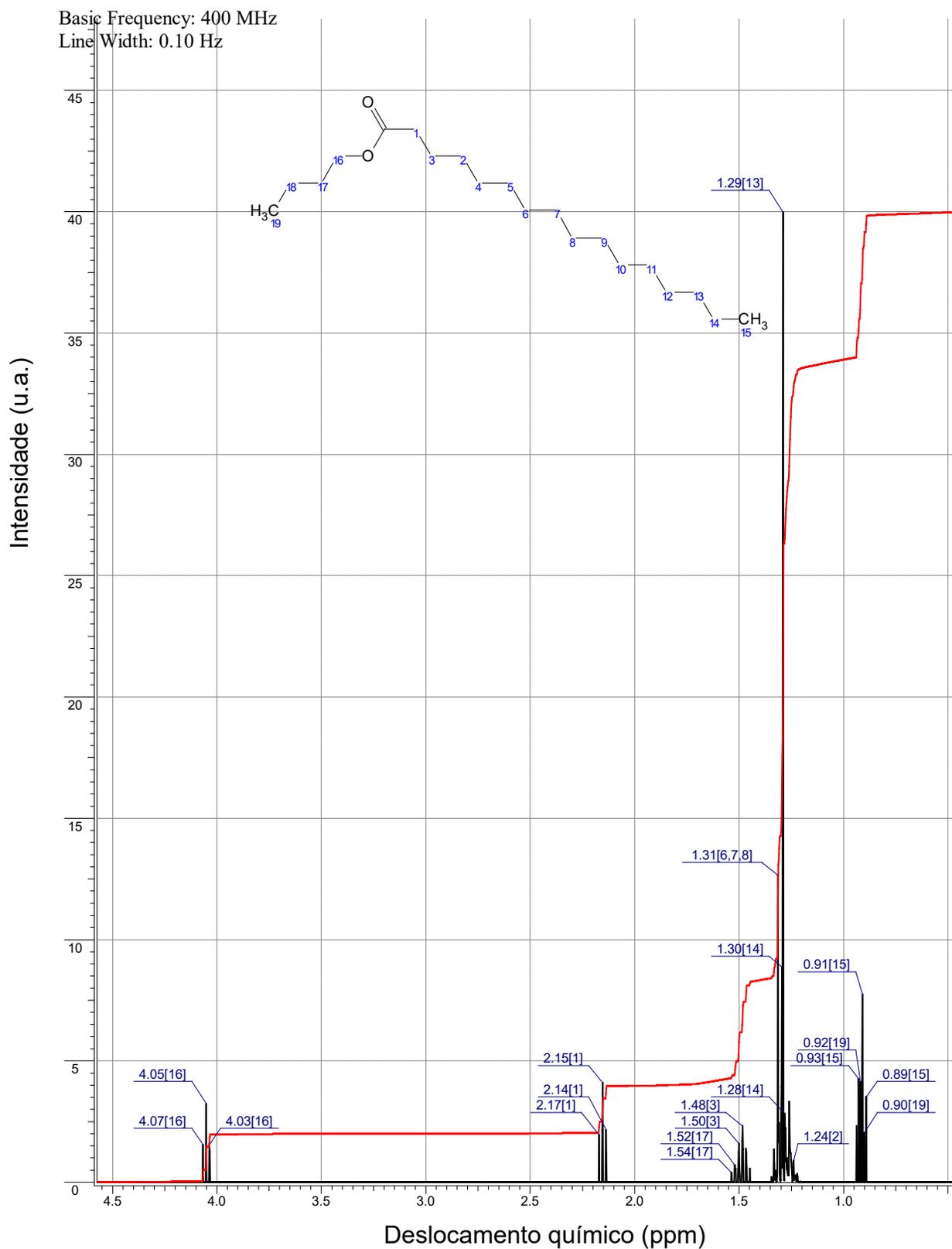


Figura 55. Espectro de RMN-<sup>1</sup>H do palmitato de butila.

