

LIDIANE CRISTINA NUNES

**MÉTODOS QUIMIOMÉTRICOS APLICADOS NA DETERMINAÇÃO
ESPECTROFOTOMÉTRICA DE MISTURAS DE AMINOÁCIDOS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Agroquímica, para obtenção do título de *Magister Scientiae*.

**VIÇOSA
MINAS GERAIS – BRASIL
2007**

LIDIANE CRISTINA NUNES

**MÉTODOS QUIMIOMÉTRICOS APLICADOS NA DETERMINAÇÃO
ESPECTROFOTOMÉTRICA DE MISTURAS DE AMINOÁCIDOS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Agroquímica, para obtenção do título de “Magister Scientiae”.

APROVADA: 28 de fevereiro de 2007

Prof. César Reis
(Conselheiro)

Prof. Benjamin Gonçalves Milagres
(Conselheiro)

Prof. Paulo Henrique Fidêncio

Prof^a. Maria Eliana L. R. de Queiroz

Prof. Efraim Lázaro Reis
(Orientador)

*A coisa mais bela que o homem pode experimentar é o mistério. É essa
emoção fundamental que está na raiz de toda ciência e toda arte.
(Albert Einstein)*

Dedico este trabalho aos meus queridos pais
Levi e Maria das Dores, pelo carinho e apoio incondicional.

AGRADECIMENTOS

À Universidade Federal de Viçosa, de maneira especial ao Departamento de Química, pela oportunidade da realização deste trabalho.

Ao professor Efraim Lázaro Reis, pela orientação, apoio e confiança.

Aos professores César Reis, Benjamin Gonçalves Milagres e Paulo Henrique Fidêncio, pelo aconselhamento e ensinamentos que muito contribuíram para realização deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, pela concessão de bolsa de estudos.

À Leandra, pela amizade e convívio.

À Glasiela pela colaboração na execução deste projeto.

A todos os amigos do Laboratório de Instrumentação e Quimiometria pelo convívio agradável e colaborações.

Aos funcionários da secretaria, Marisa e Jorge, pelo auxílio constante.

Aos meus pais, pelo apoio, carinho e dedicação em todos os momentos; serei eternamente grata.

Aos queridos Dirceu, Simone, Dirlene, Aires, Dirciene, Antônio, Leninha, Tim e a todos os familiares pelo carinho e incentivo.

Aos meus amados sobrinhos Damatta, Mariana, Lázaro, Raul e Gabriella, que sempre trouxeram muita alegria e paz.

Ao Rodrigo, pelo carinho, amor, incentivo e apoio na realização deste trabalho.

A Tamara e Juarez, que me acolheram com muito carinho.

Às minhas amigas Adriana e Anete pelo companheirismo, amizade, e alegrias.

Agradeço ao grande e poderoso Deus, minha força, meu consolo e auxílio, pela presença constante; por ter colocado todas essas pessoas maravilhosas no meu caminho e por estar me ajudando hoje e sempre tendo me revelado a sua graça...

Enfim, a todos aqueles que direta ou indiretamente contribuíram de alguma maneira para realização deste trabalho.

BIOGRAFIA

LIDIANE CRISTINA NUNES, filha de Levi Nunes Araújo e Maria das Dores Gomes Araújo, nasceu em Rio Vermelho, Minas Gerais, em 10 de novembro de 1979.

Iniciou o Curso de Licenciatura em Química em fevereiro de 2000 na Universidade Federal de São João del Rei – UFSJ, diplomando-se em fevereiro de 2005.

Em março de 2005 ingressou no programa de Pós-Graduação em Agroquímica, em nível de mestrado, na Universidade Federal de Viçosa, MG, submetendo-se à defesa de dissertação em fevereiro de 2007.

SUMÁRIO

LISTA DE FIGURAS.....	ix
LISTA DE TABELAS.....	xii
RESUMO.....	xiv
ABSTRACT.....	xvi
INTRODUÇÃO.....	1
1. Introdução.....	2
REVISÃO BIBLIOGRÁFICA.....	4
2. Revisão Bibliográfica.....	5
2.1. Aminoácidos.....	5
2.1.1. Classificação dos Aminoácidos.....	7
2.1.2. Determinação de Aminoácidos.....	11
2.2. Análise em Fluxo.....	12
2.2.1. Sistema Fia com Multicomutação e Amostragem Binária (FIA-MBS).....	16
2.3. Espectrofotometria UV/VIS.....	17
2.3.1. Espectrofotometria Derivativa.....	22
2.4. Métodos de Análise Multivariada.....	23
2.4.1. Quimiometria.....	23
2.4.2. Calibração Multivariada.....	24
2.4.3. Análise de Componentes Principais (PCA).....	35
2.4.4. Regressão por Mínimos Quadrados Parciais (PLS).....	37
2.5. Ambiente Computacional MATLAB.....	39
PARTE EXPERIMENTAL.....	41
3. Parte Experimental.....	42
3.1. Determinação de Aminoácidos.....	42

3.2. Equipamentos e Acessórios.....	42
3.3. Preparo de soluções.....	43
3.4. Parâmetros Reacionais.....	44
3.5. Procedimento Geral.....	45
3.6. Processamento do Dados.....	47
3.7. Desenvolvimento de Modelos por Espectrofotometria UV-VIS	48
3.8. Planejamento Fatorial.....	48
3.9. Análise das Componentes Principais (PCA).....	50
3.10. Utilização do PLS para Calibração Multivariada.....	50
3.11. Pré-Processamento dos Dados.....	51
3.12. Análise em Amostras Reais.....	51
RESULTADOS E DISCUSSÃO.....	52
4. Resultados e Discussão.....	53
4.1. Influência do Tempo de Leitura.....	53
4.2. Curvas Analíticas dos Aminoácidos.....	56
4.3. Análise Multivariada.....	60
4.3.1. Análise das Componentes Principais (PCA).....	63
4.3.1.1. Análise de Componentes Principais sem Pré-processamento.....	64
4.3.1.2. Análise de Componentes Principais com Dados Derivados.....	68
4.3.2. Métodos dos Mínimos Quadrados Parciais (PLS).....	72
4.3.2.1. PLS com Espectros sem pré-processamento.....	74
4.3.2.2. PLS com Dados Pré-Processados.....	83
4.4. Análise de Amostras Reais.....	92
4.4.1. Análise de Componentes Principais.....	92

CONCLUSÕES.....	95
5. Conclusões.....	96
RECOMENDAÇÕES FUTURAS.....	98
6. Recomendações Futuras.....	99
REFERÊNCIAS BIBLIOGRÁFICAS.....	100
7. Referências Bibliográficas.....	101
APÊNDICE.....	108
8. Apêndice.....	109

LISTA DE FIGURAS

Figura 1: Estrutura geral dos aminoácidos	5
Figura 2: Aminoácidos que apresentam grupo R não-polar e alifático.....	8
Figura 3. Aminoácidos que apresentam grupo R aromático.....	9
Figura 4. Aminoácidos que apresentam grupo não-carregados, mas polares.....	9
Figura 5. Aminoácidos que apresentam grupo R básico.....	10
Figura 6. Aminoácidos que apresentam grupo R ácido.....	10
Figura 7. Equação geral da reação entre o NQS e um aminoácido.....	11
Figura 8. Diagrama de fluxo de um módulo de análise. (a) Módulo de análise em linha única; (b) Módulo de análise em confluência.....	14
Figura 9. Representação da matriz de dados X.....	27
Figura 10. Representação gráfica da somatória dos produtos dos vetores escores e pesos.....	35
Figura 11. Esquema do sistema empregado para obtenção do sinal analítico baseado na reação de 1,2 naftoquinona-4-sulfonato de sódio (NQS) com aminoácidos.....	46
Figura 12. Sistema utilizado para reação de 1,2 naftoquinona-4-sulfonato de sódio (NQS) com aminoácidos em pH 10...	47
Figura 13. Sistema utilizado no processo de aquisição de dados.....	47
Figura 14. (a) Espectros obtidos para a solução de Fenilalanina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Fenilalanina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) no intervalo de 0 a 10 min.....	53
Figura 15. (a) Espectros obtidos para a solução de Histidina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Histidina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) no intervalo de 0 a 10 min.....	54

Figura 16. (a) Espectros obtidos para a solução de Triptofano ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Prolina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) no intervalo de 0 a 10 min.....	54
Figura 17. (a) Espectros obtidos para a solução de Prolina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Triptofano ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) no intervalo de 0 a 10 min.....	55
Figura 18. (a) Espectros obtidos para a solução de isoleucina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Isoleucina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) no intervalo de 0 a 10 min.....	55
Figura 19. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Fenilalanina.....	57
Figura 20. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Histidina.....	57
Figura 21. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Prolina.....	57
Figura 22. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Triptofano.....	58
Figura 23. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações Isoleucina.....	58
Figura 24. Estruturas químicas (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano e (e) Isoleucina.....	60
Figura 25. Espectros de absorção UV-VIS de Fenilalanina, Histidina, Prolina, Triptofano e Isoleucina obtidos em pH 10. Concentração dos aminoácidos: $5,00 \times 10^{-4} \text{ mol L}^{-1}$	61
Figura 26. Espectros de Absorção UV-VIS das misturas sintéticas de Fenilalanina, Histidina, Prolina, Triptofano e Isoleucina obtidos em pH 10 e utilizados no modelo multivariado.....	61

Figura 27. Espectros de absorção UV-VIS das misturas de Fenilalanina, Histidina, Prolina, Triptofano e Isoleucina, em primeira derivada, obtidos em pH 10 e utilizados no modelo multivariado.....	62
Figura 28. Autovalores em função do número de componentes principais aplicando-se PCA.....	64
Figura 29. Análise de componentes principais (PC1 versus PC2) para as 32 misturas contendo aminoácidos em solução com os dados centrados na média.....	65
Figura 30. Espectros de Absorção UV-Vis dos sete grupos da PCA .	67
Figura 31. Autovalores em função do número de componentes principais aplicando-se PCA.....	69
Figura 32. Análise das componentes principais (PC1 versus PC2) para as 32 misturas contendo aminoácidos em solução com dados da primeira derivada concomitantemente com pré-processamento de centrar na média	70
Figura 33. RMSECV versus número de variáveis latentes com dados sem pré-processamento (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano, (e) Isoleucina.....	76
Figura 34. Resíduos Estudentizados versus “Leverage” para Modelo Desenvolvido com 12 VLS e dados sem pré-processamento (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano, (e) Isoleucina.....	77
Figura 35. RMSECV versus número de variáveis latentes com dados pré-processados (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano, (e) Isoleucina.....	84
Figura 36. Resíduos Estudentizados versus “Leverage” para modelo desenvolvido com 9 VLS com dados pré-processados (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano, (e) Isoleucina.....	85
Figura 37. Espectros das 36 soluções utilizadas nos estudos por análise de componentes principais.....	93
Figura 38. Autovalores em função do número de componentes principais aplicando-se PCA.....	94
Figura 39. Escores da PC1 com PC2.....	94

LISTA DE TABELAS

Tabela 1. Planejamento fatorial 2^5 realizado para calibração multivariada.....	49
Tabela 2. Variação da absorbância em função do tempo para os cinco aminoácidos.....	56
Tabela 3. Valores de concentrações obtidas a partir das curvas analíticas de calibração para os níveis (-) e (+).....	58
Tabela 4. Composição das misturas utilizadas nas etapas de calibração e validação para o desenvolvimento de modelos multivariados.....	59
Tabela 5. Cálculo de variância e percentual de contribuição das componentes utilizadas no cálculo da PCA.....	65
Tabela 6. Cálculo de variância e percentual de contribuição das componentes utilizadas no cálculo da PCA.....	69
Tabela 7. Resultados da previsão das concentrações das misturas sintéticas de Fenilalanina para validação, utilizando 12 VLS, sem pré-processamento.....	78
Tabela 8. Resultados da previsão das concentrações das misturas sintéticas de Histidina para validação, utilizando 12 VLS, sem pré-processamento.....	79
Tabela 9. Resultados da previsão das concentrações das misturas sintéticas de Prolina para validação, utilizando 12 VLS, sem pré-processamento.....	80
Tabela 10. Resultados da previsão das concentrações das misturas sintéticas de Triptofano para validação, utilizando 12 VLS, sem pré-processamento.....	81
Tabela 11. Resultados da previsão das concentrações das misturas sintéticas de Isoleucina para validação, utilizando 12 VLS, sem pré-processamento.....	82
Tabela 12. Resultados da previsão das concentrações das misturas sintéticas de Fenilalanina para validação, utilizando 9 VLS, pré-processados com primeira derivada.....	86
Tabela 13. Resultados da previsão das concentrações das misturas sintéticas de Histidina para validação, utilizando 9 VLS, pré-processados com primeira derivada.....	87

Tabela 14. Resultados da previsão das concentrações das misturas sintéticas de Prolina para validação, utilizando 9 VLS, pré-processados com primeira derivada.....	88
Tabela 15. Resultados da previsão das concentrações das misturas sintéticas de Triptofano para validação, utilizando 9 VLS, pré-processados com primeira derivada.....	89
Tabela 16. Resultados da previsão das concentrações das misturas sintéticas de Isoleucina para validação, utilizando 9 VLS, pré-processados com primeira derivada.....	90
Tabela 17. Média dos erros de previsão (%) para os aminoácidos estudados.....	91

RESUMO

NUNES, Lidiane Cristina, M.Sc., Universidade Federal de Viçosa, fevereiro de 2007. **Métodos quimiométricos aplicados na determinação espectrofotométrica de misturas de aminoácidos.** Orientador: Efraim Lázaro Reis. Co-Orientadores: César Reis e Benjamin Gonçalves Milagres.

