

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE FARMÁCIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS
FARMACÊUTICAS

Ferramentas Quimiométricas Aplicadas no Desenvolvimento de Metodologia Analítica para Avaliação de Adulteração em Leite Bovino por Adição de Soro do Queijo por ATR-FTIR

LAYANE LENARDON

Porto Alegre, 2016

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE FARMÁCIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS
FARMACÊUTICAS

Ferramentas Quimiométricas Aplicadas no Desenvolvimento de Metodologia Analítica para Avaliação de Adulteração em Leite Bovino por Adição de Soro do Queijo por ATR-FTIR

Dissertação apresentada por **Layane Lenardon**
para a obtenção do grau de MESTRE em Ciências
Farmacêuticas

Orientadora: Prof^a. Dra. Ana Maria Bergold

Coorientador: Prof. Dr. Marco Flôres Ferrão

Porto Alegre, 2016

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Farmacêuticas, em nível de Mestrado Acadêmico da Faculdade de Farmácia da Universidade Federal do Rio Grande do Sul e aprovada em 23/03/2015 pela Banca Examinadora constituída por:

Prof. Dr. Adilson Ben da Costa

Universidade de Santa Cruz do Sul

Dr. Fabiano Barreto

Ministério da Agricultura, Pecuária e Abastecimento LANAGRO/RS

Prof. Dr. Martin Steppe

Universidade Federal do Rio Grande do Sul

CIP - Catalogação na Publicação

Lenardon, Layane
Ferramentas Quimiométricas Aplicadas no
Desenvolvimento de Metodologia Analítica para
Avaliação de Adulteração em Leite Bovino por Adição de
Soro do Queijo por ATR-FTIR / Layane Lenardon. --
2016.
109 f.

Orientadora: Ana Maria Bergold.
Coorientador: Marco Flôres Ferrão.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Faculdade de Farmácia, Programa
de Pós-Graduação em Ciências Farmacêuticas, Porto
Alegre, BR-RS, 2016.

1. Controle de qualidade. 2. Quimiometria. I.
Bergold, Ana Maria, orient. II. Ferrão, Marco Flôres,
coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Este trabalho foi desenvolvido em colaboração com o Laboratório de Resíduos de Pesticidas e Medicamentos Veterinários (RPM) do Laboratório Nacional Agropecuário/RS (LANAGRO/RS), localizado em Porto Alegre, com o Instituto de Química (IQ) da Universidade Federal do Rio Grande do Sul e com financiamento do CNPq.

AGRADECIMENTOS

Primeiramente, agradeço à Deus que sempre esteve ao meu lado, atendendo minhas orações e me dando força para superar todos os obstáculos.

Ao Programa de Pós-graduação em Ciências Farmacêuticas (PPGCF) da Universidade Federal do Rio Grande do Sul (UFRGS) pela qualidade e oportunidade do mestrado.

Ao meu noivo, Robson Willians Vinciguerra, pelo amor, carinho, companheirismo, motivação, ensinamentos, paciência, por sempre acreditar no meu potencial e buscar o melhor para mim durante todos esses anos juntos.

A Prof^ª. Dra. Ana Maria Bergold, pela orientação e oportunidade concedida, pelo exemplo de profissionalismo, pelas oportunidades de crescimento profissional e pela paciência e amizade.

Ao Prof. Dr. Marco Flôres Ferrão, pela coorientação deste trabalho, pela paciência e disponibilização do laboratório para o desenvolvimento do mesmo, pelos ensinamentos e dedicação, pela amizade e convívio durante esse trabalho.

Ao Dr. Leonardo Zanchetti Meneghini, pela paciência, dedicação, pelas ideias e oportunidades realizadas e por tudo que foi desenvolvido durante do mestrado.

A Ma. Tanara Magalhães Campos Motta, pelos ensinamentos e disponibilidade durante todas as análises no CL-EM/EM, pela amizade.

Aos colegas do laboratório D-216, em especial, a Camila Correa, Marcelo Marcelo e o Eduardo Lopes pelos ensinamentos quimiométricos, pela paciência e dedicação prestadas.

A toda equipe do LANAGRO/RS, setor RPM, pela doação de materiais, pelos equipamentos e todo suporte técnico prestado.

Enfim, agradeço a todas as pessoas que contribuíram de alguma forma para a realização deste trabalho.

“Tudo o que um sonho precisa para ser realizado é alguém que acredite que ele possa ser realizado”

Roberto Shinyashiki

RESUMO

No Brasil, o monitoramento da qualidade do leite é realizado pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) através da Instrução Normativa nº 68 (IN 68). A qualidade do leite consumido é uma constante preocupação devido às fraudes descobertas nos últimos anos. Uma das fraudes mais comuns é a adulteração de leite por adição de soro de queijo; esta adição é proibida pela legislação brasileira para o leite destinado diretamente ao consumo. Essa fraude pode ser detectada através da determinação do índice de caseinomacropéptido (CMP). O CMP é um peptídeo específico do soro do queijo; deste modo é um marcador da adulteração. Neste trabalho foi desenvolvido um método analítico capaz de quantificar o CMP no leite empregando a Espectroscopia no Infravermelho Médio com Transformada de Fourier e Reflexão Total Atenuada (ATR-FTIR) e Quimiometria. A análise exploratória dos dados foi realizada através de Análise por Componentes Principais (PCA) e Análise de Agrupamentos Hierárquicos (HCA), os quais indicaram uma semelhança entre as amostras de leite cru (LC) e leite semidesnatado (LS) contaminadas com CMP, em virtude da presença de gorduras nesses leites. Na PCA foi possível observar uma tendência de separação das amostras com distintas adições de CMP. Na análise dos modelos de regressão foram utilizados os algoritmos de Mínimos Quadrados Parciais (PLS), Mínimos Quadrados Parciais por Sinergismo de Intervalo (SIPLS) e Máquinas de Vetores de Suporte com Mínimos Quadrados (LS-SVM) para selecionar o modelo adequado para quantificação do CMP em amostra de leite. Desse modo o modelo de escolha foi o LS-SVM. Assim, o método proposto mostrou ser rápido, simples e econômico para detectar amostras de leite adulteradas com CMP, cujos resultados foram comparados com o método de referência, com boa correlação ($R^2 = 0,9984$).

Palavras-chaves: leite, adulteração, CMP, soro do queijo, ATR-FTIR, regressão multivariada.

ABSTRACT

In Brazil, the milk quality is conducted by the Ministry of Agriculture, Livestock and Supply (MAPA) through Normative Instruction No. 68 (IN 68) and it is a constant concern due to the discovered frauds in recent years. A common fraud that occurs in Brazilian milk is the milk whey addition; which is prohibited by Brazilian law for milk of direct consumption. This fraud can be detected by determining the Caseinmacropeptide index (CMP). Since the CMP is a specific peptide of whey, it can be used as marker of milk adulteration. In this work, an analytical method capable of quantifying the CMP in milk, using the Fourier Transform Infrared Spectroscopy (FTIR) and chemometrics, was developed. Exploratory data analysis was performed using Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA). These methods indicated a similarity between the samples of raw milk and semi-skimmed milk contaminated with CMP. This similarity may be due to the presence of fat in these milks. In the scores of the PCA, it was possible to observe a trend of separation of samples with different CMP concentrations. It was used Partial Least Squares algorithms (PLS), Partial Least Squares with Interval Synergism (SIPLS) and Support Vector Machines with Least Squares (LS-SVM) algorithms as regression algorithms. The model of choice was the s4i16M with LS-SVM. The proposed method proved to be fast, simple and economical to detect milk samples adulterated with CMP. These results corroborated with the reference method ($R^2 = 0,9984$).

Keywords: milk, adulteration, CMP, whey, FTIR, multivariate regression.

LISTA DE FIGURAS

- Figura 1:** Esquema de processo enzimático na produção de queijos: clivagem da κ - caseína pela ação da enzima quimosina, na qual a para- κ -caseína precipita (queijo) e o CMP permanece solúvel no soro (MOTTA *et al.*, 2014). 37
- Figura 2:** Sequência de aminoácidos das variantes genéticas (A e B) do CMP. 39
- Figura 3:** Pontos de clivagem da digestão enzimática do CMP utilizando pepsina. É gerado o peptídeo MAIPPKKNQDKTEIPTINT (19 aminoácidos e massa molecular de 2140 Daltons) oriundo do CMP. Adaptado de (MOTTA *et al.*, 2014). 43
- Figura 4:** Representação das (a) vibrações de estiramento e (b) deformações angulares (SKOOG; HOLLER & NIEMAN, 2009). 45
- Figura 5:** Representação simplificada da análise por ATR. 47
- Figura 6:** Representação em forma de rede neural para as LS-SVM empregando dados espectroscópicos (Adaptado: FERRÃO *et al.*, 2007). 62
- Figura 7:** Espectros médios de ATR-FTIR na região de 4000 a 800 cm^{-1} das 17 amostras de leite adulteradas com CMP 89
- Figura 8:** Dendrograma das amostras de leite cru (LC), leite semidesnatado (LS) e leite desnatado (LD) analisado por ATR-FTIR. 90
- Figura 9:** Gráfico dos escores PC1 x PC2 das amostras de leite cru (LC), leite semidesnatado (LS) e leite desnatado (LD). 92
- Figura 10:** Gráfico dos escores PC1 x PC3 das amostras de leite cru (LC), leite semidesnatado (LS) e leite desnatado (LD). 92
- Figura 11:** Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas pelo FTIR/PLS para determinação de CMP utilizando toda a faixa espectral 97
- Figura 12:** Gráfico dos valores do método de referência *versus* valores previstos pelo ATR-FTIR/PLS para a determinação de CMP em leite utilizando toda a

faixa espectral, sendo os losangos as amostras de calibração e os quadrados cinza as de previsão..... 97

Figura 13: Gráfico dos valores do método de referência versus valores previstos pelo ATR-FTIR/LS-SVM para a determinação de CMP em leite utilizando toda a faixa espectral, sendo os losangos vermelhos as amostras de previsão e as bolinhas cinza as de calibração..... 99

Figura 14: Gráfico de diagnóstico de *outlier* demonstrando a presença da amostra LD 30% sendo confirmado como uma amostra anômala..... 106

Figura 15: Gráfico da análise por componentes principais (PCA) com a presença da amostra *outlier*. 107

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1: Regiões espectrais do infravermelho..... | 44 |
| Tabela 2: Resultados das figuras do mérito dos modelos de regressão por siPLS para a determinação do CMP pelo método de referência..... | 95 |
| Tabela 3: Resultados da figura de mérito dos melhores modelos de regressão por siPLS em função ao número de combinações entre intervalos para a determinação de CMP. | 96 |
| Tabela 4: Resultados do LS-SVM..... | 98 |

LISTA DE ABREVIATURAS

1. ATR Reflexão Total Atenuada (do inglês *Attenuated Total Reflection*)
2. ATR-FTIR Espectroscopia no Infravermelho Médio com Transformada de Fourier e Reflexão Total Atenuada (do inglês *Fourier Transform Infrared Spectroscopy with Attenuated Total Reflection*)
3. CID Célula de Colisão (do inglês *Collision Induced Dissociation*)
4. CMP Caseinomacropéptido
5. CLAE Cromatografia Líquida de Alta Eficiência
6. CL-DAD Cromatografia Líquida Acoplada com Detector de Arranjos de Diodos
7. CL-EM Cromatografia Líquida Acoplada a Espectrômetro de Massas
8. CL-EM/EM Cromatografia Líquida Acoplada a Espectrômetro de Massas em Modo *Tandem*
9. DBO Demanda Bioquímica de Oxigênio
10. EC Eletroforese Capilar
11. EM Espectrometria de Massas
12. EM/EM Espectrometria de Massas em Modo *Tandem*
13. ESI Ionização Branda-ESI (do inglês *Eletrospray Ionization*)
14. FIR Infravermelho Distante (do inglês *Far Infrared*)
15. FTIR Espectroscopia no Infravermelho Médio com Transformada de Fourier (do inglês *Fourier Transform Infrared*)
16. HCA Análise de Agrupamentos Hierárquicos (do inglês *Hierarchical Cluster Analysis*)
17. IN Instrução Normativa
18. iPLS Mínimos Quadrados Parciais por Intervalo (do inglês *Interval Partial Least Squares*)
19. IR Infravermelho (do inglês *Infrared*)
20. KNN Método dos k -ésimos Vizinhos Mais Próximos
21. KS *Kennard-Stone*
22. LD Limites de Detecção
23. LS-SVM Máquinas de Vetor de Suporte com Mínimos Quadrados (do inglês *Last Squares Support Vector Machine*)

24. MAPA Ministério da Agricultura, Pecuária e Abastecimento
25. MIR Infravermelho Médio (do inglês *Mid Infrared*)
26. MLR Regressão Linear Múltipla (do inglês *Multiple Linear Regression*)
27. NIR Infravermelho Próximo (do inglês *Near Infrared*)
28. PC Análise de Componente Principal (do inglês: *Principal Component*)
29. PCA Análise por Componentes Principais (do inglês *Principal Component Analysis*)
30. PCR Regressão por Componentes Principais (do inglês *Principal Component Regression*)
31. PLS Mínimos Quadrados Parciais (do inglês *Partial Least Squares*)
32. RIISPOA Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal
33. RMSE Raiz Quadrada dos Erros Médios (do inglês *Root Mean Square Error*)
34. RMSEC Erro Quadrático Médio de Calibração (do inglês *Root Mean Square Error of Calibration*)
35. RMSECV Erro Quadrático Médio de Calibração Cruzada (do inglês *Root Mean Square Error of Cross Validation*)
36. RMSEP Erro Quadrático Médio de Predição (do inglês *Root Mean Square Error of Prediction*)
37. SIMCA Método *Soft Independent Modeling of Class Analogy*
38. siPLS Mínimos Quadrados Parciais por Sinergismo de Intervalos (do inglês *Sinergy Interval Partial Least Squares*)
39. SVM Máquinas de Vetor de Suporte (do inglês *Support Vector Machine*)
40. TCA Ácido Tricloroacético
41. UHT Ultra Alta Temperatura (do inglês *Ultra-High-Temperature*)
42. UV Ultravioleta
43. VLs Variáveis Latentes
44. VTs Variáveis Totais

SUMÁRIO

| | | |
|----------|---|----|
| 1 | INTRODUÇÃO | 22 |
| 2 | OBJETIVOS | 28 |
| 2.1 | Objetivo Geral | 30 |
| 2.2 | Objetivos Específicos | 30 |
| 3 | REVISÃO BIBLIOGRÁFICA | 32 |
| 3.1 | Composição e Qualidade do Leite | 34 |
| 3.2 | Produção do Queijo | 36 |
| 3.3 | Fraudes em Leite por Caseinomacropéptido (CMP) | 38 |
| 3.4 | Legislação | 40 |
| 3.4.1 | Legislação para Produção do Leite: Instrução Normativa N° 62, 29 de dezembro de 2011 | 40 |
| 3.4.2 | Legislação para Controle de Leite e Produtos Lácteos: Instrução Normativa N° 68, 12 de dezembro de 2006 | 40 |
| 3.4.3 | Legislação para Índice de CMP: Instrução Normativa N° 69, 13 de dezembro de 2006 | 41 |
| 3.4.4 | Legislação do Método Oficial para Determinação do CMP: Instrução Normativa N° 07, 10 de março de 2010 | 41 |
| 3.5 | Digestão Enzimática e Análise de Peptídeos por Cromatografia Líquida Acoplada à Espectrometria de Massas em Modo <i>Tandem</i> (CL-EM/EM) | 42 |
| 3.5.1 | Digestão Enzimática: Pepsina | 42 |
| 3.6 | Espectroscopia no Infravermelho | 44 |
| 3.6.1 | Espectroscopia por Reflexão Total Atenuada (ATR) | 46 |
| 3.7 | Quimiometria e Análise Multivariada | 48 |
| 3.8 | Análise Exploratória | 49 |
| 3.8.1 | Análise por Componentes Principais (PCA) | 49 |
| 3.8.2 | Análise de Agrupamentos Hierárquicos (HCA) | 50 |
| 3.9 | Regressão Multivariada | 51 |
| 3.9.1 | Regressão por Mínimos Quadrados Parciais (PLS) | 51 |
| 3.9.2 | Métodos de Seleção de Variáveis em Modelos de Regressão | 54 |

| | |
|--|------------|
| 3.9.3 Seleção do Conjunto de Calibração e Previsão | 55 |
| 3.9.4 Tratamentos e Pré-Processamentos dos Dados | 55 |
| 3.9.5 Avaliação dos Modelos de Regressão | 60 |
| 3.10 Máquinas de Vetores de Suporte com Mínimos Quadrados (LS-SVM) | 61 |
| | |
| 4 REFERÊNCIAS BIBLIOGRÁFICAS | 66 |
| | |
| 5 ARTIGO CIENTÍFICO - Ferramentas Quimiométricas Aplicadas no Desenvolvimento de Metodologia Analítica para Avaliação de Adulteração em Leite Bovino por Adição de Soro do Queijo por Espectroscopia no Infravermelho (FTIR)..... | 78 |
| 5.1 Apresentação..... | 80 |
| 5.2 Artigo..... | 80 |
| | |
| 6 DISCUSSÃO GERAL..... | 104 |
| | |
| 7 CONCLUSÃO | 108 |

1 INTRODUÇÃO

No Brasil o setor leiteiro vem passando por significativas mudanças nos últimos anos (ESPER, SILVA & SÁ, 2014). A qualidade do leite tornou-se preocupação para os pesquisadores de todo o mundo. Uma das fraudes econômicas mais aplicadas ao leite é a adição do soro do queijo. O soro do queijo é um subproduto de baixo custo da indústria queijeira, resultante da fração aquosa do leite separada da caseína durante a produção dos queijos (OLIVEIRA *et al.*, 2009).

A qualidade do leite é determinada a partir de diversos parâmetros que se baseiam nas condições físico-químicas do mesmo. Essas condições abrangem a composição do leite, que é constituída pela parte úmida, representada pela água, e a parte sólida, representada por dois grupos de componentes: o extrato seco total e o extrato seco desengordurado, e também a densidade. Outros fatores que envolvem a qualidade do leite são: alimentação do gado, raça do animal, ordenha, manejo do bezerro, ordem da ordenha. Os parâmetros de higiene-sanitária também são avaliados através de teste de acidez do leite, prova do álcool-alizarol, teste de redução do azul metileno, contagem total de bactérias (ESPER, SILVA & SÁ, 2014).

A legislação brasileira proíbe a adição do soro do queijo no leite por reduzir o valor nutritivo do alimento e diluir os teores de gorduras. Além disso, essa prática atinge o direito do consumidor e em casos de consumo prolongado, pode gerar sérios riscos à saúde (OLIVEIRA *et al.*, 2009).

Em 2013 no estado do Rio Grande do Sul, foi descoberto um esquema de fraude no leite bovino, tendo sido apreendidos pelos fiscais do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), milhares de litros de leite adulterado, através da operação chamada de “leite compensado” (BRASIL, 2013).

A cadeia reprodutiva do leite é realizada através do fornecedor de matéria prima que é o produtor onde uma determinada empresa de transporte realiza a coleta e é responsável pelo transporte deste até as indústrias ou cooperativas de

laticínios onde acontecerá o processamento desta matéria prima até chegar ao consumidor. As fraudes podem acontecer ainda na propriedade rural ou durante o transporte e armazenamento dessa matéria prima onde ela é adulterada em algumas dessas etapas (BRASIL, 2013).

O soro do queijo apresenta bastante semelhança com o leite; dessa forma facilita o processo de incorporação ao leite processado. A presença de Caseinomacropéptido (CMP) no leite, de acordo com a concentração, pode indicar uma adição fraudulenta de soro de queijo (MARQUES *et al.*, 2011; FRIEDRICH *et al.*, 2010).

O CMP é um composto específico do soro; dessa forma ele é um indicativo para detecção de adulteração no leite. Métodos oficiais no Brasil têm sido muito questionados, tendo em vista se fundamentar na detecção de uma fração proteica que pode também estar presente no leite. O MAPA adota a Instrução Normativa que aprova o método oficial de determinação de CMP em leite, por Espectrometria de Massas aplicada ao leite Ultra Alta Temperatura (UHT, do inglês *Ultra-High-Temperature*) e pasteurizado. Assim, é necessário desenvolver um método para substituir o método convencional adotado em nossa legislação (BRASIL, 2010).

As análises do leite são realizadas de acordo com a instrução normativa nº62/2011 e se baseiam em algumas propriedades físico-químicas do produto. Na maioria das vezes esses métodos não conseguem detectar de uma forma direta a autenticidade do leite bovino (SANTOS, 2013).

Embora existam diversos estudos feitos para a determinação do CMP em leite bovino, ainda permanece uma atual necessidade de desenvolvimento de métodos analíticos rápidos, precisos, baratos para a identificação de eventuais fraudes às quais o leite está sujeito. Esse trabalho pretende utilizar metodologia empregando a quimiometria através da aplicação de estatística multivariada, modelagem matemática e métodos baseados na lógica matemática formal aplicada à resolução de problemas, para desenvolver um método analítico com as características acima descritas (SANTOS, 2013; LAVINE, WORKMAN, 2008).