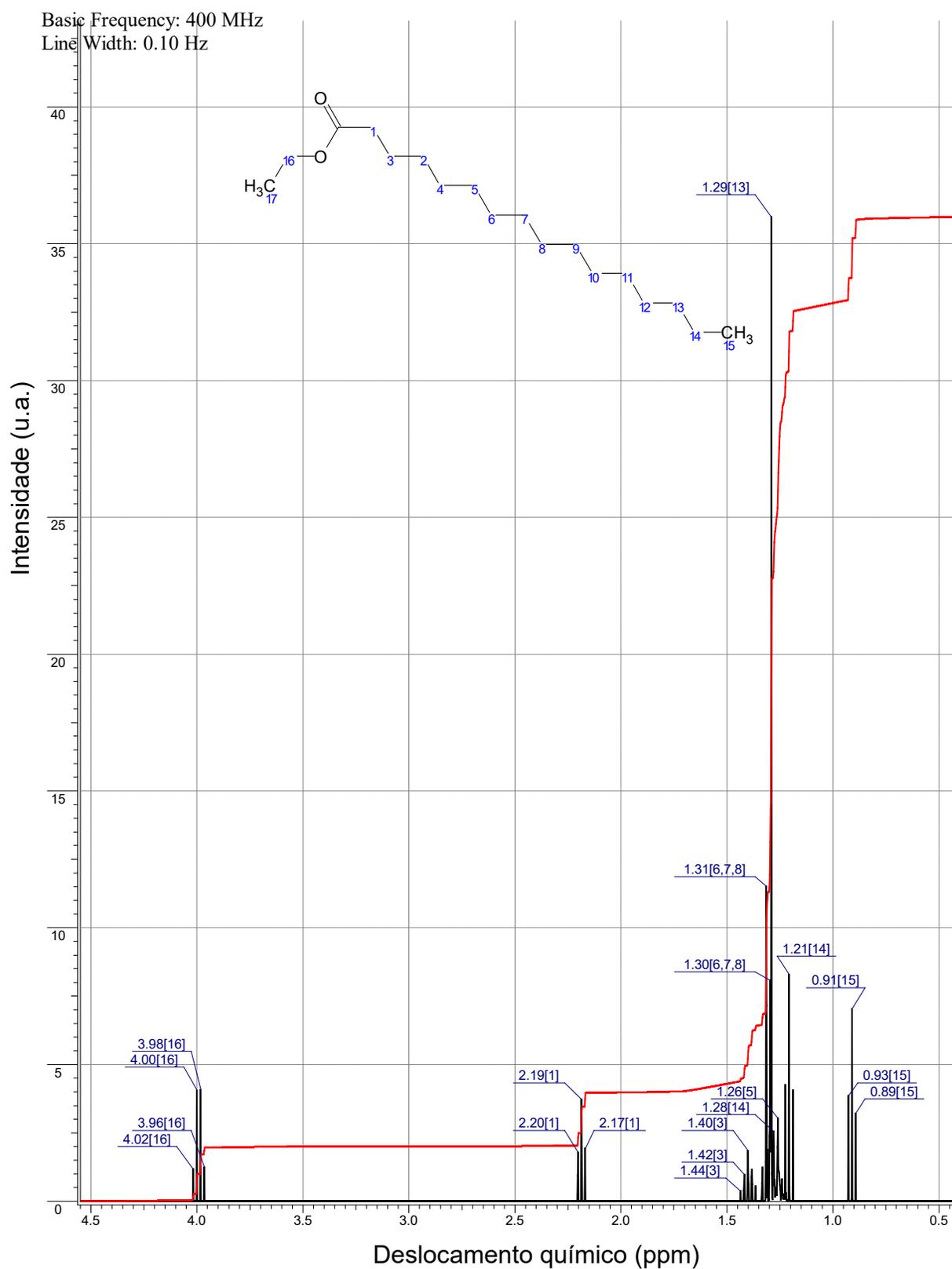


Figura 56. Espectro de RMN-<sup>1</sup>H do palmitato de etila.