O propósito deste estudo foi fazer determinações simultâneas de aminoácidos usando métodos de calibração multivariada. A metodologia é baseada na reação entre aminoácidos e 1,2 naftoquinona-4-sulfonato de sódio (NQS). Uma série de 32 misturas sintéticas com diferentes concentrações de Fenilalanina, Histidina, Prolina, Triptofano e Isoleucina foi preparada de acordo com um planejamento fatorial 2^5 . Os espectros derivativos foram obtidos pela reação em um sistema em fluxo; as soluções foram adicionadas continuamente, empregando uma bomba peristáltica com vazão fixa em $0,40 \text{ mL min}^{-1}$. Uma vez que o NQS se decompõe em meio alcalino, a reação foi desenvolvida on-line pela inserção em confluência de NQS e a solução tampão ($\text{Na}_2\text{CO}_3/\text{NaOH}$) a pH 10. Para o desenvolvimento da reação a bobina de reação foi colocada em um banho termostatizado e mantido a 70°C . A solução final foi coletada em cubeta de quartzo e a leitura foi realizada em um espectrofotômetro UV-VIS, Hitachi modelo U 2000. Duas técnicas quimiométricas foram aplicadas, análise das componentes principais (PCA) e regressão por mínimos quadrados parciais (PLS). Os modelos multivariados foram desenvolvidos a partir de espectros originais (UV-VIS), registrados entre 290 e 590 nm e pré-processados com primeira derivada. Para o PLS os melhores resultados foram conseguidos com modelos desenvolvidos a partir dos dados derivados, envolvendo o uso de 9 variáveis latentes (VLS). Nestas condições, erros de previsão inferiores a 6% foram obtidos para todas as espécies em estudo. Na análise de amostras reais, algumas inconsistências foram observadas, originadas principalmente pela presença de substâncias não modeladas. De maneira geral, os resultados apresentaram uma boa capacidade de previsão dos modelos multivariados fundamentados em espectrofotometria UV-VIS. Para análise de amostras reais, entretanto, há necessidade de considerar a presença de

outras substâncias, as quais deverão ser modeladas junto com os aminoácidos de interesse.

ABSTRACT

NUNES, Lidiane Cristina, M.Sc., Universidade Federal de Viçosa, February of 2007. **Chemometrics methods applied in the spectrophotometric determination of amino acid mixtures.** Adviser: Efraim Lázaro Reis. Co-advisers: César Reis and Benjamin Gonçalves Milagres.

The purpose of this paper was to make simultaneous amino acid determination using methods of multivariate calibration. The method is based on the reaction between amino acids and sodium 1, 2-naphthoquinone 4-sulphonate (NQS). A set of 32 synthetic mixtures with different concentrations of Phenylalanine, Histidine, Proline, Tryptophan and Isoleucine was prepared in accordance with an factorial planning 2^5 . The derivative specters were obtained by the reaction in a flow system, the solutions were added continuously, using a peristaltic pump with fixed outflow in $0,40 \text{ ml min}^{-1}$. Once that the NQS decomposes in alkaline way, the reaction was developed on-line for the insertion in confluence of NQS and the tamponade solution ($\text{Na}_2\text{CO}_3/\text{NaOH}$) pH of 10. For the development of the methodology, the reaction bobbin was placed in a 70°C thermal bath and kept at this temperature. The final solution was collected in quartz bucket and the reading was carried through in a Hitachi model U 2000 UV-VIS spectrophotometer. Two chemometric techniques were applied: Principal Component Analysis (PCA) and Partial Least Squares (PLS). The multivariate models were developed from the original specters (UV-VIS), registered between 290 and 590 nm and preprocessed with first derivative. For the PLS technique, better results were obtained with models developed from the derivated data, involving the use of 9 latent variable model (VLS). In these conditions, forecast errors below to 6% were obtained for all the species in study. In the real samples analysis, some inconsistencies originated mainly for the presence of not shaped substance were observed. In a generalized manner, the results show a good forecast capacity of the multivariate models based on a UV-VIS spectrophotometry. For real samples analysis, however, it has the need of considering the presence of other substances that should be shaped with the interest amino acids.

INTRODUÇÃO

1. INTRODUÇÃO

Os aminoácidos são unidades estruturais básicas das proteínas, a quantificação e a qualificação dos mesmos tornam-se necessárias uma vez que o principal fator determinante da qualidade da proteína é a sua composição em aminoácidos [1].

As proteínas e os aminoácidos são de fundamental importância na nutrição, porque estão intimamente relacionados com os processos vitais das células e, conseqüentemente, de todo o organismo [2].

As proteínas são constituídas de unidades de aminoácidos provenientes de fontes exógenas ou sintetizadas pelo organismo. Quando o aminoácido não pode ser sintetizado pelo organismo ou é sintetizado em quantidades insuficientes, é denominado essencial, uma vez que deve ser obtido de fontes exógenas. Consideram-se como aminoácidos essenciais para o homem a lisina, triptofano, histidina, fenilalanina, leucina, isoleucina, treonina, metionina e valina. Os demais são sintetizados a partir de produtos comuns do metabolismo intermediário, razão pela qual foram classificados como não-essenciais [3].

Como as proteínas são formadas por vários aminoácidos, o organismo animal necessita dos mesmos para sintetizar as suas próprias proteínas. Embora alguns aminoácidos mais simples sejam sintetizados pelo organismo, pelo menos 10 deles não são sintetizados na velocidade necessária para atender as necessidades orgânicas sendo, portanto, necessária sua presença na dieta [2].

O desequilíbrio dos aminoácidos pode causar aumento da síntese de gordura, aumento da sensibilidade aos contaminantes de alimentos, crescimento reduzido, redução da eficiência alimentar, reprodução ineficiente e problemas com a aparência dos pelos e da pele. As dietas precisam ser convenientemente equilibradas para manter o balanço nitrogenado que regula a intensidade da urogênese e a taxa de transformação de aminoácidos em lipídios e glicídios [2].

A bioquímica dos aminoácidos é complexa tanto na nutrição humana quanto na animal, portanto, conhecer as variações do conteúdo de

aminoácidos e proteínas em um determinado alimento são de extrema importância.

Os aminoácidos, em geral, são dificilmente analisados por espectrofotometria molecular utilizando-se métodos univariados, porque seus espectros na região UV-VIS são severamente sobrepostos [4]. Utilizando os métodos de calibração multivariada é possível determinar componentes em mistura, eliminar ou minimizar efeitos de interferência específicos ou de matriz e ruído, além de detectar amostras anômalas. Estas características também tornam possível reduzir as etapas necessárias na preparação das amostras, reduzindo o tempo de análise e custos envolvidos. Assim, o uso dos métodos de calibração multivariada apresenta-se como uma alternativa viável, prática e rápida no tratamento de problemas químicos para os quais os métodos tradicionais não oferecem respostas satisfatórias [5].

Os métodos de calibração multivariada têm sido cada vez mais utilizados em química analítica, principalmente quando os componentes presentes numa mistura necessitam ser determinados, mas a informação analítica disponível não apresenta seletividade; isto é, quando em uma mistura não é possível identificar os componentes individuais, a partir da resposta instrumental [6].

Nesse sentido, a avaliação de técnicas de calibração multivariada, para a determinação espectrofotométrica de misturas de aminoácidos baseada nas reações destes com agentes cromóforos constitui a proposta desse trabalho.

Foram empregadas duas técnicas de estatística multivariada, análise de componentes principais (PCA) e o método dos mínimos quadrados parciais (PLS) que são baseadas na concentração da informação total de uma matriz da resposta.

REVISÃO

BIBLIOGRÁFICA

2. REVISÃO BIBLIOGRÁFICA

2. 1. Aminoácidos

Os aminoácidos e as proteínas são importantes elementos nutricionais. Os aminoácidos são produtos da hidrólise das moléculas protéicas, como o nome sugere são compostos orgânicos contendo grupo ácido carboxílico (-COOH) e amino primário (-NH₂).

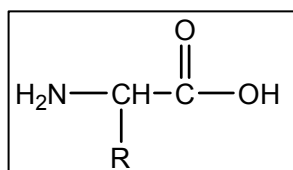


Figura 1. Estrutura geral dos aminoácidos

O primeiro aminoácido isolado foi a cistina em 1810 por Wallanston, porém só em 1820, Branconnot isolou a glicina por hidrólise ácida e demonstrou que os aminoácidos são produtos da decomposição primária das proteínas. Em 1846, Liebig isolou a tirosina pela hidrólise alcalina da caseína [2].

Todos os 20 aminoácidos encontrados nas proteínas são α -aminoácidos. Eles têm um grupo carboxila e um grupo amino, ligados ao mesmo átomo de carbono (o carbono α). Diferem entre si, por suas cadeias laterais ou grupos R, os quais variam em estrutura, tamanho e carga elétrica e influenciam a solubilidade do aminoácido em água [2].

Os 20 aminoácidos das proteínas são, freqüentemente, referidos como os aminoácidos primários para distingui-los dos aminoácidos menos comuns, que são resíduos modificados no interior das proteínas, depois que estas são sintetizadas e de muitos outros tipos de aminoácidos presentes nos organismos vivos, porém não nas proteínas [2].

Todos os aminoácidos têm nomes triviais ou comuns; em alguns casos, foram retirados da fonte de que foram isolados pela primeira vez [2]. Os aminoácidos primários, por convenção internacional, têm sido designados por abreviação de três letras (derivadas de seus nomes na

língua inglesa) ou por um símbolo de uma única letra; ambos usados como abreviaturas para indicar a composição e a seqüência dos aminoácidos nas proteínas [7].

Os aminoácidos têm várias funções como, por exemplo, composição do tecido animal e dos sistemas enzimáticos, fornecimento de energia de sistemas específicos e armazenamento de nutrientes (lipídios e minerais), efeito tampão e regulação da pressão osmótica, reprodução, imunidade e transporte de oxigênio. Portanto, os animais devem receber durante a vida uma quantidade mínima diária de proteínas para atender suas necessidades que podem ser de crescimento, recuperação dos tecidos e gestação [2].

O valor nutritivo de uma proteína está relacionado à proporção de aminoácidos que a compõem. Alguns destes devem ser fornecidos por meio da dieta aos organismos adultos, aos jovens em crescimento ou a ambos, uma vez que não têm a capacidade de sintetizá-los. Para o ser humano adulto, são essenciais os seguintes aminoácidos: isoleucina, leucina, lisina, metionina, fenilalanina, treonina, triptofano; enquanto para crianças, além destes é necessária a histidina, existindo algumas evidências de que também seja para o indivíduo adulto.

As quantidades relativas de aminoácidos presentes nas proteínas ingeridas devem ser próximas aos requerimentos do organismo em questão. Por outro lado, quando a ingestão de carboidratos é baixa, a quantidade de proteína utilizada pelo organismo para fins anabólicos reduz-se, sendo desviada para a produção de energia. Torna-se relevante, portanto, integrar as informações relativas à qualidade e quantidade da proteína ingerida com as do valor energético total da dieta. As proteínas da dieta devem ter uma composição adequada para que substituam satisfatoriamente as proteínas degradadas no catabolismo. Entretanto, com exceção do leite materno para os recém nascidos, os demais alimentos são relativamente incompletos para satisfazer as necessidades nutricionais da espécie humana.

O valor protéico de um alimento é determinado pela sua composição em aminoácidos essenciais, sendo que o aproveitamento biológico dos aminoácidos (biodisponibilidade) depende também da digestibilidade da proteína. Sendo assim, o perfil aminoacídico fornece boa indicação da qualidade de proteínas alimentares [8].

Em geral, as proteínas de origem animal têm valor biológico mais elevado que as de origem vegetal e ainda, a digestibilidade das proteínas vegetais é baixa, quando comparada às de origem animal. Em termos protéicos, entretanto, não só a falta, mas também o excesso de um ou mais aminoácidos pode ser prejudicial. Entre os casos patológicos por desequilíbrio de aminoácidos cita-se na literatura clínica infantil, correlação com distúrbios mentais (fenilalanina, histidina, valina) assim como alterações morfológicas no sistema nervoso os quais estariam associados a deficiências protéicas. Cerca de dois terços da população mundial ainda utilizam essencialmente arroz, milho, trigo e mandioca como fonte principal de proteínas, sendo aquela de origem animal, ainda limitada.

As proteínas destacam-se na hierarquia bioquímica que mantém a homeostase do organismo vivo, não apenas por suas funções de sustentação de órgãos e tecidos, como por sua atuação como hormônios protéicos e enzimas, relacionadas a quase todas as etapas do metabolismo. Assim, entre os grupos de substâncias associados à alimentação e nutrição, a carência de proteínas reflete-se mais intensamente sobre o equilíbrio dinâmico do metabolismo que outros grupos integrantes da alimentação de seres humanos e animais. Isto se torna mais evidente em estados de carência, que persistindo por certo período compromete o indivíduo levando a danos irreversíveis [9].

2.1.1. Classificação dos Aminoácidos

Os aminoácidos podem ser agrupados em cinco classes principais, tendo como base as propriedades dos seus grupos R, em particular sua polaridade, ou tendência para interagir com a água em pH biológico (próximo de 7). A polaridade dos grupos R varia amplamente, desde um comportamento totalmente não-polar ou hidrofóbico até um altamente polar ou hidrofílico [7].

