



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós-Graduação em Química

Dissertação de Mestrado

**Análise de triagem para confirmação do estado de conservação de
insulina humana NPH injetável**

Suely Fernandes da Silva

João Pessoa – PB – Brasil
Dezembro de 2020



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós-Graduação em Química

Dissertação de Mestrado

Análise de triagem para confirmação do estado de conservação de insulina humana NPH injetável

Suely Fernandes da Silva*

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal da Paraíba como requisito para obtenção do título de Mestre em Química, área de concentração em Química Analítica.

Orientador: Prof. Dr. Mário César Ugulino de Araújo

Coorientador: Dr. David Douglas de Sousa Fernandes

***Bolsista CAPES**

João Pessoa – PB – Brasil
Dezembro de 2020

Catálogo na publicação
Seção de Catalogação e Classificação

S586a Silva, Suelly Fernandes da.

Análise de triagem para confirmação do estado de
conservação de insulina humana NPH injetável / Suelly
Fernandes da Silva. - João Pessoa, 2021.

74 f. : il.

Orientação: Mário César Ugulino de Araújo.

Coorientação: David Douglas de Sousa Fernandes.

Dissertação (Mestrado) - UFPB/CCEN.

1. Química. 2. Controle de qualidade. 3. Insulina. 4.
Espectroscopia NIR. 5. Quimiometria. 6. One class. I.
Araújo, Mário César Ugulino de. II. Fernandes, David
Douglas de Sousa. III. Título.

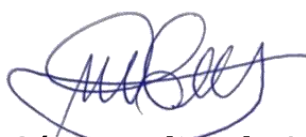
UFPB/BC

CDU 54(043)

Elaborado por MARILIA RIANNY PEREIRA COSMOS - CRB-15/862

Análise de triagem para confirmação do estado de conservação de insulina humana NPH injetável.

Dissertação de Mestrado apresentada pela aluna Suelly Fernandes da Silva e aprovada pela banca examinadora em 16 de dezembro de 2020.



Prof. Dr. Mário César Ugulino de Araújo
Orientador/Presidente



Prof. Dr. David Douglas de Sousa Fernandes
Examinador



Profa. Dra. Simone da Silva Simões
Examinadora



Profa. Dra. Liliana de Fátima Bezerra Lira Pontes
Examinadora

Assinaturas da Banca realizadas em modo Webconferência em 16/12/2020, digitalizadas e certificadas pelo Prof. Dr. Mário César Ugulino de Araújo (SIAPE 334937) em 16/12/2020



Agradecimentos

Primeiramente eu gostaria de agradecer a Deus por ter me dado força, saúde e sabedoria para superar os momentos difíceis durante a minha jornada.

Agradeço também à minha família que esteve ao meu lado em todos os momentos. Em especial aos meus pais, Sebastião e Valdete por todo amor e incentivo para continuar seguindo em frente. Às minhas irmãs Suelma e Suênia por todo apoio, e as minhas sobrinhas, Kemilly Vitória e Maria Sophia, por todo amor e alegria. Aos meus familiares e amigos, que sempre me entendiam quando eu não podia ir vê-los.

Ao professor Mário César Ugulino pela oportunidade de trabalho e por seus ensinamentos. Ao David Douglas por sua orientação, ensinamentos e paciência.

Aos professores Liliana de Fátima, Wallace Duarte e Simone Simões por terem aceitado fazer parte da banca avaliadora deste trabalho e por contribuírem com suas valiosas discussões.

A todos os professores que fizeram parte da minha caminhada, que contribuíram com seus ensinamentos e conselhos, meu MUITO OBRIGADA!

A todos os amigos ao longo do meu mestrado pelas valiosas discussões e momentos divertidos

Ao LAQA, pela oportunidade de realizar as minhas pesquisas de mestrado e pelas amizades construídas.

À Welisson de Pontes por todo amor, dedicação e paciência e por sempre está me apoiando em todos os momentos.

À Mel, minha companheira canina, pelo amor incondicional e por todas as lições diárias.

Aos funcionários da UFPB pela presteza e atendimento quando nos foi necessário.

À Universidade Federal da Paraíba pela oportunidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, pela bolsa concedida.

A todos que contribuíram direta e indiretamente para a realização deste trabalho.

A todos, meu muito OBRIGADA!!!

Sumário

RESUMO.....	VIII
ABSTRACT.....	IX
LISTA DE FIGURAS.....	X
LISTA DE TABELAS.....	XII
LISTAS DE ABREVIATURAS E SIGLAS.....	XIII
1 INTRODUÇÃO.....	15
1.1 CARACTERIZAÇÃO DO PROBLEMA.....	15
2 OBJETIVOS.....	18
2.1 OBJETIVO GERAL.....	18
2.2 OBJETIVOS ESPECÍFICOS.....	18
3 FUNDAMENTAÇÃO TEÓRICA.....	20
3.1 INSULINA.....	20
3.1.1 <i>Insulina e Diabetes Mellitus</i>	21
3.1.2 <i>Tipos de insulina</i>	23
3.1.3 <i>Composição da insulina</i>	24
3.1.4 <i>Conservação e Armazenamento</i>	25
3.2 METODOLOGIAS DE ANÁLISE DA INSULINA HUMANA INJETÁVEL.....	27
3.4 ESPECTROSCOPIA NA REGIÃO DO INFRAVERMELHO PRÓXIMO (NIR).....	30
3.5 QUIMIOMETRIA.....	31
3.5.1 <i>Pré-Processamento</i>	32
3.5.2 <i>Técnicas de reconhecimento de padrões</i>	33
3.5.2.1 <i>Análise por componentes principais (PCA)</i>	34
3.5.2.2 <i>Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados (DD-SIMCA)</i>	35
3.5.2.2 <i>Mínimos Quadrados Parciais de uma classe (OC-PLS)</i>	37
4 METODOLOGIA.....	42
4.1 AQUISIÇÃO DE AMOSTRAS.....	42
4.2 AQUISIÇÃO DOS ESPECTROS NIR.....	42
4.3 PROCEDIMENTO QUIMIOMÉTRICO E SOFTWARE.....	43
5. RESULTADOS E DISCUSSÃO.....	46
5.1 ESPECTROS NIR (BANCADA).....	46
5.1.2 <i>Pré-Processamentos</i>	48
5.1.3 <i>Análise por componentes principais</i>	49

5.1.4 Classificação	51
5.2. ESPECTROS NIR (PORTÁTIL)	54
5.2.2 Pré-processamentos	55
5.2.3 Análise por componentes principais	57
5.2.4 Classificação	58
6. CONCLUSÕES E PERSPECTIVAS.....	62
6.1 CONCLUSÕES	62
6.2 PERSPECTIVAS	62
REFERÊNCIAS.....	63

Resumo

A insulina é um hormônio produzido pelo pâncreas e é necessário para a utilização de glicose (açúcar) pelo organismo como fonte de energia. Quando o pâncreas não produz uma quantidade de insulina suficiente capaz de controlar o açúcar no sangue, ocorre o diabetes. O tratamento da diabetes pode ser utilizando vários tipos de insulina de acordo com a necessidade do paciente. As insulinas apresentam boa estabilidade e tem sua ação biológica preservada, desde que sejam seguidas as orientações de conservação e transporte. Portanto, é de grande importância a quantificação e o controle de qualidade desse medicamento, para que possa ser utilizado de forma eficaz e mais seguro, uma vez que os pacientes podem ser prejudicados com a variação da sua eficácia. Sendo assim, este trabalho propõe o desenvolvimento de uma metodologia analítica para o controle de qualidade da insulina NPH, sem a violação da embalagem, utilizando espectrometria na região do infravermelho próximo com um equipamento de bancada e um portátil, associada à métodos quimiométricos de reconhecimento de padrão. Classificadores considerados do tipo “one class” Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados (DD-SIMCA) e Mínimos Quadrados Parciais de uma classe (OC-PLS) foram empregados como ferramentas de reconhecimento de padrão. Em termos de desempenho de classificação, o melhor resultado observado foi empregando-se DD-SIMCA ao pré-processamento suavização Savitzky-Golay (com janela de 41 pontos) com a correção de linha de base *offset* alcançando nível máximo na discriminação das amostras no NIR de bancada. Para o NIR portátil, tanto DD-SIMCA quanto OC-PLS alcançaram níveis máximos de desempenho, classificando corretamente as amostras utilizando os pré-processamentos dos dados: correção de espalhamento multiplicativo (MSC) com correção de linha de base *offset* e MSC com correção de linha de base linear (LBC), respectivamente. Portanto, a metodologia proposta representa uma ferramenta promissora, não destrutiva, rápida, e de baixo custo para a triagem da insulina humana NPH.

Palavras-chave: química, controle de qualidade, insulina, espectroscopia NIR, quimiometria, “one class”.

Abstract

Insulin is a hormone produced by the pancreas and is necessary for the use of glucose (sugar) by the body as an energy source. When the pancreas does not produce enough insulin to control blood sugar, diabetes occurs. The treatment of diabetes can be using various types of insulin according to the patient's need. Insulins have stability, good stability and their biological action is preserved, as long as conservation and transport guidelines are followed. Therefore, the quantification and quality control of this medication is of great importance, so that it can be used effectively and safer since patients can be harmed by varying its effectiveness. Therefore, this work proposes the development of an analytical methodology for the quality control of NPH insulin, without the violation of the packaging, using spectrometry in the near infrared (NIR) region with a bench and a portable equipment, associated with chemometric methods of pattern recognition. Classifiers “*one class*” considered *Data Driven-Soft Independent Modeling of Class Analogy* (DD-SIMCA) and *One-Class Partial Least Squares* (OC-PLS) were used as pattern recognition tools. In terms of classification performance, the best result observed was using DD-SIMCA for Savitzky-Golay smoothing pre-processing (with a 41-point window) with baseline offset (BO) reaching maximum level in the discrimination of the samples in the bench NIR. For the portable NIR, both DD-SIMCA and OC-PLS reached maximum performance levels, correctly classifying of the samples using data pre-processing: multiplicative scatter correction (MSC) with BO correction and linear baseline correction (LBC) with MSC, respectively. Therefore, the proposed methodology represents a promising tool, non-destructive, fast, and low-cost for the screening of NPH human insulin.

Keywords: chemistry, quality control, insulin, NIR spectroscopy, chemometrics, “*one class*”.

Lista de Figuras

Figura 1- Estrutura da insulina humana. Adaptado de [16].	22
Figura 2- Espectrofotômetro FTIR (Analyzer Diamond 20, Applied Instrument Technologies).	43
Figura 3- Espectros brutos das amostras de insulina NPH na região NIR entre 3000 a 12000 cm^{-1} .	46
Figura 4- Espectros brutos das amostras de insulina NPH na região NIR entre 4800 a 10000 cm^{-1} .	47
Figura 5- Espectros pré-processados das classes MAC (verde) e MAI (azul) com: a) Suavização Savitzky-Golay (SG) com janela de 41 pontos, LBC e MSC, b) SG+BO+SNV, c) SG+BO, d) SG+LBC+SNV, e) SG+LBC, f) SG+BO+MSC, g) derivação Savitzky-Golay (dSG) com primeira derivada, janela de 115 pontos e polinômio de segunda ordem, h) dSG+SNV e i) dSG+MSC.	49
Figura 6- Gráfico dos escores de PC1 versus PC2 para as 64 amostras de insulina estudadas usando os espectros pré-processados: a) suavização Savitzky-Golay (SG) com janela de 41 pontos, LBC e MSC, b) SG, LBC e SNV, c) derivação Savitzky-Golay (dSG) com primeira derivada, janela de 115 pontos e polinômio de segunda ordem, d) SG, correção BO e SNV, e) SG e correção BO, f) dSG e MSC, g) SG, correção BO e MSC, h) dSG e i) SG e LBC, (●: MAC, ●: MAI).	50
Figura 7- Gráficos de área de aceitação para o conjunto de (a,c) treinamento e (b,d) teste empregando o DD-SIMCA para o pré-processamento SG+BO e OC-PLS para dSG+MSC, respectivamente. As bolas verdes e azuis representam amostras de insulina acondicionadas corretamente e incorretamente, respectivamente. As linhas verdes em DD-SIMCA e as linhas vermelhas em OC-PLS representam a borda entre	

as amostras regulares e extremas; a linha vermelha em DD-SIMCA é a fronteira da área de aceitação.....	53
Figura 8- Espectros brutos das amostras de insulina NPH na região NIR entre 900 a 1700 nm.	54
Figura 9- Espectros pré-processados das classes MAC (verde) e MAI (azul) com: a) Correção BO e SNV, b) correção BO, c) LBC e SNV, d) LBC, e) MSC, f) Correção BO e MSC, g) derivação Savitzky-Golay (dSG) primeira derivada, janela de 9 pontos e polinômio de segunda ordem, h) LBC e SNV e i) LBC e MSC.....	56
Figura 10- Gráfico dos escores de PC1 versus PC2 para as 50 amostras de insulina estudadas usando os espectros pré-processados: a) derivação Savitzky-Golay (dSG) primeira derivada, janela de 9 pontos e polinômio de segunda ordem, b) LBC e SNV, c) SNV, d) LBC e MSC, e) MSC, f) correção BO e SNV, g) LBC, h) Correção BO e MSC e i) Correção BO: (●: MAC, ●: MAI).	57
Figura 11- Gráficos de área de aceitação para o conjunto de (a,c) treinamento e (b,d) teste empregando o DD-SIMCA para o pré-processamento BO+MSC e OC-PLS para LBC+MSC, respectivamente. As bolas verdes e azuis representam amostras de insulina acondicionadas corretamente e incorretamente, respectivamente. As linhas verdes em DD-SIMCA e as linhas vermelhas em OC-PLS representam a borda entre as amostras regulares e extremas; a linha vermelha em DD-SIMCA é a fronteira da área de aceitação.....	59

Lista de Tabelas

Tabela 1- Resultados dos parâmetros de desempenho sensibilidade e especificidade para DD-SIMCA e OC-PLS nos conjuntos de treinamento e teste para o NIR de bancada.51

Tabela 2- Resultados dos parâmetros de desempenho sensibilidade e especificidade para DD-SIMCA e OC-PLS nos conjuntos de treinamento e teste para o NIR portátil.58

Listas de Abreviaturas e Siglas

BO	Correção de linha de base <i>offset</i>
CLAE	Cromatografia Líquida de Alta Eficência
DD-SIMCA	Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados, do inglês <i>Data Driven-Soft Independent Modeling of Class Analogy</i>
DM	Diabetes <i>Mellitus</i>
EM	Espectrometria de massas
KS	Kennard-Stone
LBC	Correção de Linha de Base Linear, do inglês <i>linear baseline correction</i>
MAC	Medicamento Acondicionado Corretamente
MAI	Medicamento Acondicionado Incorretamente
MSC <i>correction</i>	Correção de Espalhamento Multiplicativo, do inglês <i>multiplicative scatter</i>
NIR	Infravermelho próximo, do inglês <i>Near Infrared</i>
NPH	<i>Neutral Protamine Hagedorn</i>
OC-PLS	Mínimos Quadrados Parciais de uma classe, do inglês <i>One-class - Partial least squares</i>
PCA <i>Analysis</i>	Análise por Componentes Principais, do inglês <i>Principal Component</i>
SNV	Variação Normal Padrão, do inglês <i>Standard Normal Variate</i>

Capítulo 1:

Introdução

1 INTRODUÇÃO

1.1 Caracterização do problema

A diabetes *mellitus* (DM) consiste em um distúrbio metabólico hormonal caracterizado por hiperglicemia persistente, decorrente de deficiência na produção de insulina quanto da ação da insulina [1]. Atualmente a diabetes é considerada uma epidemia por possuir alta prevalência e incidência em esfera mundial [2]. No Brasil, as taxas de DM também são crescentes, com potencial de se transformar em um grave problema de saúde pública [3]. Para controlar a DM tipo 1 quanto à tipo 2, várias estratégias são utilizadas incluindo planejamento nutricional, uso de fármacos, doses de insulina para monitoração da glicemia e prática regular de exercícios [3].

O tratamento utilizando a insulina é uma necessidade médica para pessoas que vivem com DM tipo 1 mas, muitas pessoas com DM tipo 2 também a utilizam para atingir as quantidades de glicemia individualizadas [4]. A dosagem correta da insulina é muito importante para então evitar flutuações hipoglicêmicas e hiperglicêmicas [4,5].

A insulina é uma proteína, utilizada no tratamento de todos os tipos de diabetes. É degradada facilmente e esse processo pode ser acelerado por altas temperaturas [6]. Assim, o armazenamento inadequado da insulina pode diminuir a sua eficácia e prejudicar sua ação farmacológica. No entanto, pacientes e pessoas que trabalham na saúde negligenciam com frequência os problemas com o armazenamento [6].

Além disso, há relatos na literatura em que um paciente apresentou fortes reações adversas, chegando até ao coma, sendo tratado no hospital e posteriormente recuperado. Antes do incidente, o paciente tinha feito o uso da insulina que possivelmente a dosagem estava errada. Uma má dosagem pode causar sérios problemas ao indivíduo podendo levar até a morte. Meses depois, a empresa que fabricou a insulina lançou uma nota de recolhimento de vários lotes, entre eles o utilizado pelo paciente. O motivo do recolhimento foi um problema relacionado à embalagem que fez com que a dose de insulina desviasse por até mais ou menos 50% do valor especificado [5].

Diante disto, considerando que a fabricação de produtos biológicos passa por várias etapas e cada uma delas com várias oportunidades de erros, é inevitável que alguns frascos do produto final apresentem defeitos [4,5]. Por isso, cabe a inspeção

do produto capturar ou remover os frascos defeituosos antes de chegar às mãos dos consumidores [5].

É de grande importância o controle de qualidade desse medicamento antes de ser distribuído.

Para determinar todos os parâmetros de qualidade de um processo industrial, geralmente, precisa-se de muitas análises. Atualmente, a maioria destas análises é realizada empregando métodos instrumentais, sendo a cromatografia líquida de alta eficiência acoplada a detectores como a espectrometria de massas ou ultravioleta (CLAE-EM/CLAE-UV) mais utilizada. Porém apresentam desvantagens como a necessidade de preparação de amostras que torna a análise mais demorada, é dispendiosa, pois requer reagente caros e é uma técnica invasiva por precisar destruir a amostra durante a análise [4,6].

Perante o exposto, este trabalho propõe o desenvolvimento de uma metodologia analítica para triagem de insulina humana NPH em relação ao estado de conservação, sem a violação da embalagem, utilizando espectrometria na região do infravermelho próximo (NIR) utilizando um equipamento de bancada e um portátil, associada à métodos quimiométricos de reconhecimento de padrão.