Assim sendo, o objetivo deste trabalho é a aplicação de ferramentas quimiométricas para desenvolver um método rápido capaz de ser aplicado em plataforma nas indústrias de laticínios, com grande capacidade operacional, com o preparo de amostra simples, passível de transposição para análise em campo (sondas), mas que seja correlacionado com método da Instrução Normativa nº 69 (IN 69) por Cromatografia Líquida Acoplada a Espectrômetro de Massas em Modo *Tandem* (CL-EM/EM) é o objetivo deste trabalho. De acordo com a literatura consultada, não há método para esse tipo de matriz no Brasil, nem a nível mundial, que correlacione, de forma rápida e simples, eventuais fraudes com os resultados fornecidos pelo método oficial de CL-EM/EM na faixa requerida pela IN 69 (BRASIL, 2006b).

2 OBJETIVOS

2.1 Objetivo Geral

- Desenvolvimento de método multivariado por espectroscopia no infravermelho para quantificação de CMP em leite bovino.

2.2 Objetivos Específicos

- Desenvolver modelo de regressão multivariada por Espectroscopia no Infravermelho Médio com Transformada de Fourier (FTIR, do inglês *Fourier Transform Infrared*) correlacionado com o método oficial IN 69;
- Determinação de CMP em amostras fraudadas de leite bovino utilizando Espectroscopia no Infravermelho Médio com Transformada de Fourier e Reflexão Total Atenuada (ATR-FTIR, do inglês *Fourier Transform Infrared Spectroscopy with Attenuated Total Reflection*).

3 REVISÃO BIBLIOGRÁFICA

3.1 Composição e Qualidade do Leite

Em 2014 o Brasil foi o quinto colocado no ranking mundial na produção de leite bovino perdendo apenas para União Europeia, Índia, Estados Unidos e China, respectivamente. O Brasil como quinto colocado produziu 33,3 bilhões de litros de leite no ano 2014. Os principais estados que se destacam nesse setor de produção de leite bovino são Minas Gerais (27%), seguido de Rio Grande do Sul (13%), Paraná (12,6%) e Goiás (11%) (SEAB, 2014).

Em todas as partes do mundo o leite é consumido, tanto na sua forma líquida como na forma de seus mais diversos derivados (MAGALHÃES, 2008). O leite é composto por várias substâncias; sua principal função é servir como alimento fornecendo nutrientes. Ainda, o leite permite um amplo processamento industrial para diferentes produtos na alimentação humana. A preparação dos diferentes produtos lácteos depende das composições físico-químicas do leite (MOTTA, *et al.*, 2014).

O MAPA define o leite como um produto obtido da ordenha completa e ininterrupta, em condições de higiene, de vacas leiteiras sadias, bem alimentadas e em repouso. O leite de outros animais deve se designar segundo a espécie de que proceda (BRASIL, 2002).

A composição média do leite bovino é indicada como sendo água (87,30%), lactose (4,90%), gordura (3,80%), proteínas (3,30%) e sais minerais (0,72%). Na fase líquida do leite as partículas suspensas são gotículas de gordura e micelas da caseína. Porém, o leite produzido pelo animal pode variar o volume e a relação dos diferentes componentes. A composição do leite pode modificar dependendo dos fatores como: raça, espécie animal, individualidade animal, intervalo entre ordenhas, variação durante a ordenha, diferenças entre os quartos, período de lactação, influência das estações, alimentação, temperaturas, idade do animal, doenças e condições climáticas (ROMA JÚNIOR, 2008; OLIVEIRA, 2010).

No Brasil a qualidade do leite consumido vem sendo uma constante preocupação das autoridades e indústrias de laticínios. Um leite de qualidade pode ser determinado em termos de sua integridade, ou, seja, livre de adição de

substâncias e/ou retirada de componentes, de sua composição química e características sensoriais, como sabor, odor e aspecto; além disso, deve ser isento de patógenos e de deterioração microbiológica (DÜRR, 2004).

Alguma adulteração na composição do leite é muito importante, pois pode modificar significativamente o seu valor nutricional como matéria prima para a fabricação de derivados (MOTTA *et al.*, 2014). A detecção e quantificação da adição fraudulenta de soro ao leite são realizadas pela determinação de CMP (MARQUES *et al.*, 2011).

O Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal (RIISPOA, 1950), pelo Artigo 543 considera fraudado, adulterado ou falsificado o leite que for adicionado de água, tiver sofrido subtração de qualquer de seus componentes, for adicionado de substâncias conservadoras ou de quaisquer elementos estranhos à sua composição, estiver cru e for vendido como pasteurizado ou ainda for exposto ao consumo sem as devidas garantias de inviolabilidade. A fraude é definida como a adição ou subtração parcial ou total de qualquer substância na composição original do produto (BRASIL, 2008).

As fraudes têm por objetivo aumentar o volume do leite, conseqüentemente alterando as características e os componentes específicos do teor de proteínas com finalidade de receber bonificações, proporcionando assim uma concorrência desleal no mercado (VIOTTO & CUNHA, 2006; GUAN *et al.*, 2010).

No início as adulterações no leite almejavam o aumento do volume do leite, através de adição de água, e desnate para produção de creme de leite. Após foram surgindo novos tipos de adulterações, como adição de soro de queijo, de substâncias conservantes (peróxido de hidrogênio), neutralizantes (hidróxido de sódio, bicarbonato de sódio) e reconstituintes da densidade e crioscopia (sal, açúcar e amido) (ALMEIDA, 2013).

Como já foi dito, uma das fraudes econômicas mais aplicadas ao leite no Brasil é a adição do soro do queijo. O soro do queijo é uma preocupação tanto para os órgãos de fiscalização nacionais quanto os internacionais, pois trata de uma fraude que prejudica o consumidor e os concorrentes que seguem a

legislação vigente, além da fraude ser prevista como crime na legislação (OLIVEIRA *et al.*, 2009; CARVALHO *et al.*, 2007).

O soro do queijo apresenta bastante semelhança com o leite; dessa forma facilita o processo de incorporação ao leite processado. A presença de CMP no leite, de acordo com a concentração, pode indicar uma adição fraudulenta de soro de queijo (MARQUES *et al.*, 2011; FRIEDRICH *et al.*, 2010).

No Brasil ainda não existe levantamento que aponte o quanto exatamente é produzido de soro do queijo. O soro do queijo produzido pelas indústrias de laticínios se não for descartado com tratamento e em lugar apropriado pode acarretar em poluição e danos para o meio ambiente. Quando o soro é dispensado sem tratamento produz uma Demanda Bioquímica de Oxigênio (DBO) de 30.000 a 50.000 mg de oxigênio por litro de soro (BERRY & BULLERMAN, 1966).

Muitos métodos estão sendo desenvolvidos para a detecção desse tipo de fraude: composição de aminoácidos (cistina e cisteína, ácido aspártico, alanina e prolina); relação entre os constituintes (relação de nitrogênio soroproteico/nitrogênio caseínico); métodos crioscópicos; observação microscópica com luz polarizada; imunoenzimático; CLAE; determinação do ácido siálico livre; entre outros (MARQUES *et al.*, 2011).

3.2 Produção do Queijo

A indústria queijeira produz cerca de 9 kg de soro para cada quilograma do produto final. De acordo com o MAPA, queijo é o “produto fresco ou maturado que se obtém por separação parcial do soro do leite, coagulado pela ação física do coalho, utilizado em todos os tipos de queijo exceto os frescos tipo “cottage”, e precipitar a caseína presente no leite” (MOTTA *et al.*, 2014).

A enzima mais importante responsável por essa ação é a renina (quimosina), uma fosfoproteína de ação proteolítica específica que atua hidrolisando ligações peptídicas, entre os aminoácidos (aas) 105-106 da κ -caseína (169 aas), transformando-a em p- κ -caseína (105 aas) que precipita em caseinomacropéptido (64 aas) que se conserva solúvel no soro (Figura 1). A

caseína é uma fosfoproteína que estabelece ligações com o cálcio, o que lhe confere certa estabilidade, motivo pelo qual não coagula no ponto de ebulição do leite (MOTTA *et al.*, 2014; MARQUES *et al.*, 2011).

O soro do queijo possui além de caseinomacropéptido também as proteínas β - lactoglobulina e α -lactalbumina que estão presentes em grandes quantidades. A massa aproximada do caseinomacropéptido é 7000 Da e é comumente utilizado como marcador da adulteração de leite por adição de soro de queijo (MOTTA *et al.*, 2014).

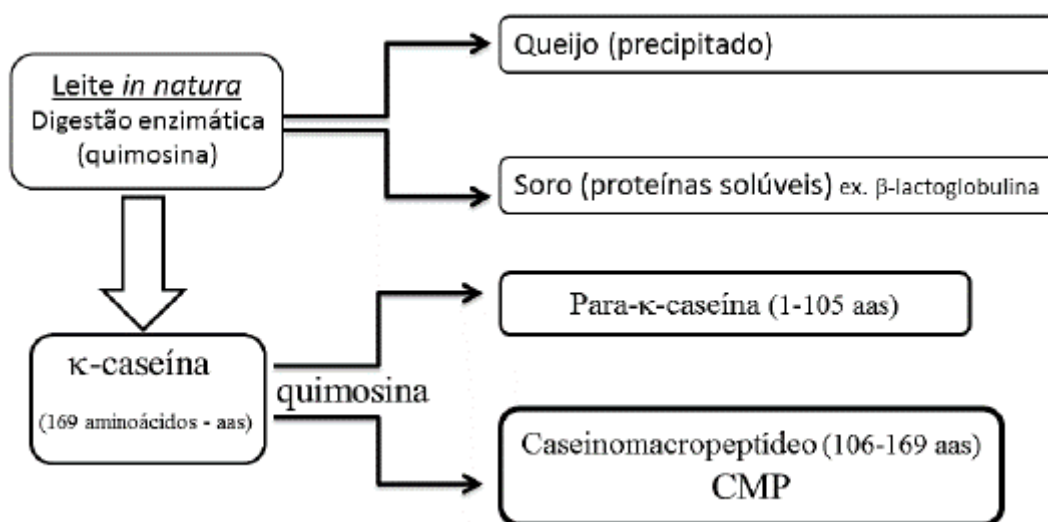


Figura 1: Esquema de processo enzimático na produção de queijos: clivagem da κ - caseína pela ação da enzima quimosina, na qual a para- κ -caseína precipita (queijo) e o CMP permanece solúvel no soro (MOTTA *et al.*, 2014).

Existe outro método de coagulação da caseína que é acrescentar ácido orgânico ao leite em quantidade suficiente para igualar o pH do meio isoelétrico da proteína (pH 4,5). As micelas da caseína neste pH agregam-se e precipitam. Porém, esse método resulta em queijos de baixa qualidade comparados com os produzidos pelo método enzimático (MOTTA *et al.*, 2014).

3.3 Fraudes em Leite por Caseinomacropeptídeo (CMP)

Uma das fraudes econômicas mais aplicadas ao leite é a adição do soro do queijo. O soro do queijo é um subproduto de baixo custo da indústria queijeira, resultante da fração aquosa do leite separada da caseína durante a produção dos queijos e corresponde a cerca de 80 a 90% do volume do leite dependendo do tipo de queijo (OLIVEIRA *et al.*, 2009; ZINSLY *et al.*, 2001).

Dependendo das técnicas de fabricação e do tipo de queijo o soro do queijo pode diferenciar-se; este soro pode ser obtido por meio de três processos fundamentais: coagulação enzimática (enzima quimosina ou renina), resultando no coágulo de caseínas que é a matéria-prima utilizada na produção de queijos e obtenção do soro doce. Outros tipos de processos são a precipitação ácida que origina o soro ácido e a separação física das micelas e das proteínas do soro, formando concentrado ou isolado protéico (ZINSLY *et al.*, 2001).

O tipo de soro mais comum no Brasil é o soro doce obtido pelo processo de coagulação enzimático (quimosina) resultado da fabricação de queijos do tipo mussarela, prato, frescal, e outros que são os mais comercializados no país. O soro ácido é proveniente da manufatura de queijos de coagulação ácida com menor consumo (ricota e requeijão) e da fabricação da caseína, rara no Brasil (OLIVEIRA *et al.*, 2009).

O soro é um produto que contém grande quantidade de proteínas com alto valor nutritivo; é constituído também por aminoácidos, lactose, vitaminas e sais minerais. No entanto, leite não pode ser substituído por soro, pois a quantidade de soro precisa ser especificada, de acordo com a IN 69; caso isso não aconteça pode ser considerado como fraude (VELOSO *et al.*, 2002; BRASIL, 2006b).

A legislação brasileira institui que o leite pasteurizado, o leite esterilizado e o leite em pó não podem conter sólidos de soro de queijo acima do especificado na IN 69 do MAPA (MOTTA *et al.*, 2014; BRASIL, 2006b). Não somente o Brasil, mas as legislações de todos os países proíbem a adição do soro do queijo no leite para consumo (BRANDÃO, 2000).

O caseinomacropeptídeo é um elemento específico do soro; dessa forma ele é um marcador para detecção de adulteração no leite (MARQUES *et al.*, 2011).

O CMP é um peptídeo específico encontrado no soro do queijo, formado a partir da clivagem da κ -caseína pela ação da quimosina; também pode ser gerado pela ação da proteólise bacteriana (Pseudo-CMP), esta é causada por contaminação no leite que pode ocorrer naturalmente devido às condições de transporte e armazenamento. O pseudo-CMP pode diferir do CMP por um único aminoácido (metionina terminal no CMP, resíduo 106 – alanina terminal do pseudo-CMP, resíduo 107). O pseudo-CMP possui um aminoácido a menos que o CMP (fragmento 106-169 da κ -caseína) (MOTTA *et al.*, 2014; RECIO *et al.*, 1997).

O leite possui quatro tipos principais de caseína: α -S1, α -S2, β e κ -caseína. A caseína é composta por cerca de 50% α -caseína 30% de β -caseína e 15% de κ -caseína. Todas são fosforiladas em resíduos específicos de serina e em consequência a κ -caseína é glicosilada em diferentes graus. Portanto, o CMP é heterogêneo e têm todos os sítios originais de modificações pós-traducionais da κ -caseína: sítios de glicosilação e fosforilação do CMP (Figura 2).

Quatro variantes genéticas do CMP foram identificadas, oriundos da κ -caseína A, B, C e E. As variantes A e B são as formas mais comuns no leite. Ambas as variantes existem em formas individual- e duplamente fosforilada e podem ser expressas de modo homozigoto ou heterozigoto (MOTTA *et al.*, 2014).

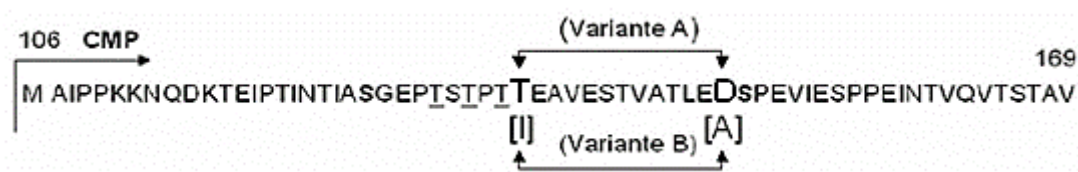


Figura 2: Sequência de aminoácidos das variantes genéticas (A e B) do CMP.

Na Figura 2, os pontos de variação genética A e I estão entre colchetes. Os resíduos de Serina (S) em negrito representam os sítios de fosforilações e os

resíduos tirosina (T) sublinhados representam os sítios de glicosilações (MOTTA *et al.*, 2014; MARQUES *et al.*, 2011).

A qualidade do leite é uma constante preocupação no Brasil. A adulteração do leite com a adição soro de queijo é uma das adulterações mais preocupantes, pois, ocorre a substituição de parte do leite por soro de queijo ao usar um derivado do próprio leite. Isso determina dificuldades para detectar a adulteração. A cada 10% de soro adicionado ocorre a redução de 8% no teor de proteínas do leite (OLIVEIRA *et al.*, 2009).

3.4 Legislação

Atualmente estão em vigor as IN 07, 62, 68 e 69 do MAPA referentes à produção e processamento do leite no Brasil.

3.4.1 Legislação para Produção do Leite: Instrução Normativa N° 62, 29 de dezembro de 2011

O MAPA pulicou, em 2002, a IN 51 e em 29 de dezembro de 2011 a atualização desta pela IN 62, onde regulamenta a Produção, Identidade, Qualidade, Coleta e Transporte do Leite Tipo A, Tipo B, Tipo C, Leite Cru, Leite Refrigerado e Leite Pasteurizado (Art. 1°) (BRASIL, 2011).

3.4.2 Legislação para Controle de Leite e Produtos Lácteos: Instrução Normativa N° 68, 12 de dezembro de 2006

Essa IN tem como finalidade oficializar os Métodos Analíticos Físico-Químicos, para Controle de leite e Produtos Lácteos, e conformidade com o anexo desta IN determinando que sejam utilizados em Laboratórios Nacionais Agropecuários (Art. 1°) (BRASIL, 2006a).

A IN 68 apresenta detalhadamente o método para detecção de fraudes por adição de soro de queijo em leite, denominada índice de CMP. Este método

baseia-se na detecção e quantificação de CMP proveniente da ação proteolítica de enzimas por meio de CLAE com separação em coluna de filtração em gel e detecção em ultravioleta (UV) (BRASIL, 2006a).

3.4.3 Legislação para Índice de CMP: Instrução Normativa N° 69, 13 de dezembro de 2006

A IN 69 institui critério de avaliação da qualidade do leite *in natura*, concentrado e em pó, reconstituídos, com base no método analítico físico-químico oficial denominado “índice de CMP”, de que trata a IN 68, de dezembro de 2006 (Art. 1º) (BRASIL, 2006b).

De acordo com o Art. 2º, somente quando o índice de CMP se for igual ou inferior a 30 mg/L, o leite, pode ser destinado para o abastecimento direto. Quando o índice de CMP estiver entre 30 mg/L e 75 mg/L, de acordo com o primeiro parágrafo desse artigo, o leite deveria ser destinado à produção de derivados lácteos (BRASIL, 2006b).

3.4.4 Legislação do Método Oficial para Determinação do CMP: Instrução Normativa N° 07, 10 de março de 2010

A IN 07 institui o método oficial para a determinação do CMP em leite, por CLAE, Eletroforese Capilar (CE, do inglês *Capillary Electrophoresis*) e Espectrometria de Massas (EM) em leite, nas formas integrais, semidesnatadas e desnatadas, tratados por processos UHT ou pasteurização (Art. 1º). O método de que trata esta IN foi adotado pelos laboratórios pertencentes à Rede Nacional de Laboratórios Agropecuários do Sistema Unificado de Atenção à Sanidade Agropecuária (BRASIL, 2010).

3.5 Digestão Enzimática e Análise de Peptídeos por Cromatografia Líquida Acoplada à Espectrometria de Massas em Modo *Tandem* (CL-EM/EM)

Para análise de proteínas pôr EM utiliza-se a digestão enzimática, pois estas possuem geralmente massa molecular alta, além de variantes genéticas e sítios de modificações traducionais. O uso de enzimas causa a clivagem em pontos específicos e conversão das proteínas em um conjunto de peptídeos (QIAN *et al.*, 1995; RECIO *et al.*, 1997; LÓPEZ & RECIO, 2006).

As proteínas são de difícil manuseio e tendem a degradar-se facilmente, por isso, o uso de peptídeos ao invés de proteínas intactas. Também, o espectrômetro de massas possui maior sensibilidade para análise de peptídeos, sendo mais eficiente para obter informação estrutural de peptídeos com até 20 aminoácidos (BOGDANOV & SMITH, 2005).

3.5.1 Digestão Enzimática: Pepsina

Para a análise de proteínas a enzima pepsina é frequentemente utilizada para realizar a digestão enzimática (LÉONIL & MOLLÉ, 1990; RECIO *et al.*, 1997; MARTIN-OURE; HENRY & BOUHALLAB, 1999; LÓPEZ & RECIO, 2006). No CMP, a pepsina (EC 3.4.23.1) cliva preferencialmente as ligações carboxila entre os aminoácidos hidrofóbicos tirosina e (iso)leucina. (ENTRIES, 1995). Esta clivagem resulta em produtos de digestão de faixa e tamanho adequados para a análise por EM, para a maior parte das proteínas. O espectro de massas originado pode ser comprado com os fragmentos obtidos em programas específicos de sequenciamento de proteínas para identificar as proteínas (ou peptídeos) de importância (MOTTA *et al.*, 2014).

Pela ação da enzima pepsina neste caso, é gerado o peptídeo com massa molecular de 2140 Daltons, conforme a Figura 3, que mostra os pontos de clivagem do CMP (MOTTA *et al.*, 2014).