Grupos R não- polares e alifáticos: Leucina, Isoleucina, Prolina e Alanina. Os grupos R nessa classe de aminoácidos são hidrofóbicos e não-polares. As volumosas cadeias laterais de alanina, valina, leucina e isoleucina, com suas formas características, são importantes na estabilização da estrutura das proteínas pela promoção de interações hidrofóbicas em seu interior.

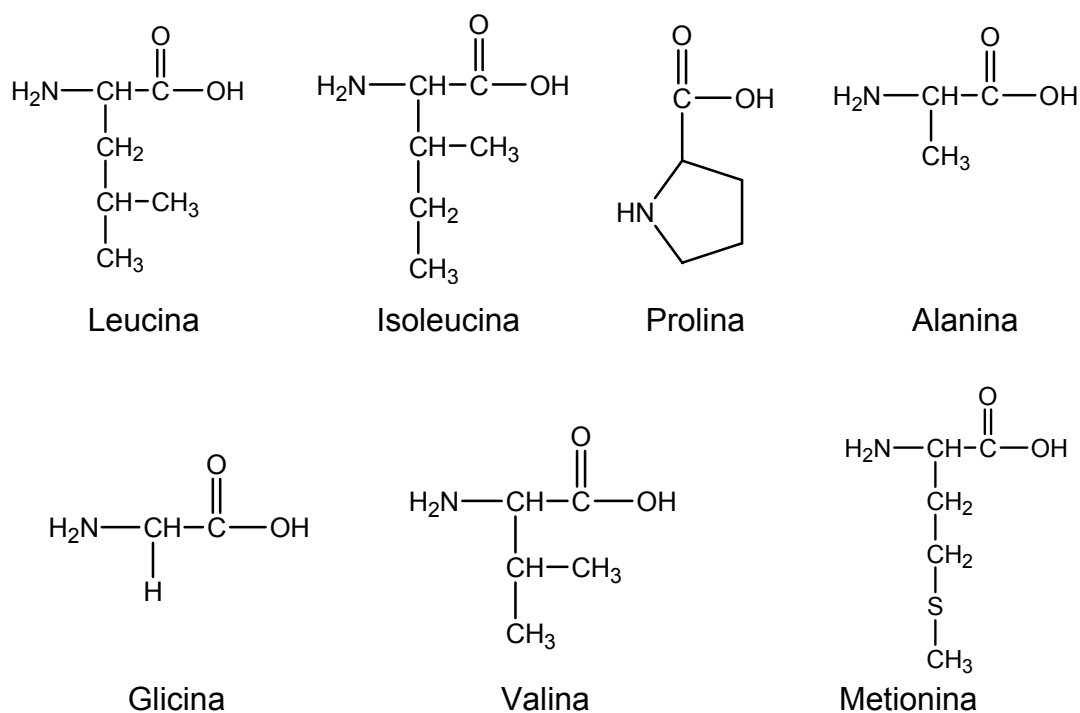


Figura 2. Aminoácidos que apresentam grupo R não-polar e alifático.

Grupos R aromáticos: Fenilalanina, Tirosina e Triptofano. Com suas cadeias laterais aromáticas, são relativamente apolares. Todos podem participar de interações hidrofóbicas.

A tirosina e o triptofano são significativamente mais polares que a fenilalanina em virtude do grupo hidroxila da tirosina e do nitrogênio do anel indol do triptofano.

O triptofano, a tirosina e, com menor intensidade, a fenilalanina absorvem luz na região ultravioleta do espectro. Esse fato é responsável pela forte e característica absorbância da luz pelas proteínas no comprimento de onda de 280 nm, propriedade muito explorada pelos pesquisadores na caracterização de proteínas [7].

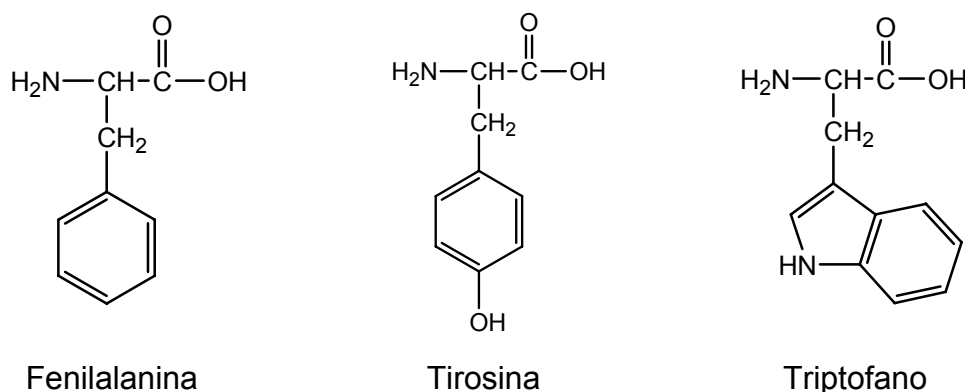


Figura 3. Aminoácidos que apresentam grupo R aromático.

Grupos R não-carregados, mas polares: Os grupos R desses aminoácidos são mais solúveis em água que os dos aminoácidos não-polares, porque contêm grupos funcionais que formam pontes de hidrogênio com a água. Essa classe de aminoácidos inclui serina, treonina, cisteína, asparagina e glutamina [7].

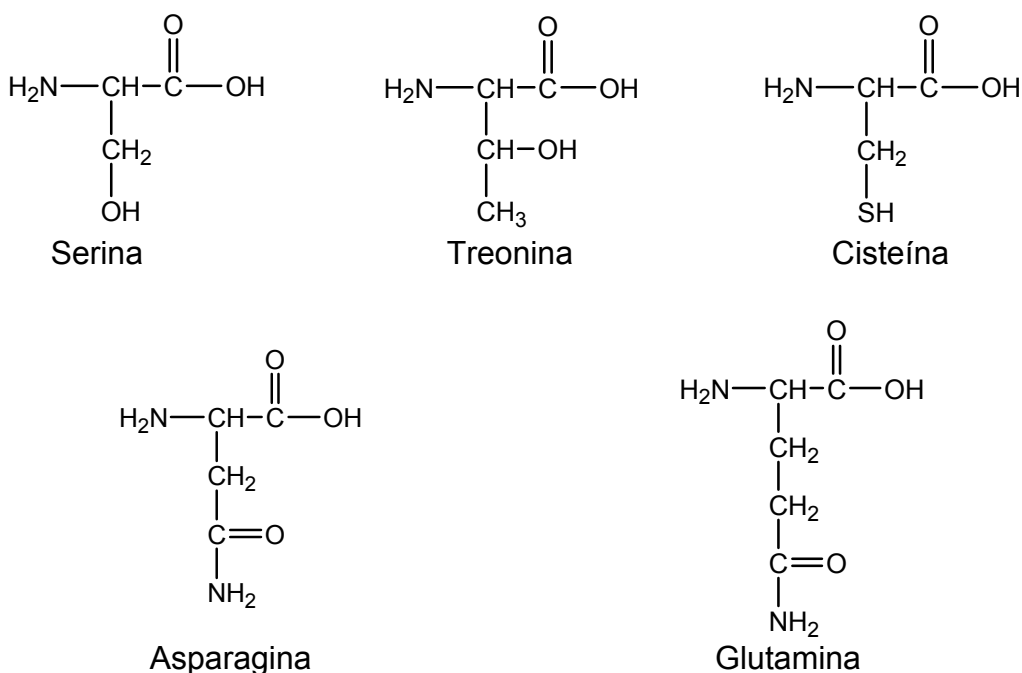


Figura 4. Aminoácidos que apresentam grupo não-carregados, mas polares.

Grupos R carregados positivamente (básicos): Os grupos R mais hidrofóbicos são aqueles que são positiva ou negativamente carregados. Os aminoácidos nos quais os grupos R têm uma carga positiva líquida em pH 7 são a lisina, a qual tem um segundo grupo na posição ε da sua cadeia

alifática; a arginina, a qual tem um grupo carregado positivamente, o grupo guanidino; a histidina que contém um grupo imidazol. A histidina é o único aminoácido primário que possui uma cadeia lateral com pK_a próximo da neutralidade. Em muitas reações catalisadas por enzimas, um resíduo de histidina facilita a reação ao servir como doador ou aceptor de prótons [7].

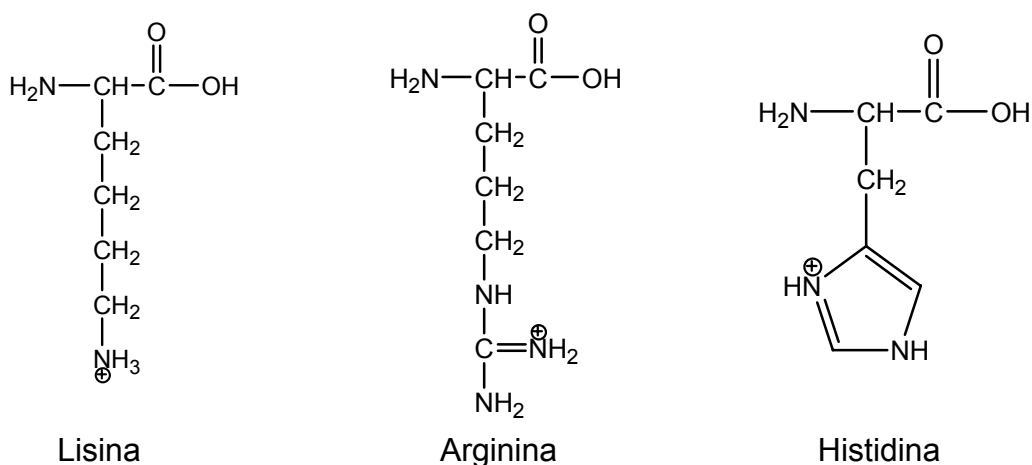


Figura 5. Aminoácidos que apresentam grupo R básico.

Grupos R carregados negativamente (ácidos): Os aminoácidos, que apresentam grupos R com uma carga líquida negativa em pH 7, são o aspartato e o glutamato; cada um deles possui na molécula um segundo grupo carboxila [7].

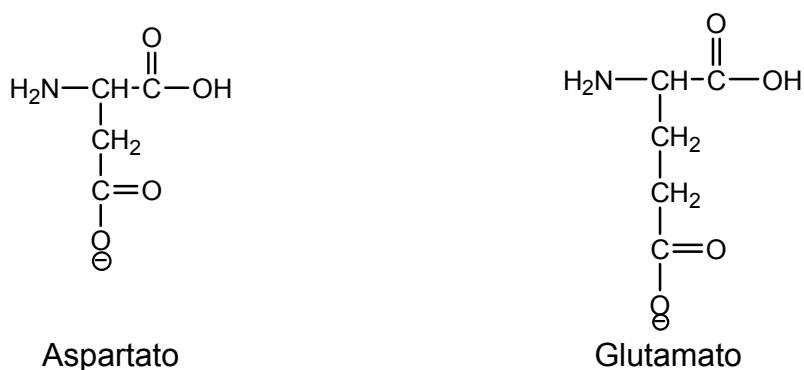


Figura 6. Aminoácidos que apresentam grupo R ácido.

2.1.2. Determinação de Aminoácidos

A determinação de aminoácidos é realizada, geralmente, por técnica de cromatografia líquida de alta eficiência (CLAE) que apresenta boa seletividade, mas o processo pode ser inadequado para analisar um grande número de amostras por causa do tempo consumido; os métodos oficiais de derivatização pré ou pós-coluna são demorados e de custos elevados.

Os aminoácidos não são detectáveis por técnicas espectroscópicas (UV-VIS e fluorimetria); um procedimento químico para derivatização destes é necessário. Os grupos amino e carboxílico oferecem possibilidades para derivatização, mas a maioria dos métodos tem sido desenvolvido para a função amino. A cromatografia fornece seletividade enquanto a derivatização realça a sensibilidade.

Um método espectrofotométrico para determinação de aminoácidos baseado na reação com 1,2-naftoquinona-4-sulfonato de sódio (NQS) em meio básico foi desenvolvido com um sistema de análise fluxo contínuo. A reação envolvida é a derivatização dos aminoácidos com o NQS como é mostrado na Figura 7.

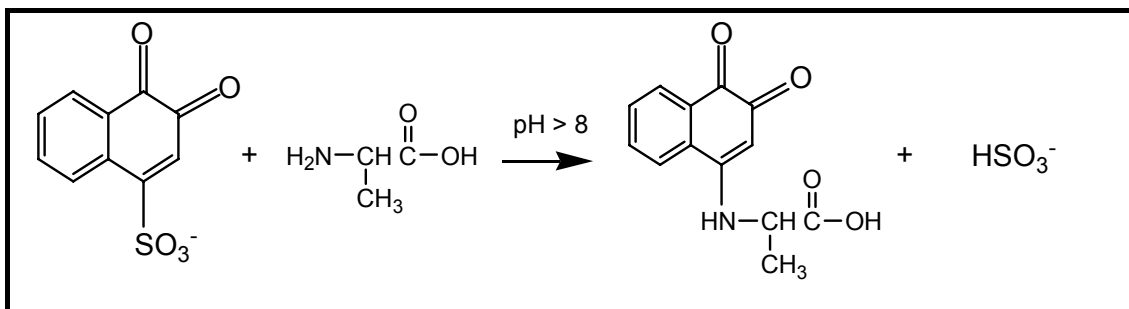


Figura 7. Equação geral da reação entre o NQS e um aminoácido.

O reagente 1,2-naftoquinona-4-sulfonato de sódio (NQS) apresenta como vantagens, solubilidade em água, reage com grupos amino primários e secundários e baixo custo. O principal problema com este reagente é a instabilidade em meio alcalino.