Capítulo 2:

Objetivos

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver uma metodologia simples, rápida, de baixo custo e não-invasiva baseada em espectrometria no infravermelho próximo e técnicas de reconhecimento de padrão, para triagem de insulina humana NPH em relação ao estado de conservação.

2.2 Objetivos Específicos

- ✓ Registrar espectros de infravermelho próximo (NIR) de amostras de insulina humana NPH acondicionadas corretamente e incorretamente utilizando um equipamento comercial de bancada e um comercial miniaturizado.
- ✓ Identificar as regiões espectrais características das amostras acondicionadas corretamente e incorretamente.
- ✓ Construir e validar diferentes modelos de classificação para amostras acondicionadas corretamente e incorretamente a partir dos espectros NIR obtidos pelos dois instrumentos comerciais.
- ✓ Utilizar os diferentes modelos de classificação validados para a previsão de amostras comerciais desconhecidas.

Capítulo 3:

Fundamentação

Teórica

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Insulina

A insulina é um hormônio produzido pelo pâncreas (glândula situada perto do estômago) necessária para a utilização de glicose (açúcar) pelo organismo como fonte de energia. Quando o pâncreas não produz insulina suficiente para suprir as necessidades do organismo, a glicose não utilizada fica acumulada no sangue, ocasionando o diabetes [7].

Depois de muitos anos de estudos, a insulina foi descoberta em 1921. A atribuição de sua descoberta é alvo de discussão [8]. Para o comitê do Prêmio Nobel a descoberta é atribuída a Frederick Banting que recebeu o prêmio de Fisiologia por este feito em 1923 [9]. Anos depois, em 1958, o cientista Frederick Sanger determinou a estrutura primária da insulina onde descobriu como os aminoácidos desta proteína se ligam nas duas cadeias polipeptídicas, e por esta descoberta ganhou o prêmio Nobel de Química de 1958 [10]. Em 1969, a determinação estrutural da insulina, ou seja, sua conformação espacial foi determinada por Dorothy Hodgkin usando difração de raios-X (mesma técnica que lhe rendeu o prêmio Nobel de Química de 1964 pela descoberta da conformação espacial da vitamina B12) [11].

A descoberta da insulina foi um dos mais importantes avanços da medicina, ainda que houvessem efeitos colaterais, devido às impurezas presentes [12,13]. Anos depois, a utilização de uma tecnologia apropriada proporcionou a produção de insulinas mais puras, diminuindo os efeitos colaterais, principalmente a alergia, com o surgimento da tecnologia de DNA recombinante, em 1978 [13]. Esta tecnologia usou uma bactéria comum e inseriu genes humanos no material genético deste microrganismo. Então, este microrganismo por sua vez produziu a proteína desejada, neste caso a insulina idêntica à humana [14].

A dose da insulina é expressa em “unidades internacionais” (UI) e sua extração pode se dar a partir do pâncreas bovino e do pâncreas suíno. Contudo, o uso de derivados animais apresenta algumas desvantagens, incluindo a imunogenicidade, dificuldades de obter suplementos suficientes e risco de contaminação. Além disso, a sequência de aminoácidos que formam a insulina é conservada de acordo com cada espécie, de forma que a distinção entre a insulina humana e a insulina suína se dá

apenas em um aminoácido e da insulina bovina se resume a três aminoácidos [15]. Entretanto, atualmente, grande parte da insulina utilizada é de origem humana e fabricada pela tecnologia do DNA recombinante onde houveram melhoras das propriedades farmacocinéticas e farmacodinâmicas específicas da proteína [16]. Nesta técnica, foi utilizado uma cepa da *Escherichia Coli*, não patogênica, onde um plasmídeo do DNA recombinante que codifica a insulina humana é introduzido na bactéria que é cultivada por fermentação, produzindo as cadeias A e B da insulina humana separadamente. Em seguida, as duas cadeias foram isoladas, purificadas individualmente e então ligadas quimicamente, constituindo uma cadeia de insulina equivalente à derivada do pâncreas humanos [15,17].

3.1.1 Insulina e Diabetes *Mellitus*

Diabetes *Mellitus* (DM) é uma doença metabólica crônica caracterizada principalmente pela hiperglicemia devido a deficiência da insulina ou aumento da resistência à insulina ou ambas. A mesma pode ser dividida em dois tipos: diabetes *mellitus* tipo 1 e diabetes *mellitus* tipo 2 [18,19]. De acordo com a Sociedade Brasileira de Diabetes, há mais de 13 milhões de pessoas vivendo com diabetes, o que representa 6,9% da população, e esse número continua crescendo [20].

A diabetes *mellitus* tipo 1 ocorre geralmente na infância ou adolescência, mas pode ser também diagnosticada em adultos. Este tipo de diabetes é caracterizada quando as células β do pâncreas responsáveis pela síntese da insulina são atacadas pelo sistema imunológico equivocadamente, provocando uma produção insuficiente do hormônio insulina ou até mesmo a não produção [21]. Em consequência, a glicose fica acumulada no sangue, em vez de ser usada como energia, resultando na hiperglicemia. Portanto, há a necessidade de reposição externa da insulina para que haja o controle das taxas glicêmicas, ou seja, evitar a hiperglicemia [20–23].

Já a diabetes *mellitus* tipo 2 é mais frequente em adultos, mas crianças também podem apresentar. Neste tipo de diabetes há produção de insulina e o organismo não consegue aproveitar adequadamente, ou seja, existe uma resistência do organismo a sua ação, como também o hormônio não é produzido suficiente para controlar a taxa de glicemia [20,21,24,25]. À medida em que a resistência à insulina se desenvolve, o corpo responde produzindo mais insulina, com isso as células β no pâncreas que

trabalharam muito para a produção, se desgastam e não conseguem mais acompanhar a demanda por insulina, ocasionando o aumento de açúcar no sangue e desenvolvendo a diabetes do tipo 2. A insulina é considerada tratamento primário para pacientes da diabetes tipo 1 mas não é considerado tratamento primário no caso da tipo 2, pois geralmente é administrada em alguns casos, por exemplo, quando os medicamentos orais, hiperglicêmicos, falham em produzir os resultados efetivos [21]. A realização de atividades físicas, planejamento alimentar, melhoram o controle glicêmico, mas também possuem um impacto positivo na prevenção deste tipo de diabetes [20,26].

A insulina foi a primeira proteína que determinaram completamente a sequência exata de aminoácidos da sua estrutura molecular. Por ter sido o primeiro a esclarecer uma sequência de aminoácidos de uma proteína, Frederick Sanger ganhou o Prêmio Nobel em Química de 1958. Desde então, a insulina começou a ser aplicada como modelo para analisar propriedades estruturais das proteínas, surgindo muitas descobertas valiosas [18]. A molécula de insulina humana consiste em 51 aminoácidos, incluindo duas cadeias peptídicas (A e B), que contém uma ligação dissulfeto dentro da subunidade e duas ligações de dissulfeto entre as subunidades. Os aminoácidos que compõem a estrutura da insulina humana são: G- glicina; V- valina; E- ácido glutâmico; Q- glutamina; C- cisteína; A- alanina; S- Serina; F- fenilalanina; I- isoleucina; N- asparagina; H- histidina; T- treonina; L- leucina; Y- tirosina; P- prolina; K- lisina; I- isoleucina [16]. Geralmente, a cadeia A é constituída de 21 resíduos de aminoácidos e a cadeia B de 30 resíduos de aminoácidos, com massa molar de aproximadamente 5.800 Da [17,27]. A **Figura 1** é apresentada a estrutura insulina humana.

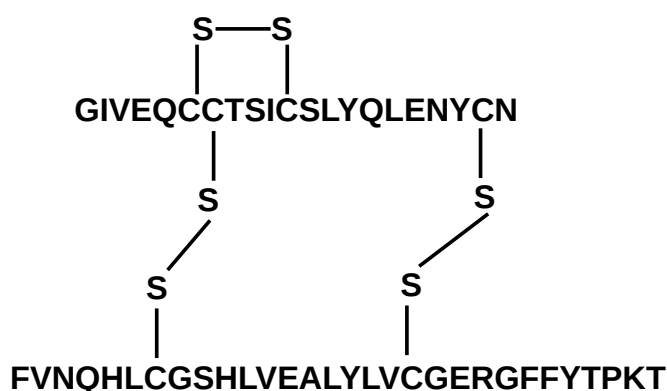


Figura 1- Estrutura da insulina humana. Adaptado de [16].

3.1.2 Tipos de insulina

Atualmente, existem vários tipos de insulina disponíveis comercialmente para o tratamento da diabetes. O modo de ação das insulinas são semelhantes, porém a sua farmacocinética é diferente, ou seja, elas se diferenciam pelo tempo em que ficam ativas no corpo, pelo tempo em que começam a agir e são aplicadas de acordo com a situação do dia em que elas venham a ser mais eficientes [20,28]. A reposição de insulina no corpo pode ser aplicada com auxílio de caneta, seringa e bomba de insulina [29].

As insulinas podem ser classificadas em análogas e humanas [20]. As insulinas mais modernas são chamadas de insulinas análogas, pois são produzidas a partir da insulina humana e então modificadas, utilizando técnicas de engenharia genética, de modo a terem tempos ação diferentes no organismo. Essa modificação se dá através da substituição de aminoácidos da estrutura da insulina humana. Entre os diversos tipos de insulinas análogas, podemos destacar a insulina glargina, insulina aspart, insulina lispro, insulina glulisina e a insulina detemir [20,28,30].

A insulina glargina e a insulina detemir são análogos da insulina humana, possuem início de ação mais lenta, porém permite uma duração mais prolongada [20,28,30]. A insulina glargina é semelhante à insulina humana na estrutura, exceto pela adição de dois resíduos de aminoácidos nas extremidades da cadeia B. São duas moléculas de argininas adicionadas nas extremidades da cadeia B nas posições 31 e 32, como também, a substituição do aminoácido asparagina por glicina na posição 21 da cadeia A da molécula [28,30]. Possui uma duração de ação de 24 horas, sem pico, e apresenta um bom controle glicêmico [20,30]. Já a remoção do aminoácido treonina da cadeia B na posição 30, com a adição do ácido mirístico a lisina na posição 29 da cadeia B, desenvolveu a insulina detemir [31].

A insulina lispro e a insulina aspart apresentam uma ação rápida, ou seja, tem um efeito imediato assim que administrada, porém possuem uma duração mais curta no organismo comparada as insulinas glargina e detemir [20,28]. Apresentam também semelhanças com a estrutura da insulina humana, porém, na insulina lispro há uma troca de posições entre as moléculas prolina e lisina da cadeia B, nas posições 28 e 29. Já na insulina aspart, há uma substituição do aminoácido prolina pelo ácido aspártico na posição 28 da cadeia B [28].

A insulina glulisina foi desenvolvida substituindo o aminoácido asparagina pela lisina na posição 3 e a substituição da lisina por ácido glutâmico da posição 29, ambos na cadeia B [31]. É um análogo considerado de ação rápida, absorvido rapidamente pelo organismo [20,31].

Como exemplos da insulina humana temos a insulina regular e a NPH (Neutral Protamine Hagedorn), também desenvolvidas em laboratório a partir da tecnologia do DNA recombinante. A insulina chamada regular é idêntica à insulina humana na sua estrutura e considerada de ação rápida com duração de quase 7 horas. Já a insulina humana recombinante NPH, possui uma ação intermediária e é considerada de longa duração em comparação a insulina regular [20].

A insulina NPH (Neutral Protamine Hagedorn), foi criada pelo cientista Hans Christian Hagedorn (H), contém concentrações de protamina que é uma proteína (P) e pH neutro (N) [32]. Foi obtida a partir da combinação da insulina humana recombinante na presença de zinco e protamina, o que fez com que a sua solubilidade diminuísse e sua liberação passou a ser de forma lenta, com duração maior no organismo. É uma suspensão isófana de insulina humana e sua administração é subcutânea [14,31]. O termo isófano denota que o preparado contém quantidades estequiométricas de insulina protamina. Essa preparação foi formulada para evitar um problema de excesso de protamina, que causa dissolução excessivamente lenta da insulina [32].

3.1.3 Composição da insulina

Neste trabalho, já foi reportado que a insulina é uma proteína. As proteínas são polímeros naturais, macromoléculas biológicas, caracterizados pela junção química de curtas unidades repetidas de aminoácidos que são denominadas de monômeros, quando isoladas [33,34].

Já são centenas de aminoácidos conhecidos, mas a partir de 20 aminoácidos diferentes, a natureza criou uma quantidade impressionante de proteínas distintas, incluindo enzimas, anticorpos, hormônios peptídicos e proteínas com outras funções [34]. Além dos 20 aminoácidos formadores de proteínas bem conhecidos, são recuperados vários aminoácidos raros a partir da hidrólise de proteínas. Por definição,

um aminoácido contém pelo menos um grupo amino ($-\text{NH}_2$), um grupo carboxila ($-\text{COOH}$) e uma cadeia lateral (grupo R) ligado a um átomo de carbono [35,36].

Estão amplamente presentes nas proteínas grupos carboxila e ânions carboxilatos. Sob condições fisiológicas, ($\text{pH} \approx 7,4$) os grupos carboxilas são geralmente desprotonados, ou seja, o grupo carboxila é ionizado, formando o íon carboxilato carregado negativamente ($-\text{COO}^-$), mas em alguns ambientes específicos eles ainda podem ser encontrados na forma protonada [37,38]. E nessas mesmas condições, o grupo amino encontra-se protonado carregado positivamente ($-\text{NH}_3^+$) [38]. A síntese das proteínas envolve a formação de ligações peptídicas formadas pela combinação desses dois grupos, liberando uma molécula de água, de maneira que quase não se tem esses tipos de íons para possíveis reações químicas, a não ser pela eventualidade de formação de pontes de hidrogênio [38,39]. Deste modo, a cadeia lateral (grupo R) é quem vai determinar o papel dos aminoácidos em uma reação química, ou seja, ela é responsável pelas propriedades físico-químicas e fisiológicas do aminoácido [36,39]. A diversidade estrutural das cadeias laterais permite a formação de proteínas com uma variedade de estruturas e funções. E essa diversidade não é apenas pelo pK_a dos aminoácidos, mas também pelo seu tamanho e hidrofobicidade. Os aminoácidos com cadeias laterais alifáticas ou aromáticas mais longas, como a isoleucina, leucina, e fenilalanina têm maior hidrofobicidade que as de cadeias laterais mais curtas, como o grupo metil encontrado na alanina. Os aminoácidos neutros com grupos polares como hidroxila ou amina em suas cadeias laterais são mais hidrofílicos. Os aminoácidos ácidos têm cadeias laterais com ácidos carboxílicos, já os aminoácidos básicos com grupos amino, guanidino e imidazol. A cadeia lateral de tiol ($-\text{SH}$) da cisteína oxida-se facilmente e pode se ligar a outras moléculas via ponte de dissulfeto [38].

3.1.4 Conservação e Armazenamento

A insulina, da mesma forma que outros medicamentos à base de hormônios peptídicos, é sensível à temperatura e diante das condições de armazenamento sua estabilidade é afetada. Isso é refletido nas recomendações de armazenamento de insulina e a importância do controle de temperatura ao longo da distribuição [29,31].

A estabilidade pode ser definida como sendo os fatores de qualidade de um produto farmacêutico específicos passíveis a alterações durante seu armazenamento e provavelmente influenciam sua qualidade, eficácia, pureza e segurança [31]. Para ter eficácia a insulina necessita de uma série de cuidados que preservem sua viabilidade em todo o seu tempo de uso. Um desses cuidados é a sua conservação, que deve ser adequada tanto nos centros de distribuições, farmácias, hospitais, veículos de entrega quanto em residências [29,40].

As pessoas com diabetes em todo o mundo carregam a insulina diariamente, em frascos, canetas ou bombas próximas ao corpo e armazenam o seu suprimento de insulina em geladeiras domésticas. Regularmente, a insulina está sujeita a fatores de risco que podem afetar a sua eficácia quando em uso, como temperatura ambiente alta, luz solar direta e agitação mecânica. Além de altas temperaturas, o congelamento também deve ser evitado, pois isso pode afetar a integridade da formulação bem como o frasco, como exemplo, o congelamento pode provocar pequenas rachaduras no vidro, e com isso, prejudicar sua higiene [40].

A temperatura de armazenamento da insulina é cuidadosamente gerenciada e controlada em todas as etapas da cadeia de fornecimentos por fabricantes, distribuidores, profissionais e autoridades de saúde. Para garantir a qualidade ideal, na distribuição, a insulina precisa ser mantida refrigerada entre 2 e 8 °C. Quando a mesma chega às farmácias e hospitais, deve ser armazenada em refrigeradores com temperatura controlada e monitorada [29,40].

No final, os portadores de diabetes são quem vão manipular a insulina quando for entregue pela farmácia e o folheto informativo da embalagem se torna a principal fonte de informações quanto sua estabilidade. Depois de entregue, a insulina é mantida resfriada até começar o uso do frasco, bomba ou caneta. E as instruções do folheto informativo são manter a insulina em uma temperatura entre 2 e 8 °C até a sua validade e não congelar [29,40]. Esta temperatura de armazenamento também é recomendada pelos fabricantes das insulinas utilizadas neste trabalho, como também os frascos de insulina quando em uso e o portador que não possuir uma geladeira, eles podem ficar em temperatura ambiente (15 °C a 30 °C) por até 4 semanas [29,40]. Lembrando que as recomendações podem ser diferentes dependendo dos fabricantes, dos tipos de insulinas e das diferentes formas de aplicação [29,40].

No entanto, muitos desses portadores de diabetes vivem em locais sem acesso à refrigeração, e informações atuais apontam que mesmo as condições de

armazenamento de insulina nos refrigeradores domésticos normalmente não atendem com êxito às recomendações dos fabricantes. Como também, a maioria desses portadores armazenam quantidades consideráveis de insulina para suprir suas necessidades na geladeira por um período de muitos meses e desse modo, isso representa um risco à propriedade da insulina que seria inaceitável pelos padrões profissionais [40].

Projetados para armazenar alimentos e otimizados para a segurança alimentar, os refrigeradores domésticos não possuem circulação de ar, o que leva a um gradiente de temperaturas e a criação de zonas quentes e frias. De outro modo, os refrigeradores farmacêuticos possuem ventiladores para a circulação de ar, não apresentando flutuações na temperatura de armazenamento [40].