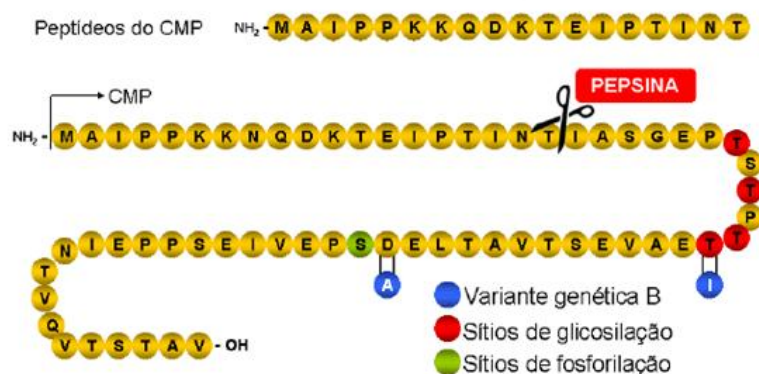


Figura 3: Pontos de clivagem da digestão enzimática do CMP utilizando pepsina. É gerado o peptídeo MAIPPCKKNQDKTEIPTINT (19 aminoácidos e massa molecular de 2140 Daltons) oriundo do CMP. Adaptado de (MOTTA *et al.*, 2014).

Depois da digestão enzimática é realizada a análise por CLAE (KREUFL *et al.*, 2014), com a detecção por UV ou fluorescência e acoplamento a EM (MOTTA *et al.*, 2014). Alguns estudos mostram que analisadores do tipo quadrupolo de Espectrometria de Massas em Modo *Tandem* (EM/EM) têm fornecido excelentes resultados quando utilizados em estudos confirmatórios (RECIO *et al.*, 1997). A separação dos íons moleculares selecionados na varredura acontece no primeiro quadrupolo, já no segundo quadrupolo, também conhecido como Célula de Colisão (CID, do inglês *Cell Collision induced dissociation*), acontece a fragmentação dos íons e no terceiro quadrupolo, a seleção dos fragmentos originados no CID é direcionada ao detector de massas (MOTTA *et al.*, 2014).

A utilização de fontes de Ionização Branda-ESI (ESI, do inglês *Eletrospray Ionization*) foi um avanço no desenvolvimento de métodos analíticos confiáveis. Na técnica CL-EM/EM, dependendo da característica da molécula aplicado na câmara de ionização é possível gerar íons positivos (modo positivo) ou negativos (modo negativo), dependendo das características da molécula. A escolha do sistema mais apropriado, dependerá dos analitos e da matriz em estudo, e é definido experimentalmente. A técnica de CL-EM/EM, por ser uma técnica

seletiva, sensível, capaz de avaliar Limites de Detecção (LD) na ordem de $\mu\text{g L}^{-1}$ e ng^{-1} e ser especialmente adequada devido à sua seletividade, se tornou indispensável na análise proteômica (MOTTA *et al.*, 2014).

3.6 Espectroscopia no Infravermelho

A região do infravermelho do espectro eletromagnético é a base para a espectroscopia no infravermelho. Cada faixa do espectro eletromagnético possui suas aplicações analíticas específicas e existem equipamentos dedicados para cada faixa. Na região do infravermelho a faixa compreende aproximadamente de 780 a 100000 nm. A Tabela 1 demonstra a porção desta faixa no espectro eletromagnético e a sua divisão em três partes: Infravermelho Próximo (NIR), Infravermelho Médio (MIR) e Infravermelho Distante (FIR) (SKOOG; HOLLER & NIEMAN, 2009).

Tabela 1: Regiões espectrais do infravermelho.

| Região | Comprimento de onda (λ), nm | Números de onda ($\bar{\nu}$), cm^{-1} | Frequência (ν), Hz |
|-----------------|---|--|---|
| Próximo | 780 a 2500 | 12800 a 4000 | $3,8 \times 10^{14}$ a $1,2 \times 10^{14}$ |
| Médio | 2500 a 5000 | 4000 a 200 | $1,2 \times 10^{14}$ a $6,0 \times 10^{12}$ |
| Distante | 5000 a 100000 | 200 a 10 | $6,0 \times 10^{12}$ a $2,0 \times 10^{13}$ |

Fonte: Adaptado de (SKOOG; HOLLER & NIEMAN, 2009).

A radiação infravermelha não produz transições eletrônicas, pois não possui energia suficiente; assim, para que tenha absorção da radiação infravermelha uma molécula precisa apresentar uma variação no momento de dipolo durante seu movimento rotacional ou vibracional. Apenas nestas ocasiões o campo elétrico alternado da radiação pode interagir com a molécula e ocasionar variações na amplitude de um de seus movimentos, sendo que a energia absorvida está associada aos modos vibracionais e rotacionais de grupos funcionais (SKOO; HOLLER & NIEMAN, 2009).

Os inúmeros tipos de vibrações e rotações das ligações da molécula variam de acordo com as posições relativas dos átomos. As vibrações possuem categorias e podem ser classificadas como estiramento ou deformação angular e simétricas ou assimétricas. Existem quatro tipos de vibrações de deformação e são caracterizados por variações no ângulo entre duas ligações: deformação simétrica no plano, deformação assimétrica no plano, deformação simétrica fora do plano e deformação assimétrica fora do plano (Figura 4) (SKOOG; HOLLER & NIEMAN, 2009).

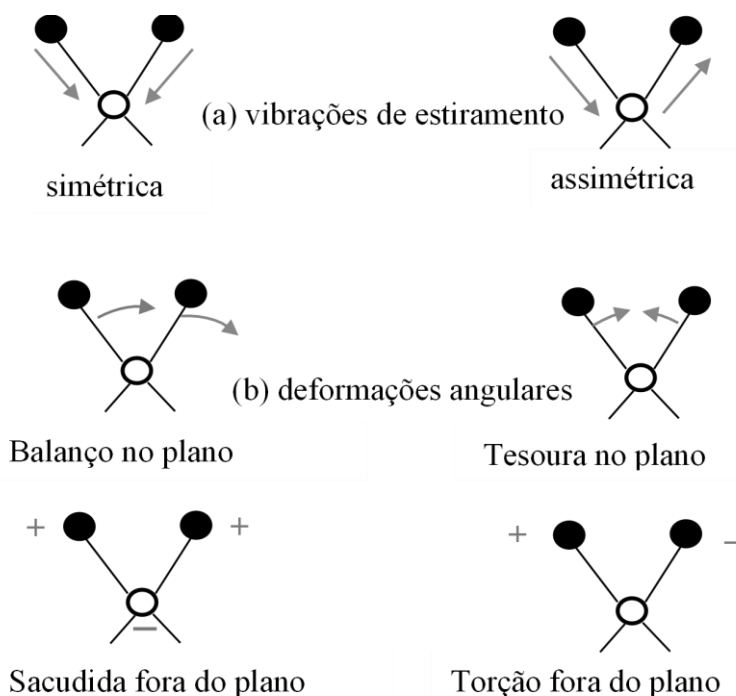


Figura 4: Representação das (a) vibrações de estiramento e (b) deformações angulares (SKOOG; HOLLER & NIEMAN, 2009).

Os espectros na região do infravermelho são caracterizados pelas bandas de absorção associadas a interações de sobretons e combinações de vibrações fundamentais de diversos grupos funcionais, tais como alifáticos (CH), aromáticos (CH), aminas (NH), amidas (NO), hidroxil (OH), carboxil (CO). As bandas de absorção na região do NIR são bem menos intensas e mais sobrepostas que no MIR (ZHANG, 2012).

Os espectros obtidos na região do MIR apresentam picos bem definidos, as bandas são intensas e os picos possuem altas absorvâncias, facilitando a interpretação dos espectros. A faixa espectral do MIR, conhecida como região de impressão digital, compreendida entre $1200 - 600 \text{ cm}^{-1}$ é de grande utilidade para identificação de compostos, visto que pequenas diferenças na estrutura e na constituição das moléculas resultam em mudanças significativas no perfil e na distribuição dos picos de absorção. Deste modo, é possível identificar, por exemplo, um composto orgânico comparando-se seu espectro com banco de dados pré-existent (SKOOG; HOLLER & NIEMAN, 2009).

As principais técnicas de reflexão no infravermelho são: Reflexão Total Atenuada (ATR, do inglês *Attenuated Total Reflection*), reflexão difusa e reflexão especular (ou externa). Neste trabalho, foi utilizada ATR-FTIR.

3.6.1 Espectroscopia por Reflexão Total Atenuada (ATR)

A espectroscopia por reflexão total atenuada é uma técnica robusta, não destrutiva e segura para estudos quantitativos e qualitativos envolvendo líquidos com o mínimo preparo de amostra ou nenhum (SKOOG, HOLLER & NIEMAN, 2009).

O princípio da espectroscopia por ATR está associado às múltiplas reflexões da radiação infravermelha que ocorrem no interior de cristais, de materiais com alto índice de refração como, por exemplo, o seleneto de zinco (ZnSe), interagindo apenas com a amostra que estiver superficialmente no cristal (FERRÃO *et al.*, 2004).

Um acessório de Reflexão Total Atenuada opera medindo as mudanças que ocorrem em um feixe de infravermelho que sofre reflexão interna total, ao entrar em contato com a amostra. Embora ocorra a reflexão, parte da radiação penetra além da superfície do cristal de ATR, interagindo com o material absorvente (amostra) (FERRÃO, 2001).

A profundidade de penetração (dp) é dada pela Equação 1:

$$dp = \frac{\lambda}{2\pi\eta_1 \left[\sin^2\theta_1 - \left(\frac{\eta_3}{\eta_2}\right)^2 \right]^{\frac{1}{2}}}, \quad (1)$$

onde θ_1 é o ângulo de incidência, η_2 é o índice de refração do elemento de ATR e η_3 é o índice de refração da amostra.

Esta radiação penetrante é chamada de onda evanescente. Se a amostra absorve determinado comprimento de onda evanescente, ocorre uma atenuação do feixe cujo comprimento de onda corresponde às bandas de absorção no infravermelho. É em função deste mecanismo que surge o espectro (FERRÃO, 2001). A Figura 5 mostra a representação simplificada da análise por ATR.

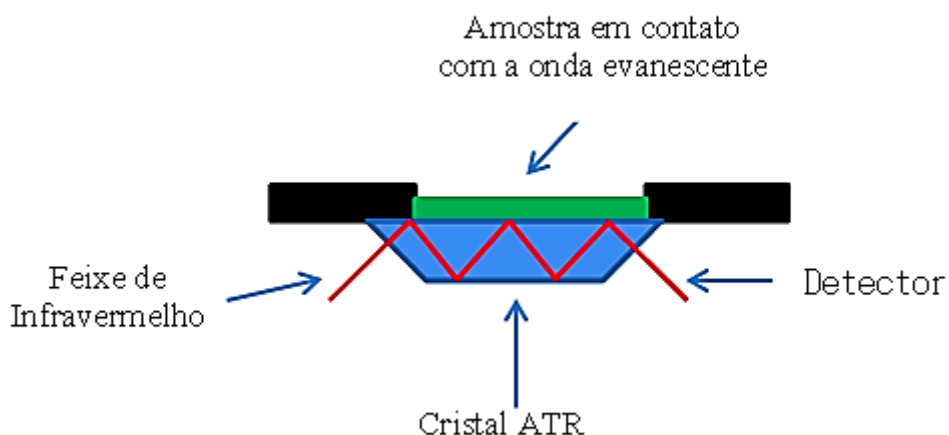


Figura 5: Representação simplificada da análise por ATR.

Recentemente a técnica de ATR tem revolucionado as análises de amostras líquidas e sólidas porque combate os aspectos mais desafiadores da análise no infravermelho, como a preparação das amostras e a reprodutibilidade espectral. Destaca-se por empregar pequena quantidade de amostra e por não envolver processos laboriosos no preparo destas (FERRÃO *et al.*, 2011; LIRA *et al.*, 2010).

3.7 Quimiometria e Análise Multivariada

A quimiometria é uma ferramenta criada pelo químico Svante Wold em 1974, que pode ser utilizada como uma alternativa para análises nas indústrias de alimentos. Essa ferramenta usa métodos matemáticos, estatísticos e computacionais para projetar ou selecionar experimentos de forma otimizada e para obter o máximo de informação pela análise de dados químicos (FERREIRA *et al.*, 1999; WORKMAN *et al.*, 2011).

A demanda em obter resultados rápidos, expressos de forma simples e conclusivos a partir de um volume grande de dados faz a quimiometria presente em praticamente todas as áreas de aplicação da química analítica (análise farmacêutica, proteômica, metabolômica, toxicologia, qualidade dos alimentos, entre outros) (BRERETON, 2003; FERREIRA & KIRALJ, 2007; WORKMAN *et al.*, 2011).

Dentro da quimiometria a calibração multivariada ganhou ênfase para tratamento de dados espectrais. Na fase de tratamento dos resultados alcançados, as ferramentas de classificação dos dados podem ser usadas com o objetivo de classificar e reconhecer padrões ou utilizar a multidimensionalidade da resposta analítica do instrumento de medição (NIR, CL-EM, CL-DAD), na forma de calibração multivariada (BRERETON, 2003; LAVINE, WORKMAN, 2008).

Para a classificação dos dados é possível utilizar algoritmos que se baseiam na distância direta entre os objetos (HCA, KNN) ou na projeção dos dados em espaços de menor dimensão (PCA, *Análise de Fatores*, SIMCA) (BRERETON, 2003; LAVINE, WORKMAN, 2008).

As principais técnicas de análise multivariada utilizadas para a elaboração de modelos são: Regressão Linear Múltipla (MLR, do inglês *Multiple Linear Regression*), Regressão por Componentes Principais (PCR, do inglês *Principal Component Regression*), Análise por Componente Principal (PCA, do inglês *Principal Component Analysis*) e Mínimos Quadrados Parciais (PLS, do inglês, *Partial Least Squares*). Dentre essas, o PLS permitiu progresso significativo na aplicação da quimiometria (BURNS & CIURCZAK, 2008).

Já na calibração multivariada há opção de utilizar a calibração via mínimos quadrados parciais (PLS), presente desde a década de 80 ou em versões mais recentes, como o N-PLS, BLLS, TLLS, entre outros (BRERETON, 2003; OLIVIERI, 2008).

Em uma abordagem mais atual, redes neurais em diversos campos da ciência estão sendo usadas para modelagens. Esse método permite optar por regressão via Máquinas de Vetor de Suporte com Mínimos Quadrados (LS-SVM, do inglês *Last Squares Support Vector Machine*), que pode ser uma técnica bastante promissora na avaliação de controle da qualidade de alimentos (FERRÃO *et al.*, 2007).

3.8 Análise Exploratória

3.8.1 Análise por Componentes Principais (PCA)

A PCA encontra-se entre as mais importantes ferramentas de análise multivariada, inclusive por constituir a base para o fundamento de vários métodos multivariados de análise de dados. Além disso, a PCA é um dos métodos quimiométricos mais utilizados de análise exploratória, classificada como não supervisionada, visto que nenhuma informação com relação à identidade das amostras é levada em consideração. Essa ferramenta permite revelar a existência ou não de amostras anômalas, de relações entre as variáveis medidas e de relações ou agrupamentos entre amostras (LYRA *et al.*, 2010).

A PCA basicamente consiste na transformação de coordenadas da matriz de dados originais. O objetivo da PCA é representar as variações presentes em muitas variáveis por meio de um número menor de variáveis. Dessa forma, o método permite a redução da dimensionalidade através da representação do conjunto de dados em um novo sistema de eixos, esses sistemas estabelecem novas variáveis ortogonais entre si, chamadas de Componentes Principais (PC) (do inglês: *Principal Component*). Essas permitem a visualização das estruturas dos dados de natureza multivariada em gráficos bidimensionais ou

tridimensionais e, também, detectar amostras anômalas (*outliers*) (SOUZA & POPPI, 2012).

Na PCA a matriz X de dados originais é decomposta em um produto de duas matrizes, denominadas escores T e pesos P , mais uma matriz de erros E , conforme a Equação 2:

$$X = TP^T + E. \quad (2)$$

Assim os dados originais são agrupados em função das correlações existentes entre as variáveis através de combinações lineares, gerando um novo conjunto de eixos (componentes principais), onde a primeira componente principal (PC1) representa a maior variação do conjunto de dados e a segunda (PC2), que é ortogonal à primeira, descreve a maior percentagem da variação não explicada pela PC1 e assim por diante (SOUZA & POPPI, 2012).

Os escores estão relacionados com a posição ocupada pelas amostras nos novos eixos, e a informação do quanto cada variável original contribui para a formação de cada novo eixo está contida nos pesos. As relações entre as amostras são expressas pelos escores enquanto os pesos mostram a relações entre as variáveis originais (MARTENS & NAES, 1989).

3.8.2 Análise de Agrupamentos Hierárquicos (HCA)

A Análise de Agrupamentos Hierárquicos (HCA, do inglês *Hierarchical Cluster Analysis*) é uma técnica de análise exploratória aglomerativa não supervisionada, pois não requer conhecimento prévio das amostras. Esse método permite o agrupamento das amostras em grupos (*clusters*) e, além disso, examina as distâncias interpontuais entre todas as amostras do conjunto de dados com base na similaridade das amostras de um mesmo grupo e nas diferenças entre os membros de grupos diferentes e representa a informação na forma de gráfico bidimensional chamado dendrograma (PARREIRA, 2003; CORREA & FERREIRA, 2007).

Os agrupamentos e similaridades entre as amostras e/ou variáveis podem ser visualizados por meio do dendrograma. A construção do dendrograma é feita com base na proximidade existente entre todas as amostras (agrupamentos) do conjunto, em pares e então definindo uma matriz de similaridade cujos elementos são chamados índices de similaridade, que variam entre zero e um. Um índice alto indica uma distância pequena entre dois agrupamentos e, assim, uma alta similaridade (PARREIRA, 2003).

O cálculo da distância entre dois vetores (amostras), d_{ab} , no espaço m -dimensional (variáveis), conhecida como distância euclidiana, é feito conforme a Equação 3:

$$d_{ab} = \sqrt{\sum_{i=1}^m (x_{a_i} - x_{b_i})^2} \quad (3)$$

sendo que, d_{ab} corresponde à distância entre os pontos a e b para um par de amostras no conjunto de dados. Depois de realizado o cálculo das distâncias entre todos os pares de amostras, agrupamentos sucessivos, de acordo com suas similaridades, são formados até não haver mais similaridade entre as amostras (BRERETON, 2007).

3.9 Regressão Multivariada

3.9.1 Regressão por Mínimos Quadrados Parciais (PLS)

Os métodos de regressão são utilizados para analisar a relação entre uma variável dependente (concentração ou propriedade de interesse) e um conjunto de variáveis independentes (sinais instrumentais), sendo empregadas para quantificar propriedades de interesse de um conjunto de amostras (BRERETON, 2007).

Entre os métodos de regressão, destaca-se o método por Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*), sendo o mais utilizado da atualidade para análise quantitativa (HUCK, 2014). Esse método não

requer um conhecimento exato de todos os componentes presentes nas amostras, podendo realizar a previsão de parâmetros de interesse mesmo na presença de interferentes, desde que esses estejam presentes na construção do modelo.

O método PLS possui uma grande vantagem que é a robustez; essa permite que os parâmetros do modelo não se alterem de maneira significativa quando novas amostras são adicionadas ou subtraídas do conjunto de calibração (GELADI & KOWALSKI, 1986).

O PLS está na decomposição de uma matriz de dados X (variáveis independentes) e Y (variáveis dependentes), simultaneamente, em uma soma de produtos de dois vetores: t (escores) e p (pesos) (Equações 4 e 5). A esses vetores, soma-se, ainda, mais uma matriz de erros E que corresponde à parte não modelada da matriz X . Na regressão por PLS, tanto a matriz de dados X como a matriz de propriedades Y é decomposta e projetada em um novo sistema de coordenadas (BRERETON, 2003; FERRÃO *et al.*, 2004).

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_h p_h^T + E \quad \text{ou} \quad X = TP^T + E = \sum t_h p_h^T + E \quad (4)$$

$$Y = UQ^T + F = \sum u_h q_h^T + F \quad (5)$$

onde, T e U são as matrizes de escores das matrizes X e Y , respectivamente; P e Q são as matrizes de pesos das matrizes X e Y , respectivamente; h corresponde ao número de variáveis latentes e E e F correspondem às matrizes de resíduos.

A relação interna dos escores das matrizes X e Y é obtida através do coeficiente de regressão linear para cada componente principal b_h de acordo com a Equação 6:

$$u_h = b_h t_h \quad (6)$$

sendo b_h obtido através da Equação 7:

$$b_h = \frac{u_h^T t_h}{t_h^T t_h} \quad (7)$$

onde, u e t são os elementos das matrizes U e T , respectivamente.

Os valores de b_h são agrupados na matriz diagonal B (matriz identidade), que contém os coeficientes de regressão entre a matriz de escores U de Y e a matriz de escores T de X . Entretanto, como cada matriz é decomposta separadamente, pode ocorrer uma relação não linear entre os escores dos dois blocos. A melhor relação linear possível entre os escores das matrizes é obtida através de pequenas rotações das variáveis latentes das matrizes X e Y , devendo existir um compromisso entre a capacidade de descrever as amostras individuais (modelagem dos blocos X e Y) e o aumento da correlação entre T e U (GELADI; MACDOUGALL & MARTENS, 1985).