O reagente NQS tem sido usado em determinação de aminas alifáticas em bebidas alcoólicas por extração líquido-líquido com sistema de injeção em fluxo [10].

2.2. Análise em Fluxo

A técnica de Análise por Injeção em Fluxo, do inglês “Flow Injection Analysis” (FIA), foi proposta em 1975 por Ruzicka & Hansen [11,12] e tem como conceito básico a injeção de uma alíquota da solução da amostra em um fluido carregador que a transporta para o detector [12, 13, 14]. O carregador pode ser uma solução de reagentes ou um fluido quimicamente inerte em relação à espécie química de interesse. Enquanto é transportada para o detector, a amostra dispersa no fluido carregador, gerando um gradiente de concentração. O grau de dispersão da amostra depende, principalmente, do comprimento do percurso analítico e da alça de amostragem e reflete-se na magnitude e formato do sinal gerado pelo detector. Portanto, controlando-se os parâmetros que governam a dispersão, pode-se ajustar a magnitude do sinal gerado dentro da faixa operacional do instrumento. É entendido como faixa operacional do instrumento, o intervalo de medidas em que o mesmo apresenta resposta linear com a concentração da espécie de interesse [14].

Inicialmente, a injeção da amostra no fluxo de reagente era feita, empregando-se uma seringa hipodérmica, o que deu origem ao nome do processo [12,15]. A amostra a ser analisada era inserida no carregador perfurando-se com a agulha da seringa um septo de borracha, colocado em um dispositivo que foi chamado de injetor. Este dispositivo tinha uma vida útil muito curta, visto que após duas ou três dezenas de injeções, o septo de borracha começava a vazar. Essa fragilidade foi superada com a invenção de um injetor que dispensava o uso da agulha. Nesse novo dispositivo o septo de borracha funcionava como uma válvula, que por pressão externa cedia, liberando o canal que dava acesso ao percurso analítico [15].

O injetor, também chamado de válvula de injeção, é o dispositivo fundamental do sistema de análise por injeção em fluxo contínuo. Além de introduzir a amostra no percurso analítico, pode ser empregado para selecionar as vazões do carregador e dos reagentes e também variar o comprimento do percurso analítico, o que aumenta a flexibilidade do processo. Vários tipos de injetores têm sido apresentados na literatura, sendo os mais comuns os de válvula rotatória (“rotary valve”), desenvolvido

por Ruzicka e Hansen, e o injetor proporcional desenvolvido pelos pesquisadores da Seção de Química Analítica do CENA/USP [16]. O emprego desses dois tipos de injetores trouxe, como primeira vantagem, maior estabilidade aos sistemas FIA, pois a comutação da posição de amostragem para a posição de injeção, e vice-versa, não criava transiente de pressão. Também a reprodutibilidade das medidas deixou de depender da habilidade do operador, pois o volume de cada alíquota de amostra injetada era definido pela dimensão da alça de amostragem [15].

Apesar de não serem utilizadas seringas para injeção de amostras no sistema, substituídas por injetores manuais, válvulas rotativas, injetores comutadores e comutadores temporizados, o termo “injeção” ainda é utilizado. Esta técnica, quando introduzida, teve grande repercussão na comunidade científica, sendo atualmente um dos métodos de automatização, baseados em fluxo, mais empregados em virtude de sua grande versatilidade de aplicação [17].

O diagrama de fluxos mais simples é o de linha única, onde o fluido carregador é o reagente (Figura 8(a)). Em um módulo de análise com esta configuração, a mistura da amostra com o reagente ocorre exclusivamente por dispersão, enquanto a mesma é transportada em direção ao detector. A dispersão pode afetar a magnitude do sinal analítico [12,16]. Quando a razão entre os volumes da alíquota de amostra e do percurso analítico é inadequada, o reagente pode não alcançar o centro da amostra, em quantidade suficiente, para satisfazer as condições para o desenvolvimento das reações químicas envolvidas. Esta limitação foi superada projetando-se os diagramas de fluxo para adição dos reagentes por confluência (Figura 8(b)), e empregando-se uma solução quimicamente inerte como carregador da amostra [12].

Quando os reagentes são adicionados por confluência, cada fração da amostra recebe a mesma quantidade do reagente, portanto não se formam gradientes de concentração de reagente dentro do percurso analítico [12,16]. Chama-se percurso analítico o espaço percorrido pela amostra do ponto de injeção até a cela de detecção. A reação química tem início a partir da confluência, em vista disso, o percurso analítico pode ser mais curto, tendo como resultado um aumento na velocidade analítica [12].

Empregando-se a adição do reagente por confluência, a vazão do mesmo pode ser menor que a do carregador, o que leva a uma diminuição do consumo de reagente [12].

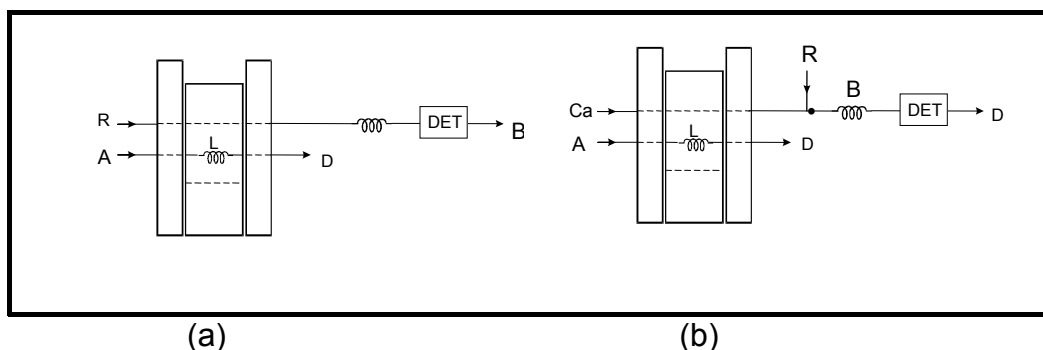


Figura 8. Diagrama de fluxo de um módulo de análise. (a) Módulo de análise em linha única; (b) Módulo de análise em confluência. A: amostra, D: descarte, R: reagente, L: alça de amostragem, B: bobina de reação, Ca: fluido carregador e DET: detector.

Desde a sua introdução em 1975, a técnica de Análise por Injeção em Fluxo vem se estabelecendo em vários campos como uma técnica analítica adequada para determinação de muitas espécies químicas em solução [18].

Visando melhorar o desempenho dos sistemas FIA, uma série de alterações do sistema original foram propostas como, por exemplo, melhor homogeneização entre amostra e reagente utilizado; adição de reagentes por confluência; economia de reagentes utilizando os processos de zonas coalescentes e fluxo intermitente; ou ainda, diluição em linha da solução da amostra para que a concentração da mesma se situasse na faixa de operação ótima do instrumento utilizando o processo de reamostragem [19].

Um sistema básico para Análise por Injeção em Fluxo é constituído de uma válvula de injeção, em que a comutação é feita através de rotação. Neste dispositivo, uma alça tubular (ou “loop”) de volume fixo é preenchida com a amostra e, após a rotação da válvula, a alça é intercalada no circuito de fluxo e descarrega por deslocamento em direção à bobina de reação e/ou outros dispositivos que precedem o detector [20]; um micro reator formado a partir dos condutores por onde passam os fluidos a serem transportados, e um sistema de detecção capaz de monitorar uma propriedade relacionada,

direta ou indiretamente, com a concentração da espécie química de interesse [18].

Os sistemas em fluxo podem ser classificados quanto ao processo de amostragem, em contínuo ou discreto, e quanto à característica do fluxo, em segmentado, não-segmentado ou monossegmentado [21].

O processo analítico divide-se, basicamente, em três estágios: pré-tratamento das amostras, determinação dos analitos e processamento e análise dos dados. O primeiro estágio envolve etapas críticas como: amostragem e preservação da amostra, dissolução, separação/pré-concentração, reações químicas e transporte até o detector, as quais irão definir a exatidão e a precisão dos resultados. É justamente nesse estágio que a versatilidade da injeção em fluxo foi, desde cedo, reconhecida, uma vez que tornava possível a execução, em linha, dessas operações, as quais eram, normalmente, realizadas em batelada.

Os sinais tipicamente obtidos em sistemas FIA são os chamados fiagramas. A altura está relacionada com a concentração da amostra. O tempo de residência, decorrido entre o ponto de inserção da amostra e o máximo do pico, é o tempo transcorrido entre a inserção e a detecção do pico transiente. Como a resposta é geralmente, rápida, na maioria dos casos, é possível realizar duas ou mais determinações por minuto [22].

Uma característica importante do sistema FIA, é que o módulo de análise é um sistema fechado e uma vez introduzida a alíquota da amostra no percurso analítico, as reações químicas e a detecção ocorrem sem nenhum contato externo [12].

O sistema FIA além de ser uma alternativa aos procedimentos convencionais de manipulação de soluções, apresenta vantagens que o recomendam como técnica muito especial no ensino, pesquisa e na indústria. É de fácil construção, baixo custo de montagem para operação e manutenção, exige pequeno consumo de amostras, padrões e reagentes, alta frequência analítica (60 a 300 determinações por hora) [23] e facilidade de associação com diferentes detectores, tais como, espectrométrico, fluorimétrico, luminescente e eletroquímico. Além disso, é mínima a manipulação de amostras e reagentes; possui grande velocidade analítica com tempo de resposta de 3 a 60 s [24]; elevadas exatidão, precisão,

sensibilidade e, principalmente, pode ser acoplado com outros métodos e técnicas [11, 18, 19].

De um modo geral, desde as primeiras contribuições, os sistemas em fluxo têm experimentado uma grande evolução e diversificação, tanto com relação às configurações estruturais como no tocante a automação de análises químicas. Neste aspecto, os sistemas em fluxo vêm se destacando como uma alternativa eficiente para análises químicas, nos mais diferentes ramos da ciência e, principalmente, em química analítica, merecendo especial atenção por serem de fácil desenvolvimento e economicamente viável [21].

2.2.1. Sistema Fia com Multicomutação e Amostragem Binária (FIA-MBS)

Comutação é a denominação da mudança de posição da parte interna de um injetor proporcional, manual ou automático, com mudança entre portas de uma válvula, para efetuar a inserção das soluções de amostra e reagentes no percurso analítico. A unidade de comutação e reação, de fabricação muito simples, inclui um injetor ou comutador, bobinas de homogeneização e/ou desenvolvimento de reações químicas e, eventualmente, conectores para a adição de fluídos confluentes [25].

Amostragem binária é o conceito designado ao processo de amostragem onde a zona de amostra é constituída de várias alíquotas da solução de amostra, intercaladas com alíquotas de solução de reagente, sendo que o processo de mistura entre as soluções ocorre em função do contato mútuo entre duas interfaces de soluções diferentes e, por isso, a denominação de binária [21].

Usualmente, o módulo de análise para implementar este sistema é constituído por válvulas solenóides, sendo o processo de amostragem controlado por programa computacional. A inserção das soluções no percurso analítico é efetuada acionando-se as válvulas em uma seqüência que possibilite a inserção das soluções individualmente em uma dada ordem pré-definida. Empregando-se esta estratégia, uma fração do percurso analítico é preenchida com alíquotas da solução da amostra separadas por

alíquotas da solução do reagente. Durante o transporte para o detector a mistura entre as soluções processa-se rapidamente, devido às várias interfaces de contato das alíquotas das soluções reagentes, agilizando o desenvolvimento da reação química.

Esse conceito propiciou o desenvolvimento de vários trabalhos de análise química, sendo que o mesmo vem sendo aplicado na automação de diferentes metodologias, tais como titulações empregando detecção fotométrica, espectrofotométrica e potenciométrica [21].

2.3. Espectrofotometria UV/VIS

A espectroscopia molecular baseada na radiação ultravioleta, visível e infravermelha é amplamente empregada para a identificação e determinação de muitas espécies inorgânicas, orgânicas e bioquímicas [26].

De um modo mais amplo, a espectrofotometria é um processo de medida analítica que, basicamente, emprega as propriedades dos átomos e moléculas de absorverem e/ou emitirem energia eletromagnética em uma das regiões do espectro eletromagnético [11].

Os métodos espectrofotométricos apresentam características importantes que incluem:

- Ampla aplicabilidade. Um número de espécies orgânicas, inorgânicas e bioquímicas absorve radiação ultravioleta ou visível sendo assim receptivas para uma determinação quantitativa direta. Muitas espécies não-absorventes podem também ser determinadas após conversão química para um derivado absorvente;
- Alta sensibilidade. Os limites de detecção para a espectroscopia estão na faixa de 10^{-4} a 10^{-5} mol L⁻¹. Com certas modificações de procedimento, essa faixa pode freqüentemente ser estendida para 10^{-6} ou até mesmo 10^{-7} mol L⁻¹;
- Seletividade entre moderada a alta. Com freqüência, pode-se encontrar um comprimento de onda no qual somente o analito absorve, tornando assim, qualquer separação preliminar desnecessária. Além disso, quando ocorre sobreposição de bandas de absorção, as correções baseadas em medidas

adicionais em outros comprimentos de onda eliminam, algumas vezes, a necessidade de uma etapa de separação;

- Boa exatidão. Os erros relativos em concentração observados para os procedimentos espectrofotométricos ou fotométricos empregando-se a radiação ultravioleta e visível estão na faixa de 1% a 5%, embora com precauções espectrais os erros possam ser reduzidos a poucos décimos por cento;

- Facilidade e conveniência de aquisição de dados [26,27].