Os portadores de diabetes passam por problemas com a insulina frequentemente. O controle glicêmico dos portadores de diabetes é mantido quando a dose exata é administrada, mas quando há uma diminuição na potência biológica da insulina ao longo do tempo devido às condições inadequadas de armazenamento, alterações como a sensibilidade à insulina são percebidas por esses portadores e precisam ajustar a dose conforme a necessidade. Logo, alterações na potência da insulina podem contribuir para essa variabilidade observada nos níveis de glicose, no entanto, atualmente esse fator não é suficientemente considerado, sendo necessária uma análise laboratorial para verificar se a potência da insulina foi afetada [40].

Por consequência, é complicado para os portadores de diabetes avaliarem quais riscos eles enfrentam ao usar insulina que foi ou pode ter sido armazenada fora das recomendações. Deste modo, uma aplicação ineficaz pode gerar danos ao usuário, uma vez que o mesmo estaria vulnerável a taxa de hiperglicemia e a complicações da doença [29,40].

3.2 Metodologias de análise da insulina humana injetável

O método oficial (farmacopeico) de análise de insulina humana é baseado na utilização da Cromatografia Líquida de Alta Eficiência (CLAE) com detector ultravioleta [41]. Apesar de ser uma técnica com alta precisão e exatidão, há necessidade de preparação de amostra que é demorada, e dispendiosa, pois, gera gastos com

reagentes químicos para as preparações das soluções padrão e requerem instrumentos sofisticados para detecção dos analitos.

Vários procedimentos são descritos na literatura enfocando determinação da insulina humana injetável em formulações farmacêuticas. Carter e Heinemann [42] levantaram preocupações sobre o potencial impacto da cadeia global de fornecimento de frio na qualidade e quantidade de insulina. Pois, existe uma falta de confirmação independente disso (por exemplo, a concentração intacta de insulina nos frascos/cartuchos) no final da cadeia de suprimentos a frio quando comprada por pacientes com diabetes nas farmácias. Em um estudo, mediram quantitativamente a insulina em frascos de insulina humana multidoses adquiridos aleatoriamente pelo método analítico padrão método UPLC-MS. Dezoito frascos de 10 mL de dois fabricantes (M1 e M2) foram adquiridos aleatoriamente. A concentração de insulina variou de 13,9 a 94,2 U/mL, com média de 40,2 U/mL. Nenhum frasco atingiu o padrão mínimo de 95 U/mL. Esses resultados sugeriram que a cadeia de suprimentos a frio afeta as concentrações de insulina em maior extensão do que o previsto e que os pacientes estão pagando altos preços pela insulina e devem receber frascos de insulina com conteúdo adequado de insulina em troca.

O artigo de Carter e Heinemann [42] levantou sérias preocupações sobre as concentrações de insulina em frascos vendidos nas farmácias dos EUA. Então, Moses e colaboradores [43] em um novo estudo avaliaram as concentrações de insulina dentro dos frascos para injetáveis de três fontes diferentes ao longo da cadeia de distribuição usando uma metodologia atualmente aceita pela Farmacopeia dos Estados Unidos com a cromatografia líquida de alta pressão de fase reversa (RP-HPLC). Neste estudo, os dados demonstraram que, sem exceção a qualidade da insulina com base nos dados de estabilidade foi mantida, que os dados sobre a qualidade da produção de insulina confirmam que a insulina fornecida às farmácias e quando utilizada pelos pacientes atende aos padrões regulamentares exigidos e não fornece evidência de preocupação com o conteúdo de insulina. Os resultados também incluíram análises de amostras de insulina que foram devolvidas à Novo Nordisk “por justa causa” e estavam fora da cadeia de frio recomendada durante seu retorno para a Novo Nordisk. Esse é exatamente o tipo de amostra que pode ser previsto conter insulina degradada, uma vez que o transporte dessas amostras ocorre por via postal, mas o conteúdo de insulina atendeu às especificações.

Neumiller e colaboradores [4] propuseram avaliar se a potência da insulina humana estaria abaixo do padrão mínimo permitido de 95 U/mL e comparar os resultados obtidos através de dois métodos diferentes, a Cromatografia Líquida de Ultra Performance acoplada com detector Ultravioleta (UPLC-UV) e Cromatografia Líquida de Ultra Performance acoplada a Espectrometria de Massas (UPLC-MS). Para a quantificação da insulina pelo método UPLC-UV as amostras foram analisadas com uma coluna C18 (Cortecs, 1,6 μ m, 2,1 x 50 mm) a 40 °C. A fase móvel, no modo eluição isocrática, de 74% de tampão A (2,7 mL de ácido fosfórico e 28,6 g de Na₂SO₄ em 1000 mL de água) e 26% de tampão B (100% de acetonitrila) por seis minutos com a taxa de fluxo definida em 1 mL/min. O comprimento de onda de detecção de UV foi fixado em 214 nm e o volume de injeção da amostra foi de 5 μ L. Já para a quantificação da insulina humana pelo método UPLC-MS que foi equipado com um gerenciador de solventes Acquity UPLC, um amostrador automático FTN, uma coluna UPLC C18 (Cortecs, 1,6 μ m, 2,1 x 50 mm) e um espectrômetro de massa Xevo-G2S QToF. A temperatura no gerenciador de amostras foi ajustada em 15 °C e a temperatura da coluna foi de 60 °C. Devido à alta sensibilidade do espectrômetro de massa, o volume de injeção da amostra foi ajustado em um valor baixo de 0,2 μ L. A taxa de fluxo de LC foi de 0,25 mL/min. A condição LC inicial era de 85% de tampão A (0,1% de ácido fórmico) e 15% de tampão B (0,1% de ácido fórmico em acetonitrila), com um gradiente linear para 40% de B em quatro minutos, seguido de um gradiente isocrático a 40% de B por um minuto e reequilibrar com 85% de tampão A por um minuto antes da injeção da próxima amostra. A avaliação mostrou que a grande maioria das amostras estava acima do limite mínimo de concentração de 95 U/mL quando medido por UPLC-UV. Quando analisadas por UPLC-MS, as concentrações medidas de insulina foram ligeiramente inferiores, com uma concentração média medida de 91,5 U/mL para 60 frascos de insulina testados.

Algumas técnicas não destrutivas também vêm sendo utilizadas na análise de medicamentos. Briggs e colaboradores [5] utilizaram RMN no domínio de tempo (RMN de bancada) para o controle de qualidade de insulina. Um analisador de RMN de bancada foi utilizado para a inspeção do medicamento com uma taxa de relaxação transversal do próton da água (¹H₂O), R₂ (¹H₂O), para acessar a uniformidade do conteúdo das canetas de insulina e explorar seu potencial para inspecionar quantitativamente, mas não destrutivamente, todos os frascos em um lote de um medicamento. A intensidade do campo magnético foi de 0,5 T, com uma frequência

de ressonância de ^1H de 23,8 MHz. A temperatura do ímã era de 25 °C e o tempo de inspeção para cada frasco foi de 30 a 60 s. A sonda 1193 Pro 26 mm acomodava facilmente frascos da insulina, com um diâmetro externo de 25 mm ou menos. Esta metodologia proporcionou a análise do medicamento sem a necessidade de sua destruição e ao mesmo tempo, a análise de todas as amostras do lote em poucos segundos sem a necessidade de um preparo de amostras que consequentemente deixaria a análise mais dispendiosa e demorada.

Como visto, a método de referência para análise da insulina, a CLAE, exige a violação da embalagem dos medicamentos como também o seu consumo durante a análise, além de necessitar de muita atenção na sua manipulação analítica.

Então, para superar estes problemas, é que se propõe o desenvolvimento de métodos analíticos usando a espectrometria NIR para a análise deste medicamento. Esta metodologia permite a realização de análises das amostras diretamente em suas embalagens, evitando a destruição das amostras e diminuindo a sua manipulação pelos analistas.

3.4 Espectroscopia na região do infravermelho próximo (NIR)

A espectroscopia na região do infravermelho é uma técnica analítica que envolve basicamente a interação da luz com a matéria. Onde para que as moléculas de uma amostra ou de compostos absorvam energia de radiação infravermelho é necessário uma variação no momento de dipolo relacionado ao movimento rotacional ou vibracional das moléculas. Os perfis espectrais estão fundamentados na combinação de bandas de vibrações moleculares e por sobretons [44].

O infravermelho pode ser dividido em três regiões: o infravermelho próximo (NIR, do inglês *Near Infrared*) cuja faixa é de 12800 – 4000 cm^{-1} , o infravermelho médio (MIR) onde a faixa compreende 4000 – 200 cm^{-1} e o infravermelho distante (FIR) com faixa de 200 – 10 cm^{-1} [45].

A radiação NIR apresenta baixa energia e a sua interação com a matéria produz transições no mesmo estado fundamental de vibração. As bandas de absorção com maior intensidade que aparecem na região do NIR estão relacionadas às alterações do momento dipolar de uma molécula (bandas de combinação, sobretons) [46]. Os

sobretons estão relacionados as variações nos níveis vibracionais e apresentam intensidade fraca, cerca de 10 a 100 vezes inferiores que as observadas para as transições fundamentais observadas no MIR e FAR. Já as combinações ocorrem quando dois estados vibracionais diferentes são excitados simultaneamente por um fóton. De tal forma que, à soma ou à diferença entre as duas frequências das bandas fundamentais que interagem para a sua formação é aproximadamente igual a frequência da banda de combinação. Dessa forma o NIR apresenta menor definição de bandas com um perfil mais voltado à análise quantitativa [46].

Como as absorções no NIR estão relacionados a alterações do momento dipolar de uma molécula, moléculas diatômicas homonucleares não são detectadas num espectro de NIR enquanto compostos com ligações químicas heteronucleares são facilmente detectados. Os perfis espectrais presentes no NIR tem aplicações bastante abrangentes, compreendendo moléculas que contenham ligações C-H, N-H, O-H, S-H, por exemplo [47].

O NIR é uma técnica não destrutiva e o fato de não necessitar de preparação de amostras poupa reagentes e tempo. No entanto, os espectros de NIR são até certo ponto complexos e dependendo da amostra não são de interpretação direta pela possibilidade de apresentar sobreposição de bandas de absorção. Porém o uso de pré-processamentos e modelos de calibração podem ser usados para extrair as diferenças espectrais e auxiliar na identificação. Dessa maneira, métodos quimiométricos multivariados são extremamente importantes para o uso da espectroscopia do infravermelho, sobretudo na faixa do NIR [48].

3.5 Quimiometria

Em 1971, o termo “Quimiometria” foi introduzido por Svante Wold. Posteriormente, em 1974, em companhia de Bruce Kowalski, lançaram a Sociedade Internacional da Quimiometria (*International Chemometrics Society*), que definia a quimiometria como a ciência relacionada às medidas efetuadas em um sistema ou processo químico, buscando informações sobre o estado do sistema através da aplicação de métodos matemáticos e estatísticos [49].

O surgimento da quimiometria tem proporcionado o desenvolvimento de vários métodos que permitem a extração de informações, identificação de certos padrões e

características de interesse, reduzindo uma grande quantidade de dados que antes seriam impossíveis de serem analisados [50,51]. Sendo assim, melhorando a confiabilidade das análises e facilitando a interpretação dos resultados [50]. Como exemplo, dados produzidos por um espectrofotômetro, na região do infravermelho próximo (NIR) são complexos dificultando a interpretação espectral e a previsão de parâmetros quantitativos, principalmente quando ocorre a presença de interferentes. Diante disto, para melhorar a utilização e a interpretação dos espectros NIR, abordagens estatísticas ou análises quimiométricas multivariadas são desenvolvidas e empregadas [51].

Neste trabalho, técnicas de reconhecimentos de padrões são empregadas. É de suma importância avaliar os diferentes tipos de pré-processamentos dos dados adquiridos antes mesmo da aplicação das técnicas quimiométricas pois, a privação desta etapa pode comprometer a execução dos modelos construídos.

3.5.1 Pré-Processamento

Para que as técnicas quimiométricas sejam aplicadas, é necessário a escolha de um pré-processamento adequado. O pré-processamento nos dados é uma etapa muito importante para limpar e homogeneizar os dados antes da análise, visto que, a aquisição dos dados pode ser afetada por variações nas condições operacionais, instabilidade instrumental, ruídos instrumentais, manuseio do operador e espalhamento da luz produzido pelo instrumento durante a análise [52,53].

A fim de minimizar variações obtidas durante a aquisição dos dados, diferentes ferramentas de pré-processamentos são utilizadas, dentre as quais destacam-se as derivadas Savitzky-Golay, suavização, correção de espalhamento multiplicativo (MSC, do inglês *multiplicative scatter correction*), variação normal padrão (SNV, do inglês *Standard Normal Variate*) e correção de linha de base [51,52].

O Polinômio de Savitzky é o mais utilizado para suavizar e derivar os dados [54]. A suavização dos dados reduz o ruído aleatório e aumenta a relação sinal/ruído presente no conjunto de dados antes da modelagem [55]. As derivadas são pré-processamentos bastante utilizados, sendo que, a primeira derivada é muito empregada quando se tem problemas associados variações de linha de base ou espalhamento da radiação, promovendo uma melhor separação das bandas

existentes nos sinais originais [51]. Já a segunda derivada, por sua vez, remove variação linear de linha de base associadas aos espalhamentos [54]. As derivadas são usadas principalmente para resolver sobreposição de picos e eliminar alterações constantes e lineares da linha de base entre as amostras. A derivação por Savitzky Golay inclui dois parâmetros, onde, na suavização há o ajuste do polinômio à um tamanho de janela [56]. Recomenda-se ter atenção com a qualidade dos espectros que pretende empregar o cálculo da derivada, dado que não só os sinais espectrais, como também os ruídos, torna-se mais acentuados [56].

O MSC é outro método de pré-processamento que tem sido utilizado para reduzir os efeitos multiplicativos da dispersão de luz na amostra e os efeitos de linha de base [57,58]. Este método opera com base na reversão do espectro medido em comparação com o espectro de referência, corrigindo o espectro medido usando a inclinação (e, possivelmente a interceptação). Portanto, um espectro corrigido é semelhante ao espectro de referência, bem como, com relação à escala e deslocamento de linha de base [59].

Outro método de pré-processamento regularmente usado para diminuir os efeitos multiplicativos é o SNV [52]. Semelhante ao MSC, o SNV também é útil para minimizar a variação nos espectros causada pelo tamanho das partículas e espalhamento [57]. No método SNV é necessário calcular a média e desvio padrão de todo o espectro. Em seguida, faz-se a subtração do espectro bruto pela média e divide pelo desvio padrão para obter o espectro corrigido [60].

3.5.2 Técnicas de reconhecimento de padrões

As técnicas de reconhecimento de padrões (RP) têm como objetivo analisar uma possível existência de semelhanças e dissemelhanças entre grupos de amostras que estão vinculadas a um determinado tipo de diagnóstico [61,62]. Essas técnicas podem ser classificadas em dois tipos: não supervisionadas ou supervisionadas.

1. Não Supervisionadas: efetuada para verificar os agrupamentos característicos apenas com os dados medidos, sem levar em consideração rótulos e informações sobre as classes [63]. A análise de componentes principais (PCA, do inglês *Principal Component Analysis*)

e análise de agrupamento hierárquico (HCA, do inglês Hierarchical Cluster Analysis) são técnicas de RP não supervisionadas [64]. Apenas a PCA foi avaliada, neste trabalho.

2. Supervisionadas: usadas para construir modelos a partir de um conjunto de amostras conhecidas, designadas conjunto de treinamento, para identificar as informações responsáveis por dividir as amostras em diferentes grupos [63]. São amplas as aplicações das técnicas de reconhecimento de padrões supervisionadas, tais como impressões digitais [53], avaliação da qualidade de alimentos [65], detecção de falsificação bem como na indústria farmacêutica [66], etc. São exemplos de técnicas RP supervisionada bem difundidas na literatura: Modelagem Independente e Flexível por Analogia de Classes (SIMCA, do inglês *Soft Independent Modeling of Class Analogy*); Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados (DD-SIMCA, do inglês *Data Driven - Soft Independent Modeling of Class Analogy*); Modelagem por Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*); Mínimos Quadrados Parciais de uma classe (OC-PLS, do inglês *One-class Partial least squares*); Análise Discriminante Linear (LDA, do inglês *Linear Discriminant Analysis*); Algoritmo das Projeções Sucessivas acoplado com Análise Discriminante Linear (SPA-LDA, do inglês *Linear Discriminant Analysis with wavelength selection by the Successive Projections Algorithm*) [67–72].

3.5.2.1 Análise por componentes principais (PCA)

A análise de componentes principais (PCA) é a técnica principal para análise exploratória dos dados e é regularmente utilizada para identificar padrões nos dados medidos [51]. A PCA é uma técnica que reduz principalmente a dimensionalidade do conjunto de dados sem perder nenhuma informação dos dados originais [51,73]. Então, a PCA substitui as variáveis originais correlacionadas, em novas variáveis não correlacionadas, denominadas componentes principais (PC's) que contêm toda a

informação. Essas novas variáveis são ortogonais entre si e são classificadas em ordem decrescente de importância [51,74]. As PC's são obtidas a partir de uma combinação linear das variáveis originais medidas [53]. A partir dessas componentes principais, a primeira componente principal (PC1) explica a máxima variância presente no conjunto de dados, enquanto a segunda (PC2) explica o máximo variância restante e não tem nenhuma correlação com a PC1, sendo ortogonal à mesma [75]. E as seguintes PC's são ortogonais às anteriores e direcionadas de tal modo que representam sempre a máxima variância restante [76]. As informações da componente principal de menor variância contém principalmente ruído e removê-la é semelhante a filtrar o ruído do conjunto de dados. Por essa razão, as primeiras componentes principais geralmente são selecionadas para análise de problemas [77].

Os coeficientes lineares da relação inversa das combinações lineares são chamados de pesos (*loadings*), isto é, os coeficientes de correlação entre as componentes principais e as variáveis originais. Os valores que representam as amostras no espaço definido pelas componentes principais são os escores (*scores*). Os escores podem ser usados como entrada para outras técnicas de análise multivariada, em vez das variáveis originais medidas [74,76] .

3.5.2.2 Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados (DD-SIMCA)

A Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados (DD-SIMCA) é considerada um dos tipos de classificadores “*one class*” (de uma classe) [78]. Então, nesses tipos de classificadores vai ser determinado se uma amostra pertence ou não à classe alvo [79].