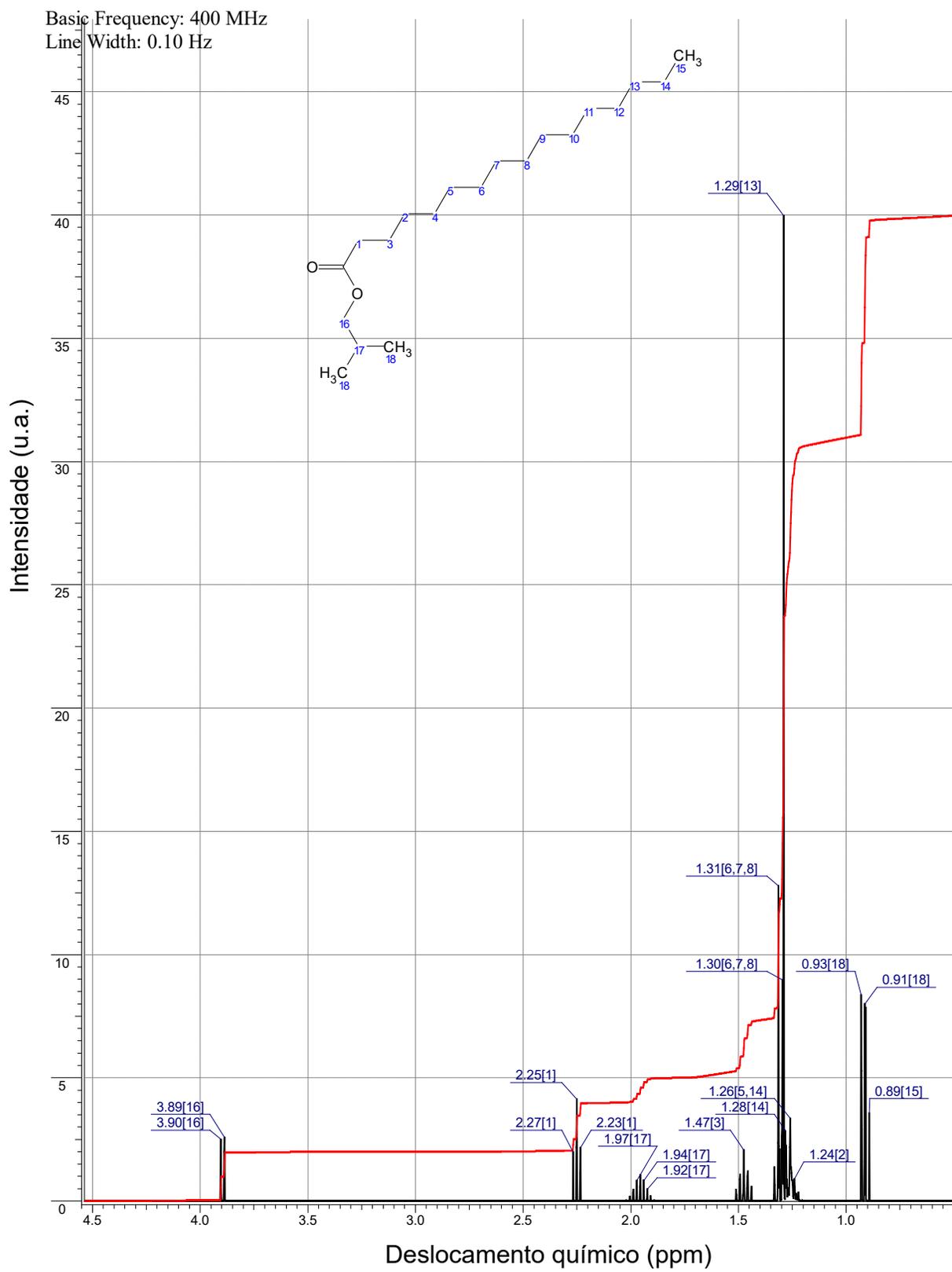


Figura 57. Espectro de RMN-<sup>1</sup>H do palmitato de isobutila.