A espectrofotometria é fundamentada na lei de Lambert-Beer, conhecida também como lei de Beer, que é a base matemática para medidas de absorção de radiação para amostras no estado sólido, líquido ou gasoso, nas regiões ultravioleta, visível e infravermelho do espectro eletromagnético. A lei de Beer diz quantitativamente como a grandeza da atenuação depende da concentração das moléculas absorventes e da extensão do caminho sobre o qual ocorre a absorção. À medida que a luz atravessa um meio contendo um analito, um decréscimo de intensidade ocorre na proporção que o analítico é excitado. Para uma solução do analito de determinada concentração, quanto mais longo for o comprimento do caminho do meio através do qual a luz passa (caminho óptico), mais centros absorventes estarão no caminho, e maior será a atenuação. Também, para um dado caminho óptico, quanto maior for a concentração de absorventes, mais forte será a atenuação [26].

A absorbância A de uma solução está relacionada com a transmitância de forma logarítmica. Para medidas de absorção de radiação em determinado comprimento de onda, tem-se:

$$A = -\log T = \log \frac{I_0}{I} = \varepsilon bc \quad \text{Equação 1}$$

onde A é a absorbância, I_0 é a intensidade da radiação monocromática que incide na amostra e I é a intensidade da radiação que emerge da amostra. O coeficiente de absorvidade (ε) é uma grandeza característica da espécie absorvente, cuja magnitude depende do comprimento de onda da radiação

incidente. O termo c é a concentração da espécie absorvente e b é a distância percorrida pelo feixe através da amostra.

A constante de proporcionalidade, que relaciona a absorbância com o caminho óptico e a concentração das espécies absorventes, denominada de coeficiente de absortividade (ϵ), empiricamente assume valores máximos da ordem de $10^5 \text{ L mol}^{-1} \text{ cm}^{-1}$, para absorção de radiação na região UV-VIS. A magnitude de ϵ depende da seção transversal de captura (S) da espécie e da probabilidade de ocorrência da transição eletrônica (P), relacionada à absorção de radiação, da maneira descrita pela equação 2:

$$\epsilon = 8,7 \times 10^{19} P \times S \quad \text{Equação 2}$$

Considerando-se esta equação, podem-se planejar modificações estruturais nos reagentes, de modo a maximizar a seção transversal de captura e/ou a probabilidade de ocorrência de transições eletrônicas, visando o aumento da absortividade molar da espécie.

As medidas espectrofotométricas de absorbância são normalmente feitas em um comprimento de onda correspondente a um máximo de absorção porque a variação na absorbância por unidade de concentração é a maior possível neste ponto; é obtida, assim, a máxima sensibilidade. Nessa região, a curva de absorção apresenta tangente nula; nessas circunstâncias, uma boa concordância com a lei de Beer pode ser esperada. As medidas são menos sensíveis a incertezas que surgem de falhas em se reproduzir precisamente o ajuste de comprimento de onda do instrumento [27].

A espectrofotometria UV-VIS é a técnica analítica mais empregada nos laboratórios de rotina em função de robustez, custo relativamente baixo das análises, da facilidade de operação, da sensibilidade atingida e grande número de aplicações desenvolvidas [28,29].

O acoplamento desta técnica a sistemas de injeção em fluxo (FIA) tem se mostrado versátil em função da seletividade e sensibilidade atingidas, da facilidade de operar etapas de separação e/ou pré-concentração em condições altamente repetitivas e do gerenciamento de soluções não-equilibradas: a detecção, etapas de separação, dispersão, tempos de

residência da amostra e adição de reagentes. Vários componentes dos analisadores em fluxo tais como amostradores, filtros, reatores, pontos de adição de reagentes, colunas de troca-iônica, detectores óticos e eletroquímicos, podem ser deslocados de sua posição original até outra previamente selecionada do percurso analítico mediante comutação [28].

Para a determinação espectrofotométrica de espécies na região UV-VIS, normalmente é necessário o uso de reagentes para a conversão da espécie de interesse em uma forma que permita a medida de absorção de radiação com maior sensibilidade e/ou seletividade. Sendo assim, espécies que absorvem fracamente podem ser convertidas em compostos com maior absorvidade, visando a determinação de menores quantidades do analito [29].

O comprimento de onda no qual uma molécula orgânica absorve, depende de quão fortemente seus elétrons estão ligados. Os elétrons compartilhados em ligações simples carbono-carbono ou carbono-hidrogênio estão tão fortemente presos que suas excitações requerem energias correspondentes ao comprimento de onda da região do ultravioleta de vácuo, abaixo de 180 nm. Os espectros de ligações simples não têm sido amplamente explorados para as finalidades analíticas em razão das dificuldades experimentais de se trabalhar nessa região. Essas dificuldades ocorrem porque tanto o quartzo como os componentes da atmosfera absorvem nessa região, o que requer o uso de espectrofotômetros mantidos sob vácuo com óptica de fluoreto de lítio.

Os elétrons envolvidos em ligações duplas e triplas das moléculas orgânicas não estão tão fortemente presos sendo, portanto, mais fáceis de serem excitados pela radiação; assim, as espécies com ligações insaturadas, geralmente, exibem picos de absorção úteis.

A sensibilidade e a seletividade de reagentes orgânicos usados para determinação espectrofotométrica podem ser modificadas pela incorporação de diferentes substituintes na molécula do reagente cromogênico [29].

Um cromóforo é um grupo funcional que tem absorção característica na região do ultravioleta e/ou do visível. Estes grupos têm invariavelmente ligações duplas ou triplas e incluem a ligação C=C (e, portanto, o anel de

benzeno), a ligação $C\equiv C$, os grupos nitro e nitroso, o grupo azo e os grupos carbonila e tiocarbonila [30].

A análise multicomponente por espectrofotometria de absorção molecular UV-VIS, usando reagentes cromogênicos, é tradicionalmente realizada utilizando-se um reagente específico para cada componente ou separando-se o elemento de interesse de seus interferentes. Todavia, na prática, nem sempre se dispõe de um reagente específico para cada espécie, e a separação de elementos interferentes presentes na amostra, em geral, é trabalhosa e demorada. Por esse motivo a análise espectrofotométrica simultânea de misturas normalmente se limita à análise parcial de apenas um ou dois componentes usando bandas de absorção características, quando essas aparecem isoladas. Os sistemas com mais de dois componentes são usualmente analisados por etapas, determinando-se um componente de cada vez. Portanto, as pesquisas sobre análise multicomponente simultânea por espectrofotometria de absorção molecular UV-VIS (AMS-EAM) desenvolvem-se no sentido de evitar a exaustiva tarefa de separação de interferentes e de permitir a determinação simultânea de um número cada vez maior de componentes, favorecendo, conseqüentemente, a redução do tempo e do custo das análises. Grande parte das análises multicomponente simultânea por espectrofotometria de absorção molecular UV-VIS vêm sendo realizadas usando diferentes métodos de estatística multivariada. Esses métodos levam à construção de modelos matemáticos que se ajustam aos dados obtidos experimentalmente e incluem métodos de classificação, de otimização, e também métodos de análise qualitativa e/ou quantitativa, que são de especial interesse para os químicos analíticos [31].

Nos últimos anos, a combinação entre métodos quimiométricos e técnicas de espectroscopia molecular apresentou grandes contribuições para a química analítica, tornando possível, por exemplo, a determinação de espécies sem resolução de sinal analítico [32].

2.3.1. Espectrofotometria Derivativa

A espectrofotometria derivativa consiste na representação gráfica das derivadas da absorbância em relação ao comprimento de onda, em função do comprimento de onda. A diferenciação da Lei de Lambert-Beer permite obter as equações 3, 4, 5 que mostram que as derivadas $d^n A/d\lambda^n$ são sempre proporcionais às concentrações do analito, sendo as aplicações analíticas baseadas neste fato.

$$dA/d\lambda = c b (d\epsilon/d\lambda) \quad \text{Equação 3}$$

$$d_2 A/d\lambda_2 = c b (d_2 \epsilon/d\lambda_2) \quad \text{Equação 4}$$

$$d_n A/d\lambda_n = c b (d_n \epsilon/d\lambda_n) \quad \text{Equação 5}$$

A derivação dos espectros não aumenta o conteúdo de informações do espectro original, no entanto permite uma interpretação mais completa do mesmo, pois evidencia com maior clareza seus máximos, mínimos e pontos de inflexão, desde que seja possível localizá-los.

Os métodos de derivação espectral envolvem diferenciação eletrônica ou por software e são disponíveis em muitos espectrofotômetros UV-VIS. Em geral, a derivação espectral visa determinações simultâneas, bem com aumento de seletividade. Entretanto, freqüentemente observa-se aumento de sensibilidade e melhoria do limite de detecção. O aumento da sensibilidade observado na espectrometria derivativa é baseado na observação de que a amplitude da derivada n da absorbância em relação ao comprimento de onda ($D_N = d^n A/d\lambda^n$) é inversamente proporcional à largura da banda do espectro ordinário (W), de acordo com a equação 5, por esta razão, a sensibilidade em métodos derivados depende, não somente dos parâmetros instrumentais e da forma medida do sinal (método da tangente, pico-a-pico ou zero-a-pico), mas também das características do espectro de absorção ordinário. A ordem da derivada deve ser cuidadosamente selecionada, visto que usualmente verifica-se um aumento do nível de ruído com o aumento da ordem de derivação.

$$D_n \propto 1/W^n \quad \text{Equação 6}$$

Esta técnica possibilita a separação do sinal sobreposto; a eliminação do sinal de fundo causado pela presença de outros compostos e a quantificação de um ou mais analito sem prévia separação e purificação.

Baseado na literatura científica, as seguintes aplicações podem ser discutidas:

- Análise de multielementos: Este método consiste na determinação de um ou mais analito em matrizes de amostras complexas. Há processos que permitem o aumento da seletividade, sensibilidade e precisão.
- Determinação de equilíbrio de reações, cálculos de constantes físico-química e de complexação.
- Investigações de cinética de reação [33].

2.4. Métodos de Análise Multivariada

2.4.1. Quimiometria

A aquisição de dados, principalmente na área de química analítica, atingiu um ponto bastante sofisticado com o interfaceamento de instrumentos aos computadores produzindo uma enorme quantidade de informação, muitas vezes complexa e variada. Uma das características mais interessantes dos modernos instrumentos é o número das variáveis que podem ser medidas em uma única amostra. Um exemplo notável é a intensidade de absorção em mil ou mais comprimentos de onda que é rotineiramente registrada em um único espectro. De posse de tal quantidade de dados, a necessidade de ferramentas novas e mais sofisticadas para tratá-los e extrair informações relevantes cresceu muito rapidamente, dando origem à Quimiometria, que é uma área especificamente destinada à análise de dados químicos de natureza multivariada [34].

A quimiometria, do inglês “chemometric”, pode ser definida como a pesquisa e utilização de métodos matemáticos e estatísticos para o tratamento de dados químicos de forma a extrair uma maior quantidade de informações e melhores resultados analíticos.

Os métodos utilizados na quimiometria, a princípio, foram desenvolvidos em outras disciplinas que com a aplicação e pesquisas voltadas para o tratamento de dados químicos acabou dando origem a uma nova área dentro da química analítica. Isso se deu após a segunda metade dos anos 60, com o surgimento de métodos instrumentais computadorizados para a análise química que promoveu a geração de uma grande quantidade de dados. Até este período, os químicos baseavam suas decisões em uma pequena quantidade de dados que, na maioria das vezes, eram obtidos de forma lenta e dispendiosa. A partir dos anos 60, com a grande quantidade de dados de obtenção rápida e com menor esforço, foi preciso analisar todos esses dados e extrair maior quantidade de informações relevantes. Foi então, que teve início à pesquisa e utilização dos métodos matemáticos e estatísticos que acabaram resultando nessa nova área conhecida como quimiometria.

Com o crescimento da quimiometria foram desenvolvidas novas ferramentas para tratamento de dados encontrando aplicações distintas conforme o objetivo do estudo como, por exemplo, a otimização de processos, a classificação de dados e as determinações quantitativas.

A principal linha de pesquisa da quimiometria aplicada à química analítica tem sido a construção de modelos de regressão a partir de dados de primeira ordem, ou seja, dados que podem ser representados através de um vetor para cada amostra, sendo a construção desses modelos denominada calibração multivariada [35].

2.4.2. Calibração Multivariada

Um dos mais importantes problemas na química analítica é a estimativa do número e a concentração das espécies em misturas através de espectros ou cromatogramas. Uma série de técnicas estatísticas tem sido utilizada para desenvolver metodologias multivariadas e para extrair informação dos espectros e cromatogramas, com o objetivo de identificar as espécies presentes e determinar quantitativamente as concentrações de algumas ou todas elas [4].

A aplicabilidade de cada metodologia depende do conjunto de dados (informação experimental) submetidos à análise. Um requisito importante para aplicação dessas metodologias é que os espectros das misturas devem ser uma combinação linear dos espectros das espécies, ponderados por suas concentrações. Em outras palavras, que a Lei de Beer ou alguma expressão semelhante seja obedecida [4].

Os métodos de estatística multivariada levam à construção de modelos matemáticos que se ajustam aos dados obtidos experimentalmente e incluem métodos de classificação, de otimização, e também métodos de análise qualitativa e/ou quantitativa, que são de especial interesse para os químicos analíticos [31].

A calibração multivariada é uma parte importante da quimiometria, que inclui métodos de análise estatística, modelamento matemático e inteligência artificial, com o objetivo de projetar e/ou selecionar procedimentos e experimentos de forma otimizada para a aquisição de dados químicos, obtendo o máximo de informação através da análise destes [5]. O uso dos métodos de análise estatística tanto na química, como em outras ciências experimentais, está intimamente ligado ao desenvolvimento do computador e de equipamentos capazes de realizar diversas medidas em uma amostra simultânea ou seqüencialmente, permitindo assim a aquisição e o tratamento multivariado destes dados [36].