Em DD-SIMCA, para cada classe de interesse, denominada de classe alvo, é construído um modelo de classificação. A classe alvo é definida como sendo o conjunto representativo das amostras (MAC) como utilizadas neste trabalho. Essas amostras participam tanto do conjunto de treinamento quanto do conjunto de teste que são utilizados para desenvolver um modelo e para estimar a sensibilidade deste modelo, que é a parcela de amostras corretamente identificadas da classe alvo [80]. Na primeira etapa, O DD-SIMCA aplica a análise por componentes principais (PCA) nos dados de treinamento da classe alvo. O número de componentes principais (PC's)

determina a complexidade do modelo e influencia principalmente na qualidade da classificação. Quanto maior for o número de PC's, maior será a parte da variação dos dados de treinamento explicada pela decomposição de PCA. Ao mesmo tempo, a inserção de componentes desnecessárias é capaz de originar-se um modelo que não considera apenas as principais características da classe, mas também o ruído irrelevante. Sendo assim, o princípio da parcimônia é regularmente empregado para selecionar a complexidade do modelo [80,81].

Já na segunda etapa, o DD-SIMCA calcula as distâncias de escores (SD) e as distâncias ortogonais (OD) para cada objeto do conjunto de treinamento com base na dimensão de PC's significativas. Essas distâncias são empregadas para estabelecer uma distância crítica em um determinado nível de confiança utilizado como limite ou área de aceitação [80,81]. Cada tipo de distância é representado usando uma distribuição qui-quadrada, cujos graus de liberdade são estimados com os dados [79]. Então, essas duas distribuições qui-quadradas são utilizadas para definir uma área de aceitação para um determinado erro, erro tipo I [81]. Desse modo, área de aceitação definida é empregada para a classificação do novo conjunto de dados (amostras de teste) [81]. Caso haja a possibilidade de uma nova classe disponível, o DD-SIMCA fornece a chance de calcular o erro tipo II, com relação à esta classe [80].

Para a construção de modelos de uma classe, duas abordagens podem ser aplicadas. A primeira é chamada de “rigorosa” e é criada com base apenas no conjunto de dados de treinamento. E as condições ótimas são obtidas usando o erro tipo I, que consiste de uma amostra não ser colocada dentro da sua própria classe [79]. Já a segunda é chamada de “concordante”, em que informações complementares sobre as amostras não alvo são utilizadas, nos casos em que um ou mais conjuntos de dados de classes não alvo estão à disposição. O erro tipo II, consiste de uma amostra “estranha” ser atribuída à classe alvo (falso positivo) [79]. Nessa abordagem “concordante”, o modelo ideal é determinado considerando os dois tipos de erro, I e II, para selecionar a máxima eficiência [79].

Na construção do modelo, cada amostra do conjunto de treinamento é caracterizada por sua posição no gráfico de aceitação e pode ser classificada como uma amostra “regular”, ou seja, uma amostra atribuída à classe alvo, ou amostra “extrema”, que está localizada fora da área de aceitação, e é estabelecida como amostra “*outliers*” (não membro) [82]. Em seguida, é realizada a etapa de classificação de teste ou novos objetos empregando o modelo estabelecido. Os resultados da

classificação são apresentados no gráfico de aceitação correspondente onde poderemos ver amostras regulares, amostras extremas e *outliers* (caso apareçam no modelo) [82].

Os termos frequentemente utilizados para apresentar os resultados da classificação são sensibilidade e especificidade [82–85]. E são usualmente chamados de parâmetros de mérito ou parâmetros de desempenho [83,85].

A sensibilidade é definida como a porcentagem de amostras da classe alvo, que são corretamente reconhecidas como membros dela [68,84]. Também pode ser definida como a proporção de verdadeiros positivos [83]. Para o conjunto de treinamento, a sensibilidade é matematicamente representada pela **Equação 1** [82]:

$$\text{Sensibilidade} = 100\% (\text{Amostras} - \text{Extremos})/\text{Amostras} \quad (1)$$

onde: **Amostras** representa o número total de amostras do conjunto de treinamento e **Extremos** representa o número de amostras que parecem ser extremos ou *outliers* baseado no conjunto atual dos parâmetros do modelo [82]. A sensibilidade para o conjunto de teste pode ser calculada utilizando a mesma equação [82].

Já a especificidade, é definida como a porcentagem de amostras de outras classes, que são corretamente atribuídas como não membros da classe alvo, ou seja, é a fração de amostras de uma classe não alvo que foi identificada corretamente como membro desta classe [82–84]. Também pode ser definida como sendo a porcentagem de verdadeiros negativos [83]. A especificidade para o conjunto de teste pode ser calculada através da **Equação 2**:

$$\text{Especificidade} = 100\% (\text{Objetos externos}/\text{Amostras}) \quad (2)$$

onde: **Objetos externos** representa o número de amostras da classe não alvo que são classificadas corretamente nesta classe, ou seja, não pertencente a classe alvo e **Amostras** representa o número total de amostras da classe não alvo do conjunto de teste [82].

3.5.2.2 Mínimos Quadrados Parciais de uma classe (OC-PLS)

O OC-PLS é um classificador que foi proposto como uma nova técnica de modelagem de classe baseada em mínimos quadrados parciais (PLS) [70,86–90]. As técnicas de modelagem de classe ou classificadores de uma classe visam caracterizar a organização de uma classe e responder à pergunta sobre a possibilidade de uma amostra futura ser aceita ou rejeitada pela classe alvo [87].

Em OC-PLS, um modelo de mínimos quadrados parciais é ajustado correlacionando os espectros, ou seja, as medidas instrumentais com um vetor composto de valores 1 [91]. Ao contrário do DD-SIMCA, cujas componentes principais explicam a maioria das variações dos dados, em OC-PLS é considerado as variações explicadas, através das componentes chamadas de variáveis latentes (VL's) e a compactação das amostras dentro classe alvo, simultaneamente [87,89].

Primeiramente, OC-PLS executa uma regressão de acordo com a **Equação 3** abaixo [70,87,89]:

$$\mathbf{1} = \mathbf{X}\mathbf{b}_{\text{PLS}} + \mathbf{e} \quad (3)$$

onde, $\mathbf{1}$ é o vetor de resposta, $\mathbf{X}$ é o conjunto de dados originais e $\mathbf{b}_{\text{PLS}}$ contém os coeficientes de regressão e $\mathbf{e}$ é o vetor de resíduos do modelo.

O número de variáveis latentes (VL's) podem ser estimadas por validação cruzada [87,89]. Após a construção do modelo duas medidas de distância são calculadas, o valor de *Hotelling* T^2 , baseado na distância dos escores (SD) e os resíduos absolutos centralizados (ACR) são calculados, conforme as **Equações 4 e 5** [70,89]:

$$ACR = |1 - y_j - \mu_e| \quad (4)$$

em que y_j é a resposta ajustada da amostra j e μ_e é a média dos erros de treinamento.

$$T^2 = \sum_{i=1}^K \frac{(t_i - t_{i,avg})^2}{S_{t,i}^2} \quad (5)$$

em que t_i, avg e $S_{t,i}^2$ são a média e a variância da amostra da i -ésima variável latente, t_i , respectivamente, e K é o número de variáveis latentes significativas.

O ACR pode assumir uma distribuição normal com média 0. O desvio padrão do modelo residual pode ser determinado por validação cruzada, de acordo com a **equação 6** [70,89]:

$$\sigma_e = \sqrt{\sum_{i=1}^N \frac{(1-y_i-\mu_e)^2}{(N-1)}} \quad (6)$$

onde N é o número total de amostras omitidas durante a validação cruzada e y_i é a resposta prevista da i -ésima amostra omitida.

Atribuído um nível de confiança, α , os valores limite para ACR e T^2 determinados, conforme as **equações 7 e 8** [70]:

$$ACR_L = Z_{\alpha_2} \times \sigma_e \quad (7)$$

onde Z_{α_2} é o ponto crítico superior da distribuição normal padrão.

$$T_L^2 = \frac{(n^2-1)}{n(n-K)} F_{\alpha(K,n-K)} \quad (8)$$

em que $F_{\alpha(K,n-K)}$ é o ponto crítico superior da distribuição F com $(K, n - K)$ graus de liberdade.

SD é uma medida de distância de uma amostra ao centro da classe no espaço ocupados pelas VL's significativas de OC-PLS, enquanto o ACR é uma medida de dispersão (dos resíduos) da projeção do vetor dos coeficientes de regressão OC-PLS [89]. Um valor de SD excessivamente grande ou ACR indicará que uma amostra afastou-se da maior parte da classe [89]. Em conformidade com os valores de SD e ACR, existem quatro possibilidades de atribuições de grupos: amostra regular (pequeno SD e pequeno ACR); pontos de alavancagem bons (grande SD e pequeno ACR); *outliers* de classe (pequeno SD e grandes valores de ACR) e pontos de alavancagem ruins (grande SD e grande ACR) [91]. Pontos de alavancagens bons,

outliers de classes e pontos de alavancagem ruins podem ser reconhecidos como diferentes tipos de *outliers* [91].

Capítulo 4:

Metodologia

4 METODOLOGIA

4.1 Aquisição de amostras

O fármaco escolhido para as análises foi a insulina humana NPH injetável 100 UI/mL (100 Unidades de insulina em cada mL), comercializada em frasco ampola de 10 mL de coloração transparente.

Para este trabalho foram utilizadas amostras de insulina humana NPH de duas marcas distintas e diferentes lotes. As amostras foram adquiridas nas farmácias das cidades de João Pessoa-PB e Campina Grande-PB e em seguida foram acondicionadas em temperatura recomendada pelo fabricante.

Para análise no NIR de bancada foram utilizadas 64 amostras de insulina que foram divididas em duas classes distintas, sendo a classe alvo denominada de Medicamento Acondicionado Corretamente (MAC) com 36 amostras e a classe Medicamento Acondicionado Incorretamente (MAI) com 28 amostras. Já para o nanoNIR foram utilizadas 50 amostras, onde 22 amostras fizeram parte da classe alvo e 28 amostras da classe não alvo.

As amostras da classe MAC foram adquiridas e em seguida acondicionadas em temperatura entre 2 e 8 °C recomendada pelo fabricante e as amostras da classe MAI foram expostas a temperatura ambiente (23 ± 1 °C) por vários dias.

4.2 Aquisição dos espectros NIR

Os espectros NIR ($3000 - 12000 \text{ cm}^{-1}$) das amostras de insulina humana NPH estudadas foram adquiridos utilizando um espectrofotômetro FTIR, modelo *Analyzer Diamond 20*, marca *Applied Instrument Technologies*, mostrado na **Figura 2**.



Figura 2- Espectrofotômetro FTIR (Analyzer Diamond 20, Applied Instrument Technologies).

Também foi utilizado para a aquisição dos espectros NIR um equipamento NIR portátil (900 a 1700 nm), modelo DLP *NIRscan Nano Evaluation Module*, marca *Texas Instruments nanoNIR*.

Os espectros NIR das amostras comerciais foram registrados incidindo a radiação infravermelho diretamente sobre a ampola de vidro do medicamento injetável devidamente fechada, em triplicata para cada amostra. Estes espectros foram utilizados na análise e para construção dos modelos de classificação.

Todas as análises foram feitas a temperatura ambiente (23 ± 1 °C). Cada espectro NIR registrado foi obtido a partir da média de 16 varreduras, empregando uma resolução de 4 cm^{-1} para o NIR de bancada. Para o NIR portátil, cada espectro foi obtido a partir de uma média de 32 varreduras. O espectro de uma pastilha de politetrafluoretileno (PTFE) foi usado como branco para a aquisição dos espectros em ambos os equipamentos.

4.3 Procedimento Quimiométrico e Software

Tendo em vista que o método a ser utilizado neste trabalho constitui-se em usar técnicas de reconhecimento de padrões em que apenas a classe alvo é modelada, foram consideradas apenas duas classes para fins de classificação. Para o modelo de classificação construído com os espectros registrados no instrumento de bancada, a classe alvo foi composta por 36 amostras e a outra classe por 28 amostras. Já para o modelo construído utilizando os espectros registrados utilizando o NIR portátil, a classe alvo foi composta por 22 amostras enquanto a outra classe por 28 amostras.

A classe alvo no NIR de bancada foi composta por 36 amostras e a outra classe por 28 amostras. Enquanto no NIR portátil esses valores foram de 22 e 28, respectivamente.

Após obter-se os espectros, foi realizada uma avaliação espectral por meio de um exame visual com o objetivo de descartar regiões não informativas. Após a escolha da região de trabalho, foram realizados vários tipos pré-processamentos adequados a natureza dos dados.

Os pré-processamentos utilizados foram: SNV, MSC, suavização Savitzky-Golay, correção de linha de base linear e *offset*, derivação Savitzky-Golay. Diferentes combinações de pré-processamentos foram avaliados.

Para a construção dos modelos de classificação DD-SIMCA e OC-PLS, as amostras foram divididas em conjuntos de treinamento (70%) e teste (30%) empregando o algoritmo Kennard-Stone (KS) [92] tanto para as amostras analisadas no NIR de bancada quanto para o NIR portátil.

Para o NIR de bancada, 26 amostras da classe MAC, classe alvo, foram selecionadas como conjunto de treinamento. As 10 amostras restantes foram incluídas no grupo de teste com as 28 amostras da classe MAI, considerada a classe adulterada, para avaliar a robustez dos modelos construídos. Já para o NIR portátil, 17 amostras foram selecionadas para o conjunto de treinamento e as 5 amostras restantes foram incluídas no grupo de teste com as 28 amostras da classe MAI, para avaliar também a robustez dos modelos construídos.

A qualidade dos modelos de classificação foi avaliada em termos de sensibilidade e especificidade [79,82]. Esses parâmetros foram calculados a partir das equações 1 e 2, descritas na fundamentação teórica [82].

Para DD-SIMCA, o nível de significância de *outliers* foi de 0,05, a área de aceitação foi baseada na distribuição qui-quadrado no modo robusto. Em OC-PLS, os níveis de significância para distância de escores (SD) e para o resíduo absoluto centralizado (ACR) foram de 0,05. O número adequado de componentes principais (PC's) para o DD-SIMCA e de variáveis latentes (VL's) para o OC-PLS foi escolhido observando o comportamento da sensibilidade do modelo no conjunto de treinamento utilizando validação cruzada completa.

Os procedimentos de pré-processamento, os cálculos para PCA, DD-SIMCA, OC-PLS e o algoritmo KS foram realizados utilizando o *Software* Matlab® 2019^a.

Capítulo 5:

Resultados e

Discussão

5. RESULTADOS E DISCUSSÃO

5.1 Espectros NIR (bancada)

Na **Figura 3** são apresentados os espectros brutos na faixa espectral de 3000 a 12000 cm^{-1} utilizando o equipamento NIR de bancada das 64 amostras do medicamento insulina NPH injetável comercial, acondicionadas corretamente (MAC) e incorretamente (MAI).

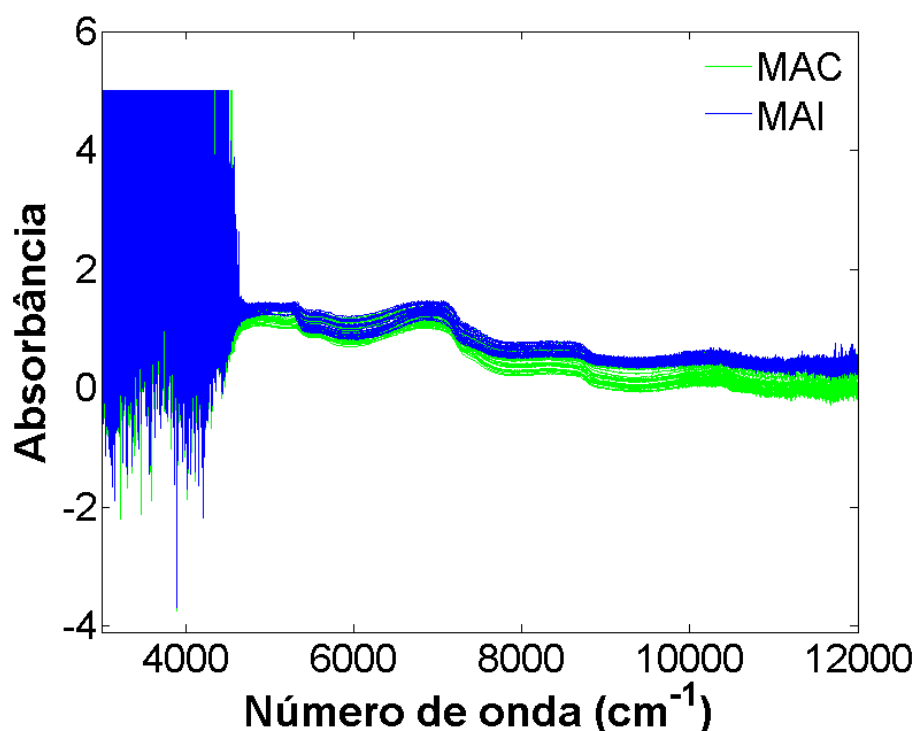


Figura 3- Espectros brutos das amostras de insulina NPH na região NIR entre 3000 a 12000 cm^{-1} .

Pode-se observar que, os valores de absorvância na região compreendida de 3000 a 4800 cm^{-1} apresenta um elevado ruído devido a saturação do detector, e uma região não informativa e ruidosa na faixa de 10000 a 12000 cm^{-1} . Estas regiões foram descartadas por seleção *a priori* na definição da região espectral de trabalho.

A região selecionada para a análise foi a faixa espectral entre 4800 a 10000 cm^{-1} , que se encontra apresentada na **Figura 4**.