A matriz Y pode ser calculada por meio das informações contidas em u_h (6), conforme a Equação 8:

$$Y = TBQ^T + F \quad (8)$$

onde, T são os escores da matriz X , B é a matriz identidade de b_h , Q^T são os pesos da matriz Y e F é a matriz residual de Y . Dessa forma, a concentração das novas amostras pode ser prevista a partir dos novos escores de X , dado por T^* , substituído na Equação 9:

$$Y = T^*BQ^T. \quad (9)$$

Para se obter o melhor modelo referente aos dados sob investigação devem ser avaliados o número de variáveis latentes (VLs) e a Raiz Quadrada dos Erros Médios (*RMSE*, do inglês *Root Mean Square Error*). O PLS usa o termo de variáveis latentes para designar as componentes principais; isso se deve ao fato da construção das mesmas ser feita a partir de informações contidas no vetor das variáveis dependentes. Durante o processo de validação cruzada é determinado o número de variáveis latentes que será utilizado (PARREIRA, 2003).

3.9.2 Métodos de Seleção de Variáveis em Modelos de Regressão

A calibração multivariada no PLS pode ser realizada empregando a informação de toda a faixa espectral de trabalho para construir um modelo de regressão correlacionando com a propriedade de interesse. Dessa forma, quando se utiliza toda a faixa espectral, a grande quantidade de variáveis pode interferir na modelagem e também deixar os tratamentos dos dados mais lentos. Para que isso não ocorra, tem sido utilizados nas técnicas de calibração multivariada, procedimentos apropriados para seleção das regiões espectrais associadas (FERREIRA, 2015).

O objetivo dos métodos de seleção de variáveis é escolher determinadas regiões do espectro que permitam ao modelo de regressão diminuir o número de variáveis e assim minimizar os erros de previsão. Dessa forma, é importante observar e selecionar apenas as regiões que tenham correlação com a propriedade de interesse. A vantagem desse método é que permite produzir um modelo mais robusto, simples de interpretar e com menores erros de previsões (OLIVEIRA *et al.*, 2004).

Atualmente os métodos mais utilizados para esse tipo de seleção de variáveis são os Mínimos Quadrados Parciais por Intervalo (iPLS, do inglês *Interval Partial Least Squares*) e os Mínimos Quadrados Parciais por Sinergismo de Intervalos (siPLS, do inglês *Synergy Interval Partial Least Squares*).

O iPLS é uma extensão desenvolvida para o PLS, onde é feita uma regressão por mínimos quadrados parciais em cada intervalo equidistante ao longo de toda a extensão do espectro. Desse modo, é avaliada a importância da informação nas diferentes subdivisões espectrais, de onde é possível identificar e selecionar o intervalo cujas variáveis exibem informações mais importantes. Um modelo de PLS é construído para cada intervalo, sendo apresentados os resultados na forma de gráficos, dessa forma, facilitando a comparação com toda a faixa espectral (LEARDI & NORGAARD, 2004; NORGAARD, *et al.*, 2000).

O siPLS é uma extensão do algoritmo iPLS. Esse algoritmo consiste na divisão do espectro em regiões equidistantes (intervalos) seguido da combinação

desses intervalos, combinando 2 a 2, 3 a 3 até 4 a 4 subregiões do espectro, possibilitando a obtenção de modelos que podem apresentar melhores coeficientes de correlação (R) e menores erros de calibração e previsão, que aqueles encontrados por iPLS (ZOU *et al.*, 2010).

3.9.3 Seleção do Conjunto de Calibração e Previsão

Para o conjunto de calibração e previsão, a seleção das amostras que irá compor este conjunto é feita através do algoritmo *Kennard-Stone* (KS). Este algoritmo começa selecionando as duas amostras com a maior distância euclidiana entre si no espaço X . Para cada uma das amostras restantes, calcula-se a distância mínima em relação às amostras já selecionadas. Feito isso, a amostra que será escolhida é a que apresentar a maior distância em relação à amostra média. A segunda amostra a ser selecionada será a que apresentar maior distância em relação à primeira amostra escolhida. A próxima amostra a ser selecionada apresentará maior distância em relação à última amostra separada e o procedimento é repetido até que um determinado número de amostras seja selecionado (HONORATO *et al.*, 2007; DASZYKOWSKI; WALCZAK & MASSART, 2002).

3.9.4 Tratamentos e Pré-Processamentos dos Dados

Os dados experimentais devem, se necessário, ser pré-tratados antes da análise quimiométrica. Esse é um procedimento muito importante em qualquer análise e, em geral, vários métodos são testados para garantir que o pré-tratamento mais adequado seja utilizado (FERREIRA, 2015).

O pré-tratamento tem como objetivo reduzir matematicamente as variações indesejáveis que não foram removidas durante a aquisição dos dados e que não são eliminadas naturalmente durante a análise, mas que podem influenciar os resultados finais. O tratamento dos dados é aplicado na fase de calibração para melhorar a eficiência do método, produzindo um modelo de regressão com maior

capacidade de previsão das amostras; também têm por objetivo remover variações sistemáticas não desejadas do espectro, como mudanças na linha de base, efeitos de espalhamento e fatores externos não controláveis (FERREIRA, 2015).

Para uma análise adequada, as transformações dos dados espectrais são necessárias, consistindo em tratamentos e pré-processamentos destes dados. O pré-processamento é aplicado quando se deseja comparar variáveis com diferentes dimensões e consistem, basicamente, em centrar os dados na média ou autoescalar os dados (FERREIRA, 2015).

Centragem dos dados na média é um dos pré-processamentos mais comuns. Para centrar os dados na média, calcula-se primeiro o valor médio de cada coluna da matriz de dados e, a seguir, esse valor é subtraído de cada um dos valores da respectiva coluna. Uma vez coletados, os dados são organizados em linhas e colunas, constituindo uma matriz X , em que cada linha corresponde a uma amostra i , como mostrado na equação 10, na qual $i = 1, 2, 3, \dots, I$ e $j = 1, 2, 3, \dots, J$:

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_I^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1J} \\ x_{21} & x_{22} & \dots & x_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \dots & x_{IJ} \end{bmatrix} = [x_1 \quad x_2 \cdots \cdots x_J]. \quad (10)$$

Cada amostra i é associada ao respectivo vetor-linha, x_i^T , cujos elementos são os J valores das medidas feitas para descrever suas propriedades. Esse vetor pode ser associado a um ponto no espaço de dimensão J , ou seja:

$$x_i^T = [x_{i1} \quad x_{i2} \quad x_{i3} \cdots x_{iJ}].$$

Cada coluna x_j da matriz X se refere a uma variável, ou seja, a uma medida j realizada para todas as amostras e que também pode ser associada a um vetor ou a um ponto no espaço de dimensão I , em que:

$$x_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Ij} \end{bmatrix}.$$

Portanto, uma representação compacta da matriz de dados é dada por $X (I \times J)$, com um total de I linhas (amostras) e J colunas (variáveis), cujo elemento x_{ij} é a resposta para a variável j referente à amostra i . Assim para fazer o cálculo para centrar os dados na média utiliza-se a seguinte expressão:

$$x_{ij(cm)} = x_{ij} - \bar{x}_j$$

onde,

$$\bar{x}_j = \frac{1}{I} \sum_{i=1}^I x_{ij}$$

é a média da j -ésima coluna dos dados.

Existem várias técnicas matemáticas comuns que podem ser aplicadas com o objetivo de reduzir tanto as variações aleatórias quanto as informações sistemáticas indesejáveis. A componente aleatória (ruído experimental) pode ser reduzida por meio das técnicas de alisamento (do inglês *Smoothing*) (FERREIRA, 2015).

O alisamento tem grande aplicação na área da espectroscopia para a redução da componente aleatória dos dados e o consequente aumento da razão sinal ruído do sinal analítico. Também, o alisamento pode ser utilizado para destacar pequenos picos que estão escondidos numa linha de base onde o ruído é alto e, ainda, como um método de compressão de dados (FERREIRA, 2015).

As técnicas de alisamento mais empregadas consideram um pequeno segmento ou uma “janela” do espectro que se desloca percorrendo-o totalmente. Inicia-se o procedimento definindo uma faixa com um número ímpar de comprimentos de onda adjacentes (uma janela) em que as intensidades, para

todos eles, são usadas para determinar uma única resposta, que será atribuída ao centro da janela. Em seguida, a janela se move e o processo é repetido, percorrendo, dessa forma, todo o espectro (FERREIRA, 2015).

O alisamento pelo método de Savitzky-Golay é um dos pré-processamentos mais utilizados. O processo envolve a seleção e o deslocamento de uma janela móvel ao longo de todo espectro. Uma janela de tamanho $2m + 1$ contém $2m + 1$ variáveis designadas por $-m, -m + 1, \dots, -1, 0, 1, \dots, m - 1, m$ e cujo centro se localiza no zero. A resposta alisada, x_{alis} , e que será atribuída ao centro da janela, é uma média das respostas das $2m + 1$ variáveis da respectiva janela, como representa a seguinte equação:

$$x_{alis} = \sum_{k=-m}^{+m} c_k x_k = \frac{1}{(2m + 1)} \sum_{k=-m}^{+m} x_k$$

já que $c_k = \frac{1}{(2m+1)}$.

O alisamento por Savitzky-Golay utiliza uma média ponderada, em que os pesos são obtidos por um ajuste polinomial. Para isso, basta ajustar um polinômio de grau n às respostas da primeira janela, que contém $2m + 1$ pontos, indicados aqui como $k = -m, -m + 1, \dots, -1, 0, 1, \dots, m - 1, m$, onde $k = 0$ representa o seu centro. A seguir, a resposta no ponto central da janela é substituída pelo valor do polinômio ajustado. A janela se move, e o valor do polinômio é calculado no novo centro. O processo é repetido para a janela seguinte, onde a janela se desloca de uma em uma variável. A cada deslocamento, apenas a intensidade à esquerda é excluída e uma nova é incluída à direita, e assim por diante até que todo o espectro tenha sido alisado. Esta metodologia não permite o alisamento no início e no final do espectro, pois as respostas alisadas são alocadas no centro das janelas (FERREIRA, 2015).

Dois parâmetros são importantes nesse método: o número de pontos da janela (m à esquerda e m à direita do centro) e o grau do polinômio ajustado, n . Ambos podem afetar o resultado do alisamento. Em um polinômio de baixo grau,

filtra-se mais ruído, mas por outro lado, o risco de distorção no sinal é maior. Já um polinômio de alto grau tende a se ajustar melhor aos dados e a descrever melhor a variação do sinal, mas com certeza modela o ruído. O tamanho da janela deve ser selecionado cuidadosamente para que não haja perda de informação, pois uma janela muito grande pode causar a perda de resolução no espectro alisado, ocorrendo assim um superalisamento. Para evitar o mesmo deve-se iniciar com uma janela menor e realizar o aumento gradativamente, mantendo a tensão na forma e abertura do pico estreito do espectro (FERREIRA, 2015).

Outro pré-processamento muito empregado é a normalização. Nessa transformação os valores de cada uma das variáveis de uma dada amostra i são divididos por um fator de normalização, como, por exemplo, pela norma dessa amostra, representada por $\|x_i\|$. Dessa forma, todas as amostras estarão em uma escala pré-determinada. A expressão abaixo representa a transformação de normalização feita em cada elemento de uma linha da matriz de dados:

$$x_{ij(norm)} = \frac{x_{ij}}{\|x_i\|} \quad j = 1, 2, \dots, J.$$

Existem vários tipos de normas, neste trabalho foi utilizada a seguinte norma:

$$\|x_i\|_\infty = \max_{1 \leq j \leq J} |x_{ij}|,$$

denominada norma *sup* ou norma infinita, ι_∞ .

Quando os dados são normalizados pela norma *sup*, a resposta máxima de cada uma das amostras se torna igual a 1; após a normalização, o espectro terá comprimento igual a 1 (FERREIRA, 2015).

3.9.5 Avaliação dos Modelos de Regressão

Na quimiometria, os dados consistem principalmente de informação e ruído e esta ferramenta pode ser utilizada para extrair informações importantes sobre os dados, criando um modelo matemático, o qual sempre deve ser validado antes de sua aplicação (GEMPERLINE, 2006).

Os métodos de calibração multivariada utilizam diversas estratégias para determinar o nível do erro de predição de um modelo. Para que um modelo seja satisfatório, este deve explicar o máximo da variância dos dados com um mínimo de Variáveis Latentes (VLs), minimizando a influência do ruído para obter um modelo de fácil interpretação (GEMPERLINE, 2006).

Para descrever um modelo é necessário um número ótimo de VLs; o modelo será o número de VLs para qual um erro mínimo da previsão seja adquirido (BRERETON, 2003).

Assim, um modelo resultará em subajuste quando um número de VLs for inferior ao ideal; dessa forma não irá considerar a totalidade da informação contida no conjunto de dados. Também, o subajuste ocorre com mais frequência quando se tem uma quantidade maior de VLs, portanto, aumentará o ruído e os erros do modelo resultante. Essa maior quantidade de VLs origina resultados satisfatórios para a previsão da propriedade de amostras do conjunto de calibração, contudo, não adequados para a previsão de amostras de um conjunto de validação ou previsão devido à informação não necessária contida no modelo (GEMPERLINE, 2006; GOMEZ-CARRACEDO, 2007).

Por meio do número de VLs pode ser feito o cálculo da avaliação da Raiz Quadrada dos Erros Médios (RMSE, do inglês *Root Mean Square Error*), que é utilizada para avaliação dos modelos construídos. Este erro pode ser estimado através dos conjuntos: Raiz Quadrada do Erro Médio Quadrático de Calibração (RMSEC, do inglês *Root Mean Square Error of Calibration*), Raiz Quadrada do Erro Médio de Predição (RMSEP, do inglês *Root Mean Square Error of Prediction*) e Raiz Quadrada do Erro Médio de Validação Cruzada (RMSECV, do inglês *Root Mean Square Error of Cross Validation*), sendo esse último erro

utilizado para se determinar o número de VLs do modelo (BRERETON, 2000; EASTMENT, 1982). O cálculo da RMSE é dado pela Equação 11:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (11)$$

onde, y_i é o valor de referência para a i -ésima amostra e \hat{y}_i é o valor de previsão para esta mesma amostra, sendo n o número total de amostras utilizadas para a construção do modelo.

A validação cruzada está baseada na avaliação da magnitude do erro de predição de um dado modelo de calibração e consiste na remoção de uma ou um bloco de amostras do conjunto de calibração e construção do modelo sem as mesmas. As amostras removidas são previstas no modelo e calcula-se o erro de predição. Esse procedimento ocorre até que todas as amostras do modelo sejam retiradas e previstas. Este erro é denominado de Erro Quadrático Médio de Calibração Cruzada (RMSECV do inglês *Root Mean Square Error of Cross Validation*) (BRERETON, 2003).

Através do cálculo da RMSEP, conforme a Equação 11, é realizada a avaliação da capacidade de predição do modelo de calibração para as amostras de predição.

3.10 Máquinas de Vetores de Suporte com Mínimos Quadrados (LS-SVM)

As Máquinas de Vetor de Suporte (SVM, do inglês *Support Vector Machine*) são novas técnicas quimiométricas aplicadas em calibração multivariada e em problemas de reconhecimento de padrões e classificação de dados. A SVM, em alguns casos, possui um tempo computacional elevado e sua implementação não é muito simples. Recentemente, foi sugerida uma extensão da teoria das SVM utilizando método dos mínimos quadrados (LS-SVM). Este tem como objetivo a soma do erro médio quadrático; assim é possível manter as

características do SVM original e diminuir os custos computacionais para a realização das operações do método, além de possuir elevada capacidade de generalização (FERRÃO *et al.*, 2007; SUYKENS, 2002; GILSBERTS, 2009).

As LS-SVM são análogas às redes neurais de avanço, no sentido em que processam a informação de modo sequencial por meio de camadas e podemos usar funções de base análogas às funções de transferência utilizadas nas redes neurais (FERRÃO *et al.*, 2007).

A Figura 6 representa a LS-SVM como uma rede neural, onde a informação que entra na rede neural são as absorvâncias lidas em diversos comprimentos de onda do espectro no infravermelho, logo em seguida processadas pelos vetores suporte, que são obtidos pelo método de Lagrange. Quando necessário, as funções são linearizadas e, depois disso, aplicam-se a função do núcleo e os pesos, que são os multiplicadores de Lagrange. Assim, temos a somatória e a função y prevista é comparada com o valor real \hat{y} . Na maior parte dos casos é necessário otimizar os valores de σ^2 e γ e, dessa forma, propagar o sinal pela rede neural até que se alcance a convergência desejada para os dados de treinamento (FERRÃO *et al.*, 2007).

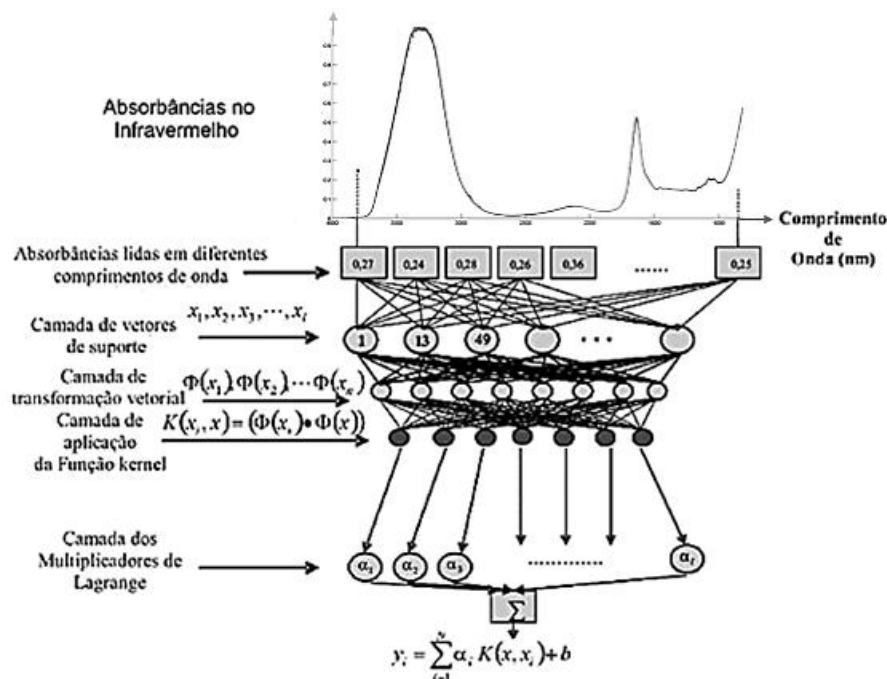


Figura 6: Representação em forma de rede neural para as LS-SVM empregando dados espectroscópicos (Adaptado: FERRÃO *et al.*, 2007).

Os problemas de calibração multivariada lineares e não lineares podem ser tratados rapidamente utilizando as LS-SVM (FERRÃO *et al.*, 2007). A SVM aproxima a relação entre a saída e a entrada utilizando a equação:

$$\gamma = \omega * \Phi(x) + b \quad (12)$$

onde, b é um limite escalar (*threshold*), ω um coeficiente de ponderação e $\Phi(x)$ uma não linearidade mapeada a partir da entrada. Os coeficientes ω e b serão estimados pelo algoritmo de otimização tendo como função objetiva a minimização da função de risco J , dada por:

Minimizar

$$J = \frac{1}{2} \|W\|^2 + \gamma \frac{1}{2} \sum_{i=1}^N \epsilon(y_i, f(x_i)) \quad (13)$$

tal que

$$\epsilon(y_i, f(x_i)) = \begin{cases} 0, & \|y_i, f(x_i)\| < \epsilon \\ \|y_i, f(x_i) - \epsilon\|, & \text{outros} \end{cases} \quad (14)$$

onde, w é o vetor de ponderações e γ o parâmetro de regularização que estabelece um equilíbrio entre a complexidade e o erro de treinamento do modelo.

A normalização dos pesos é feita pela primeira parte da Equação 13, convergindo para valores menores. Este procedimento é adotado para reduzir a variação imposta no modelo por pesos demasiadamente grandes o que deteriora a capacidade de generalização do modelo acentuando o problema de sobreajuste da função aos dados de entrada. A segunda parte da Equação 13 representa os erros de regressão para o conjunto de dados de treinamento. A restrição de igualdade imposta pela Equação 14 fornece a definição do erro de regressão. Quando aplicado a padrões não linearmente separáveis, são adicionadas variáveis de

folga, ξ_i e ξ_i^* sendo então possível transformar a Equação 14 em uma função objetivo primal dada por:

Minimizar

$$J = \frac{1}{2} \|W\|^2 + \gamma \frac{1}{2} \sum_{i=1}^N \epsilon(\xi_i + \xi_i^*) \quad (15)$$

sujeito a

$$y_i - W * \phi(x_i) - b \leq \epsilon + \xi_i \quad (16)$$

$$W * \phi(x_i) + b - y_i \leq \epsilon + \xi_i^* \quad (17)$$

onde, $i = 1, \dots, N$ e $\xi_i, \xi_i^* \geq 0$.