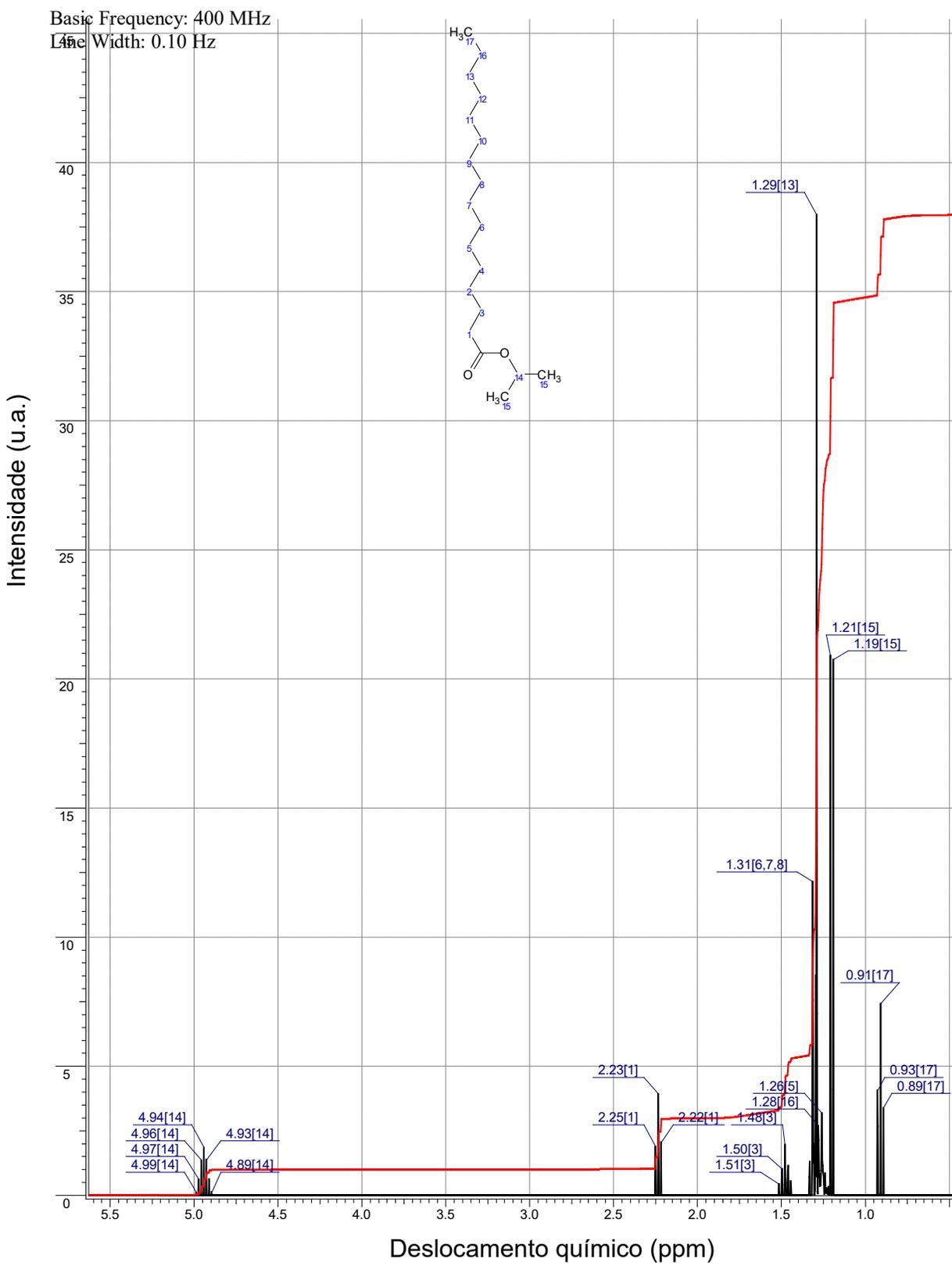


Figura 58. Espectro de RMN-<sup>1</sup>H do palmitato de isopropila.