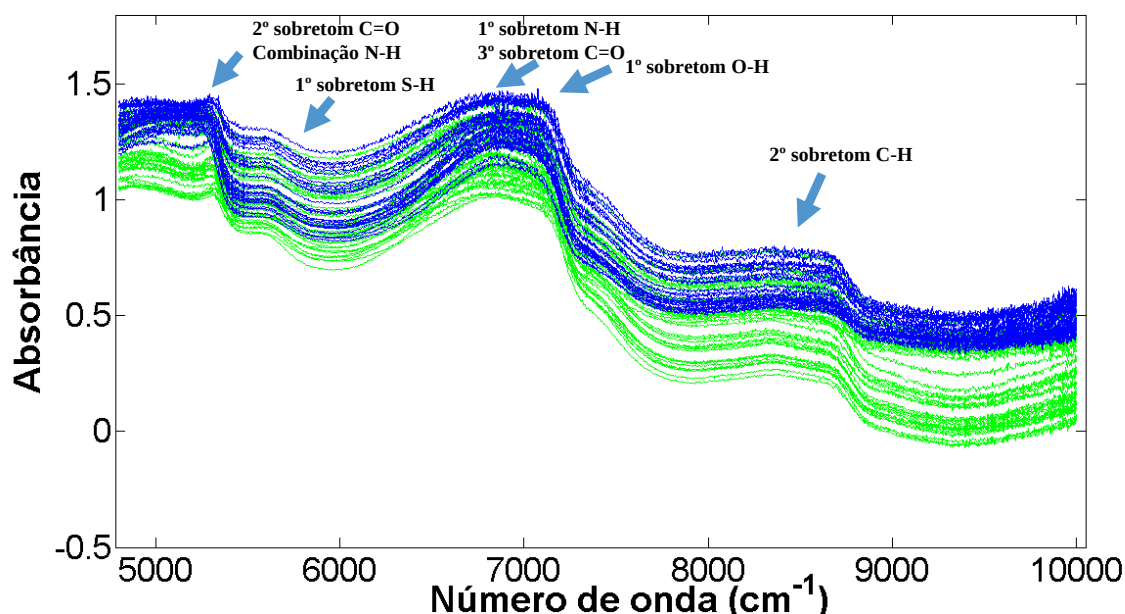


Figura 4- Espectros brutos das amostras de insulina NPH na região NIR entre 4800 a 10000 cm^{-1} .

Esta região caracteriza-se pela ocorrência principalmente de sobretons e combinações de transições fundamentais. Depois de selecionada a faixa espectral de trabalho, observa-se a presença de ruído aleatório e variação de linha de base, como também o espalhamento dos espectros que pode estar relacionado com o tipo de amostra, pois a insulina é uma suspensão e isso faz com que haja um espalhamento de luz na aquisição dos espectros. É possível observar também uma grande semelhança entre os espectros de absorção na região NIR para as amostras, medicamento acondicionado corretamente (MAC, em verde) e medicamento acondicionado incorretamente (MAI, em azul).

Investigando os espectros apresentados na **Figura 4** podemos verificar que existem bandas de absorção em torno de 5200 cm^{-1} e 6900 cm^{-1} ambas associadas à ligação O-H da água [93]. A banda em 5200 cm^{-1} é decorrente da combinação de estiramento assimétrico do O-H e vibração de flexão do grupo hidroxila (H-O-H), e em 7000 cm^{-1} é atribuída ao primeiro sobretom da vibração de estiramento de O-H [58,94]. As bandas espectrais específicas para a presença de proteínas na faixa dos espectros NIR são as mais difíceis de observar, pois sua posição é mascarada por bandas mais intensas originadas de outros constituintes químicos. Os sobretons vibracionais N-H, indicando conteúdo de proteína, estão localizados em torno de 5917, 7326 e 8425 cm^{-1} [86]. Outras designações são as seguintes: aproximadamente 8333 cm^{-1} (segundo sobretom de estiramento de C-H) e de 5747 a 5814 cm^{-1} (primeiro sobretom de

estiramento C-H) [86], 5500 a 6000 cm^{-1} (primeiro sobretom de estiramento de C-H para os grupos funcionais CH_2 , CH_3 e outros grupos) [54], 4864 a 5100 cm^{-1} (vibração de estiramento de N-H) [95,96], 6000 a 7000 cm^{-1} (combinação de estiramento simétrico e assimétrico de O-H e primeiro sobretom de estiramento N-H) e 4200 a 4900 cm^{-1} (sobreposição das bandas de combinação N-H, C-N e C-H) [94].

A banda de 5700 a 5814 cm^{-1} está relacionada ao primeiro sobretom de estiramento S-H na ligação S-H presente no aminoácido cisteína; 5070 a 5264 cm^{-1} refere-se ao segundo sobretom do estiramento C=O, combinação do estiramento N-H e a flexão angular do NH_2 , da ligação - CONH_2 ; 6738 a 6944 cm^{-1} (primeiro sobretom de estiramento N-H e terceiro sobretom de C=O, da ligação das amidas RCONH e CONH_2); 6493 a 6729 cm^{-1} (primeiro sobretom de estiramento N-H, de RNH_2); 4677 a 4916 cm^{-1} (segundo sobretom de estiramento C=O, estiramento e banda de combinação N-H e combinação de estiramento C-N, da ligação química HCONH_2) [97]. Nota-se que as absorbâncias das amostras acondicionadas corretamente tem as mesmas bandas que as acondicionadas incorretamente, porém com intensidades bem menores. Pode-se relacionar com a presença de mais grupos funcionais livres devido a quebra das ligações peptídicas, e dissulfeto nas amostras acondicionadas incorretamente, estiramento S-H.

5.1.2 Pré-Processamentos

Para eliminar a presença do ruído aleatório, variações na linha de base e espalhamento dos espectros foram aplicados diversos tipos de pré-processamentos nos dados originais (**Figura 4**).

Foi realizada uma suavização Savitzky-Golay (com janela de 41 pontos) associada com diferentes métodos de pré-processamentos, sendo eles: correção de linha de base linear (LBC, do inglês *Linear baseline correction*), correção de linha de base *offset* (BO, do inglês *Baseline Offset*), correção de espalhamento multiplicativo (MSC) e variação normal padrão (SNV), antes da construção dos modelos. Foi utilizada a derivação Savitzky-Golay (primeira derivada, janela de 115 pontos e polinômio de segunda ordem), como também em conjunto com os pré-processamentos MSC e o SNV.

Na **Figura 5** é apresentado os espectros NIR utilizando os diferentes tipos de pré-processamentos e suas combinações.

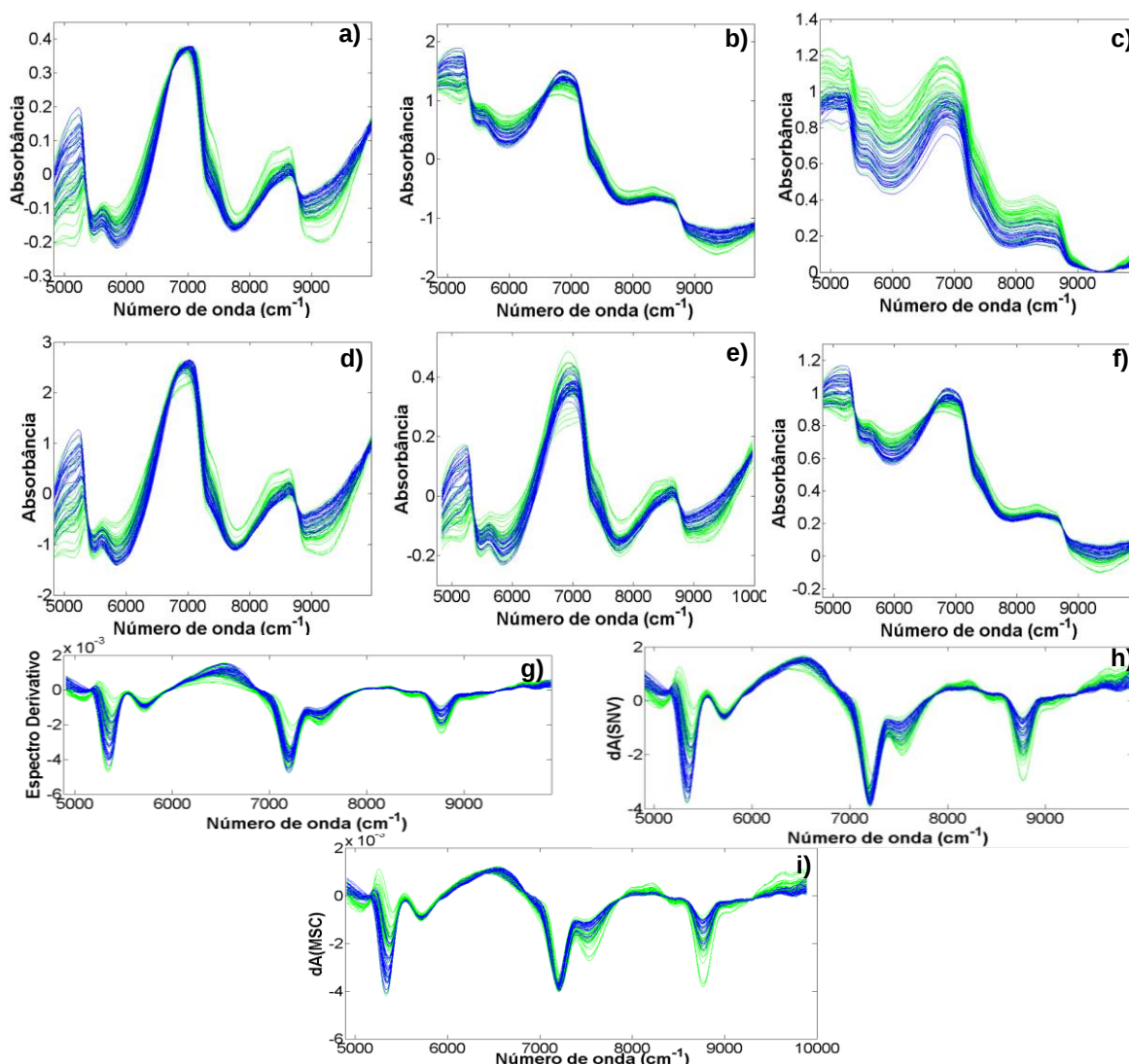


Figura 5- Espectros pré-processados das classes MAC (verde) e MAI (azul) com: **a)** Suavização Savitzky-Golay (SG) com janela de 41 pontos, LBC e MSC, **b)** SG+BO+SNV, **c)** SG+BO, **d)** SG+LBC+SNV, **e)** SG+LBC, **f)** SG+BO+MSC, **g)** derivação Savitzky-Golay (dSG) com primeira derivada, janela de 115 pontos e polinômio de segunda ordem, **h)** dSG+SNV e **i)** dSG+MSC.

Pode-se observar que após a aplicação dos diversos tipos de pré-processamentos nos espectros das duas classes, uma diminuição do ruído, a correção da linha de base e a diminuição do espalhamento dos espectros. Em seguida, foi realizado uma análise exploratória dos dados utilizando PCA.

5.1.3 Análise por componentes principais

Com a intenção de realizar uma avaliação exploratória dos dois grupos de amostras (MAC e MAI), uma análise por componentes principais (PCA) foi realizada no conjunto dos espectros NIR pré-processados. Na **Figura 6** é apresentado o gráfico de escores de PC1 *versus* PC2 resultante da aplicação de PCA aos espectros pré-processados.

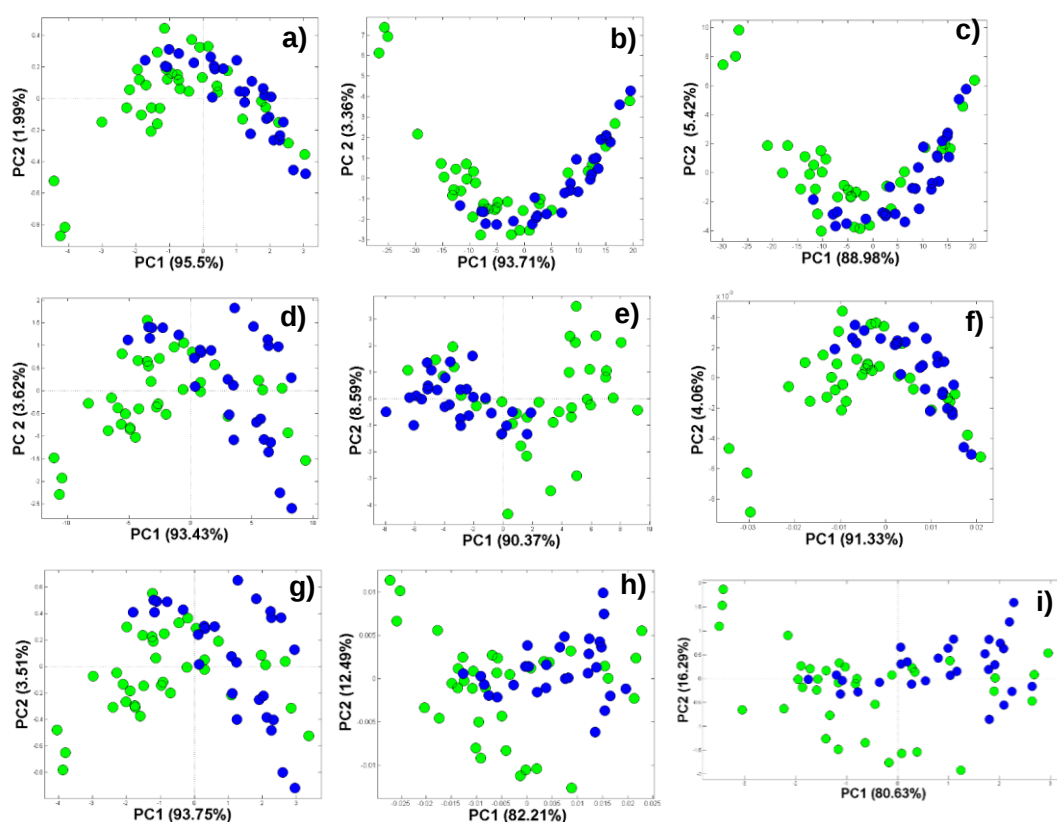


Figura 6- Gráfico dos escores de PC1 *versus* PC2 para as 64 amostras de insulina estudadas usando os espectros pré-processados: **a)** suavização Savitzky-Golay (SG) com janela de 41 pontos, LBC e MSC, **b)** SG, LBC e SNV, **c)** derivação Savitzky-Golay (dSG) com primeira derivada, janela de 115 pontos e polinômio de segunda ordem, **d)** SG, correção BO e SNV, **e)** SG e correção BO, **f)** dSG e MSC, **g)** SG, correção BO e MSC, **h)** dSG e **i)** SG e LBC, (●: MAC, ●: MAI).

Nota-se na **Figura 6** uma formação de agrupamentos naturais. Contudo, não é possível discriminar com total clareza a separação entre as amostras das duas classes. Verifica-se que, embora uma sobreposição entre os grupos seja evidente, o pré-processamento dos dados pode conduzir a uma classificação mais adequada das amostras do medicamento insulina acondicionada corretamente e incorretamente. Desta forma, apenas com a aplicação de PCA não foi possível a identificar as

amostras nas suas respectivas classes, o que requer a aplicação de técnicas de classificação “one class”, como DD-SIMCA e OC-PLS, aos espectros pré-processados das amostras e os resultados são apresentados na seção seguinte.

5.1.4 Classificação

Na abordagem “one class”, apenas a classe MAC foi modelada. Dois métodos de classificação de modelagem de uma classe foram testados, DD-SIMCA e OC-PLS. Na **Tabela 1** é apresentado os resultados obtidos para os modelos de classificação DD-SIMCA e OC-PLS em termos de sensibilidade e especificidade, tanto para o conjunto de treinamento quanto de teste, usando os diversos tipos de pré-processamentos e suas combinações nos espectros NIR.

Tabela 1- Resultados dos parâmetros de desempenho sensibilidade e especificidade para DD-SIMCA e OC-PLS nos conjuntos de treinamento e teste para o NIR de bancada.

MÉTODOS DE CLASSIFICAÇÃO		OC-PLS			DD-SIMCA		
		VL's	Sens.	Espec.	PC's	Sens.	Espec.
SG+LBC	Treinamento	12	100	-	4	96,15	-
	Teste		100	57,14		100	100
SG+BO	Treinamento	8	100	-	3	100	-
	Teste		100	60,71		100	100
SG+BO+MSC	Treinamento	5	100	-	4	89,66	-
	Teste		100	64,29		90	100
SG+BO+SNV	Treinamento	7	100	-	4	92,86	-
	Teste		100	67,86		90	100
SG+LBC+SNV	Treinamento	6	100	-	2	92,86	-
	Teste		100	57,14		90	100
SG+LBC+MSC	Treinamento	6	100	-	4	100	-
	Teste		100	57,14		100	100
dSG+MSC	Treinamento	5	100	-	3	100	-
	Teste		100	82,14		100	85,7
dSG+SNV	Treinamento	6	100	-	3	100	-
	Teste		100	57,14		100	75
dSG	Treinamento	8	100	-	3	100	-
	Teste		100	57,14		100	100

VL's: variáveis latentes; PC's: componentes principais; SG: suavização Savitsky-Golay; dSG: derivada Savitsky-Golay; SNV: variação normal padrão; MSC: correção de espalhamento multiplicativo; BO: correção de linha de base *offset*; LBC: correção de linha de base linear.

Observa-se que, para o DD-SIMCA tanto utilizando a derivação Savitzky-Golay (primeira derivada, janela de 115 pontos e polinômio de segunda ordem) com MSC, quanto a combinação suavização Savitzky-Golay (com janela de 41 pontos) com a

correção de linha de base *offset* como pré-processamento dos dados, alcançou níveis máximos de sensibilidade para o conjunto de treinamento e níveis máximos de sensibilidade e especificidade para o conjunto de teste, utilizando apenas 3 componentes principais. No entanto, o melhor modelo considerado foi a combinação suavização Savitzky-Golay (com janela de 41 pontos) com a correção de linha de base *offset* pois, a derivada pode promover uma ênfase no nível de ruído, além disso necessita de cálculos mais complexos e cuidados na escolha do tamanho da janela que pode provocar perda de informações importantes [55].

Para o OC-PLS, ainda de acordo com a **Tabela 1** verifica-se que, apesar dos resultados de sensibilidade atingirem os valores mais elevados para o conjunto de treinamento em todos os tipos de pré-processamentos utilizados, apenas para a derivação Savitzky-Golay (primeira derivada, janela de 115 pontos e polinômio de segunda ordem) combinada com MSC foi observado uma especificidade maior que 80%. Os valores menores correspondem a um maior número de erros por inclusão de amostras acondicionada incorretamente na classe.

Na **Figura 7** são apresentados os gráficos de área de aceitação para DD-SIMCA e OC-PLS nos conjuntos de treinamento e teste para os melhores pré-processamentos nos quais os melhores desempenhos foram observados (SG+BO e dSG+MSC, respectivamente).