A calibração multivariada estabelece uma relação entre dois blocos de dados de informação química disponível, o das medidas instrumentais e o da propriedade calibrada [35]. No entanto, a aquisição de um grande volume de dados, mesmo que de alguma forma os experimentos tenham sido planejados, não significa necessariamente que estes contenham a informação necessária para descrever a propriedade ou propriedades de interesse nas amostras, nem que o modelo construído descreva este sistema de forma otimizada. Assim, o planejamento dos experimentos e a seleção das variáveis a ser consideradas nestes experimentos também são alvos de estudo da quimiometria, e devem ser levados em conta quando se pretende utilizar métodos mais sofisticados para obter informações acerca dos dados obtidos [5].

O desenvolvimento de modelos com calibração multivariada consiste de duas etapas: o desenvolvimento (ou calibração) e a validação (previsão) [34, 37].

Na etapa de calibração, as relações entre sinais analíticos (Matriz $\mathbf{X}$) e concentrações dos analitos (Matriz $\mathbf{Y}$) são deduzidas a partir de padrões representativos para o conjunto de amostras a serem analisadas; os dados utilizados nesta etapa constituem o conjunto de treinamento [38]. Nesta etapa, deve-se preparar um conjunto de padrões, a partir dos quais possam ser obtidas medidas instrumentais indiretas, constituídas de mais de uma reposta, relacionadas à propriedade de interesse. A propriedade de interesse nesses padrões deve ser quantificada através de um método de referência, independente, de precisão e exatidão conhecidas.

Antes de iniciar os cálculos, esses devem ser digitalizados e agrupados em matrizes [5]. A matriz das variáveis independentes $\mathbf{X}_{(n, m)}$ é composta por espectros ou registros de sinais analíticos [38], onde as linhas representam as n -amostras e as colunas representam as m -variáveis [38], como mostra o esquema apresentado na Figura 9 [34,38]. Na matriz, cada variável é representada por um vetor coluna, e cada amostra é representada por um vetor linha, sendo que é possível descrever cada vetor linha, ou vetor resposta, como um ponto no espaço de dimensão m , isto é, em que os eixos das coordenadas são descritos pelas variáveis [39].

O outro conjunto de dados, constituídos pelas variáveis dependentes, compõe a matriz $\mathbf{Y}_{(n, p)}$, onde as colunas (p) correspondem às concentrações dos analitos e as linhas (n) a combinação entre elas.

As matrizes de dados originais $\mathbf{X}$ e $\mathbf{Y}$ são decompostas em vetores de escores e pesos, mais uma matriz de resíduos de dados não modelados (ruídos estatísticos). O produto destes vetores origina os denominados autovetores, componentes principais ou ainda, variáveis latentes, que são capazes de representar as amostras em poucas dimensões e descrever a direção de máxima variância correlacionada com a propriedade de interesse.

Os escores são as coordenadas das amostras no novo sistema de eixos. Seu gráfico permite a identificação destas, bem como a análise de semelhanças, agrupamentos e “outliers”. Analogamente, os pesos se referem às variáveis presentes nos conjuntos de dados, e seu gráfico

permite a identificação das variáveis importantes, a classificação e a seleção das mesmas. Contêm ainda, informações sobre a relevância de cada variável original na formação dos novos eixos [38].

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

Figura 9: Representação da matriz de dados X

Além de digitalizar e agrupar os dados em matrizes pode ser necessário realizar também algum tipo de manipulação dos dados do conjunto de calibração para a obtenção de um modelo de calibração melhor. Algumas destas operações, conhecidas como pré-processamento, são usualmente aplicadas tanto nas variáveis independentes (matriz **X**) quanto nas variáveis dependentes (matriz **Y**) [40]. Estas operações podem ter uma série de grandezas matemáticas para ponderar sua significância ou corrigir não linearidades de fontes diversas e efeitos instrumentais, como, por exemplo, o espalhamento de radiação em medidas na região do infravermelho. Estes efeitos podem ser conseguidos centrando-se os dados na média, escalonando-os pela sua variância e com auto-escalonamento. Todos estes métodos são sensíveis à presença de amostras anômalas, que são amostras que apresentam um comportamento muito diferente do restante do conjunto [39].

O método de dados centrados na média é o tipo de pré-processamento mais simples e comum, amplamente utilizado nos dados de espectrofotometria [40]. É utilizado quando há presença de sinais caracterizados como ruído e é normalmente utilizado na construção de modelos de calibração [39]. Basicamente, subtrai-se o valor de cada elemento da coluna (x_{ij}) pelo valor médio dos elementos dessa coluna ($\bar{x}_j$), como mostrado na equação 7; obtém-se como resultado, uma matriz onde todas as colunas têm média zero [39]. Desta forma, a origem dos eixos na

qual os dados se encontram é deslocada de modo a colocar os dados em uma forma mais conveniente à análise e visualização [39].

$$X_{ij(cm)} = X_{ij} - \bar{X}_j \quad \text{Equação 7}$$

onde

$X_{ij(cm)}$ = valor centrado na média para a variável j na amostra i;

X_{ij} = valor da variável j na amostra i;

$\bar{X}_j$ = média dos valores das amostras na coluna j, calculado pela equação 8:

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad \text{Equação 5}$$

onde n é o número de amostras.

No método de escalonamento pela variância, cada valor da variável j é dividido pelo seu desvio padrão, como mostrado na equação 6. É utilizado quando as variáveis possuem dimensões muito discrepantes entre si. Desta forma, o peso das variáveis em diferentes escalas é considerado equivalente, minimizando o risco de perda de informações relevantes [39].

$$X_{IJ(ev)} = \frac{X_{ij}}{S_j} \quad \text{Equação 6}$$

onde

$X_{ij(ev)}$ = valor escalado pela variância para a variável j na amostra i;

X_{ij} = valor da variável j na amostra i;

S_j = desvio padrão dos valores da variável j, calculado a partir da variância

S^2 , dada pela equação 9:

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_{ij})^2 \quad \text{Equação 9}$$

onde

n = número de amostras;

X_{ij} = valor da variável j amostra i;

$\bar{X}_j$ = média dos valores das amostras na coluna j, calculado pela equação 5.

O método de auto-escalamento aplica as técnicas de dados centrados na média e escalonamento pela variância de uma só vez, de modo que a transformação realizada sobre o conjunto original de dados permite que cada variável apresente média zero e variância igual a um (auto-escalamento). Desta forma será dada a mesma importância para todas as variáveis, independente da sua dimensão [39].

Matematicamente, isto é feito centrando-se os dados na média e, em seguida, dividindo-os pelo seu respectivo desvio padrão [41], como mostrado na equação 10.

$$X_{ij(as)} = \frac{\bar{X}}{S_j} \quad \text{Equação 10}$$

onde

$X_{ij(as)}$ = valor autoescalado da variável j para a amostra i;

$\bar{X}_j$ = média dos valores das amostras na coluna j, calculado pela equação 5;

S_j = desvio padrão dos valores da variável j, calculado a partir da variância.

Usualmente, medidas espectrofotométricas apresentam uma correlação significativa entre as variáveis e, portanto, os padrões de variações para variáveis com alto e baixo valores são similares. Em decorrência destas características, o grande inconveniente da utilização deste procedimento, nestes tipos de dados, está representado pelo mesmo peso dado ao sinal analítico de interesse e os ruídos [41].

Quando operações de pré-processamento dos dados forem utilizadas durante a calibração, elas também devem ser utilizadas sempre que forem feitas previsões [5].

Depois de construir o modelo, é necessário validá-lo [5]. A validação consiste na otimização da relação que descreve as espécies de interesse e na avaliação da habilidade preditiva do modelo construído [38]. Nesta etapa, as variáveis independentes obtidas, para um outro conjunto de amostras,

são utilizadas em conjunto com o coeficiente de regressão, para calcular os valores previstos para a variável dependente.

No conjunto de validação utilizam-se amostras cujas variáveis dependentes sejam conhecidas para que seja possível estabelecer uma comparação entre os valores previstos pelo modelo e os valores conhecidos previamente através de uma metodologia padrão, o que permitirá a avaliação sobre o desempenho do modelo de calibração proposto [35]. Para isso, medidas instrumentais cujo valor de referência da propriedade de interesse é conhecido devem ter esses valores determinados com o modelo construído e comparados entre si. Diferentes estratégias podem ser utilizadas para reunir os dados que farão parte desse novo conjunto de dados que será utilizado na etapa de teste do modelo, mas recomenda-se que o assim chamado conjunto de validação seja constituído por soluções cujos espectros não tenham sido usados no cálculo modelo.

Quando os objetos do conjunto de validação forem preparados junto com os padrões utilizados na calibração, mas não fazem parte do cálculo do modelo, o processo é conhecido como validação externa. Recomenda-se utilizar validação externa sempre que possível, pois os modelos obtidos usualmente fornecem melhores resultados [5].

Quando os elementos necessários para construir um conjunto de validação para realizar uma validação externa não estão disponíveis, pode-se realizar uma validação interna. Neste caso, todos os padrões preparados são empregados no cálculo de modelo e são utilizados também como amostras para o teste do modelo construído. Uma forma de fazer isso é selecionar subconjuntos do conjunto de calibração que, alternadamente, não serão utilizados no cálculo modelo. No final do processo, denominado validação cruzada, que consiste no procedimento utilizado para a escolha do número de variáveis latentes baseada na magnitude dos erros de previsão do modelo [38], calcula-se o erro associado às determinações em cada caso [5]. Este é um procedimento freqüentemente utilizado quando não se dispõem de um número razoável de amostras para compor ambos os conjuntos.

Um caso particular muito empregado é validação cruzada do tipo “leave one out” (LO) [34] que consiste na remoção de uma amostra do

conjunto de treinamento, construindo o modelo com as demais, e assim sucessivamente. Este procedimento se repete até que todas as amostras sejam excluídas. Ao final obtém-se m modelos com $m-1$ amostras, e o erro de previsão é avaliado através da somatória dos quadrados dos erros de previsão de todas as amostras (PRESS, do inglês “prediction residual sum of squares”) [42].

O parâmetro empregado para avaliar a capacidade preditiva dos modelos, através da comparação entre os valores previstos com os valores de referência é o erro padrão de previsão (SEP), calculado como:

$$SEP = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad \text{Equação 11}$$

onde y_i é a concentração conhecida do analito, $\hat{y}_i$ é a concentração prevista pelo modelo construído e N é o número de experimentos neste conjunto. O SEP permite estimar, em média, qual a performance do modelo e pode ser usado como uma estimativa do desvio padrão das previsões e é obtido nas unidades utilizadas nas determinações. Quando mais de uma propriedade está sendo calibrada, pode-se calcular o SEP para cada uma delas ou apenas um SEP que indicará a capacidade global de previsão do modelo [5]. A escolha do número adequado de variáveis latentes que devem compor o modelo será o que corresponde ao menor erro de previsão [38].

Alternativamente, também pode-se empregar o erro de previsão relativo (%RMSE), que é independente da unidade utilizada, dado por:

$$\%RMSE = \frac{100}{\bar{y}_r} \sqrt{\frac{\sum_{j=1}^N (y_j - \hat{y}_j)^2}{N}} \quad \text{Equação 12}$$

na qual define-se $\bar{y}_r$ como a média das concentrações conhecidas do analito no conjunto em questão [5].

A detecção de amostras anômalas ou “outliers” no conjunto de calibração consiste em uma das etapas mais importantes no processo de modelagem, pois nela verifica-se a qualidade do conjunto de amostras [42].

Este deve ser constituído por um conjunto homogêneo, sendo que as amostras que se distanciam da origem do conjunto ou apresentam-se acima dos limites estabelecidos devem ser cuidadosamente examinadas e se necessário removidas. Para a identificação de “outliers” utiliza-se a análise gráfica de duas grandezas: “leverage” e resíduos de “Student”.

Pela medida de “leverage” é possível conhecer a influência de uma determinada amostra no conjunto de regressão quando comparada com as restantes. As amostras que apresentam valores pequenos influenciam pouco no modelo, porém as que apresentam altas medidas provavelmente apresentarão uma influência alta, freqüentemente inapropriada, uma vez que podem representar um “outlier” [38].

O resíduo de “Student” consiste na estimativa do resíduo para uma amostra assumindo que esta não tenha sido usada na construção do modelo [33] e apresenta o quanto da informação é mantido após o cálculo das variáveis latentes. O resíduo de “Student” para amostra i é dado como:

$$\text{Res.}y_i = \sqrt{\frac{(y-y_i)^2}{(n-1)(1-h_i)}} \quad \text{Equação 13}$$

A análise do gráfico dos resíduos de “Student” versus “leverage” para cada amostra é a melhor maneira de se determinar as amostras anômalas. Amostras com altos resíduos, mas com pequena “leverage” provavelmente têm algum erro no valor da concentração que deve, de preferência, ser medida novamente. Outra opção será a exclusão de tal amostra do conjunto de calibração. Amostras com resíduo e “leverage” altos devem sempre ser excluídas e o modelo de calibração reconstruído [34].

Depois de construídos e otimizados, os modelos desenvolvidos podem ser utilizados para realizar determinações em espectros de amostras com concentrações desconhecidas, obtidas seguindo os mesmos processos empregados na preparação dos padrões utilizados nas etapas anteriores. Para isso, é necessário, mais uma vez, reuni-los em uma matriz de espectros e introduzi-los no computador, submetendo-os à determinação das concentrações de interesse usando o modelo construído [5].