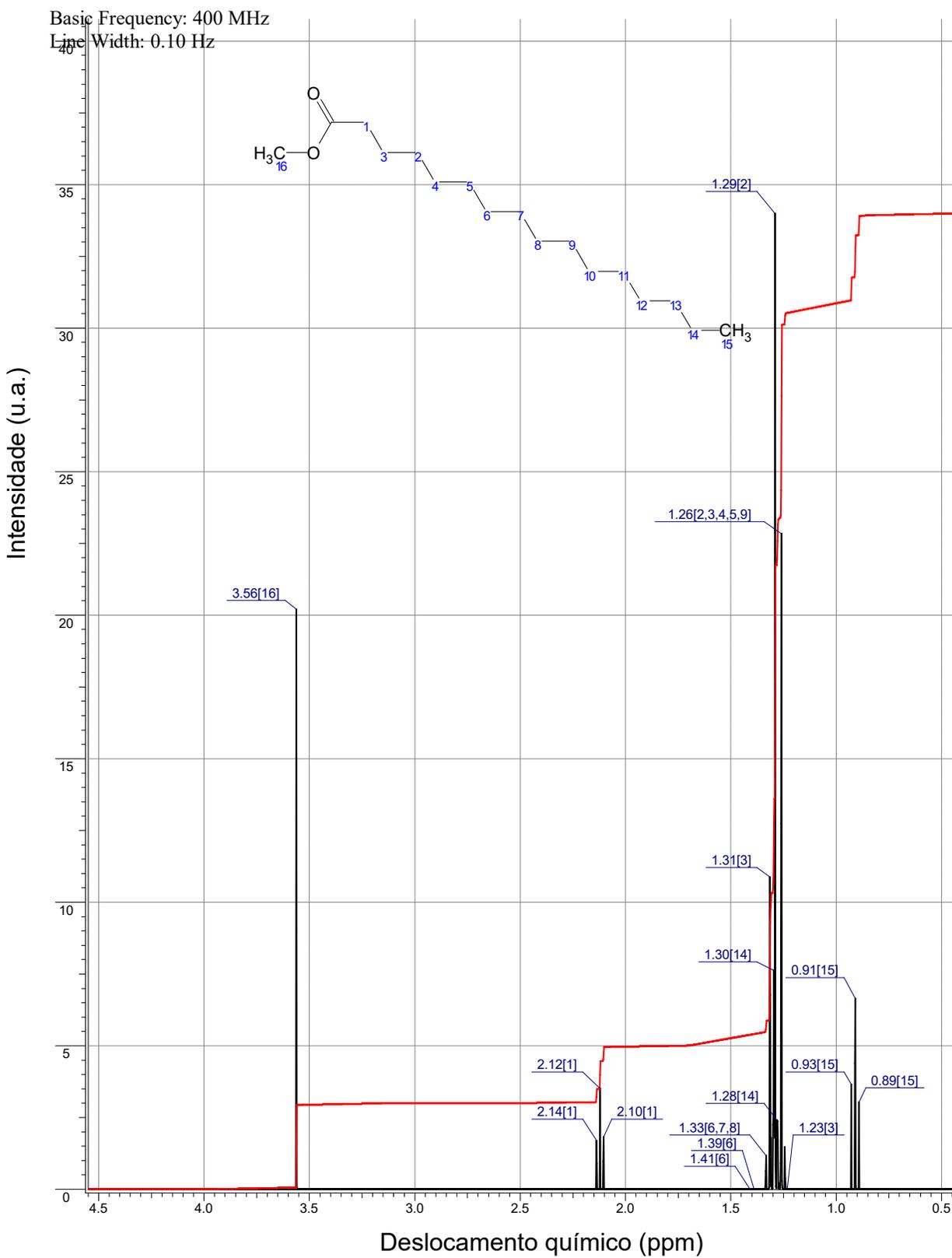


Figura 59. Espectro de RMN-<sup>1</sup>H do palmitato de metila.