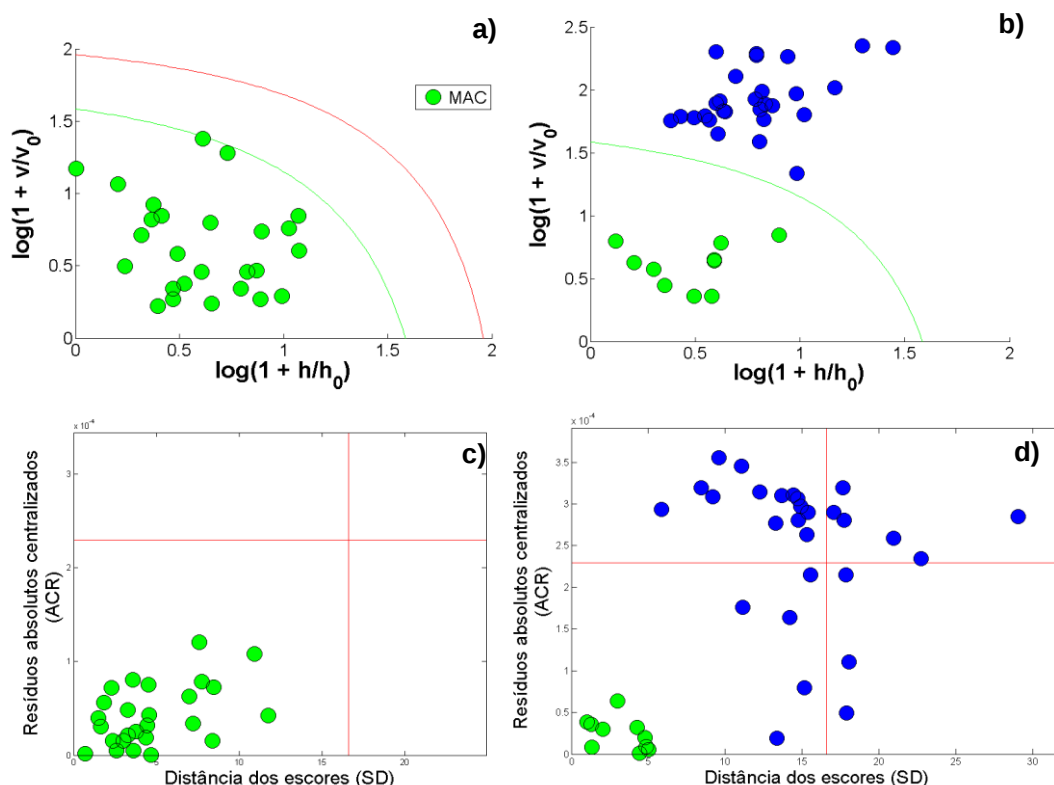


Figura 7- Gráficos de área de aceitação para o conjunto de **(a,c)** treinamento e **(b,d)** teste empregando o DD-SIMCA para o pré-processamento SG+BO e OC-PLS para dSG+MSC, respectivamente. As bolas verdes e azuis representam amostras de insulina acondicionadas corretamente e incorretamente, respectivamente. As linhas verdes em DD-SIMCA e as linhas vermelhas em OC-PLS representam a borda entre as amostras regulares e extremas; a linha vermelha em DD-SIMCA é a fronteira da área de aceitação.

Para o DD-SIMCA, os resultados de classificação são frequentemente mostrados usando gráficos de distâncias dos escores (SD) *versus* distâncias ortogonais (OD), juntamente com a curva limite ou limiar que delinea as amostras regulares, extremas e *outliers*. Com a combinação dessas duas distâncias, as amostras são descritas em termos da variável c , distância total, que assume ter uma distribuição qui-quadrado e os dois limiares são estabelecidos. Em ocorrência de amostras apresentarem valores de c menores que o primeiro limiar (linha verde na **Figura 7a e 7b**) serão consideradas como pertencentes da classe alvo, enquanto amostras com valores de c maiores que o primeiro limiar e menores que o segundo limiar são pertencentes à classe alvo, mas exibem comportamento extremo. Enquanto que para as amostras com valores de c maiores que o segundo limiar (linha vermelha na **Figura 7b**) são consideradas *outliers* [68,82]. Para o conjunto de treinamento e

teste não foram detectadas amostras extremas ou *outliers* utilizando DD-SIMCA, ou seja, o modelo obteve desempenho máximo.

Já para o OC-PLS, as amostras são representadas em termos de distâncias de escores (SD) e resíduos absolutos centralizados (ACR), e seus limiares são determinados em termos de cada parâmetro especificamente, estabelecendo uma área de aceitação retangular (linhas vermelhas nas **Figuras 7c e 7d**) [86,91]. Na **Figura 7c**, observa-se que todas as amostras estão dentro dos limites estabelecidos para o conjunto de treinamento. Já no conjunto de teste (**Figura 7d**) foram observados 5 erros, nos quais as amostras da classe não alvo foram identificadas como sendo da classe alvo, ou seja, amostras de insulina acondicionadas incorretamente foram consideradas como acondicionadas corretamente, o que representaria um risco para saúde dos consumidores.

5.2. Espectros NIR (portátil)

Na **Figura 8** são apresentados os espectros brutos das 50 amostras do medicamento insulina NPH injetável comercial, acondicionadas corretamente (MAC) e incorretamente (MAI) na faixa espectral de 900 a 1700 nm, utilizando o equipamento NIR portátil.

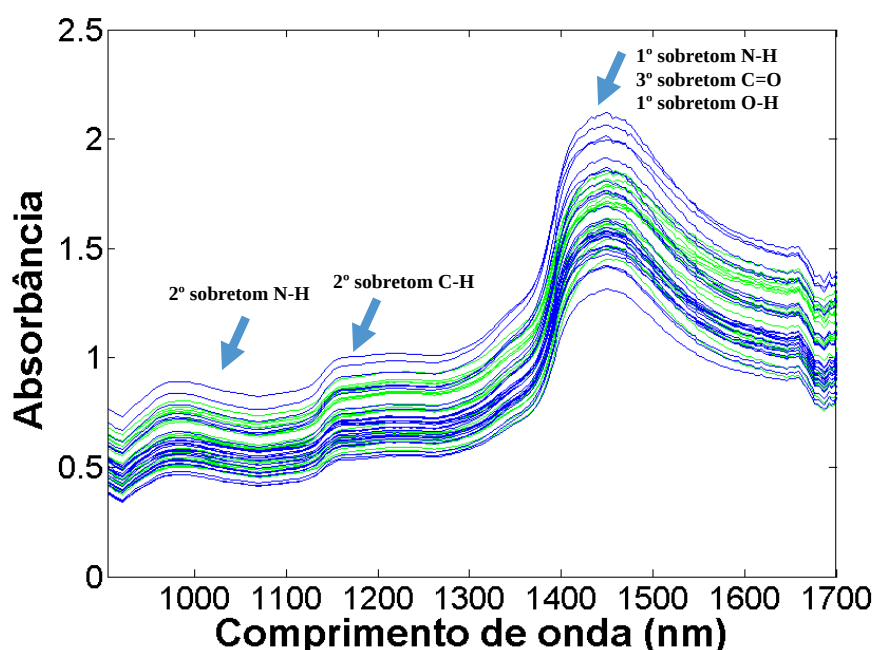


Figura 8- Espectros brutos das amostras de insulina NPH na região NIR entre 900 a 1700 nm.

Pode-se observar que, os espectros obtidos para o NIR portátil na região compreendida de 900 a 1700 nm comparado ao NIR de bancada não apresenta regiões de saturação do detector. Deste modo, por seleção *a priori* esta região foi definida como a região espectral de trabalho (900 a 1700 nm).

Ainda de acordo com a **Figura 8**, verifica-se a presença de variação de linha de base e espalhamento dos espectros, que pode estar relacionado à amostra insulina ser uma suspensão fazendo com que haja um espalhamento de luz na aquisição dos espectros. Sendo assim, há uma necessidade da aplicação de pré-processamentos antes da construção dos modelos.

Esta região de trabalho definida caracteriza-se pela ocorrência principalmente de sobretons e combinações de transições. Assim como no NIR de bancada, é possível observar também uma grande semelhança entre os espectros de absorção na região NIR para as amostras de insulina NPH, medicamento acondicionado corretamente (MAC, em verde) e medicamento acondicionado incorretamente (MAI, em azul).

Como já mencionado, as bandas espectrais específicas da presença de proteínas são as mais difíceis de observar porque sua posição é mascarada por faixas intensas originadas de outros constituintes químicos. Às vezes os sobretons vibracionais da ligação N-H são localizados por volta de 1187, 1365 e 1690 nm. Podemos ter outras atribuições, de 1440 a 1484 nm (primeiro sobreton de estiramento N-H e terceiro sobreton de C=O, da ligação das amidas RCONH e CONH₂); 1486 a 1540 nm (primeiro sobreton de estiramento N-H, de RNH₂) [97]. Nota-se que, os espectros apresentam as mesmas bandas de absorção para as duas classes, e assim como no NIR de bancada, maiores absorbâncias são relacionadas às amostras acondicionadas incorretamente que a acondicionada corretamente.

5.2.2 Pré-processamentos

Para eliminar a presença variações na linha de base e espalhamento dos espectros foram aplicados diversos tipos de pré-processamentos nos dados originais bem como suas combinações (**Figura 8**).

Diferentes métodos de pré-processamentos foram utilizados, sendo eles: correção de linha de base linear (LBC), correção de linha de base *offset* (BO), correção

de espalhamento multiplicativo (MSC) e variação normal padrão (SNV), antes da construção dos modelos. Foi utilizada também a derivação Savitzky-Golay (primeira derivada, janela de 9 pontos e polinômio de segunda ordem).

Na **Figura 9** é apresentado os espectros NIR utilizando alguns dos tipos de pré-processamentos escolhidos e suas combinações.

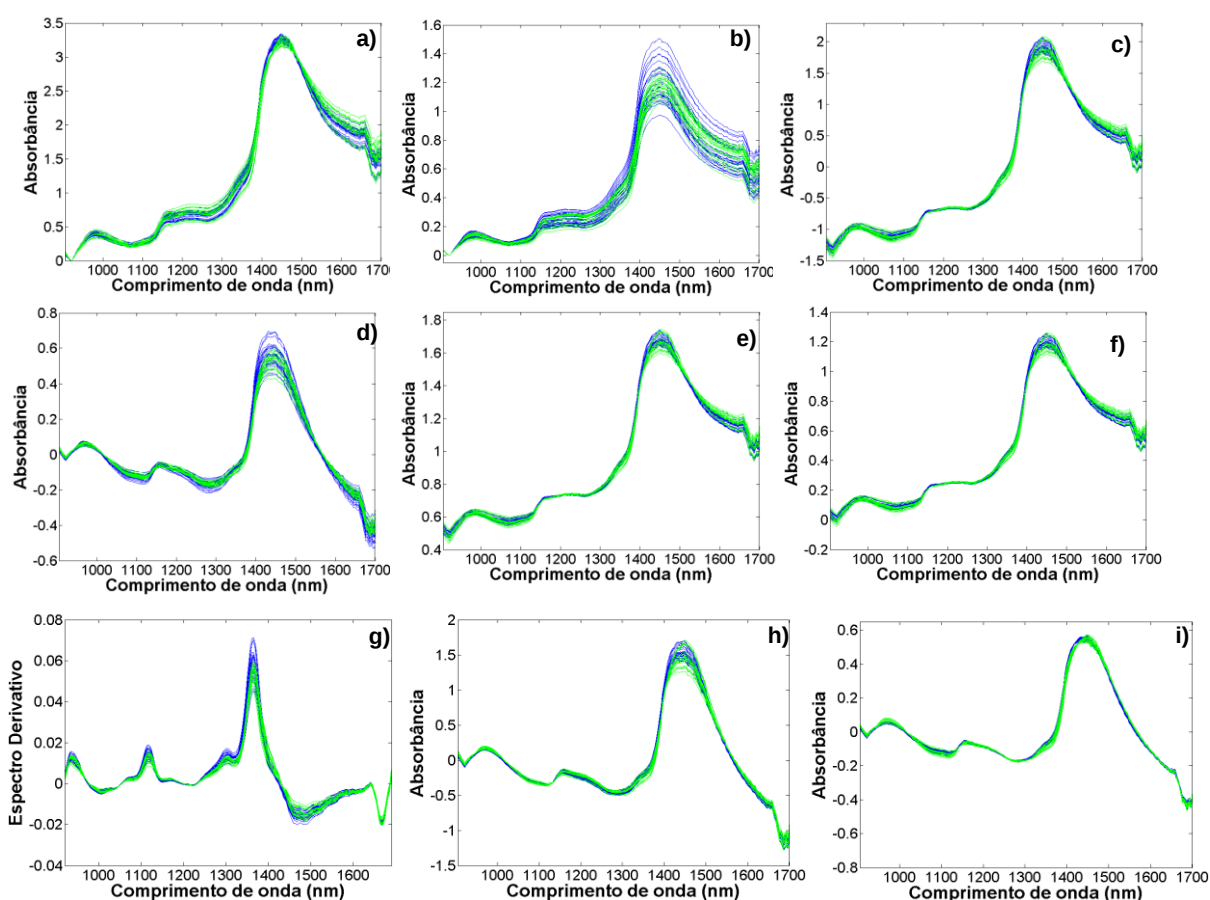


Figura 9- Espectros pré-processados das classes MAC (verde) e MAI (azul) com: **a)** Correção BO e SNV, **b)** correção BO, **c)** LBC e SNV, **d)** LBC, **e)** MSC, **f)** Correção BO e MSC, **g)** derivação Savitzky-Golay (dSG) primeira derivada, janela de 9 pontos e polinômio de segunda ordem, **h)** LBC e SNV e **i)** LBC e MSC.

Pode-se observar que após a aplicação dos diversos tipos de pré-processamentos nos espectros das duas classes, a diminuição do espalhamento dos espectros e a correção da linha de base. Em seguida, foi realizado uma análise exploratória dos dados utilizando PCA.

5.2.3 Análise por componentes principais

Com a intenção de realizar uma avaliação exploratória dos dois grupos de amostras (MAC e MAI), uma análise por componentes principais (PCA) foi realizada no conjunto dos espectros NIR pré-processados. Na **Figura 10** é apresentado os gráficos de escores de PC1 x PC2 resultante da aplicação de PCA aos espectros pré-processados com o método os diferentes pré-processamentos e suas combinações.

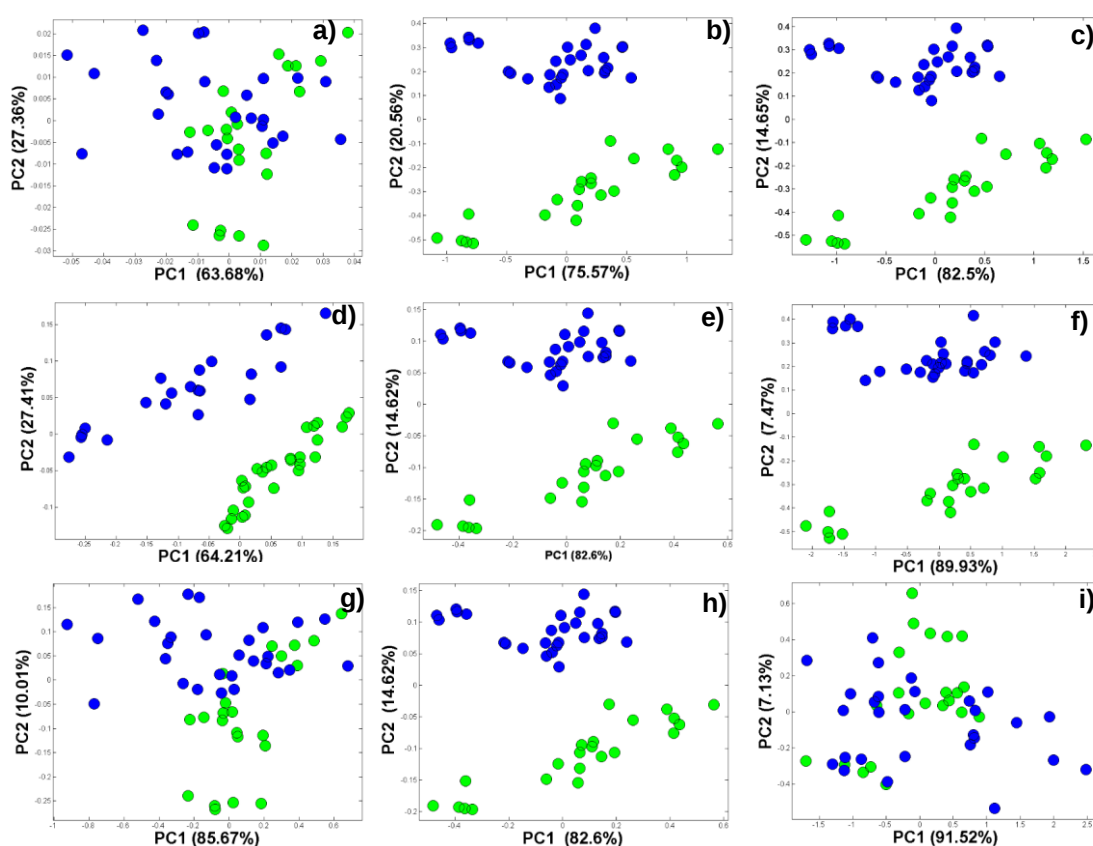


Figura 10- Gráfico dos escores de PC1 versus PC2 para as 50 amostras de insulina estudadas usando os espectros pré-processados: **a)** derivação Savitzky-Golay (dSG) primeira derivada, janela de 9 pontos e polinômio de segunda ordem, **b)** LBC e SNV, **c)** SNV, **d)** LBC e MSC, **e)** MSC, **f)** correção BO e SNV, **g)** LBC, **h)** Correção BO e MSC e **i)** Correção BO: (●: MAC, ●: MAI).

É possível observar a partir dos gráficos dos escores nas **Figuras 10b, 10c, 10e, 10f e 10h** presença de agrupamentos das classes em torno de PC2. Já nos demais, a PCA mostrou uma tendência de separação entre as amostras de insulina das duas classes, indicando pouca discriminação entre as amostras de insulina acondicionadas corretamente e incorretamente.

5.2.4 Classificação

Assim como no NIR de bancada, para o NIR portátil também foi realizada uma abordagem “*one class*” onde, apenas a classe MAC foi modelada. O DD-SIMCA e OC-PLS também foram os métodos de classificação de modelagem de uma classe testados. A **Tabela 2** apresenta os resultados obtidos através das técnicas de classificação “*one class*” para a identificação de alteração nas amostras de insulina em termos de sensibilidade e especificidade, tanto para o conjunto de treinamento quanto de teste, usando os diversos tipos de pré-processamentos e suas combinações nos espectros NIR.

Tabela 2- Resultados dos parâmetros de desempenho sensibilidade e especificidade para DD-SIMCA e OC-PLS nos conjuntos de treinamento e teste para o NIR portátil.