Introduzindo-se os vetores de multiplicadores de Lagrange α_i e α_i^* , denominados vetores de suporte, a função núcleo e maximizando-se a função dual da Equação 15, a função de regressão dada pela Equação 12 apresenta a seguinte explícita:

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) * K(x_i, x_j) + b \quad (18)$$

onde, $K(x_i, x_j)$ é a função núcleo. Os vetores α_i são obtidos por meio da solução do sistema linear de equações, seguindo as condições de Karush- Kuhn- Tucker. O valor de $K(x_i, x_j)$ determina o produto interno de dois vetores x_i e x_j no espaço característico, $\phi(x_i)$ e $\phi(x_j)$, assim $K(x_i, x_j) = \phi^T(x_i)\phi(x_j)$. Utilizar a função núcleo tem como objetivo o cálculo de $\phi(x_i)$ e $\phi(x_j)$, que apresenta uma complexidade muito alta, de forma aproximada e mais simples. Esta função núcleo gera um mapeamento entre o espaço de entrada e o espaço de alta dimensionalidade, dito característico.

4 REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, T. V. **Detecção da adulteração em leite: análises de rotina e espectroscopia de infravermelho**. 2013. 23 f. Seminários Aplicados (Mestrado). (Ciência Animal). Escola de Veterinária e Zootecnia da Universidade Federal de Goiás, Goiás.

BERRY, E. C.; BULLERMAN, L. B. Use of cheese whey for vitamin B₁₂ production II. Cobalt, precursor and aeration limits. **Appl Microbiol**, v. 14, n. 3, p. 356-357, 1996.

BOGDANOV, B.; SMITH, R. D. Proteomics by FTICR mass spectrometry: Top down and bottom up. **Mass. Spectrom. Rev**, v. 24, n. 1, p. 168-200, 2005.

BURNS, D. A.; CIURCZAK, E. W. **Handbook of near-infrared analysis**. 3 Ed. Boca Raton: CRC, 2008. p. 808.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Instrução Normativa Nº 42, de 20 de dezembro de 1999 - ANEXO I. *Altera o Plano Nacional de Controle de Resíduos em Produtos de Origem Animal - PNCR e os Programas de Controle de Resíduos em Carne - PCRC, Mel – PCRM, Leite – PCRL e Pescado – PCRP*. **Diário Oficial da União**, Brasília, 22 de dezembro de 1999.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Instrução Normativa Nº 07, de 2 de março de 2010. ANEXO I. *Aprovar o Método Oficial de Determinação de CMP (caseinomacropéptido) em leite, por HPLC, Eletroforese Capilar e Espectrometria de Massas em leite, em apresentações integrais, semidesnatadas e desnatadas, tratados por processos de UHT ou pasteurização*. **Diário Oficial da União**, Brasília, 03 de março 2010. Seção 1, p. 1-3.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Instrução Normativa Nº 69, de 13 de dezembro de 2006 – ANEXO I. *Institui critério da qualidade do leite in natura, concentrado e em pó, reconstituídos, com base no método analítico oficial físico-químico denominado “índice de CMP”, de que trata Instrução normativa nº68, de 12 de dezembro de 2006*. **Diário Oficial da União**, Brasília, 15 de dezembro de 2006b.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Instrução Normativa Nº 11, de 22 de maio de 2012 – ANEXO I. *Aprova o Escopo Analítico para o Monitoramento dos Produtos de Origem Animal*. **Diário Oficial da União**, Brasília, 2012.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento (2013). “Operação descobre fraude em transportadora de leite no RS”. Disponível em: <[http://www.agricultura.gov.br/comunicacao/noticias/2013/11/operacao-](http://www.agricultura.gov.br/comunicacao/noticias/2013/11/operacao)

REFERÊNCIAS BIBLIOGRÁFICAS

descobre-fraude-em-transportadora-de-leite-no-rs>. Acesso em: 23 de maio 2014.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. Instrução Normativa nº68, de 12 de dezembro de 2006. ANEXO I. *Institui Métodos Analíticos Oficiais Físico-químicos para controle de leite e Produtos Lácteos*. **Diário Oficial da União**, Brasília, 12 de dezembro de 2006a.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Instrução Normativa N° 62, de 29 de dezembro de 2011. ANEXO II. *Institui regulamentos técnicos de produção, Identidade e Qualidade do Leite tipo A, do Leite tipo B, Leite C, do leite Pasteurizado e do Leite Cru Refrigerado e o Regulamento Técnico da Coleta de Leite Cru Refrigerado e seu Transporte a Granel*. **Diário Oficial da União**, Brasília, 29 de dezembro 2011.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regulamento de Inspeção Industrial e Sanitária de Produtos Origem Animal - RIISPOA. Decreto nº 30.691, 29 março de 1952. Alterado pelos Decretos nº 1236 de 02 de setembro de 1994, nº 1812 de 08 de fevereiro de 1996, nº 2244 de 04 junho de 1997, nº 6385 de 27 fevereiro de 2008.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regulamento de Inspeção Industrial e Sanitária de Produtos Origem Animal - RIISPOA. Decreto nº 30.691, 29 março de 1952. Art. 14 da Lei nº1.283, de 18 de dezembro de 1950.

BRANDÃO, S.C.C. O futuro da qualidade do leite brasileiro. **Indústria de Laticínios**, São Paulo, v. 5, n. 28, p. 68-71, 2000.

BRERETON, R. G. **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. England: Chichester. v. 8, n. 471, p. 1-489, 2013.

BRERETON, R. G. **Applied Chemometrics for Scientists**; John Wiley & Sons Ltd: Chichester, UK, 2007.

BRERETON, R. G. **Chemometrics: data analysis for the laboratory and chemical plant**. John Wiley & Sons Ltd: Chichester, UK, 2003.

BRERETON, R. G. **Introduction to multivariate calibration in analytical chemistry**. *Analyst*, v. 125, p. 2125-2154, 2000.

CARVALHO, M. E.; CARVALHO, L. M.; ALCÂNTRA, L. A. P.; BONOMO, R. C. F. Métodos de detecção de fraude em leite por adição de soro de queijo. **Revista Eletrônica de Veterinária**, v. 8, n. 6, p. 7, 2007.

CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends. Food Science and Technology**, v. 18, p. 72-83, 2007.

CORREIA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticas. **Química Nova**, v. 30, p. 481-487, 2007.

DASZYKOWSKI, M.; WALCZAK, B. & MASSART, D. L. Representative subset selection. **Analytica Chimica Acta**, v. 468, n. 1, p. 91-103, 2002.

DESMARCHELIER, A.; GUILLAMON CUADRA, M.; DELATOUR, T.; MOTTIER, P. Simultaneous quantitative determination of melamine and cyanuric acid in cow's milk and milk-based infant formula by liquid chromatography-electrospray ionization tandem mass spectrometry. **Journal of Agricultural and Food Chemistry**, v. 57, n. 16, p. 7186-93, 2009.

DÜRR, J.W. Programa Nacional de Melhoria da Qualidade do leite: uma oportunidade única. In: DÜRR, J.W.; CARVALHO, M.P.; SANTOS, M.V. **O compromisso com a qualidade do leite no Brasil**. Passo Fundo: Editora Passo Fundo, 2004. p. 38-55.

EASTMENT, H. T. & KRZANOWSKI, W. J. Cross-validatory choice of the number of components from a principal components analysis. **Technometrics**, v. 24, n. 1, p. 73-77, 1982.

ENTRIES, N. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC- IUBMB). **Eur. J. Biochem**, v. 233. p. 139. 2005.

ESPER, K. C. P. SILVA, B. C. SÁ, O. R. Avaliação da qualidade do leite através de análises físicas, químicas, microbiológicas e pesquisa de fraudes frente as Normativas 51/2002 e 62/2011. **X Fórum Ambiental da Alta Paulista**, v. 10, n. 7, p. 127-141, 2014.

FARRELL, H.M.; JIMENEZ-FLORES, R.; BLECK, G.T.; BROWN, E.M.; BUTLER, J.E.; CREAMER, L.K.; HICKS, C.L.; HOLLAR, C.M.; NG-KWAIHANG, K.F.; SWAISGOOD, H.E. Nomenclature of the proteins of cows-milk sixth revision. **Journal of Dairy Science, Champaign**, v. 87, p. 1641-1674, 2004.

FERRÃO, M. F.; CARVALHO, C. W.; MULLER, E. I.; DAVANZO, C. U. Determinação simultânea dos teores de cinza e proteína em farinha de trigo empregando NIR-PLS e DRIFT-PLS. **Ciência e Tecnologia de Alimentos**, v. 24, n. 3, p. 333-340, 2004.

REFERÊNCIAS BIBLIOGRÁFICAS

FERRÃO, M. F.; MELLO, C.; BORIN, A.; MARETTO, D. A. & POPPI, R. J. LS-SVM: uma nova ferramenta quimiométrica para regressão multivariada. Comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR. **Química Nova**, v. 30, n. 4, p. 852-859, 2007.

FERRÃO, M. F.; VIERA, M. S. PAZOS, R. E. P.; FACHINI, D.; GERBASE, A. E. MARDER, L. Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. **Elsevier**. v. 90, n. 2, p. 701-706, 2011.

FERREIRA, M. C.; ANTUNES, A. M., MELGO, M. S., VOLPE, P. L. O. Quimiometria I: Calibração Multivariada, um Tutorial. **Química Nova**, v. 22, n. 5, p. 724-731, 1999.

FERREIRA, M. C.; KIRALJ, R. The past, present and future of chemometrics worldwide: some etymological, linguistic, and bibliometric investigations. **Journal of Chemometrics**, v. 20, p. 247-272, 2007.

FERREIRA, M. M. C. Multivariate QSAR. **Journal of the Brazilian Chemical Society**, v. 13, n. 6, p. 742-753, 2002.

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas, SP: Editora: Unicamp. 2015.

FUKUDA, S. P. **Estudo da correlação entre o método da ninidrina ácida e a cromatografia líquida de alta eficiência para a dosagem de glicomacropéptido e caseinomacropéptido em leite**. 2003. 149 f. Tese (Doutorado em Tecnologia de Alimentos) – Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Campinas.

FRIEDRICH, M. T.; FRANKEN, R. B. C.; AZEVEDO, M. S.; PRESTA, M. A.; AGNOL. C. D. Avaliação da estabilidade do leite *in natura* e UHT quanto ao índice de CMP. **Revista de Ciências Exatas Aplicadas e Tecnológicas**, n 1, v. 2, p. 21-27, 2010.

GELADI, P.; MACDOUGALL, D.; MARTENS, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. **Applied Spectroscopy**, v. 39, p. 491-500, 1985.

GELADI, P.; KOWALSKI, B. R. Partial least squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, n. 1, p. 1-17, 1986.

GEMPERLINE, P. J. **Practical guide to chemometrics**. 2ªEd., Nova York: CRC Press Taylor & Francis. 2006, p. 140-145.

GOMEZ-CARRACEDO, M. P.; ANDRADE, J. M.; RUTLEDGE, D. N.; FABER, N. M.; Selecting the optimum number of partial least squares components for the calibration of attenuated total reflectance-mid-infrared spectra of undersigned. kerosene samples. **Analytica Chimica Acta**, v. 585, p. 253-265, 2007.

GONZALÉZ-SISO, M. I. The biotechnological utilization of cheese whey: a review. **Bioresource Technology**, v. 57, n. 1, p. 1-11, 1996.

GUAN, R. LIU, D.; YE, X.; YANG, K. Use of fluometry for determination os skim milh powder adulteration in fresh milk. **Journal of Zhejiang University**, v. 6, n. 11, p. 1101-1106, 2005.

GUO, Z.; GAI, P.; HAO, T. WANG, S.; WEI, D.; GAN, N. Determination of melamine in dairy products by an electrochemiluminescent method combined with solid-phase extraction. **Talanta**, v. 83, n. 5, p. 1736-41, 2011.

HONORATO, F. A; NETO, B. B. MARTINS, M. N.; GALVÃO, R. K. H.; PIMENTEL, M. F. Transferência de calibração em métodos multivariados. **Química Nova**, v. 30, n. 5, p. 1301-1312, 2007.

HUCK, C. W. Advances of vibrational spectroscopic methods in phytomics and bioanalysis. **Journal of Pharmaceutical and Biomedical Analysis**, v, 87, p. 26-35, 2014.

KAROUI, R.; DEBAERDEMAEKER, J. A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. **Food Chemistry**, v. 102, n. 3, p. 621-640, 2007.

KLAMPFL, C. W.; ANDERSEN, L.; HAUNSCHMIDT, M.; HIMMELSBACH, M.; BUCHBERGER, W. Analysis of melamine in milk powder by CZE using UV detection and hyphenation with ESI quadrupole/TOF MS detection. **Electrophoresis**, v. 30, n. 10, p. 1743-6, 2009.

KREUFFL, M.; KRAUSE, I.; KULOZIK, U. Separation of a glycosylated and non-glycosylated fraction of caseinomacropetide using different anion-exchange stationary phases. **Journal of Chromatography A**, v. 1208, p. 126-132, 2008.

LEARDI, R.; NORGDARD, L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. **Journal of Chemometrics**, v. 18, n. 11, p. 486-497, 2004.

LÉONIL, J.; MOLLÉ, D. Liberation of tryptic fragments from caseinomacropetide of bovine kappa-casein involved in platelet function: Kinetic study. **Biochemical Journal**, v. 271, n. 1, p. 247, 1990.

REFERÊNCIAS BIBLIOGRÁFICAS

LIRA, B. L. d. F.; VASCONCELOS, F. V. C.; PEREIRA, C. F.; SILVEIRA PAIM, A. P.; STRAGEVITCH, L.; PIMENTEL, M. F. Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. **Fuel** (Guildford), v. 89, n. 1, p. 405-409, 2010.

LÓPEZ, E. I.; RECIO, I. Antibacterial activity of peptides and folding variants from milk proteins. **International Dairy Journal**, v. 16, n. 11, p. 1294-1305, 2006.

LYRA, W. S.; SILVA, E. C.; ARAÚJO, M. C. U.; FRAGOSO, W. D. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. **Química Nova**, v. 33, n. 7, p. 1594-1597, 2010.

MAGALHÃES, M.A. **Determinação de fraude de leite com soro de leite pela análise de CMP e pseudo-CMP por cromatografia líquida de alta eficiência em fase reversa com detecção por espectrometria de massa**. 2008. 74 f. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa, Viçosa.

MARTENS, H.; NAES, T. **Multivariate calibration**. John Wiley & Sons, Nova York. 1989.

MARTIN-ORUE, C.; HENRY, G.; BOUHALLAB, S. Typtic hydrolysis of [kapp]-caseinomacropetide: control of the enzymatic reaction in a continuous membrane reactor. **Enzyme and microbial technology**, v. 24, n. 1, p. 173-180, 1999.

MARQUES, F. M.; SÁ, J. F. O.; SANTOS, M. C.; MARTINS, M. F.; FURTADO, M. A. M. Caracterização de leite em pó, soro de leite em pó e suas misturas por eletroforese em gel de poliacrilamida. **Revista do Instituto Adolfo Lutz**, v. 40, n° 4, p. 473-479, 2011.

MORGANO, M. A.; QUEIROZ, S. C. N.; FERREIRA, M. M. C. Aplicação de análise exploratória na diferenciação de vegetais. **Brazilian Journal of Food Technology**, n. 2, v. 1/2. p. 73-79, 1999.

MOTTA, T. M. C.; HOFF, R. B.; BARRETO, F.; ANDRADE, R. B.; LORENZINI, D. M.; MENEGHINI, L. Z.; PIZZOLATO, T. M. Detection and confirmation of milk adulteration with cheese whey using proteomic-like sample preparation and liquid chromatography-electrospray-tandem mass spectrometry analysis. **Talanta**, n, 1, v, 120, p. 498-505, 2014.

NATALINO, R. **Caracterização de açúcar mascavo aplicando análise das componentes principais a dados espectrométricos**. 2006. 52 f. Tese de doutorado (Magister Scientiae). Universidade Federal de Viçosa, Minas Gerais.

NORGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. **Applied Spectroscopy**, v. 54, p. 413-419, 2000.

OLIVEIRA, G. B.; GATTI, M. D. S.; VALADÃO, R. C.; MARTINS, J. F. P.; LUCHESE, R. H. Detecção da adição fraudulenta de soro de queijo em leite: interferência da atividade de proteases bacterianas. **Revista de instituo laticínios cândido testes**, v. 64, n 367/368, p. 56-65, 2009.

OLIVEIRA, R. R. **Aplicação da espectroscopia de infravermelho próximo para a determinação do caseinomacropéptido em Leite UAT**. 2010. 76 f. Dissertação de mestrado (Ciência Animal). Universidade Federal de Goiás, Goiânia.

OLIVEIRA, F. C. C.; de SOUZA, A.; Dias, J. A.; DIAS, S. C. L.; RUBIM, J. C. **Química Nova**, v 27, p. 218-225, 2004

OLIVIERI, A. C. Analytical advantages of multivariate data processing. One, two, three, infinity. **Analytical Chemistry**, v. 80, n. 15, p. 5713-20, 2008.

PARISOTTO, G. **Determinação do número de acidez total em resíduo de destilação atmosférica e de vácuo do petróleo empregando a espectroscopia no infravermelho (ATR-FTIR) e calibração multivariada**. Dissertação de mestrado (Química). 2007. 102 f. UFSM. Santa Maria, Rio Grande do Sul, Brasil.

PARREIRA, T. F. **Utilização de métodos quimiométricos em dados de natureza multivariada**. 2003. 106 f. Dissertação de mestrado (Química). Universidade Estadual de Campinas, São Paulo.

PASQUINI, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. Campinas-SP: **Journal of the Brazilian Chemical Society**, v. 14, p. 198-219, 2003.

POPPI, F. A.; COSTA, M. R.; DE RENSIS, C.M. V. B., SIVIERI, K. Soro de Leite e Suas Proteínas: Composição e Atividade Funcional. **Revista UNOPAR Científica de Ciências Biológicas e da Saúde**, v. 12, n. 2, p. 31-7, 2010.

QIAN, Z. Y.; JOLLÉZ, P.; MIGLIORE-SAMOUR, D.; SCHOENTGEN, F. ; FIAT, A. M.; Sheep [κ] – casein peptides inhibit platelet aggregation . **Biochimica et Biophysica Acta (BBA)-General subjects**, v. 1244, n. 1, p. 411-417, 1995.

REFERÊNCIAS BIBLIOGRÁFICAS

RECIO, I.; AMIGO, L.; RAMOS, M. & LOPES-FANDIÑO, R. Application of capillary electrophoresis to the study of proteolysis of caseins. **Journal of Dairy Research**. v. 64, n. 2, p. 221-230, 1997.

ROMA JÚNIOR, L. C. Características quantitativa e qualitativas da proteína do leite produzido na região sudeste. 2008. p. 150. Tese (Doutorado em Agronomia), Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

SANTOS, P. M. **Aplicação de imagens digitais e técnicas espectro analíticas combinadas com quimiometria para a detecção e quantificação de adulteração em leite bovino**. 2013. 97 f. Tese de doutorado (Química analítica). Universidade Federal de São Carlos, São Carlos.

SANTOS, L. T.; FILHO, E. L. S.; COELHO, L. S. Identificação e previsão de séries temporais utilizando LS-SVM otimizado pelo algoritmo de cardumes. **Learning and Nonlinear Models**, v. 10, p. 119-126, 2012.

SEAB, Secretária de Estado da Agricultura e do Abastecimento. Departamento de Economia rural (DERAL). **Análise da Conjuntura Agropecuária**. 2014. Disponível em:
<http://www.agricultura.pr.gov.br/arquivos/File/deral/Prognosticos/bovinocultura_leite_14_15.pdf>. Acesso em: 01 de março de 2016.

SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. **Princípios de Análise Instrumental**. 6 Ed., Editora Bookman, Porto Alegre, 2009.

SOUZA, A. M.; POPPI, R. J. Experimentos didáticos de Quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: Um tutorial, parte 1. **Química nova**, v. 35, n.1, p. 223-229, 2012.

SUYKENS, S. A.; GESTEL, T. V.; BRABANTER, J.; MOOR, BART de.; VANDEWALLE, J. **Least squares support vector machines**. World Scientific. 2002.

TULLIO, L. T.; KARKLE, E. N. L.; CÂNDIDO, L. M. B. Review: isolation and purification of milk whey glycomcopeptide. **B. CEPPA**, Curitiba, v. 25, n. 1, p. 121-132, 2007.

VELOSO, A. C. A.; TEIXEIRA, N.; FERREIRA, I. M. P. V. L. O.; FERREIRA, M. A. Detecção de adulterações em produtos alimentares contendo leite e/ou proteínas lácteas. **Química Nova**, v.25, n.4, p.1-13, 2002.

VIANA. L. C. **Uso de espectroscopia no infravermelho próximo para predição de propriedades tecnológicas da madeira de *eucalyptus***. 2008.

Dissertação de mestrado (Engenharia Floresta). Universidade Federal de Lavras. Minas Gerais. Brasil.

VIOTTO, W. H.; CUNHA, C. R. **Teor de sólidos do leite e rendimento industrial. Perspectivas e avanços da qualidade do leite no Brasil**. Goiânia: Talento, 2006, p. 241-258.