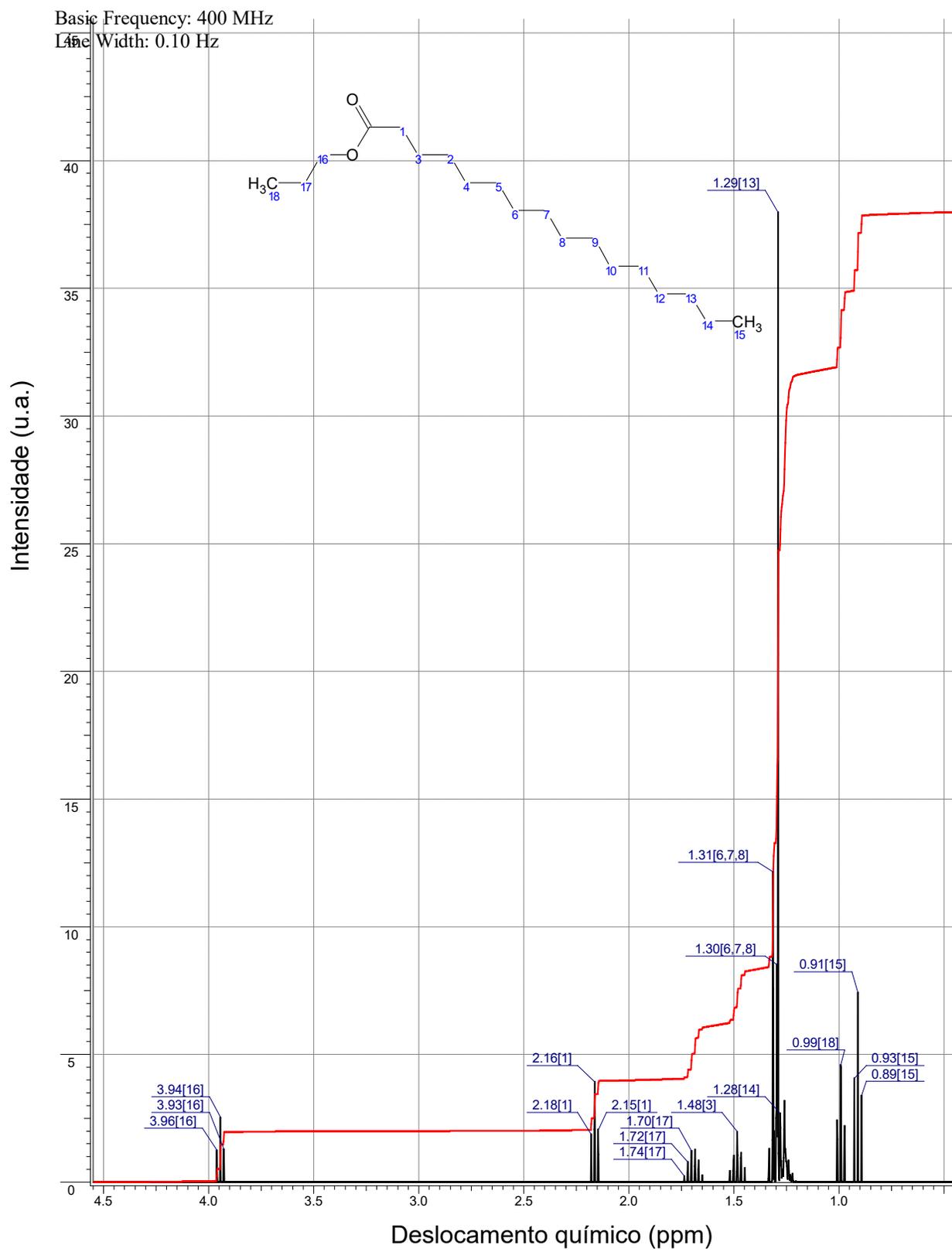


Figura 60. Espectro de RMN- $^1\text{H}$  do palmitato de propila.