MÉTODOS DE CLASSIFICAÇÃO		VL's	OC-PLS		PC's	DD-SIMCA	
			Sens.	Espec.		Sens.	Espec.
BO+SNV	Treinamento	5	100	-	3	94,44	-
	Teste		100	100		100	100
SNV	Treinamento	6	100	-	5	94,44	-
	Teste		100	92,86		100	100
dSG	Treinamento	7	100	-	6	100	-
	Teste		100	100		100	100
BO	Treinamento	4	100	-	2	94,12	-
	Teste		100	100		40	100
MSC	Treinamento	3	100	-	6	100	-
	Teste		100	100		100	100
BO+MSC	Treinamento	2	94,12	-	5	100	-
	Teste		100	100		100	100
LBC	Treinamento	5	100	-	3	88,24	-
	Teste		100	100		100	100
LBC+SNV	Treinamento	2	94,12	-	6	100	-
	Teste		100	100		100	100
LBC+MSC	Treinamento	2	100	-	6	100	-
	Teste		100	100		100	100

VL's: variáveis latentes; PC's: componentes principais; dSG: derivada Savitsky-Golay; SNV: variação normal padrão; MSC: correção de espalhamento multiplicativo; BO: correção de linha de base *offset*, LBC: correção de linha de base linear.

Observa-se que, para o DD-SIMCA apesar de outros pré-processamentos atingirem elevados valores de sensibilidade, o melhor desempenho obtido foi utilizando em conjunto os pré-processamentos, correção de linha de base *offset* e MSC, alcançando nível máximo de sensibilidade para o conjunto de treinamento, e

níveis máximos de sensibilidade e especificidade para o conjunto de teste, utilizando apenas 5 componentes principais.

Para o OC-PLS, ainda de acordo com a **Tabela 2** verifica-se que vários pré-processamentos atingiram resultados elevados de sensibilidade, tanto para o de conjunto de treinamento quanto para o de teste. O melhor resultado apresentado foi usando o pré-processamento dos dados, correção de linha de base linear (LBC) e correção de espalhamento multiplicativo (MSC), utilizando menos variáveis latentes, 2 apenas, para atingir os valores máximos em termos de sensibilidade e especificidade..

Na **Figura 11** é apresentado os gráficos de área de aceitação para DD-SIMCA e OC-PLS nos conjuntos de treinamento e teste para os melhores pré-processamentos nos quais os melhores desempenhos foram observados (BO+MSC e LBC+MSC, respectivamente).

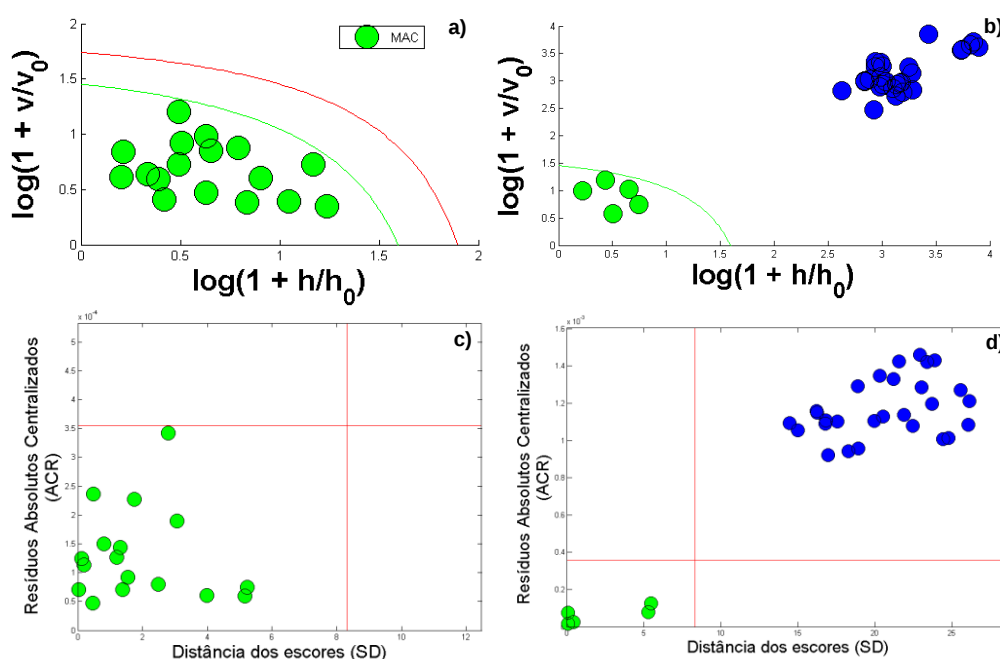


Figura 11- Gráficos de área de aceitação para o conjunto de **(a,c)** treinamento e **(b,d)** teste empregando o DD-SIMCA para o pré-processamento BO+MSC e OC-PLS para LBC+MSC, respectivamente. As bolas verdes e azuis representam amostras de insulina acondicionadas corretamente e incorretamente, respectivamente. As linhas verdes em DD-SIMCA e as linhas vermelhas em OC-PLS representam a borda entre as amostras regulares e extremas; a linha vermelha em DD-SIMCA é a fronteira da área de aceitação.

Como já mencionado, para o DD-SIMCA, os resultados de classificação são frequentemente mostrados usando gráficos de distâncias dos escores (SD) *versus* distâncias ortogonais (OD), juntamente com a curva limite ou limiar que delinea as amostras regulares, extremas e *outliers*. [68,82]. Observa-se nas **Figuras 11a e 11b** para o conjunto de treinamento e teste que não foram detectadas amostras extremas ou *outliers* utilizando DD-SIMCA para o melhor pré-processamento. Sendo assim, o modelo obteve desempenho máximo.

Já para o OC-PLS, as amostras são representadas em termos de distâncias de escores (SD) e resíduos absolutos centralizados (ACR), e seus limiares são determinados em termos de cada parâmetro especificamente, estabelecendo uma área de aceitação retangular (linhas vermelhas nas **Figuras 11c e 11d**) [86,91]. Na **Figura 11c**, observa-se que todas as amostras estão dentro dos limites estabelecidos para o conjunto de treinamento. Para o conjunto de teste, observa-se também que todas as amostras estão dentro dos limites estabelecidos (**Figura 11d**). Não foram observados erros, as amostras da classe não alvo foi identificada como sendo da classe alvo, ou seja, classe das amostras de insulina consideradas acondicionadas corretamente.

Capítulo 6:

Conclusões e

Perspectivas

6. CONCLUSÕES E PERSPECTIVAS

6.1 CONCLUSÕES

Diante dos resultados obtidos, uma nova metodologia para a análise de insulina NPH injetável diretamente na ampola foi desenvolvida. Ela se fundamenta no uso da espectroscopia NIR aliada a técnicas quimiométricas de reconhecimento de padrões (DD-SIMCA e OC-PLS).

O modelo DD-SIMCA se mostrou eficaz ao classificar corretamente todas as amostras do estudo utilizando os espectros obtidos pelo equipamento NIR de bancada quanto no NIR portátil. Já o OC-PLS apresentou resultados satisfatórios apenas para as análises no NIR portátil, classificando corretamente todas as amostras.

Portanto, a metodologia proposta representa uma ferramenta promissora, não destrutiva, rápida e de baixo custo para a triagem da confirmação do estado de conservação da insulina humana NPH.

6.2 PERSPECTIVAS

Como propostas de continuidade deste trabalho as seguintes etapas serão desenvolvidas:

- Obter os espectros de massas das amostras acondicionadas corretamente (MAC) e incorretamente (MAI);
- Propor uma rota de degradação a partir dos espectros de massas;
- Correlacionar as informações dos espectros de massas com as dos espectros NIR obtidos nos equipamentos de bancada e portátil.

Referências

- [1] R. Lyra, M. Oliveira, D. Lins, N. Cavalcanti, J.L. Gross, F.F.R. Maia, L.R. Araújo, M. Yafi, F.P.D.M. Guimarães, A.M.M. Takayanagui, J.B.D.S. Lucena, A. Golbert, M.A. a. Campos, M. da Saúde, D.S. Sartorelli, L.J. Franco, E. Prevenção, J. a. M. Marcondes, Sociedade Brasileira de Diabetes, 2020.
- [2] A.R. Streb, L.D.S. Leonel, C.S. da Silva, R.P. da Silva, G.F. Del Duca, Association between the practice of physical activity of different types and the use of insulin in adult and elderly diabetics in brazil, *Cienc. e Saude Coletiva*. 25 (2020) 4615–4622. <https://doi.org/10.1590/1413-812320202511.02332019>.
- [3] A.C. Schroder, A.P. Vanz, C. Geremia, C.S. Trindade, S.T. Canabarro, Telessaúde em um centro de referência em Diabetes Mellitus: uma análise transversal, *Esc. Anna Nery*. 25 (2021) 1–7. <https://doi.org/10.1590/2177-9465-ean-2020-0046>.
- [4] J.J. Neumiller, G. Chen, C. Newsome, S. Hughes, P. Lazarus, J.R. White, Assessment of Regular and NPH Insulin Concentration via Two Methods of Quantification: The Washington State Insulin Concentration Study (WICS), *J. Diabetes Sci. Technol.* (2019). <https://doi.org/10.1177/1932296819883291>.
- [5] K.T. Briggs, M.B. Taraban, W. Wang, Y.B. Yu, Nondestructive Quantitative Inspection of Drug Products Using Benchtop NMR Relaxometry — the Case of NovoMix® 30, *Am. Assoc. Pharm. Sci.* (2019). <https://doi.org/10.1208/s12249-019-1405-0>.
- [6] T. kongmalai, L. preechasuk, S. junnu, S. manocheewa, C. srisawat, A. sriwijitkamol, The Effect of Temperature on the Stability of In-Use Insulin Pens, *Diabetes*. 67 (2018) 1008-P. <https://doi.org/10.2337/db18-1008-p>.
- [7] P.M. Misun, B. Yesildag, F. Forschler, A. Neelakandhan, N. Rousset, A. Biernath, A. Hierlemann, O. Frey, In Vitro Platform for Studying Human Insulin Release Dynamics of Single Pancreatic Islet Microtissues at High Resolution, *Adv. Sci. News*. 1900291 (2020) 1–14. <https://doi.org/10.1002/adbi.201900291>.
- [8] J.S. Flier, Starvation in the midst of plenty: Reflections on the history and biology of insulin and leptin, *Endocr. Rev.* 40 (2018) 1–16.

<https://doi.org/10.1210/er.2018-00179>.

- [9] K. Zuber, J.S. Davis, A Century of Discovery: The Centennial of Insulin, Physician Assist. Clin. 5 (2020) xv–xvii. <https://doi.org/10.1016/j.cpha.2019.11.001>.
- [10] I. Hargittai, M. Hargittai, Frederick Sanger, Nature. 505 (2014) 73–84. https://doi.org/10.1142/9781860944543_0005.
- [11] M.D. Vargas, Dorothy Crowfoot Hodgkin: A Life Dedicated to Science, Rev. Virtual Química. 4 (2012) 85–100. <https://doi.org/10.5935/1984-6835.20120007>.
- [12] J.E. Gallegos, C. Boyer, E. Pauwels, W.A. Kaplan, J. Peccoud, The Open Insulin Project: A Case Study for ‘Biohacked’ Medicines, Trends Biotechnol. (2018) 1–8. <https://doi.org/10.1016/j.tibtech.2018.07.009>.
- [13] K. Zuber, J.S. Davis, A Century of Discovery: The Centennial of Insulin, Physician Assist. Clin. 5 (2020) xv–xvii. <https://doi.org/10.1016/j.cpha.2019.11.001>.
- [14] L. Hart, Insulin Rationing, J. Nurse Pract. 16 (2020) 213–217. <https://doi.org/10.1016/j.nurpra.2019.10.023>.
- [15] A.C. Silva, C.P. Costa, H. Almeida, J.N. Moreira, J.M.S. Lobo, Hormones, Blood Products, and Therapeutic Enzymes, Adv. Biochem. Eng. Biotechnol. (2019). <https://doi.org/10.1007/10>.
- [16] R.J. Qasem, A.S. Aldawsari, F.E. Almutairi, A.S. Alsadoon, Identification of recombinant human insulin and biosynthetic insulin analogues by multiplexed targeted unlabeled mass spectrometry of proteotypic tryptic peptides, J. Pharm. Biomed. Anal. 172 (2019) 357–363. <https://doi.org/10.1016/j.jpba.2019.04.052>.
- [17] K. Govender, T. Naicker, J. Lin, S. Baijnath, N.S. Chuturgoon, Anil Amichund Abdul, T. Docrat, H.G. Kruger, T. Govender, A novel and more efficient biosynthesis approach for human insulin production in Escherichia coli (E. coli), AMB Express. (2020). [https://doi.org/10.1016/S1734-1140\(13\)71466-X](https://doi.org/10.1016/S1734-1140(13)71466-X).
- [18] R. Mao, Y. Chen, Z. Chi, Y. Wang, Insulin and its single-chain analogue, Appl. Microbiol. Biotechnol. (2019).
- [19] M. Mobasser, M. Shirmohammadi, T. Amiri, N. Vahed, H.H. Fard, M. Ghojzadeh, Prevalence and incidence of type 1 diabetes in the world: a

- systematic review and meta-analysis, *Heal. Promot. Perspect.* 10 (2020) 98–115. <https://doi.org/10.34172/hpp.2020.18>.
- [20] SOCIEDADE BRASILEIRA DE DIABETES, Diretrizes da Sociedade Brasileira de Diabetes: 2019-2020, Sociedade Brasileira de Diabetes. -São Paulo: Clannad, (2019) .
- [21] R. Khursheed, S.K. Singh, S. Wadhwa, B. Kapoor, M. Gulati, R. Kumar, A.K. Ramanunni, A. Awasthi, K. Dua, Treatment strategies against diabetes: Success so far and challenges ahead, *Eur. J. Pharmacol.* 862 (2019) 172625. <https://doi.org/10.1016/j.ejphar.2019.172625>.
- [22] A.K. Rodrigues, A.E. Melo, C.P. Domingueti, Association between reduced serum levels of magnesium and the presence of poor glycemic control and complications in type 1 diabetes mellitus: A systematic review and meta-analysis, *Diabetes Metab. Syndr. Clin. Res. Rev.* (2020). <https://doi.org/10.1016/j.dsx.2020.01.015>.
- [23] A. Iqbal, P. Novodvorsky, S.R. Heller, Recent Updates on Type 1 Diabetes Mellitus Management for Clinicians, *Diabetes Metab. J.* (2018) 3–18.
- [24] W. Bin Hwang, J.H. Kim, S.M. Cho, Two cases of ketosis-prone diabetes mellitus in Korean adolescents, *Ann. Pediatr. Endocrinol. Metab.* (2019) 257–261.
- [25] A.S. Goldstein, B.J. Janson, J.M. Skeie, J.J. Ling, M.A. Greiner, The Effects of Diabetes Mellitus on the Corneal Endothelium: A Review, *Surv. Ophthalmol.* (2020). <https://doi.org/10.1016/j.survophthal.2019.12.009>.
- [26] Y.C. Kudva, K.S. Nair, Diabetes Mellitus : A Perspective on the Post-Insulin Era, *Mayo Clin. Proc.* 95 (2021) 15–21. <https://doi.org/10.1016/j.mayocp.2019.11.016>.
- [27] M. Zieliński, A. Romanik-Chruścielewska, D. Mikiewicz, N. Łukasiewicz, I. Sokołowska, J. Antosik, A. Sobolewska-Ruta, A. Bierczyńska-Krzysik, P. Zaleski, A. Płucienniczak, Expression and purification of recombinant human insulin from *E. coli* 20 strain, *Protein Expr. Purif.* 157 (2019) 63–69.
- [28] O. Wahl, S. Jorajuria, Development and validation of a new UHPLC method for related proteins in insulin and insulin analogues as an alternative to the European Pharmacopoeia RP-HPLC method, *J. Pharm. Biomed. Anal.* (2018).

<https://doi.org/10.1016/j.jpba.2018.12.034>.

- [29] A.B.A.S. Oliveira, L.F. Silva, S.D.P. Mello, M.S. Ferreira, J.C.S. Silva, Conhecimento de portadores de diabetes mellitus acerca da conservação da insulina, *Saúde (Santa Maria)*. 45 (2019) 10. <https://doi.org/10.5902/2236583437342>.
- [30] H. Hwang, K. Kim, S. Lee, C. Kim, C. Min, J. Yun, S.U. Lee, Y. Son, Recombinant Glargine Insulin Production Process Using *Escherichia coli*, *J. Microbiol. Biotechnol.* 26 (2016) 1781–1789.
- [31] A.K. Sharma, G. Taneja, A. Kumar, M. Sahu, G. Sharma, A. Kumar, S. Sardana, A. Deep, Insulin analogs: Glimpse on contemporary facts and future prospective, *Life Sci.* 219 (2019) 90–99. <https://doi.org/10.1016/j.lfs.2019.01.011>.
- [32] P. Lucidi, F. Porcellati, A.M. Andreoli, P. Candeloro, P. Cioli, G.B. Bolli, C.G. Fanelli, Different insulin concentrations in resuspended vs . unsuspended NPH insulin : Practical aspects of subcutaneous injection in patients with diabetes, *Diabetes Metab.* (2017) 2–6. <https://doi.org/10.1016/j.diabet.2017.05.004>.
- [33] I.R. Calori, G. Braga, P. da C.C. Jesus, H. Bi, A.C. Tedesco, Polymer Scaffolds as Drug Delivery Systems, *Eur. Polym. J.* (2020) 109621. <https://doi.org/10.1016/j.eurpolymj.2020.109621>.
- [34] M.W.T. Werten, G. Eggink, M.A.C. Stuart, F.A. De Wolf, Production of protein-based polymers in *Pichia pastoris*, *Biotechnol. Adv.* 37 (2019) 642–666. <https://doi.org/10.1016/j.biotechadv.2019.03.012>.
- [35] H.M.N. Akhtar, M.Q. Ehsan, Electrochemical studies on the interactions of iron (ii) with some essential amino acids, *J. King Saud Univ. - Sci.* (2019) 4–9. <https://doi.org/10.1016/j.jksus.2019.11.020>.
- [36] W. Song, X. Chen, J. Wu, J. Xu, W. Zhang, J. Liu, J. Chen, L. Liu, Biocatalytic derivatization of proteinogenic amino acids for fine chemicals, Elsevier Inc, 2019. <https://doi.org/10.1016/j.biotechadv.2019.107496>.
- [37] O.O. Sofronov, G. Giubertoni, A. Pérez de Alba Ortíz, B. Ensing, H.J. Bakker, Peptide Side-COOH Groups Have Two Distinct Conformations under Biorelevant Conditions, *J. Phys. Chem. Lett.* 11 (2020) 3466–3472. <https://doi.org/10.1021/acs.jpcllett.0c00711>.