Existem diversos métodos para a construção de modelos de calibração, sendo que a função que ajusta as variáveis dependentes e independentes pode ser linear ou não, dependendo da complexidade do sistema em estudo [35].

Dentre os métodos de calibração existentes os mais difundidos são ainda os métodos de calibração univariada que também são conhecidos como calibração de ordem zero, ou seja, tem-se apenas uma medida instrumental para cada uma das amostras de calibração. No entanto, a aplicação da calibração univariada é restrita, visto que, quando a amostra não é livre de interferentes e a medida é realizada diretamente na metodologia instrumental, isso provavelmente provocará desvios na determinação da propriedade de interesse e a aplicação deste método de calibração torna-se inviável [35]. Sendo assim, para a análise de dados, os métodos multivariados são os mais adequados, porque permitem um estudo com várias espécies presentes ao mesmo tempo, não importando a existência ou ausência de diferenças espectrais marcantes entre elas nem a existência de alta correlação nos dados. É possível, também, a identificação de problemas eventuais com linha base ou interferentes nas amostras usadas na calibração e nas novas amostras, de previsão [42].

Em calibração multivariada, mais de uma resposta instrumental é relacionada com a propriedade de interesse. Esses métodos de calibração possibilitam a análise mesmo na presença de interferentes, desde que estes estejam presentes nas amostras utilizadas para a construção do modelo de calibração [35].

Com a calibração multivariada facilitou-se as análises com sinais analíticos superpostos, as determinações simultâneas, a otimização de controle de qualidade além de possibilitar uma grande melhora na obtenção da informação quantitativa, com aumento da seletividade e confiabilidade [36].

Os métodos tradicionais de calibração CLS (método clássico de mínimos quadrados) e MLR (regressão linear múltipla), têm suas vantagens e desvantagens quando aplicados a problemas químicos. Ambos utilizam toda a informação contida na matriz de dados $\mathbf{X}$ para modelar a concentração, isto é, toda a informação espectral, incluindo informações

irrelevantes (fazem pequena remoção de ruído). O CLS tem como principal problema a necessidade de se conhecer as concentrações de cada espécie espectroscopicamente ativa no conjunto de calibração, o que em geral é impossível nos problemas práticos. Já o método MLR sofre do problema de colinearidade: o número de amostras deve exceder o número de variáveis, que por sua vez devem fornecer predominantemente informação única. Tem-se neste caso a opção de selecionar certo número de variáveis que seja menor que o número de amostras e que produzam informação “única”, o que pode ser demorado e tedioso. Mais interessante então é a utilização de algum método que, como o CLS, use o espectro inteiro para análise, e como o MLR, requeira somente a concentração do analito de interesse no conjunto de calibração. Dois métodos que preenchem estes requisitos são PCR e PLS [34].

Sendo assim, os métodos de calibração multivariada mais utilizados em determinações simultâneas são: Regressão em Componentes Principais (PCR) e Regressão por Mínimos Quadrados Parciais (PLS) [39]. Todas as variáveis relevantes são incluídas nos modelos via PCR ou PLS, o que implica que a calibração pode ser realizada eficientemente mesmo na presença de interferentes, não havendo necessidade do conhecimento do número e natureza dos mesmos [34]. Uma importante propriedade destes métodos é a robustez, ou seja, os parâmetros do modelo praticamente não se alteram com a inclusão de novas amostras no conjunto de calibração [34,39]. Em especial o método PLS tem se tornado uma ferramenta extremamente útil e importante em muitos campos da química, como a físico-química, a química analítica, a química medicinal, ambiental e ainda no controle de inúmeros processos industriais [34].

Estes métodos de calibração são baseados na compressão dos dados originais e tem como ponto de partida a Análise por Componentes Principais (PCA) [39].

2.4.3. Análise das Componentes Principais (PCA)

A análise das componentes principais (PCA), do inglês “Principal Component Analysis” foi introduzida na química por Malinowski no final dos anos 60, com o nome de Análise de Fatores, e a partir da década seguinte um grande número de aplicações foi desenvolvido, o que a tornou muito conhecida e explorada [39, 43].

O objetivo da análise de componentes principais é transformar dados complexos para que as informações mais importantes e relevantes se tornem mais fáceis de visualizar [39], preservando tanto quanto for possível da informação original. Isso é feito utilizando uma combinação linear das variáveis originais que geraram esses dados levando em consideração a variância a elas associada [5]. Para isso, a matriz original **X**, que é uma tabela onde os dados químicos multivariados chamados objetos são dispostos em linhas e variáveis em colunas, é aproximada para 2 matrizes menores que são as matrizes de escores (T) e de pesos (P), mais uma matriz de resíduos que indica a parte não modelada (Figura 10) [37, 44].

$$\begin{array}{c} n \\ \boxed{\mathbf{X}} \\ m \end{array} = \begin{array}{c} 1 \\ \boxed{t_1} \\ m \end{array} \begin{array}{c} n \\ \boxed{P'_1} \\ 1 \end{array} + \begin{array}{c} 1 \\ \boxed{t_2} \\ m \end{array} \begin{array}{c} n \\ \boxed{P'_2} \\ 1 \end{array} + \dots + \begin{array}{c} 1 \\ \boxed{t_h} \\ m \end{array} \begin{array}{c} n \\ \boxed{P'_h} \\ 1 \end{array}$$

Figura 10: Representação gráfica da somatória dos produtos dos vetores escores e pesos.

As combinações lineares das variáveis originais que geram cada componente principal podem ser representadas pela equação 14:

$$PC_i = a_{i1}v_1 + a_{i2}v_2 + \dots + a_{im}v_m \quad \text{Equação 14}$$

onde v são as variáveis originais e a são os coeficientes que medem a importância de cada variável na componente principal e m é o número de variáveis originais [37]. As componentes principais resultantes são ortogonais entre si e a variância descrita por elas, em relação aos dados

originais, é decrescente. Cada componente principal é dada pelo produto de uma coluna da matriz de escores e a respectiva linha da matriz de pesos, obtidas através da decomposição da matriz de dados. Estas características tornam possível utilizar a análise de componentes principais para simplificar e reduzir a dimensionalidade dos dados, selecionar as variáveis e amostras de interesse. Isso pode ser feito examinando-se quais as componentes principais importantes para descrever os dados analisados, quais as variáveis de maior contribuição relativa para o cálculo de cada componente principal e como as amostras estão relacionadas com as componentes.

A redução da dimensionalidade dos dados e sua simplificação são as propriedades mais significativas para os métodos de calibração multivariada. Isso é possível, pois pode-se selecionar quais as colunas da matriz de escores contém informação estatística relevante para a calibração da propriedade de interesse [5].

Os componentes principais são determinados baseados no critério de variância máxima. Cada componente principal subsequente descreve um máximo de variância, que não é modelado pelos componentes anteriores. De acordo com isso, o primeiro componente principal (PC1), tem a direção que descreve a máxima dispersão das amostras [34], sendo responsável pela explicação de grande parte da variância apresentada pelo conjunto. No entanto, apenas esta componente principal pode não ser suficiente para explicar o comportamento global do conjunto de dados, sendo necessária uma segunda componente (PC2), a qual deve ser ortogonal à primeira e responsável pela explicação de uma parcela importante de variância observada [34,41]. No segundo componente, há mais informação que o terceiro e, assim, sucessivamente. [5]

Em princípio, é possível extrair tantas componentes principais quanto o número de variáveis. No entanto, dada a baixa variabilidade dos dados em torno de alguma componente principal, é possível que esta seja descartada. Este procedimento atende às duas grandes premissas do PCA, ou seja, a redução de variáveis, sem perda de informação relevante [41].

2.4.4. Regressão por Mínimos Quadrados Parciais (PLS)

A regressão por mínimos quadrados parciais (PLS), do inglês “Partial Least Squares”, é considerada o método de regressão mais utilizado para a construção de modelos de calibração multivariada a partir de dados de primeira ordem. Este método, assim como o PCR, não requer um conhecimento exato de todos os componentes presentes nas amostras podendo realizar a previsão de amostras mesmo na presença de interferentes, desde que estes também estejam presentes por ocasião da construção do modelo [35].

O método PLS pode ser dividido em três etapas básicas. A primeira consiste em montar as matrizes de dados **X** a partir dos dados espectrais (variáveis independentes) e **Y** a partir das propriedades de interesse (variáveis dependentes). A segunda etapa consiste em eliminar informações que não são relevantes ao sistema, através de combinações lineares das colunas da matriz **X** feita pela análise das componentes principais e a terceira etapa, consiste em obter-se a relação entre as matrizes **X** (reduzida pela PCA) e a matriz **Y** que possui a propriedade de interesse [45].

Esta técnica multivariada tem sido utilizada para análises quantitativas e, algumas vezes, é capaz de modelar variações não lineares. O cálculo do modelo previsto é realizado diretamente e não requer consumo de tempo e nem parâmetros de aprendizagem escolhidos arbitrariamente ou arquiteturas em rede [43].

A regressão PLS é um método de análises de fatores em que a matriz de espectros é decomposta em matrizes de variações do espectro (pesos) e posição das amostras (escores). Os espectros originais podem ser considerados como combinações lineares das variações dos espectros (pesos) onde os escores representam a contribuição de cada [43].

No PLS tanto a matriz das variáveis independentes **X**, como das variáveis dependentes **Y** são representadas pela Análise das Componentes Principais [39]:

$$X = TP^t + E = \sum t_h p_h' + E$$

Equação 15

$$Y = UQ^t + F = \sum u_h q_h' + F \quad \text{Equação 16}$$

onde T e U são as matrizes de escores das matrizes **X** e **Y** respectivamente, P e Q são as matrizes de pesos das matrizes **X** e **Y** respectivamente; **E** e **F** são os resíduos. O sobrescrito t indica matriz transposta [45]. O intuito do modelo é encontrar a melhor relação entre **X** e **Y** e que ao mesmo tempo resulte num valor do resíduo F mais próximo de zero possível [39].

Uma relação entre duas matrizes pode ser realizada correlacionando-se os escores para cada componente de cada vez (variáveis independentes e dependentes), utilizando-se um modelo linear:

$$u_h = bt_h \quad \text{Equação 17}$$

onde b é o vetor de coeficientes de regressão entre a matriz de escores U de **Y** e a matriz de escores T de **X**.

Este algoritmo geralmente converge muito rápido, realizando pequenas rotações das variáveis latentes dos blocos de X e Y, para que se obtenha a melhor relação linear possível entre os escores desses dois blocos.

A matriz **Y** pode ser calculada de u_h , como pode ser verificado pela equação 18:

$$Y = TBQ' + F \quad \text{Equação 18}$$

Sendo que, a concentração de novas amostras pode ser prevista a partir dos novos escores, T^* , substituídos na equação 19:

$$Y = T^*BQ \quad \text{Equação 19}$$

Nesse processo o ideal é encontrar o número de variáveis latentes que irá fornecer uma melhor previsão das novas amostras e que forneça um menor erro de previsão [45].

Uma das limitações deste modelo é que os escores são calculados separadamente para **X** e **Y**, e somente depois é construída a relação entre eles, o que pode resultar num modelo insatisfatório, por exemplo, numa relação não linear entre **X** e **Y**. Existe um tratamento alternativo que permite

calcular os escores dos dois blocos de forma a produzir a melhor relação linear entre eles. Isto é feito fazendo-se uma rotação dos eixos das componentes principais. Como os escores são as projeções dos pontos nestes eixos, conseqüentemente seus valores são mudados, de maneira a permitir que a melhor relação linear entre os escores X e de Y seja construída [39].

O PLS se destaca por sua robustez e os parâmetros do modelo resultante praticamente não se alteram com a inclusão de novas amostras no conjunto de calibração, proporcionando um incremento na capacidade preditiva do modelo à medida que se aumenta a quantidade de informação [46]. Entretanto, a inclusão de um grande número de variáveis leva à incorporação de ruídos (“overfitting”). Por outro lado, o número reduzido de parâmetros resulta em um modelo não ajustado (“underfitting”), inadequado para a previsão de novas amostras [34].

2. 5. Ambiente Computacional MATLAB

O MATLAB é um ambiente computacional interativo de alta performance voltado para o cálculo numérico [47,48]. Permite implementar e resolver problemas matemáticos muito mais rápido e eficiente que através de outras linguagens como C, Basic, Pascal ou Fortran [49].

MATLAB integra análise numérica, alto nível de linguagem de programação, cálculo com matrizes, processamento de sinais e construção de gráficos em ambiente fácil de usar onde problemas e soluções são expressos como são escritos matematicamente.

O nome MATLAB vem de “Matrix Laboratory”, e foi originalmente desenvolvido para fornecer um acesso fácil ao tratamento de vetores e matrizes. É um sistema interativo cujo elemento básico de informação é uma matriz que não requer dimensionamento. As matrizes podem ser reais ou complexas, e o MATLAB faz uso de objetos diversos, como sinais (simples vetores), imagens, polinômios, dados estatísticos multivariados e sistemas lineares [47].

O MATLAB dispõe de uma biblioteca abrangente de funções matemáticas, geração de gráficos e manipulação de dados. E ainda possui uma vasta coleção de bibliotecas denominadas ToolBoxes [50] que são coleções de funções usadas para resolver determinados problemas como, por exemplo, otimização, manipulação algébrica, redes neurais, processamento de sinais, simulação de sistemas dinâmicos [49].