VOORT, F. R. V. Fourier transform infrared spectroscopy applied to food analysis. **Food Research International**, v. 25, n.1, p.397-403, 1992.

WORKMAN, J.; LAVINE, B.; CHRISMAN, R.; KOCH, M. Process Analytical Chemistry. **Analytical Chemistry**, v. 83, p. 4557-4578, 2011.

YANG, S.; DING, J.; ZHENG, J.; HU, B.; LI, J.; CHEN, H.; ZHOU, Z. QIAO, X. Detection of Melamine in Milk Products by Surface Desorption Atmospheric Pressure Chemical Ionization Mass Spectrometry. **Analytical Chemistry**, v. 81, n. 5, p. 2426-2436, 2009.

ZHANG, W.-B. Review on analysis of biodiesel with infrared spectroscopy. **Renewable and Sustainable Energy Reviews**, v. 16, p. 6048-6058, 2012.

ZHANG, J.; WU, M.; CHEN, D.; SONG, Z. Journal of Food Composition and Analysis Ultrasensitive determination of melamine in milk products and biological fluids by luminol-hydrogen peroxide chemiluminescence. **Journal of Food Composition and Analysis**, v. 24, n. 7, p. 1038-1042, 2011.

ZHU, L.; GAMEZ, G.; CHEN, H.; CHINGIN, K.; ZENOBI, R. Rapid detection of melamine in untreated milk and wheat gluten by ultrasound-assisted extractive electrospray ionization mass spectrometry (EESI-MS). **Chemical Communications**, n. 5, p. 559-61, 2009.

ZINSLY, P.F.; SGARBIERI, V.C.; PEREIRA DIAS, N.F.G.; JACOBUCCI, H.B.; PACHECO, M.T.B.; BALDINI, V.L.S. Produção piloto de concentrados de proteínas de leite bovino: composição e valor nutritivo. **Brazilian Journal Food Technology**, v.4, n. 1, p. 1-8, 2001.

ZOU, X.; ZHAO, J.; POVEY, M. J. W.; HOLMES, M.; MAO, H. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v. 667, n. 1, p. 14-32, 2010.

5 ARTIGO CIENTÍFICO - Ferramentas Quimiométricas Aplicadas no Desenvolvimento de Metodologia Analítica para Avaliação de Adulteração em Leite Bovino por Adição de Soro do Queijo por Espectroscopia no Infravermelho (FTIR)

5.1 Apresentação

As inspeções realizadas pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA), auxiliam no controle de qualidade do leite, porém a adulteração de leite por adição de soro de queijo é ainda uma prática muito comum, onde existe pouca produção de leite comparado com a grande demanda no Brasil.

Para auxiliar o monitoramento da qualidade do leite foi desenvolvido um método analítico capaz de quantificar o peptídeo marcador da adulteração (CMP) no leite através da Espectroscopia no Infravermelho Médio com Transformada de Fourier com Reflexão Total Atenuada (ATR-FTIR) e regressão multivariada.

Esse método para análise de adulteração com soro de queijo foi desenvolvido, a fim de garantir o controle de qualidade do leite para dos consumidores de forma rápida e com baixo custo.

5.2 Artigo

Este artigo será submetido ao periódico Food Analytical Methods tratando do desenvolvimento de um método analítico para quantificação da adulteração em leite por caseinomacropéptido (CMP) através da Espectroscopia no Infravermelho Médio com Transformada de Fourier com Reflexão Total Atenuada (ATR-FTIR) e regressão multivariada.

Ferramentas Quimiométricas Aplicadas no Desenvolvimento de Metodologia Analítica para Avaliação de Adulteração em Leite Bovino por Adição de Soro do Queijo por Espectroscopia no Infravermelho (FTIR)

Layane Lenardon^{a1}, Ana Maria Bergold^{a2}, Marco Flores Ferrão^{b3}

^aFaculdade de Farmácia, Universidade Federal do Rio Grande do Sul; Av. Ipiranga, 2752, Porto Alegre, RS, Brasil.

^bInstituto de Química, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500. Bairro Agronomia. Porto Alegre, RS, Brasil.

Abstract

Brazilian law forbids the addition of cheese whey in pasteurized milk, sterilized milk and powdered milk. However, adulteration with whey is one of the most applied fraud due to its low cost. The detection of this fraud is the quantification of Caseinmacropeptide (CMP). The CMP is a specific element of the whey that can be used as adulteration marker. In this sense, it was developed an analytical method capable of identifying CMP by Fourier Transform Infrared Spectra (FTIR) associated with chemometrics. Exploratory tools as Principal Components Analysis (PCA) and Hierarchical Cluster Analysis (HCA) indicated a similarity between the samples of raw milk and semi-skimmed milk. Moreover, in the scores of PCA, it was possible to observe a trend of separation between samples with different CMP concentrations. Regression models were performed through Partial Least Squares (PLS), Partial Least Squares with Interval Synergism (siPLS) and Support Vector Machines with Least Squares (LS-SVM). The LS-SVM s4i16M model was considered robust to quantify CMP in the milk sample.

Keywords: milk, adulteration, CMP, whey, FTIR, multivariate regression.

INTRODUÇÃO

Nos últimos anos no Brasil foram descobertas várias fraudes envolvendo adulterações no leite bovino. Uma das fraudes econômicas mais aplicadas ao leite é a adição do soro do queijo. A legislação brasileira proíbe a adição do soro do queijo no leite por reduzir o valor nutritivo do alimento e diluir os teores de gorduras (OLIVEIRA *et al.*, 2009).

A adulteração do leite com a adição soro de queijo é uma das adulterações mais preocupantes, pois, ocorre a substituição de parte do leite por soro de queijo ao usar um derivado do próprio leite. Isso determina dificuldades para detectar a adulteração. A cada 10% de soro adicionado ocorre a redução de 8% no teor de proteínas do leite (OLIVEIRA *et al.*, 2009).

A detecção e quantificação da adição fraudulenta de soro ao leite são realizadas pela determinação de caseinomacropeptídeo (CMP). O CMP é um elemento específico do soro do queijo; dessa forma ele é um marcador para detecção de adulteração no leite. Esse peptídeo é formado a partir da clivagem da κ -caseína pela ação da quimosina (MARQUES *et al.*, 2011; MOTTA *et al.*, 2014).

A adulteração por soro do queijo é uma preocupação para os órgãos de fiscalização, pois trata de uma fraude que prejudica o consumidor e os concorrentes que seguem a legislação vigente, além dessa fraude ser prevista como crime na legislação brasileira. Essa fraude é facilitada pela ausência de métodos rápidos e da logística complexa devido à grande extensão territorial do Brasil (OLIVEIRA *et al.*, 2009; CARVALHO *et al.*, 2007)

A legislação brasileira institui que o leite pasteurizado, o leite esterilizado e o leite em pó não podem conter sólidos de soro de queijo acima do especificado na Instrução Normativa n° 69 (IN 69) do Ministério da Agricultura Pecuária e Abastecimento (MAPA) (BRASIL, 2006a).

O leite bovino possui quatro tipos principais de caseína: α -S1, α -S2, β e κ -caseína. A caseína é composta por cerca de 50% α -caseína 30% de β -caseína e 15% de κ -caseína. Todas são fosforiladas em resíduos específicos de serina e em

consequência a κ -caseína é glicosilada em diferentes graus. Portanto, o CMP é heterogêneo e têm todos os sítios originais de modificações pós-traducionais da κ -caseína: sítios de glicosilação e fosforilação do CMP (MOTTA *et al.*, 2014).

Embora possua diversos estudos realizados para a determinação do CMP em leite bovino, ainda existe uma atual necessidade de desenvolvimento de métodos analíticos rápidos, precisos e econômicos para a identificação de eventuais fraudes, às quais o leite está sujeito. Esse trabalho utiliza-se metodologia empregando a quimiometria através da aplicação de estatística multivariada, modelagem matemática e métodos baseados na lógica matemática formal aplicada à resolução de problemas (SANTOS, 2013; LAVINE, WORKMAN, 2008).

Através de ferramentas quimiométricas aplicadas aos dados de ATR-FTIR propõem-se desenvolver um método rápido capaz de ser aplicado nas indústrias de laticínios, com o preparo de amostra simples, comparando com a Instrução Normativa n° 68 (IN 68) por cromatografia líquida acoplada a espectrometria de massas (CL-EM/EM) (BRASIL, 2006b).

MATERIAIS E MÉTODOS

Preparo das Amostras

Para a preparação das amostras foi utilizado o leite Ultra Alta Temperatura (UHT) semidesnatado e desnatado e o leite cru (*in natura*) mantido à temperatura ambiente. Preparou-se uma solução de CMP para cada tipo de leite. A preparação da solução de CMP (concentração 3 g L⁻¹), foi realizada pela pesagem de aproximadamente 75 mg do padrão do CMP e transferidos para balão volumétrico de 25 mL tendo o volume ajustado com o respectivo leite. A partir dessa solução foram adicionadas alíquotas em balões volumétricos de 10 mL, em triplicatas, correspondentes aos valores de adulteração com soro (5%, 10%, 15%, 20%, 25% e 30%). Assim, obteve um total de 18 amostras de leite adulteradas com a solução de CMP.

Reagentes

Reagentes: água ultrapura (Milli-Q) com resistividade controlada em 18,2 MΩm, metanol (J.T Baker, grau de pureza 100%) e acetonitrila (Tedia, grau de pureza 99,9%). Todos os solventes utilizados neste trabalho possuem grau de pureza para análise de resíduo. Pepsina da mucosa gástrica suína (Sigma-Aldrich), glicina (VETEC), ácido acético glacial (J.T Baker), ácido tricloroacético (Merck, grau de pureza $\geq 99,5\%$), ácido fórmico (Sigma-Aldrich, grau de pureza $\geq 99,5\%$).

Padrões analíticos utilizados no desenvolvimento do trabalho: caseinomacropéptido (CMP), obtido da DaviscoFoods (MN, EUA), com grau de pureza 91,3%; MAIPPKKNQDKTEIPTINT peptídeos sintéticos da digestão com pepsina, obtidos da Mimotopes (Austrália), com graus de pureza 96,0. Todos os padrões foram conservados em temperaturas entre -30°C e -10°C .

Todos os solventes, reagentes e padrões foram doados pelo Laboratório Nacional Agropecuário do Rio Grande do Sul (LANAGRO/RS).

Preparo das Amostras para a Digestão Enzimática com Pepsina para Análise em Cromatografia Líquida Acoplado Espectrometria de Massas em Modo Tandem (CL-EM/EM)

Precipitação e Extração das Proteínas

A tubos de centrífuga foi adicionado 1 mL de cada amostra adulterada com CMP. Para a precipitação das proteínas e extração dos analitos adicionaram-se 500 μL de ácido tricloroacético (TCA) na concentração final de 8% e agitou-se no *vortex* por 1 minuto. Os tubos foram colocados em banho de ultrassom por 30 minutos e centrifugados a 1200 rpm por 10 minutos. Após a centrifugação alíquotas de 200 μL do sobrenadante foram transferidas para *vials* de 1,5 mL para posterior digestão.

Digestão Enzimática com Pepsina

Preparou-se uma solução de glicina 1 mol L^{-1} , dissolvendo 7,5 g de glicina em um balão volumétrico de 100 mL com água ultrapura. A partir desta solução foi preparada diluição de forma a conter $200 \text{ } \mu\text{g mL}^{-1}$ de glicina. Esta diluição foi adicionada para aumentar o pH e em proporção de 1:1 (amostra : glicina). Para a digestão foram empregados $50 \text{ } \mu\text{L}$ de pepsina ($10 \text{ } \mu\text{g mL}^{-1}$), à temperatura de 37°C , com tempo de digestão de 24 horas.

Análise dos Produtos da Digestão com Pepsina por CL-EM/EM.

Para análise por cromatografia a líquido acoplada a espectrometria de massas, os principais parâmetros de fragmentação foram definidos na digestão teórica por meio de software específico (SKYLINE). Os parâmetros da espectrometria de massas foram otimizados no modo de ionização *electrospray* positivo (ESI+) utilizando o software *Analyst* e o peptídeo sintético e os produtos de digestão com pepsina do CMP. A quantificação foi realizada no modo MRM (do inglês *Multiple Reaction Monitoring*), com duas (ou mais) transições: quantificador e qualificador. Neste método a fase móvel foi constituída de água ultrapura (fase A) e acetonitrila (fase B) ambas com 0,1% (v/v) de ácido fórmico, empregando reagentes de grau analítico e utilizando uma Coluna PLRP-S, 300A 10-15 μm . 150 x 4.6 mm, Varian.

Metodologia para Quantificação do CMP em Leite por Cromatografia Líquida Acoplada à Espectrometria de Massas em Modo Tandem (CL-EM/EM)

O equipamento utilizado no método de referência (CL-EM/EM), para análise dos produtos da digestão da pepsina, foi o sistema de CL-EM/EM modelo API 5000, marca Applied Biosystems/MDS SCIEX (Foster City, CA USA) acoplado a um cromatógrafo líquido 1100 Series da marca Agilent

Techonologies. Para identificação e quantificação dos dados foi utilizado o software *Analyst* (Applied Biosystems).

Para a determinação em g L^{-1} do teor de CMP por CL-EM/EM foi realizada a curva de calibração com seis pontos, preparada com as amostras de leite adulteradas com CMP em diferentes concentrações após a digestão enzimática. As amostras foram analisadas em triplicatas acompanhadas por uma amostra branca de cada tipo de leite.

Os parâmetros de espectrometria de massas que foram definidos para os peptídeos sintéticos da digestão com pepsina foram obtidos na etapa de infusão dos analitos usando FIA (*Flow Injection Analysis*), sendo otimizados automaticamente pelo software *Analyst* (MOTTA *et al.*, 2014).

Espectroscopia no Infravermelho Médio

Os espectros ATR-FTIR foram adquiridos mediante espectrofotômetro Cary 630 (Agilent Technologies), acoplado a um acessório de refletância total atenuada (ATR) com cristal de seleneto de zinco (ZnSe) empregando laser hélio/neon. Foi utilizado a faixa de 4000 a 800 cm^{-1} na obtenção dos espectros com resolução de 4 cm^{-1} e 32 varreduras. Todas as amostras foram analisadas em duplicatas em temperatura ambiente, resultando 6 espectros para cada concentração de CMP. Para cada duplicata analisada foi realizado um “background”.

Análise Multivariada

As análises multivariadas das amostras de leite adulterado com CMP foram realizadas destinando os dados de ATR-FTIR para a análise exploratória por HCA e PCA e para a regressão multivariada das concentrações referentes aos valores das adulterações empregando os algoritmos PLS, siPLS e LS-SVM. Como método de referência foram utilizados os dados de CL-EM/EM. Por meio dessa estratégia, buscou-se realizar uma caracterização quantitativas das amostras.

Regressão Multivariada do Leite Adulterado com CMP por ATR-FTIR

Para obtenção dos modelos por regressão multivariada foram utilizados os algoritmos PLS, siPLS e LS-SVM através do software Matlab 8.1 (MathWorks Inc. Natick, MA) e do pacote iToolbox (NORGAARD, 2000).

Considerando o total de 18 amostras de cada leite adulterado com CMP, 12 amostras (36 espectros) foram selecionadas para a construção dos modelos de calibração e 6 amostras (17 espectros) foram utilizadas para avaliar a predição dos modelos. A seleção das amostras de calibração e previsão foi feita utilizando o algoritmo *Kennard Stone* (MARCELO *et al.*, 2014).

Para a construção dos modelos de regressão por PLS e por siPLS, foram avaliadas duas alternativas de tratamento: apenas centrar na média ou o tratamento de correção do espalhamento de luz (MSC) aliado a centrar na média. Na obtenção dos modelos de regressão por siPLS, foi aplicado apenas o pré-processamento centrar na média, em virtude do alto custo computacional.

A construção dos modelos de regressão foi realizada com os espectros médios das amostras. Os modelos de regressão por PLS foram construídos utilizando toda a informação espectral. Já os modelos de siPLS foram construídos utilizando 2, 3 e 4 combinações entre as divisões espectrais, realizando 8, 16, 32, 64 e 128 divisões equidistantes do espectro, para cada caso, do total de divisões. Escolheu-se apenas um modelo, sendo aquele em que se obtém a regressão com menor erro de validação cruzada (RMSECV). É importante ressaltar que na escolha dos melhores modelos para o algoritmo siPLS, também foram comparados o número de variáveis latentes (VLs), pois essas estão relacionadas com o grau de complexidade dos modelos multivariados.

O número de VLs de cada modelo de regressão foi determinado pelo procedimento de validação cruzada (*leave-one-out*), escolhendo-se o número de VLs em que os modelos de regressão apresentam menor RMSECV.

Os modelos de regressão por LS-SVM foram construídos utilizando os espectros médios das amostras, toda a informação espectral e, também, 3 e 4 combinações entre as divisões espectrais, realizando 8, 16, 32 e 64 divisões equidistantes do espectro, para cada caso, do total de divisões. Escolheu-se

apenas um modelo, sendo aquele que apresentou menor erro de previsão (RMSEP).

Após determinados os modelos de regressão por PLS, siPLS e LS-SVM, o melhor modelo foi definido sendo aquele que apresentou o menores valores de RMSECV, RMSEC e RMSEP, ou seja, o modelo de regressão multivariada que consegue prever a propriedade de interesse com o menor erro possível entre os modelos comparados.

Análise Exploratória dos Dados ATR-FTIR

Para análise exploratória de dados do ATR-FTIR, utilizou-se o software ChemoStat 1.0 (HELPER *et al.*, 2015).

Previamente, a análise exploratória dos espectros, foram submetidas aos tratamentos de alisamento por meio do algoritmo Savistky-Golay (empregando 10 pontos e polinômio de primeiro grau) de alinhamento dos picos, utilizando-se a normalização pelo máximo (máximo=1).

Para a análise de agrupamentos hierárquicos (HCA), utilizou-se a distância Euclidiana para a medida da distância entre as amostras no espaço n -dimensional e o método Ward para os agrupamentos de amostras.

Para a análise por componentes principais (PCA), os sinais espectrais das amostras foram também centrados na média.

RESULTADOS E DISCUSSÃO

Espectroscopia no Infravermelho

Os espectros de ATR-FTIR, obtidos na região de 4.000 a 800 cm^{-1} , das amostras de leite são apresentados na Figura 7.

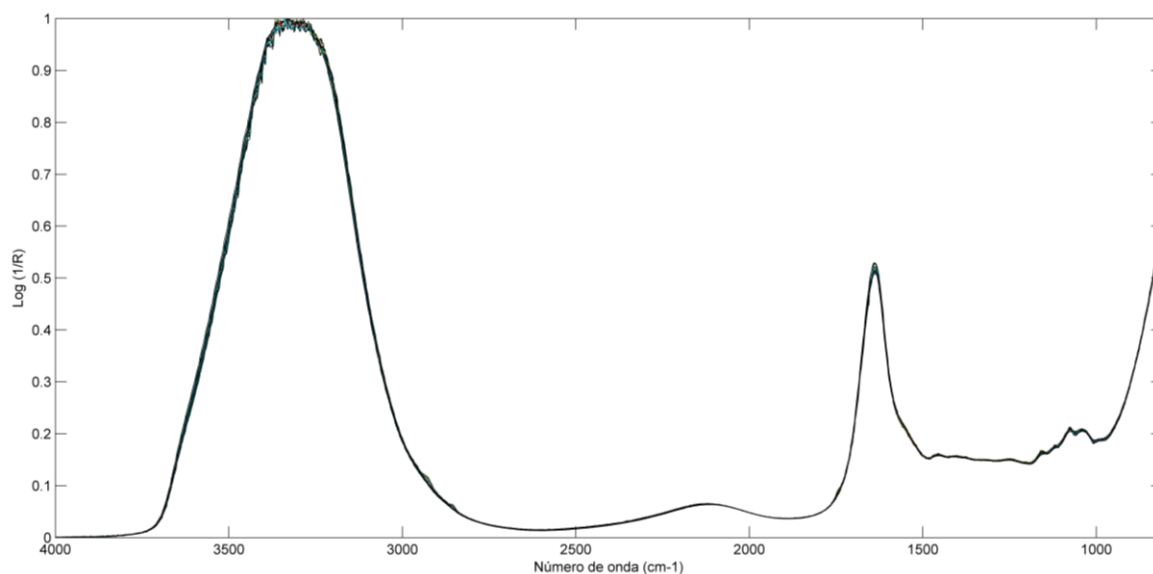


Figura 7: Espectros médios de ATR-FTIR na região de 4000 a 800 cm^{-1} das 17 amostras de leite adulteradas com CMP.

Esses espectros apresentam modos de combinações vibracionais de grupos funcionais das aminas presentes na composição química das proteínas, com bandas de absorção de maior intensidade na região de 1690 a 1650 cm^{-1} , que podem ser atribuídas às intensidades de estiramento de ligações de amidas ($\text{C}=\text{O}$); essas são encontradas nos aminoácidos.