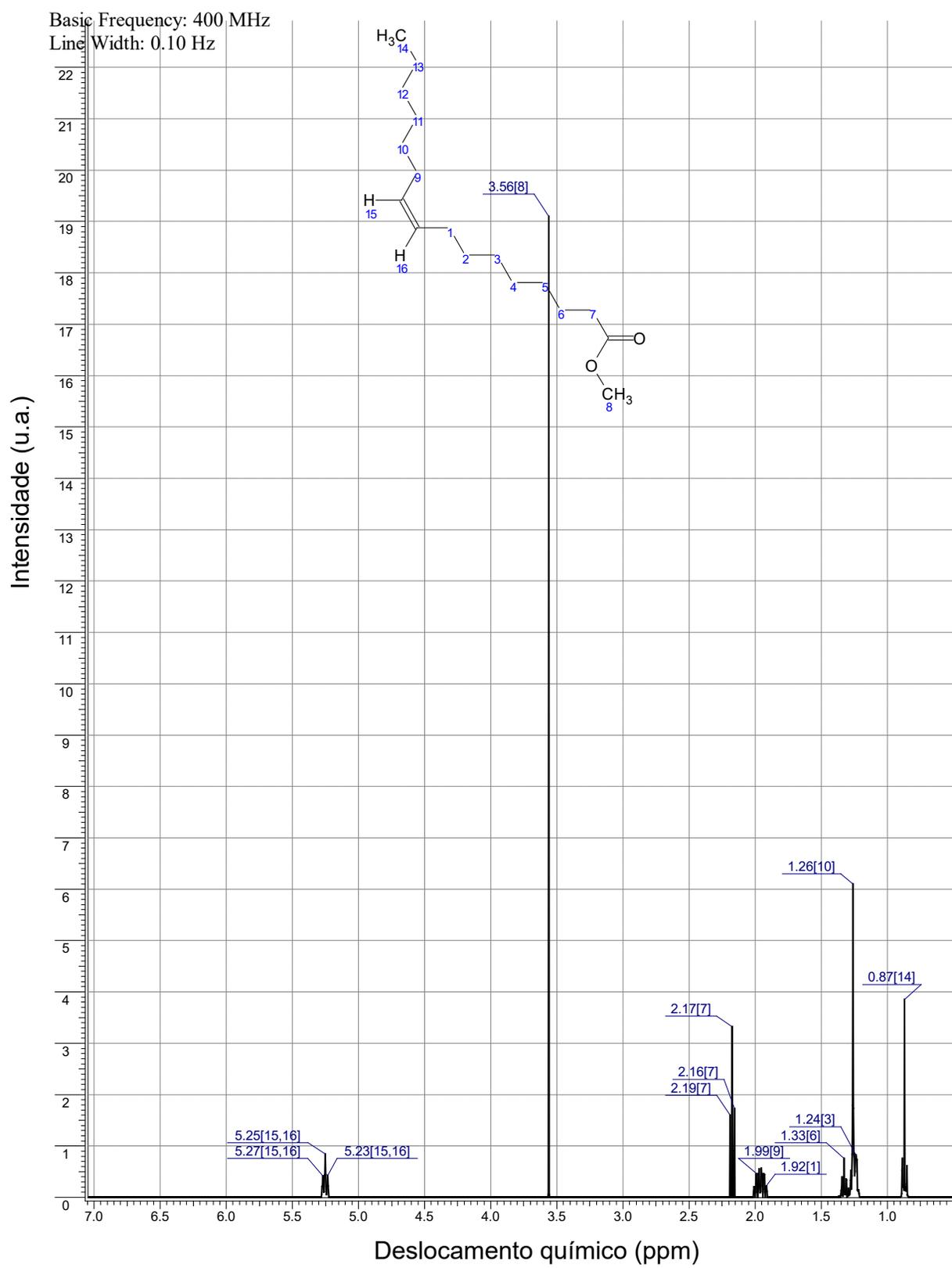


Figura 61. Espectro de RMN-<sup>1</sup>H do palmitoleato de metila.