- [38] D.J. Dietzen, Amino Acids, Peptides, and Proteins, Elsevier Inc., 2018. <https://doi.org/10.1016/b978-0-12-816061-9.00013-8>.
- [39] L.J. Hoffer, Human Protein and Amino Acid Requirements, Jounal Parenter. Enter. Nutr. (2016). <https://doi.org/10.1177/0148607115624084>.
- [40] L. Heinemann, K. Braune, A. Carter, A. Zayani, L.A. Krämer, Insulin Storage: A Critical Reappraisal, J. Diabetes Sci. Technol. (2020). <https://doi.org/10.1177/1932296819900258>.
- [41] FARMACOPEIA BRASILEIRA. 5ª edição, Primeiro Suplemento (2016).
- [42] A.W. Carter, L. Heinemann, Insulin Concentration in Vials Randomly Purchased in Pharmacies in the United States: Considerable Loss in the Cold Supply Chain, J. Diabetes Sci. Technol. 12 (2018) 839–841. <https://doi.org/10.1177/1932296817747292>.
- [43] A. Moses, J. Bjerrum, M. Hach, L.H. Wæhrens, A.D. Toft, Concentrations of Intact Insulin Concurs With FDA and EMA Standards When Measured by HPLC in Different Parts of the Distribution Cold Chain, J. Diabetes Sci. Technol. (2018). <https://doi.org/10.1177/1932296818783783>.
- [44] S.R. Skoog, D. A.; Holler, F. J.; Crouch, Princípios de Análise Instrumental, 6ª Ed. Porto Alegre Ed. Bookm. (2009) 2009.
- [45] K.M.G. Lima, I.M.R. Júnior, A.M.S. Silva, M.F. Pimentel, Sensores Ópticos com detecção no Infravermelho Próximo e Médio, Quim. Nova. 32 (2009) 1635–1643.
- [46] C. Pasquini, Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications, J. Braz. Chem. Soc. 14 (2003) 198–219. <https://doi.org/10.1590/S0103-50532003000200006>.
- [47] A. Lima, J. Bakker, Espectroscopia no infravermelho próximo para a monitorização da perfusão tecidual, Rev. Bras. Ter. Intensiva. 23 (2011) 351.
- [48] L. Mazur, P.G. Peralta-Zamora, B. Demczuk, R.H. Ribani, Application of multivariate calibration and NIR spectroscopy for the quantification of methylxanthines in yerba mate (*Ilex paraguariensis*), J. Food Compos. Anal. 35 (2014) 55–60. <https://doi.org/10.1016/j.jfca.2014.04.005>.

- [49] A.A. Cornejo-báez, L.M. Peña-rodríguez, R. Alvarez-zapata, M. Vazquez-hernández, A. Sánchez-medina, *Chemometrics : a complementary tool to guide the isolation of pharmacologically active natural products*, *Drug Discov. Today*. 00 (2019). <https://doi.org/10.1016/j.drudis.2019.09.016>.
- [50] G.C. Andrade, C.M. Medeiros Coelho, V.G. Uarrota, *Modelling the vigour of maize seeds submitted to artificial accelerated ageing based on ATR-FTIR data and chemometric tools (PCA, HCA and PLS-DA)*, *Heliyon*. 6 (2020) e03477. <https://doi.org/10.1016/j.heliyon.2020.e03477>.
- [51] S. Farres, L. Srata, F. Fethi, A. Kadaoui, *Argan oil authentication using visible/near infrared spectroscopy combined to chemometrics tools*, *Vib. Spectrosc.* 102 (2019) 79–84. <https://doi.org/10.1016/j.vibspec.2019.04.003>.
- [52] P. Mishra, J.M. Roger, D.N. Rutledge, A. Biancolillo, F. Marini, A. Nordon, D. Jouan-Rimbaud-Bouveresse, *MBA-GUI: A chemometric graphical user interface for multi-block data visualisation, regression, classification, variable selection and automated pre-processing*, *Chemom. Intell. Lab. Syst.* 205 (2020) 104139. <https://doi.org/10.1016/j.chemolab.2020.104139>.
- [53] A. Bajoub, S. Medina-Rodríguez, M. Gómez-Romero, E.A. Ajal, M.G. Bagur-González, A. Fernández-Gutiérrez, A. Carrasco-Pancorbo, *Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics*, *Food Chem.* 215 (2017) 245–255. <https://doi.org/10.1016/j.foodchem.2016.07.140>.
- [54] Q.W. Du, M.T. Zhu, T. Shi, X. Luo, B. Gan, L.J. Tang, Y. Chen, *Adulteration detection of corn oil, rapeseed oil and sunflower oil in camellia oil by in situ diffuse reflectance near-infrared spectroscopy and chemometrics*, *Food Control*. 121 (2021) 107577. <https://doi.org/10.1016/j.foodcont.2020.107577>.
- [55] W. Zhou, P. Guo, J. Chen, Y. Lei, *A rapid analytical method for the quantitative determination of the sugar in acarbose fermentation by infrared spectroscopy and chemometrics*, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 240 (2020) 118571. <https://doi.org/10.1016/j.saa.2020.118571>.
- [56] A. Shirzadifar, S. Bajwa, S.A. Mireei, K. Howatt, J. Nowatzki, *Weed species discrimination based on SIMCA analysis of plant canopy spectral data*, *Biosyst. Eng.* 171 (2018) 143–154.

<https://doi.org/10.1016/j.biosystemseng.2018.04.019>.

- [57] D.D. Silalahi, H. Midi, J. Arasan, M.S. Mustafa, J.P. Caliman, Robust generalized multiplicative scatter correction algorithm on pretreatment of near infrared spectral data, *Vib. Spectrosc.* 97 (2018) 55–65. <https://doi.org/10.1016/j.vibspec.2018.05.002>.
- [58] X. Sun, S. Cheng, H. Yuan, C. Song, L. Wang, W. Tian, X. Li, Rapid drying-free determination of pure cashmere content in scoured cashmere using a novel method of NIR spectroscopy combined with moisture elimination and spectral reconstruction strategy, *Vib. Spectrosc.* 106 (2020) 103006. <https://doi.org/10.1016/j.vibspec.2019.103006>.
- [59] M. Juybar, M. Khanmohammadi Khorrami, A. Bagheri Garmarudi, S. Zandbaaf, Determination of acidity in metal incorporated zeolites by infrared spectrometry using artificial neural network as chemometric approach, Elsevier B.V., 2020. <https://doi.org/10.1016/j.saa.2019.117539>.
- [60] L. Li, Y. Peng, Y. Li, F. Wang, A new scattering correction method of different spectroscopic analysis for assessing complex mixtures, *Anal. Chim. Acta.* 1087 (2019) 20–28. <https://doi.org/10.1016/j.aca.2019.08.067>.
- [61] R. Khodabakhshian, M.H. Abbaspour-fard, Pattern recognition-based Raman spectroscopy for non-destructive detection of pomegranates during maturity, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* (2020) 118127. <https://doi.org/10.1016/j.saa.2020.118127>.
- [62] Z. Basati, B. Jamshidi, M. Rasekh, Y. Abbaspour-gilandeh, Detection of sunn pest-damaged wheat samples using visible/near-infrared spectroscopy based on pattern recognition, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* (2018) . <https://doi.org/10.1016/j.saa.2018.05.123>.
- [63] W. Lijun, H. Yin, J. Kun, Y. Guo, W. Jue, Y. Yan, W. Yang, L. Jing, W. Ping, B. Kaishun, W. Tiejie, Potential of Near Infrared Spectroscopy and Pattern Recognition for Rapid Discrimination and Quantification of *Gleditsia Sinensis* Thorn Powder with Adulterants, *J. Pharm. Biomed. Anal.* (2018). <https://doi.org/10.1016/j.jpba.2018.07.036>

- [64] S. Rahi, H. Mobli, B. Jamshidi, A. Azizi, M. Sharifi, Different supervised and unsupervised classification approaches based on visible/near infrared spectral analysis for discrimination of microbial contaminated lettuce samples: Case study on *E. coli* ATCC, *Infrared Phys. Technol.* 108 (2020) 103355. <https://doi.org/10.1016/j.infrared.2020.103355>.
- [65] S.R. Karunathilaka, B.J. Yakes, K. He, J.K. Chung, M. Mossoba, Non-targeted NIR spectroscopy and SIMCA classification for commercial milk powder authentication: A study using eleven potential adulterants, *Heliyon.* 4 (2018) e00806. <https://doi.org/10.1016/j.heliyon.2018.e00806>.
- [66] L. Wang, Y. Hui, K. Jiang, G. Yin, J. Wang, Y. Yan, Y. Wang, J. Li, P. Wang, K. Bi, T. Wang, Potential of near infrared spectroscopy and pattern recognition for rapid discrimination and quantification of *Gleditsia sinensis* thorn powder with adulterants, *J. Pharm. Biomed. Anal.* 160 (2018) 64–72. <https://doi.org/10.1016/j.jpba.2018.07.036>.
- [67] R. Gurbanov, A.G. Gozen, F. Severcan, Rapid classification of heavy metal-exposed freshwater bacteria by infrared spectroscopy coupled with chemometrics using supervised method, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 189 (2018) 282–290. <https://doi.org/10.1016/j.saa.2017.08.038>.
- [68] O.Y. Rodionova, A.L. Pomerantsev, Chemometric tools for food fraud detection: The role of target class in non-targeted analysis, *Food Chem.* 317 (2020) 126448. <https://doi.org/10.1016/j.foodchem.2020.126448>.
- [69] A.A. Agyekum, F.Y.H. Kutsanedzie, V. Annavaram, B.K. Mintah, E.K. Asare, B. Wang, Q. Chen, FT-NIR coupled Chemometric methods rapid prediction of K-value in fish, *Vib. Spectrosc.* (2020) 103044. <https://doi.org/10.1016/j.vibspec.2020.103044>.
- [70] S.D. Rodríguez, M. Gagneten, A.E. Farroni, N.M. Percibaldi, M.P. Buera, FT-IR and untargeted chemometric analysis for adulterant detection in chia and sesame oils, *Food Control.* 105 (2019) 78–85. <https://doi.org/10.1016/j.foodcont.2019.05.025>.
- [71] H. Chen, Z. Lin, C. Tan, Classification of different animal fibers by near infrared spectroscopy and chemometric models, *Microchem. J.* 144 (2019) 489–494. <https://doi.org/10.1016/j.microc.2018.10.011>.

- [72] D.D. de Sousa Fernandes, V.E. de Almeida, M.M. Fontes, M.C.U. de Araújo, G. Vêras, P.H.G.D. Diniz, Simultaneous identification of the wood types in aged cachaças and their adulterations with wood extracts using digital images and SPA-LDA, *Food Chem.* 273 (2019) 77–84. <https://doi.org/10.1016/j.foodchem.2018.02.035>.
- [73] K. Kucharska-Ambrożej, J. Karpinska, The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices, *Microchem. J.* 153 (2020). <https://doi.org/10.1016/j.microc.2019.104278>.
- [74] A. Popovic, M. Morelato, C. Roux, A. Beavis, Review of the most common chemometric techniques in illicit drug profiling, *Forensic Sci. Int.* 302 (2019) 109911. <https://doi.org/10.1016/j.forsciint.2019.109911>.
- [75] D. Granato, J.S. Santos, G.B. Escher, B.L. Ferreira, R.M. Maggio, Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective, *Trends Food Sci. Technol.* 72 (2018) 83–90. <https://doi.org/10.1016/j.tifs.2017.12.006>.
- [76] R. Kumar, V. Sharma, Chemometrics in forensic science, *Trends Anal. Chem.* 105 (2018) 191–201. <https://doi.org/10.1016/j.trac.2018.05.010>.
- [77] N. Basha, M. Nounou, H. Nounou, Multivariate fault detection and classification using interval principal component analysis, *J. Comput. Sci.* 27 (2018) 1–9. <https://doi.org/10.1016/j.jocs.2018.04.017>.
- [78] S.J. Mazivila, R.N.M.J. Páscoa, R.C. Castro, D.S.M. Ribeiro, J.L.M. Santos, Detection of melamine and sucrose as adulterants in milk powder using near-infrared spectroscopy with DD-SIMCA as one-class classifier and MCR-ALS as a means to provide pure profiles of milk and of both adulterants with forensic evidence: A short communic, *Talanta.* 216 (2020) 120937. <https://doi.org/10.1016/j.talanta.2020.120937>.
- [79] M.D.G. Neves, R.J. Poppi, Authentication and identification of adulterants in virgin coconut oil using ATR/FTIR in tandem with DD-SIMCA one class modeling, *Talanta.* 219 (2020). <https://doi.org/10.1016/j.talanta.2020.121338>.

- [80] O.Y. Rodionova, K.S. Balyklova, A. V. Titova, A.L. Pomerantsev, Application of NIR spectroscopy and chemometrics for revealing of the 'high quality fakes' among the medicines, *Forensic Chem.* 8 (2018) 82–89. <https://doi.org/10.1016/j.forc.2018.02.004>.
- [81] P.H. Ciza, P.Y. Sacre, C. Waffo, L. Coïc, H. Avohou, J.K. Mbinze, R. Ngono, R.D. Marini, P. Hubert, E. Ziemons, Comparing the qualitative performances of handheld NIR and Raman spectrophotometers for the detection of falsified pharmaceutical products, *Talanta.* 202 (2019) 469–478. <https://doi.org/10.1016/j.talanta.2019.04.049>.
- [82] Y. V. Zontov, O.Y. Rodionova, S. V. Kucheryavskiy, A.L. Pomerantsev, DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach, *Chemom. Intell. Lab. Syst.* 167 (2017) 23–28. <https://doi.org/10.1016/j.chemolab.2017.05.010>.
- [83] O.Y. Rodionova, A.L. Pomerantsev, Chemometric tools for food fraud detection: The role of target class in non-targeted analysis, *Food Chem.* 317 (2020) 126448. <https://doi.org/10.1016/j.foodchem.2020.126448>.
- [84] M. Fidelis, J.S. Santos, A.L.K. Coelho, O.Y. Rodionova, A. Pomerantsev, D. Granato, Authentication of juices from antioxidant and chemical perspectives: A feasibility quality control study using chemometrics, *Food Control.* 73 (2017) 796–805. <https://doi.org/10.1016/j.foodcont.2016.09.043>.
- [85] O.Y. Rodionova, J.A. Fernández Pierna, V. Baeten, A.L. Pomerantsev, Chemometric non-targeted analysis for detection of soybean meal adulteration by near infrared spectroscopy, *Food Control.* 119 (2021) 107459. <https://doi.org/10.1016/j.foodcont.2020.107459>.
- [86] L. Pieszczyk, H. Czarnik-Matusiewicz, M. Daszykowski, Identification of ground meat species using near-infrared spectroscopy and class modeling techniques – Aspects of optimization and validation using a one-class classification model, *Meat Sci.* 139 (2018) 15–24. <https://doi.org/10.1016/j.meatsci.2018.01.009>.
- [87] L. Xu, S.M. Yan, C.B. Cai, X.P. Yu, One-class partial least squares (OCPLS) classifier, *Chemom. Intell. Lab. Syst.* 126 (2013) 1–5. <https://doi.org/10.1016/j.chemolab.2013.04.008>.
- [88] Z. Yuan, L. Zhang, D. Wang, J. Jiang, P. de B. Harrington, J. Mao, Q. Zhang, P. Li, Detection of flaxseed oil multiple adulteration by near-infrared spectroscopy

- and nonlinear one class partial least squares discriminant analysis, *Lwt.* 125 (2020) 109247. <https://doi.org/10.1016/j.lwt.2020.109247>.
- [89] L. Xu, M. Goodarzi, W. Shi, C.B. Cai, J.H. Jiang, A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers, *Chemom. Intell. Lab. Syst.* 139 (2014) 58–63. <https://doi.org/10.1016/j.chemolab.2014.09.005>.
- [90] J. Jiang, X. Dou, L. Zhang, J. Mao, L. Yu, P. Li, O.C. Science, Rapid authentication of sesame oil using ion mobility spectrometry and chemometrics, *Oil Crop Sci.* (2020). <https://doi.org/10.1016/j.ocsci.2020.07.002>.
- [91] C.S.W. Miaw, P.M. Santos, A.R.C.S. Silva, A. Gozzi, N.C.C. Guimarães, M.P. Callao, I. Ruisánchez, M.M. Sena, S.V.C. de Souza, Comparison of Different Multivariate Classification Methods for the Detection of Adulterations in Grape Nectars by Using Low-Field Nuclear Magnetic Resonance, *Food Anal. Methods.* 13 (2020) 108–118. <https://doi.org/10.1007/s12161-019-01522-7>.
- [92] P. Taylor, R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, *Technometrics.* 11 (1969) 137–148.
- [93] J.L. da P. Teixeira, E.T. dos S. Caramês, D.P. Baptista, M.L. Gigante, J.A.L. Pallone, Vibrational spectroscopy and chemometrics tools for authenticity and improvement the safety control in goat milk, *Food Control.* 112 (2020). <https://doi.org/10.1016/j.foodcont.2020.107105>.
- [94] M. Fan, W. Cai, X. Shao, Investigating the structural change in protein aqueous solution using temperature-dependent near-infrared spectroscopy and continuous wavelet transform, *Appl. Spectrosc.* 71 (2017) 472–479.
- [95] O. Monago-Maraña, C.E. Eskildsen, T. Galeano-Díaz, A. Muñoz de la Peña, J.P. Wold, Untargeted classification for paprika powder authentication using visible – Near infrared spectroscopy (VIS-NIRS), *Food Control.* 121 (2021). <https://doi.org/10.1016/j.foodcont.2020.107564>.
- [96] E.S. Shin, Y. Sang Park, S.H. Nam, Non-invasive assessments of the advanced glycation end products in human skin using reflectance NIR spectroscopy, *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS.* (2019) 5506–5509. <https://doi.org/10.1109/EMBC.2019.8857867>.

- [97] J. Zhou, R. Wang, X. Wu, B. Xu, Fiber-Content Measurement of Wool–Cashmere Blends Using Near-Infrared Spectroscopy, *Appl. Spectrosc.* 71 (2017) 2367–2376. <https://doi.org/10.1177/0003702817713480>.