Os aminoácidos têm as estruturas formadas por grupamento carboxila (COOH), um grupamento amina (NH_2) e um radical que determina cada um dos vinte aminoácidos (CAMPBELL, 2001). Os aminoácidos do CMP são de cadeia ramificada como a valina e isoleucina e aminoácidos aromáticos (fenilalanina, triptofano e tirosina) e dos aminoácidos sulfurados cistina e cisteína. Existem quatro aminoácidos que são encontrados em maior quantidade no CMP que são: ácido glutâmico, treonina, prolina e isoleucina (OLIVEIRA, 2010).

Análise Exploratória

Foram tratados os espectros obtidos no ATR-FTIR das seis amostras de leite (LC, LS e LD) em triplicata (A, B e C) com diferentes níveis de

adulterações (5%, 10%, 15%, 20%, 25% e 30%), utilizando-se o software MATLAB 8.1. Após o tratamento dos dados, foram realizadas as análises por Componentes Principais (PCA) e análise por agrupamentos hierárquicos (HCA).

Análise por Agrupamentos Hierárquicos (HCA)

Através do software ChemoStat 1.0 utilizou-se para realizar a HCA das médias das amostras, obtendo-se um dendrograma, conforme a Figura 8. É possível observar que a HCA apresenta três grupos, de acordo com os diferentes tipos de leite e suas similaridades. A amostra LD 30% foi constatada como *outlier* e removida para as análises de PCA, PLS, siPLS e LS-SVM.

O primeiro grupo (Grupo I), em vermelho, é composto pelas amostras LD 15%, LD 10%, LD, 5%, LD 20% e LD 25%, referentes ao leite desnatado. O segundo grupo (Grupo II), em verde, é composto pelas amostras LC 30%, LC 25%, LC 15%, LC 10%, LC 20% e LC 5%, referentes ao leite cru. O terceiro grupo (Grupo III), em azul, é composto pelas amostras LS 20%, LS 15%, LS 25%, LS 30%, LS 10% e LS 5%, referentes ao leite semidesnatado.

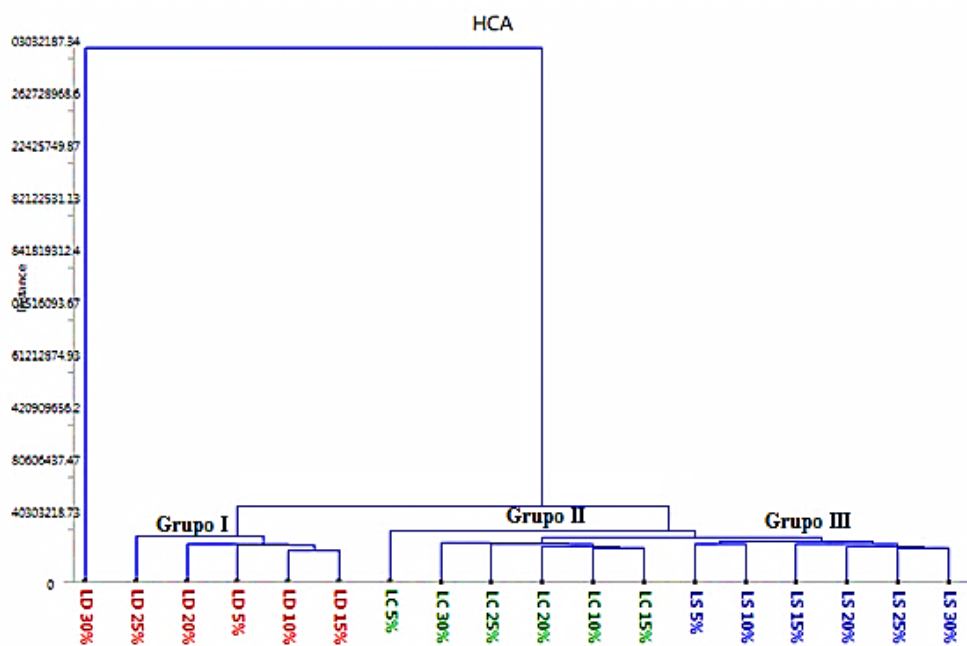


Figura 8: Dendrograma das amostras de leite cru (LC), leite semidesnatado (LS) e leite desnatado (LD) analisado por ATR-FTIR.

Análise por Componentes Principais (PCA)

Através do software ChemoStat 1.0, foi realizada Análise de Componentes Principais (PCA) para as amostras adulteradas com diferentes concentrações de CMP em leite cru (LC), leite semidesnatado (LS) e leite desnatado (LD) em triplicata (A, B e C). Empregando as médias dos espectros das amostras obteve-se um acúmulo de 62,64% da variância dos dados nas duas primeiras componentes principais (PC1 e PC2) (HELFER et al., 2015).

Através dessas PCs foi possível separar somente as amostras LD com suas respectivas concentrações (5%, 10%, 15%, 20% e 25%) com escores positivos para ambas as PCs. Já as amostras LS e LC permaneceram juntas, devido à composição desses tipos de leite ser semelhante.

A figura 9 representa o gráfico dos escores PC1 *versus* PC2 das amostras analisadas. Observa-se que a PC1 descreve a natureza da presença da gordura no leite sendo os leites LS e LC com escores positivos e o LD com escores negativos.

Já a PC2 está relacionada com a concentração de CMP apresentando um gradiente de menor para maior, com escores negativos para as menos concentradas e positivos para as mais concentradas. Por sua vez a PC3 é responsável pela segregação das amostras LS com escores negativos e das amostras LC com escores positivos (Figura 10).

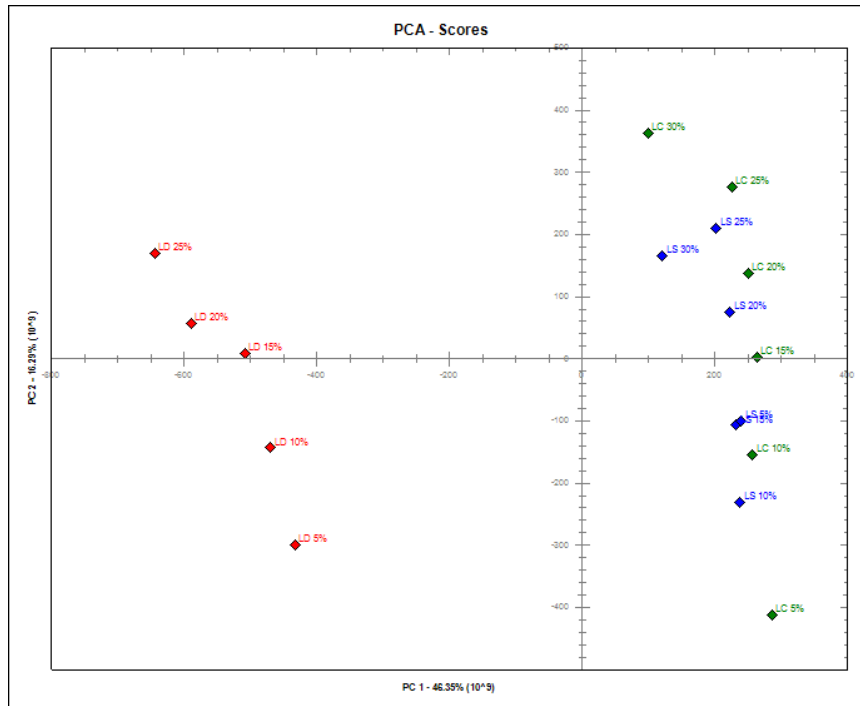


Figura 9: Gráfico dos escores PC1 x PC2 das amostras de leite cru (LC), leite semidesnatado (LS) e leite desnatado (LD).

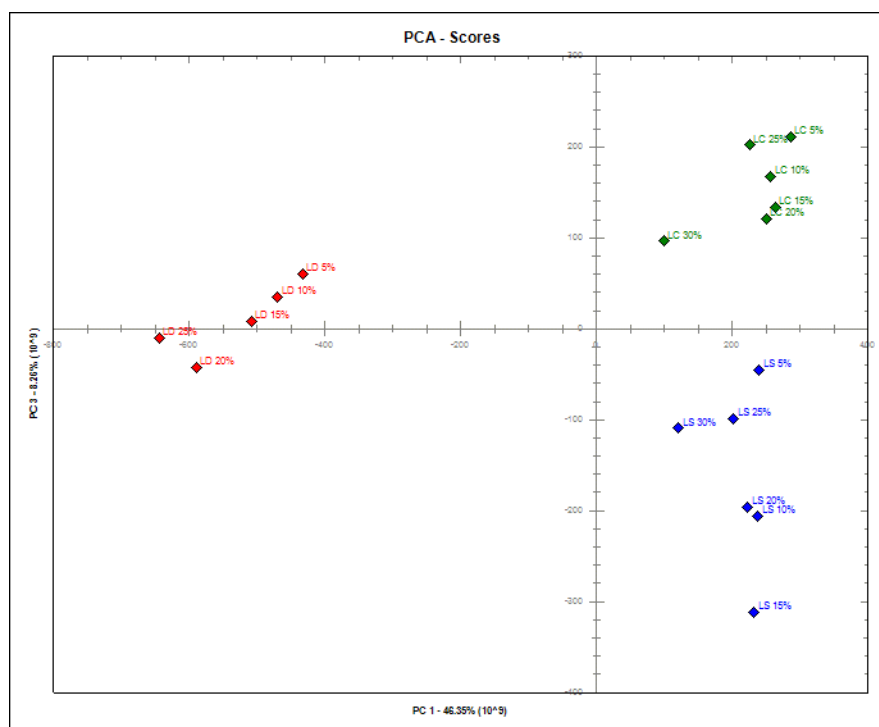


Figura 10: Gráfico dos escores PC1 x PC3 das amostras de leite cru (LC), leite semidesnatado (LS) e leite desnatado (LD).

Regressão Multivariada

Avaliação dos Modelos de Regressão por siPLS para Quantificação de CMP no Leite

Com as amostras selecionadas pelo algoritmo de *Kennard Stone* para o conjunto de calibração, foram construídos modelos de regressão por PLS utilizando toda a informação espectral e com seleções de regiões com siPLS para a predição na quantificação de CMP, utilizando os tratamentos alisamento e normalização e como pré-processamento os dados centrados na média. Os resultados estão apresentados na Tabela 2.

O modelo de regressão por PLS modelo GlobalM, possui uma pequena diferença entre o RMSECV ($\text{RMSECV}_{\text{GlobalM}} = 0,2503 \text{ mg L}^{-1}$) e RMSEP ($\text{RMSEP}_{\text{GlobalM}} = 0,5423 \text{ mg L}^{-1}$), além de apresentar o menor número de variáveis latentes ($\text{VLs} = 7$), dessa forma mostrar-se um modelo desejável. Pode-se avaliar também que o Modelo Global M, possui um bom índice de coeficiente de determinação R^2_{cal} (0,9910).

Os resultados dos modelos de siPLS construídos por validação cruzada *leave-one-out* para a predição das concentrações de CMP são apresentadas na Tabela 2, os quais estão organizados por número de divisões feitas no espectro e número de combinações entre intervalos.

Nos modelos de dois intervalos notou-se que foram obtidos dois modelos de regressão com figuras de mérito semelhantes, que são os modelos s2i8M (combinação de 2 intervalos, 8 divisões espectrais equidistantes, dados centrados na média) e s2i128M (combinação de 2 intervalos, 128 divisões, dados centrados na média). Embora o modelo s2i8M apresenta o menor RMSEC ($\text{RMSEC}_{\text{s2i8M}} = 0,0703 \text{ mg L}^{-1}$ e $\text{RMSEC}_{\text{s2i128M}} = 0,0923 \text{ mg L}^{-1}$) e maior R^2 ($R^2_{\text{cal,s2i8M}} = 0,9955^1$ e $R^2_{\text{cal,s2i128M}} = 0,9923$), ele possui um maior sobreajuste em razão da diferença entre os erros de calibração e previsão ($\text{RMSEC}_{\text{s2i8M}} - \text{RMSEP}_{\text{s2i8M}} = 0,4982 \text{ mg L}^{-1}$) ser maior que a diferença observada no modelo s2i128M ($\text{RMSEC}_{\text{s2i128M}} - \text{RMSEP}_{\text{s2i128M}} = 0,3736 \text{ mg L}^{-1}$). Por outro lado, embora o modelo s2i128M possuir o maior número de variáveis latentes, o modelo s2i128M é o

mais adequado, dentre os modelos de regressão que se combinam duas regiões espectrais, para realizar o siPLS.

Com respeito aos modelos de três intervalos observa-se que foram obtidos dois modelos de regressão com figuras de mérito muito semelhantes, que são os modelos s3i8M (combinação de 3 intervalos, 8 divisões espectrais equidistantes, dados centrados na média) e s3i16M (combinação de 3 intervalos, 16 divisões, dados centrados na média). Apesar do modelo s3i16M apresenta o menor RMSEC ($RMSEC_{s3i8M} = 0,0811 \text{ mg L}^{-1}$ e $RMSEC_{s3i16M} = 0,0370 \text{ mg L}^{-1}$) e maior R^2 ($R^2_{cal,s3i8M} = 0,9940$ e $R^2_{cal,s3i16M} = 0,9980$), observa-se que ele tem um maior sobreajuste em razão da diferença entre os erros de calibração e previsão ($RMSEC_{s3i16M} - RMSEP_{s3i16M} = 0,5894 \text{ mg L}^{-1}$) ser maior que a diferença observada no modelo s3i8M ($RMSEC_{s3i8M} - RMSEP_{s3i8M} = 0,4589 \text{ mg L}^{-1}$), além de utilizar duas variáveis latentes a menos e ter menor erro de previsão ($RMSEP_{s3i8M} = 0,5400 \text{ mg L}^{-1}$). Deste modo, o modelo s3i8M é o mais adequado, dentre os modelos de regressão que se combinam três regiões espectrais, para realizar a regressão por siPLS.

No que se refere aos modelos de regressão por siPLS por combinações de quatro intervalos espectrais, conforme a Tabela 2, os melhores modelos obtidos foram os modelos s4i8M (combinação entre 4 intervalos, 8 divisões espectrais equidistantes, dados centrados na média), s4i16M (combinação entre 4 intervalos, 16 divisões espectrais equidistantes, dados centrados na média) e s4i64M (combinação entre 4 intervalos, 64 divisões espectrais equidistantes, dados centrados na média). Embora os modelos s4i8M, s4i16M e s4i64M apresentem menor RMSEC ($RMSEC_{s4i8M} = 0,0682 \text{ mg L}^{-1}$, $RMSEC_{s4i16M} = 0,0363 \text{ mg L}^{-1}$ e $RMSEC_{s4i64M} = 0,0436 \text{ mg L}^{-1}$) e maior R^2 ($R^2_{cal,s4i8M} = 0,9958$ mg L^{-1} , $R^2_{cal,s4i16M} = 0,9988$ e $R^2_{cal,s4i64M} = 0,9983$), nota-se que os modelos s4i16M e s4i64M tem um maior sobreajuste em razão da diferença entre os erros de calibração e previsão ($RMSEC_{s4i16M} - RMSEP_{s4i16M} = 0,5812 \text{ mg L}^{-1}$ e $RMSEC_{s4i64M} - RMSEP_{s4i64M} = 0,5839 \text{ mg L}^{-1}$) ser maior que a diferença observada no modelo s4i8M ($RMSEC_{s4i8M} - RMSEP_{s4i8M} = 0,4970 \text{ mg L}^{-1}$), além de utilizar variáveis latentes a menos e ter menor erro de previsão ($RMSEP_{s4i8M} =$

0,5652 mg L⁻¹). Portanto, o modelo s4i8M é o mais adequado, dentre dos modelos de regressão em que se combinam quatro regiões espectrais, para realizar a regressão por siPLS.

Tabela 2: Resultados das figuras do mérito dos modelos de regressão por siPLS para a determinação do CMP pelo método de referência.

| <i>Modelo^a</i> | <i>VT^b</i> | <i>Intervalos selecionados</i> | <i>VLs^c</i> | <i>RMSECV^d</i> | <i>RMSEC^d</i> | <i>R²_{cal}</i> | <i>RMSEP^d</i> |
|---------------------------|-----------------------|--------------------------------|------------------------|---------------------------|--------------------------|------------------------------------|--------------------------|
| GlobalM | 1720 | Todo o espectro | 5 | 0,2503 | 0,0999 | 0,9910 | 0,5423 |
| s2i8M | 430 | 2, 6 | 9 | 0,2144 | 0,0703 | 0,9955 | 0,5685 |
| s2i16M | 215 | 1,7 | 10 | 0,1956 | 0,0437 | 0,9983 | 0,6270 |
| s2i32M | 108 | 1,9 | 11 | 0,1662 | 0,0307 | 0,9991 | 0,5987 |
| s2i64M | 54 | 14,15 | 8 | 0,1891 | 0,0806 | 0,9941 | 0,5646 |
| s2i128M | 27 | 26, 28 | 12 | 0,1971 | 0,0923 | 0,9923 | 0,4659 |
| s3i8M | 645 | 2, 4, 6 | 8 | 0,2039 | 0,0811 | 0,9940 | 0,5400 |
| s3i16M | 323 | 4, 7, 12 | 11 | 0,1850 | 0,0370 | 0,9980 | 0,6264 |
| s3i32M | 161 | 1, 2, 9 | 13 | 0,1627 | 0,0178 | 0,9997 | 0,5543 |
| s3i64M | 81 | 14, 15, 64 | 10 | 0,1679 | 0,0519 | 0,9976 | 0,5817 |
| s3i128M | 40 | 47, 71, 89 | 10 | 0,1760 | 0,0633 | 0,9964 | 0,5175 |
| s4i8M | 860 | 1, 2, 4, 6 | 9 | 0,2041 | 0,0682 | 0,9958 | 0,5652 |
| s4i16M | 326 | 4, 6, 7, 12 | 11 | 0,1805 | 0,0363 | 0,9988 | 0,6175 |
| s4i32M | 215 | 1, 2, 16, 18 | 10 | 0,1728 | 0,0449 | 0,9982 | 0,7065 |
| s4i64M | 108 | 14, 15, 43, 45 | 10 | 0,1493 | 0,0436 | 0,9983 | 0,6275 |

^as: número de intervalos combinados; i: número de divisões do espectro; M: dados centrados na média. ^bVT: variáveis totais. ^cVLs: variáveis latentes; ^dUnidade de medida – mg L⁻¹. Formatação em negrito aos melhores modelos de regressão com respeito ao mesmo número de intervalos combinados.

Na Tabela 3 são apresentados os resultados das figuras de mérito para a comparação dos melhores modelos de regressão por siPLS por combinação de 2, 3 e 4 faixas espectrais equidistantes. Apesar de, em valores absolutos, o modelo s2i18M apresentar menor RMSEP, é possível observar que não há diferença entre os valores de R^2 dos três modelos. No entanto, ao se comparar os valores de RMSEC e REMSEP dos modelos construídos, observa-se uma alta diferença entre os valores nos modelos s2i128M com os modelos s3i8M e s4i8M. Mesmo que o modelo s2i128M apresente menor valor, ele tem maior sobreajuste em função do elevado número de VLS ($VLS_{s2i128M} = 12$) e também as suas respectivas faixas espectrais não explicam o modelo desejado; desse modo não é o mais adequado. Já os modelos s3i8M e s4i8M, além de possuírem regiões espectrais adequadas, também têm um menor número de variáveis latentes, o que diminui o sobreajuste do modelo. Portanto, nesse caso o melhor modelo do siPLS é o s3i8M devido ao seu menor número de variáveis ($VLS_{s3i8M} = 8$) (Figura 11) quando comparado com o outro modelo ($VLS_{s4i8M} = 9$), além de possuir também a menor diferença entre RMSEC e REMSEP com o modelo s4i8M ($RMSEC - RMSEP_{s3i8M} = 0,4589 \text{ mg L}^{-1}$ e $RMSEC - RMSEP_{s4i8M} = 0,4970 \text{ mg L}^{-1}$).

Tabela 3: Resultados da figura de mérito dos melhores modelos de regressão por siPLS em função ao número de combinações entre intervalos para a determinação de CMP.

| <i>Modelo^a</i> | <i>VT^b</i> | <i>Intervalos selecionados</i> | <i>VLS^c</i> | <i>RMSECV^d</i> | <i>RMSEC^d</i> | <i>R²_{cal}</i> | <i>RMSEP^d</i> |
|---------------------------|-----------------------|--------------------------------|------------------------|---------------------------|--------------------------|------------------------------------|--------------------------|
| Global M | 1720 | Todo o espectro | 7 | 0,2503 | 0,0999 | 0,9910 | 0,5423 |
| s2i128M | 430 | 26, 28 | 12 | 0,1971 | 0,0923 | 0,9923 | 0,4659 |
| s3i8M | 645 | 2, 4, 6 | 8 | 0,2039 | 0,0811 | 0,994 | 0,5400 |
| s4i8M | 860 | 1, 2, 4, 6 | 9 | 0,2041 | 0,0682 | 0,995 | 0,5652 |

^aM – dados centrados na média, ^bVT – Variáveis totais; ^cVariáveis latentes – ^dUnidade de medida – mg L^{-1} .

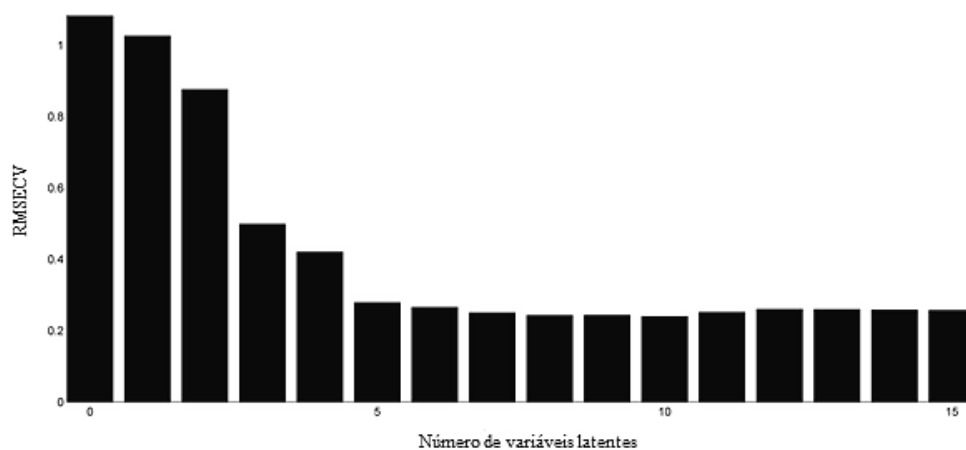


Figura 11: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas pelo FTIR/PLS para determinação de CMP utilizando toda a faixa espectral

È apresentada na Figura 12 comparação dos valores da concentração de CMP que foram determinados pelo método de referência e entre os valores previstos de ATR-FTIR/PLS, isto é, pelo modelo GlobalM. Nota-se uma boa correlação entre ambos os métodos, com desempenho satisfatório na predição na determinação de CMP para a maioria das amostras testadas.

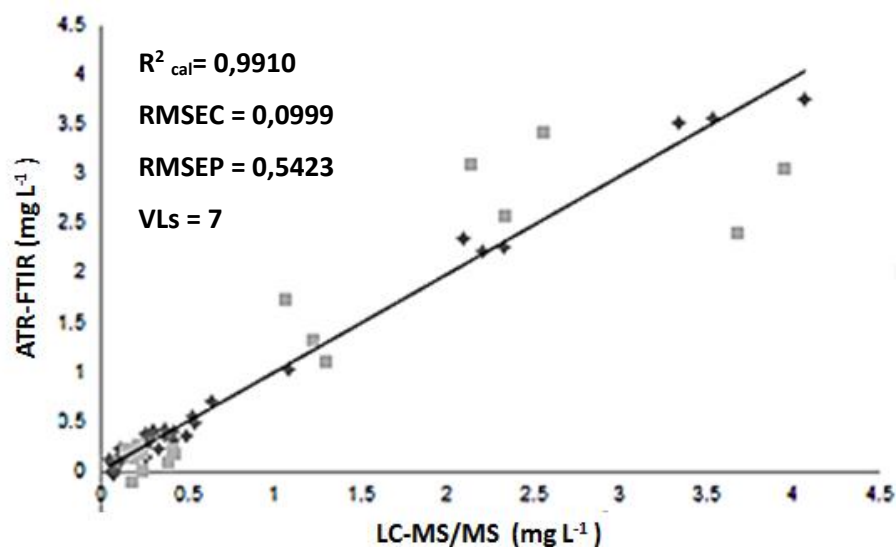


Figura 12: Gráfico dos valores do método de referência *versus* valores previstos pelo ATR-FTIR/PLS para a determinação de CMP em leite utilizando toda a faixa espectral, sendo os losangos as amostras de calibração e os quadrados cinza as de previsão.

Avaliação dos Modelos de Máquinas de Vetores de Suporte com Mínimos Quadrados (LS-SVM)

Antes da análise com LS-SVM, foi utilizado como pré-tratamento alisamento e normalização e como pré-processamento os dados centrados na média.

Os resultados para os modelos de LS-SVM são apresentados na Tabela 4. Conforme pode ser observado, o melhor modelo é o GlobalM que utiliza toda a faixa espectral apresentando menores erros de previsão para as amostras externas (RMSEP = 0,4092 mg L⁻¹) comparadas aos demais modelos que utilizam seleções de regiões espectrais. Também o modelo GlobalM apresentou-se equivalente aos demais modelos nos valores de correlação do método de referência *versus* valores previstos pelo ATR-FTIR para a determinação de CMP em leite utilizando as amostras de previsão e de calibração (Figura 13).

Tabela 4: Resultados do LS-SVM

| Modelos ^a | RMSECV ^b | RMSEP ^b | R ² _{cv} | R ² _{pred} |
|----------------------|---------------------|--------------------|------------------------------|--------------------------------|
| 3i8M | 0,3280 | 0,5217 | 0,9330 | 0,9109 |
| 3i16M | 0,3144 | 0,4625 | 0,9250 | 0,8696 |
| 4i8M | 0,3270 | 0,5418 | 0,9359 | 0,9073 |
| 4i16M | 0,3065 | 0,4150 | 0,9184 | 0,8888 |
| 4i64M | 0,1629 | 0,6600 | 0,9764 | 0,7279 |
| GlobalM | 0,2602 | 0,4092 | 0,9567 | 0,9030 |

^as: número de intervalos combinados; i: número de divisões do espectro; M: dados centrados na média. ^bUnidade de medida – mg L⁻¹. Formatação em negrito aos melhores modelos de regressão com respeito ao mesmo número de intervalos combinados.

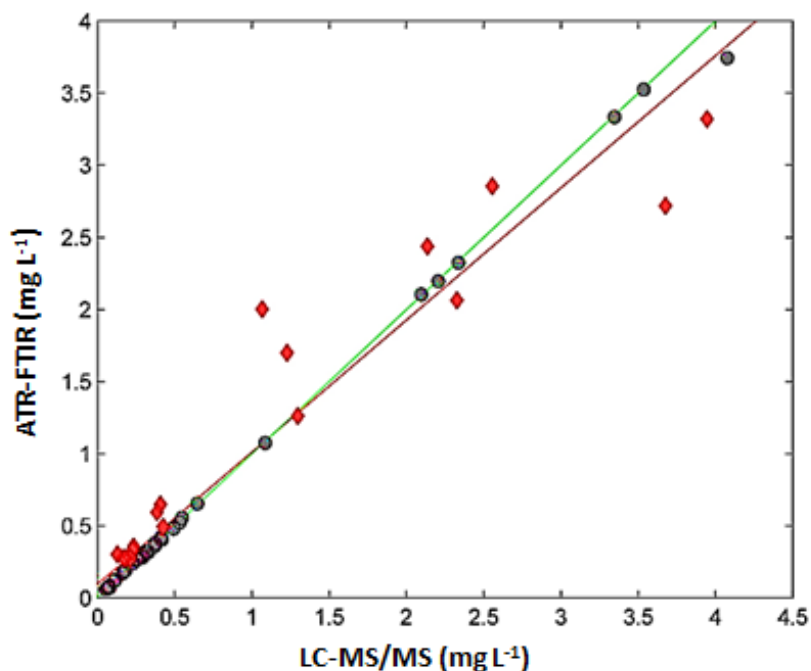


Figura 13: Gráfico dos valores do método de referência versus valores previstos pelo ATR-FTIR/LS-SVM para a determinação de CMP em leite utilizando toda a faixa espectral, sendo os losangos vermelhos as amostras de previsão e as bolinhas cinza as de calibração.

Comparando com os modelos PLS e siPLS, a LS-SVM apresentou melhores valores de previsão estando de acordo com alguns resultados relatados na literatura (MARRETO, 2011), nos quais se observa que a LS-SVM tem maior capacidade de generalização evitando assim o sobreajuste do modelo.

CONCLUSÃO

As ferramentas exploratórias de Análise de Agrupamentos Hierárquicos (HCA) e Análise por Componentes Principais (PCA) indicaram uma semelhança entre as amostras de leite cru (LC) e leite semidesnatado (LS) em virtude da presença de gordura nesses leites. Também pela PCA foi possível observar uma tendência de separação das amostras com distintas adições de caseinomacropéptido (CMP).

Através dos modelos de regressão realizados com os algoritmos de Mínimos Quadrados Parciais (PLS) notou-se que utilizando toda a faixa (GlobalM) obtém-se um menor sobreajuste no modelo, apresentando o menor número de variáveis latentes ($VL_{S_{GlobalM}} = 7$) comparado com o melhor modelo (s3i8M) com seleção de região pelo siPLS ($VL_{S_{s3i8M}} = 8$). Portanto para o presente conjunto de dados o PLS empregando a faixa espectral 4000-800 cm^{-1} constitui-se um modelo adequado para a quantificação da CMP.

Já quando comparamos o modelo de Máquinas de Vetores de Suporte com Mínimos Quadrados (LS-SVM) com os modelos empregando PLS, onde ambos utilizam a faixa espectral 4000-800 cm^{-1} , o modelo GlobalM do LS-SVM apresenta menor RMSEP ($RMSEP_{GlobalM} = 0,4092 \text{ mg L}^{-1}$) do que o modelo GlobalM do PLS ($RMSEP = 0,5423 \text{ mg L}^{-1}$). Em virtude disso o modelo GlobalM do LS-SVM, foi considerado o modelo mais robusto para quantificação do CMP em amostra de leite.

REFERÊNCIAS

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. Instrução Normativa n°68, de 12 de dezembro de 2006. ANEXO I. *Institui Métodos Analíticos Oficiais Físico-químicos para controle de leite e Produtos Lácteos. Diário Oficial da União*, Brasília, 12 de dezembro de 2006a.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Instrução Normativa N° 69, de 13 de dezembro de 2006 – ANEXO I. *Institui critério da qualidade do leite in natura, concentrado e em pó, reconstituídos, com base no método analítico oficial físico-químico denominado*

“*índice de CMP*”, de que trata Instrução normativa nº68, de 12 de dezembro de 2006. **Diário Oficial da União**, Brasília, 15 de dezembro de 2006b.

CARVALHO, M. E.; CARAVALHO, L. M.; ALCÂNTRA, L. A. P.; BONOMO, R. C. F. Métodos de detecção de fraude em leite por adição de soro de queijo. **Revista eletrônica de veterinária**, v. 8, n. 6, p. 7, 2007.

HELPER, G. A.; BOCK, F.; MARDER, L.; FURTADO, J. C.; COSTA, A. B.; FERRÃO, M. F. Chemostat, um software gratuito para análise exploratória de dados multivariados. **Química Nova**, v. 38, n. 4, p. 575-579, 2015.

LAVINE, B.; WORKMAN, J. Chemometrics. **Analytical chemistry**, v. 80, n. 12, p. 4519-4531, 2008.

MARCELO, M. C. A.; MARTINS, C. A.; POZEBON, D.; DRESSLER, V. L.; FERRÃO, M. F. Classification of yerba mate (*Ilex paraguariensis*) according to the country of origin based on element concentrations. **Microchemical Journal**, v. 177, n. 1, p. 164-171, 2014.

MARRETO, D. A. **Aplicação de máquinas de suporte para desenvolvimento de modelos de classificação e calibração multivariada em espectroscopia no infravermelho**. 2011. 133 f. Tese de doutorado (Química Analítica). Universidade Estadual de Campinas, São Paulo.

MARQUES, F. M.; SÁ, J. F. O.; SANTOS, M. C.; MARTINS, M. F.; FURTADO, M. A. M. Caracterização de leite em pó, soro de leite em pó e suas misturas por eletroforese em gel de poliacrilamida. **Revista do Instituto Adolfo Lutz**, v. 40, nº 4, p. 473-479, 2011.

MOTTA, T. M. C.; HOFF, R. B.; BARRETO, F.; ANDRADE, R. B.; LORENZINI, D. M.; MENEGHINI, L. Z.; PIZZOLATO, T. M. Detection and

confirmation of milk adulteration with cheese whey using proteomic-like sample preparation and liquid chromatography-electrospray-tandem mass spectrometry analysis. **Talanta**, n, 1, v, 120, p. 498-505, 2014.

NORGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. **Applied Spectroscopy**, v. 54, p. 413-419, 2000.

OLIVEIRA, G. B.; GATTI, M. D. S.; VALADÃO, R. C.; MARTINS, J. F. P.; LUCHESE, R. H. Detecção da adição fraudulenta de soro de queijo em leite: interferência da atividade de proteases bacterianas. **Revista de instituo laticínios cândido tostes**, v. 64, n 367/368, p. 56-65, 2009.

OLIVEIRA, R. R. **Aplicação da espectroscopia de infravermelho próximo para a determinação do caseinomacropéptido em Leite UAT**. 2010. 76 f. Dissertação de mestrado (Ciência Animal). Universidade Federal de Goiás, Goiânia.

SANTOS, P. M. **Aplicação de imagens digitais e técnicas espectro analíticas combinadas com quimiometria para a detecção e quantificação de adulteração em leite bovino**. 2013. 97 f. Tese de doutorado (Química analítica). Universidade Federal de São Carlos, São Carlos.

Na Análise de Agrupamentos Hierárquicos (HCA) foi possível diferenciar os três tipos de leite (leite cru, leite semidesnatado e leite desnatado) e também se constatou a presença de uma amostra *outlier*. Para confirmação deste *outlier* foi realizado o gráfico de diagnóstico de *outlier* (Figura 14) que confirma que a amostra LD 30% é realmente uma amostra anômala, pois o seu espectro apresenta elevado valor para o Hotelling T^2 , quando comparada com as demais.

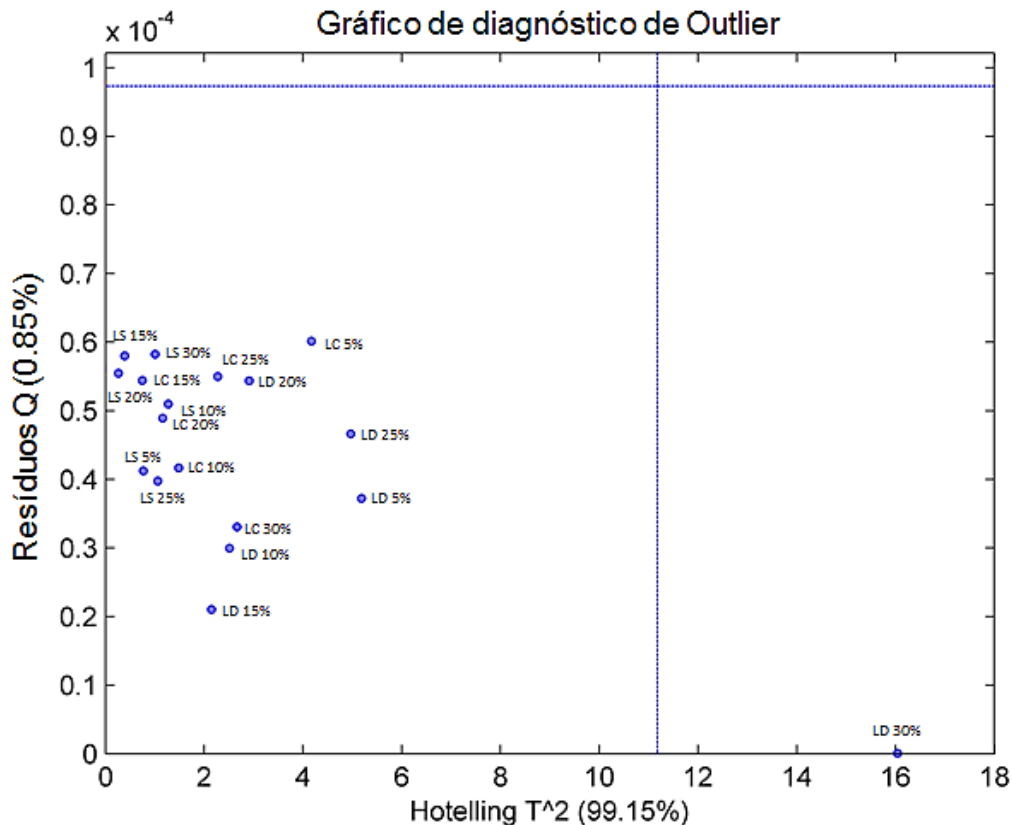


Figura 14: Gráfico de diagnóstico de *outlier* demonstrando a presença da amostra LD 30% sendo confirmado como uma amostra anômala.

Na Análise por Componentes Principais (PCA) apresentada na Figura 15, o *outlier* apresenta grande influência na 1ª componente principal, podendo interferir nos demais resultados. Em virtude disso, essa amostra foi removida para que não influencie, juntamente com os espectros das demais amostras, sendo realizada uma nova PCA. Desse modo podemos perceber uma melhora nos resultados depois da remoção desta amostra (Figura 9). Para as análises da

regressão multivariada também foi removido o espectro considerado como *outlier*, conforme consta em Regressão Multivariada do Artigo Científico (p. 94).

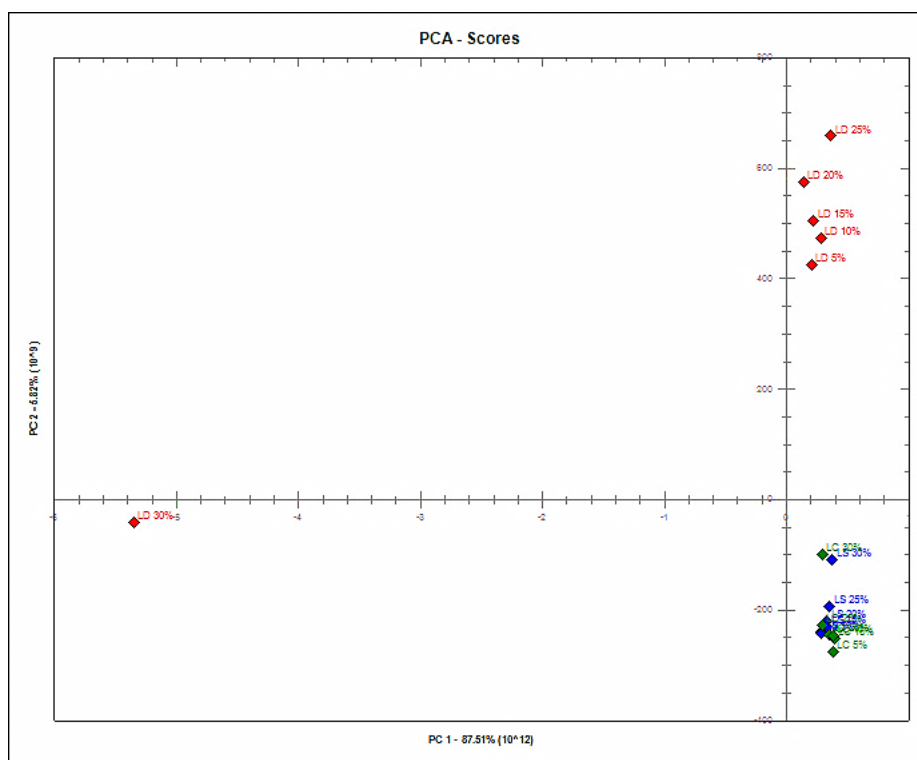


Figura 15: Gráfico da análise por componentes principais (PCA) com a presença da amostra *outlier*.

7 CONCLUSÃO

Na primeira parte do trabalho foi feita a análise exploratória dos dados de Infravermelho Médio com Transformada de Fourier (FTIR), obtidos com acessório de Reflexão Total Atenuada (ATR), foi possível interpretar as médias das 18 amostras de leite cru, leite semidesnatado e leite desnatado adulteradas com caseinomacropéptido (CMP). Através das amostras verificou-se que pelo emprego da Análise de Agrupamentos Hierárquicos (HCA) houve a formação de três grupos de acordo com o tipo de leite e também foi constatada e removida uma amostra *outlier*. Na Análise de Componentes Principais (PCA) foi possível observar que a variância de duas componentes principais foi importante para segregar os três grupos em função da gordura que compõe os tipos de leite.

Na segunda parte foram desenvolvidos modelos de regressão multivariada utilizando os dados de ATR-FTIR das amostras de leite adulteradas com CMP para quantificação por meio da construção de modelo de regressão multivariada por mínimos Quadrados Parciais (PLS), por Mínimos Quadrados Parciais por Sinergismo de Intervalos (siPLS) e por Máquinas de Vetores de Suporte por Mínimos Quadrados (LS-SVM).

Na quantificação de CMP foi possível obter por meio de ambos os algoritmos PLS e siPLS, modelos de regressão equivalentes, sendo considerado o melhor modelo o PLS, pois além de possuir um número de variáveis latentes menor, também tem o RMSECV e RMSEP semelhantes. Cabe salientar que o PLS não necessita de seleção de regiões espectrais e dessa forma reduz o tempo computacional tornando assim a análise mais rápida. Já quando comparamos o modelo realizado pela LS-SVM com os modelos empregando PLS, onde ambos utilizam toda a faixa espectral ($4000-800\text{ cm}^{-1}$), o modelo de LS-SVM apresenta menores valores de RMECV e RMSEP. Sendo assim foi considerado modelo mais robusto para quantificação do CMP em amostras de leite deste trabalho.