Um dos aplicativos das “caixas de ferramentas” do MATLAB, o PLS_ToolBox, é largamente utilizado por profissionais de química. O PLS_ToolBox é uma pacote que contém rotinas de análises de componentes principais (PCA), regressão de componentes principais (PCR), método dos mínimos quadrados parciais (PLS), regressão contínua (CR) e algumas rotinas que aplicam métodos para alguns problemas específicos; como uma rotina que usa PLS para identificar modelos dinâmicos de processos químicos. Também, estão incluídas funções de escala de dados de rotinas de conduzir predição com PLS e uma visão do PLS que usa uma relação não-linear [51].

A linguagem e o ambiente de programação MATLAB permitem ainda que o usuário possa escrever suas próprias bibliotecas em MATLAB. Assim, o usuário pode enriquecer a linguagem, incorporando a ela novas funções [50].

PARTE EXPERIMENTAL

3. PARTE EXPERIMENTAL

3.1. Amostras de Aminoácidos

Neste trabalho, misturas sintéticas de cinco aminoácidos foram utilizadas para o estudo com a reação com NQS: Fenilalanina (Fluka Biochemika) e Triptofano (Sigma Aldrich) como aminoácidos aromáticos, Prolina (Fluka Biochemika) e Isoleucina (Fluka Biochemika) como aminoácidos secundários e Histidina (Sigma Aldrich) como um aminoácido básico.

3.2. Equipamentos e Acessórios

Nos experimentos realizados, foi utilizada como unidade propulsora uma bomba peristáltica Ismatec, modelo ICP de 8 canais, provida de tubos de bombeamento de Tygon[®]. Para a introdução das alíquotas de amostras no sistema de análises, foi utilizado um injetor-comutador.

Para confecção das linhas de transmissão, bem como das alças de amostragem e dos reatores tubulares helicoidais, foram utilizados tubos de polietileno com diâmetro interno de 0,35 mm.

Espectros de absorção UV-VIS foram obtidos em espectrofotômetro HITACHI U 2000 e cubetas de quartzo de 1 cm de caminho óptico.

Todas as pesagens foram realizadas em balança analítica digital da marca Bel, modelo 210A.

As medidas de pH foram realizadas utilizando-se um pHmetro MS Tecnologia, modelo mPA210 com eletrodo combinado para pH.

O banho para termostatização do sistema reacional, mantido a 70 °C, foi obtido com um frasco, com os lados isolados termicamente, sobre um agitador magnético marca Tecnal, modelo Te-0853, que possui sensor de temperatura acoplado, para controle do mesmo.

3.3. Preparo de soluções

Todas as soluções foram preparadas com água deionizada e reagentes de grau analítico.

Solução de 1,2-naftoquinona-4-sulfonato de sódio (NQS) $0,00100 \text{ mol L}^{-1}$: Esta solução foi preparada pesando-se em um béquer de 10 mL 0,0260g de $\text{C}_{10}\text{H}_5\text{NaO}_5\text{S}$. Esta massa foi dissolvida em solução de ácido clorídrico $0,1 \text{ mol L}^{-1}$ e transferida para um balão volumétrico de 100 mL. Aferindo-se o volume.

Solução de ácido clorídrico $0,1 \text{ mol L}^{-1}$: Foram medidos 8,30 mL de ácido clorídrico concentrado e adicionou-se cuidadosamente em água deionizada. Após resfriamento da solução, transferiu-se para um balão volumétrico de 1000 mL e completou-se o volume com água deionizada.

Solução tampão de carbonato de sódio $0,1 \text{ mol L}^{-1}$ / hidróxido de sódio $0,75 \text{ mol L}^{-1}$: Foram medidos 93,75 mL de uma solução de NaOH $8,00 \text{ mol L}^{-1}$ transferiu-se para um béquer de 1000 mL e adicionou-se água deionizada. Pesaram-se 10,6000 g de Na_2CO_3 e dissolveu-se na solução de NaOH. O pH foi ajustado para 10,00 pela adição de ácido clorídrico concentrado. Transferiu-se para um balão volumétrico de 1000 mL e completou-se o volume com água deionizada.

Solução de Fenilalanina $0,01000 \text{ mol L}^{-1}$: Foram pesados 0,82595 g de $\text{C}_9\text{H}_{11}\text{NO}_2$ e dissolveu-se em água deionizada. Transferiu-se a solução para um balão volumétrico de 500,0 mL e completou-se o volume. As demais soluções foram preparadas a partir desta solução estoque.

Solução de Histidina $0,01000 \text{ mol L}^{-1}$: Foram pesados 0,7758 g de $\text{C}_6\text{H}_9\text{N}_3\text{O}_2$ e dissolveu-se em água deionizada. Transferiu-se a solução para um balão volumétrico de 500,0 mL e completou-se o volume. As demais soluções foram preparadas a partir desta solução estoque.

Solução de Prolina 0,01000 mol L⁻¹: Foram pesados 0,5756 g de C₉H₉NO₂ e dissolveu-se em água deionizada. Transferiu-se a solução para um balão volumétrico de 500,0 mL e completou-se o volume. As demais soluções foram preparadas a partir desta solução estoque.

Solução de Triptofano 0,01000 mol L⁻¹: Foram pesados 1,0211 g de C₁₁H₁₂N₂O₂ e dissolveu-se em água deionizada. Transferiu-se a solução para um balão volumétrico de 500,0 mL e completou-se o volume. As demais soluções foram preparadas a partir desta solução estoque.

Solução de Isoleucina 0,01000 mol L⁻¹: Foram pesados 0,6559 g de C₆H₁₃NO₂ e dissolveu-se em água deionizada. Transferiu-se a solução para um balão volumétrico de 500,0 mL e completou-se o volume. As demais soluções foram preparadas a partir desta solução estoque.

3.4. Parâmetros Reacionais

O desenvolvimento da reação de 1,2-naftoquinona-4-sulfonato de sódio (NQS) com aminoácidos é altamente dependente de variáveis como temperatura, pH, concentração de reagentes e volume de fluxo. Um estudo preliminar para avaliar o efeito dessas variáveis foi realizado por Saurina et al. [10].

pH e concentração da solução de tampão: O valor de pH foi estudado no intervalo de 8 a 12. O valor máximo de absorbância foi alcançado a pH 10. O efeito da concentração da solução tampão foi estudado a pH 10 para soluções contendo entre 0,01 a 1,00 mol L⁻¹ de carbonato de sódio. A absorbância aumenta significativamente quando a concentração de carbonato de sódio diminui. A solução tampão com concentração de 0,10 mol L⁻¹ de Na₂CO₃ foi selecionada.

Concentração do reagente (NQS): A concentração do reagente NQS foi variada no intervalo de 5,0 x 10⁻⁴ a 2,5 x 10⁻³ mol L⁻¹. Apesar do valor de absorbância aumentar com o valor crescente de concentração, um valor de 0,001 mol L⁻¹ foi selecionado uma vez que o ruído de fundo (background)

aumenta significativamente com a crescente concentração do reagente, apesar do máximo de absorbância não ser alcançado.

Variação do fluxo e temperatura: A combinação do comprimento e o diâmetro interno do reator e a taxa de fluxo total determina o tempo de residência da amostra na bobina da reação. O tempo de residência e temperatura são fatores de grande influência no valor de absorbância por causa da reação cinética entre NQS e aminoácidos. A influência desses parâmetros na absorbância foi simultaneamente estudada. O sinal de máximo foi alcançado com um tempo de residência longo (17-20 min) e baixa temperatura (25-30 °C), mas nestas condições o processamento de amostra se deu muito lentamente. Para baixos tempos de residência, a absorbância aumenta consideravelmente com a temperatura crescente. Os valores finais selecionados foram taxa de fluxo total de $1,10 \text{ mL min}^{-1}$; bobina de reação $10,0 \text{ m} \times 0,35 \text{ mm}$ e temperatura de 70 °C [10].

3.5. Procedimento Geral

O sistema mostrado nas Figuras 11 e 12 foi empregado para obtenção do sinal analítico baseado na reação de 1,2 naftoquinona-4-sulfonato de sódio (NQS) com aminoácidos. As soluções foram adicionadas, continuamente, empregando-se uma bomba peristáltica e tubos capilares de polietileno. A vazão de todas as soluções foi fixada em $0,40 \text{ mL min}^{-1}$.

Inicialmente, o reagente (NQS) e a solução tampão ($\text{Na}_2\text{CO}_3/\text{NaOH}$) foram inseridos por confluência gerando assim, a solução alcalina pela mistura destas em uma bobina de reação ($1,00 \text{ m} \times 0,35 \text{ mm}$). A reação foi desenvolvida em linha uma vez que a solução de NQS decompõe em meio alcalino. A amostra foi inserida no fluxo transportador e direcionada ao ponto de confluência z, misturando-se com a solução alcalina (NQS + tampão).

A reação analítica foi desenvolvida em uma bobina de reação ($10,0 \text{ m} \times 0,35 \text{ mm}$) emergida em um banho termostatizado mantido a 70 °C. A solução final foi coletada diretamente na cubeta de quartzo para leitura no espectrofotômetro.

Antes de iniciar a leitura no espectrofotômetro, foram ajustadas as condições de trabalho. A solução tampão ($\text{Na}_2\text{CO}_3/\text{NaOH}$) foi usada inicialmente para estabelecer a linha de base.

Os espectros foram registrados no intervalo de 290 a 590 nm com uma resolução espectral de 1 nm e velocidade de varredura de 300 nm min^{-1} obtendo-se 301 variáveis por espectro.

Para a obtenção dos espectros de absorção referente as espécies químicas envolvidas no ambiente reacional, o espectrofotômetro foi interfaceado a um microcomputador (Figura 11) através de uma porta serial e um cabo bidirecional RS232C, de acordo com as instruções do fabricante. O controle do espectrofotômetro e a aquisição de dados foram feitos através de um programa desenvolvido em Quickbasic; os dados foram arquivados no formato ASCII. As soluções de aminoácidos foram estudadas individualmente e em misturas.

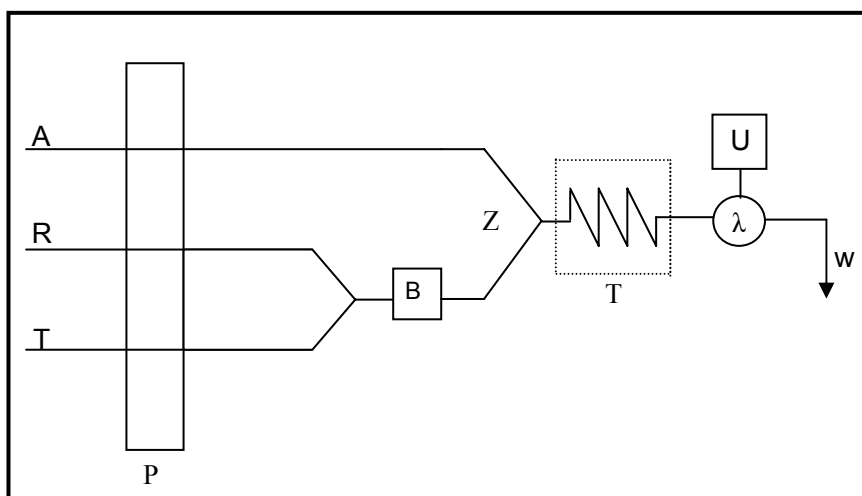


Figura 11. Esquema do Sistema empregado para obtenção do sinal analítico baseado na reação de 1,2 naftoquinona-4-sulfonato de sódio (NQS) com aminoácidos. A: amostra; R: reagente: $1,00 \times 10^{-3} \text{ mol L}^{-1}$ em $0,10 \text{ mol L}^{-1} \text{ HCl}$, T: tampão: $0,10 \text{ mol L}^{-1} \text{ Na}_2\text{CO}_3 / 0,75 \text{ mol L}^{-1} \text{ NaOH}$; P: bomba peristáltica; B: bobina de reação ($1 \text{ m} \times 0,80 \text{ mm}$); Z: ponto de confluência da amostra com solução alcalina; T: banho termostaticado; R: bobina de reação ($10 \text{ m} \times 0,8 \text{ mm}$); λ : detector; U: unidade de aquisição de dados e w: lixo.



Figura 12. Sistema utilizado para reação de 1,2 naftoquinona-4-sulfonato de sódio (NQS) com aminoácidos em pH 10.



Figura 13. Sistema utilizado no processo de aquisição de dados

3.6. Processamento dos Dados

Os arquivos com os dados espectrais armazenados no formato ASCII foram dispostos numa matriz de dados que continham os valores das leituras de absorbância a cada comprimento de onda usando o programa Origin 6.1. Em seguida, os dados foram transferidos para o ambiente MatLab 7.0.

3.7. Desenvolvimento de Modelos de Calibração por Espectrofotometria UV-VIS.

Para o desenvolvimento de modelos de calibração, foram elaboradas curvas analíticas para cada aminoácido no comprimento de onda de máxima absorção. Cada curva analítica foi obtida com 6 valores de concentração, cobrindo-se a faixa de $3,00$ a $7,00 \times 10^{-4} \text{ mol L}^{-1}$ para fenilalanina; $3,00$ a $6,00 \times 10^{-4} \text{ mol L}^{-1}$ para Histidina, $1,50$ a $7,50 \times 10^{-4} \text{ mol L}^{-1}$ para Isoleucina; $0,50$ a $9,00 \times 10^{-4} \text{ mol L}^{-1}$ para prolina e $1,00$ a $9,00 \times 10^{-4} \text{ mol L}^{-1}$ para Triptofano. As soluções foram preparadas pela dissolução da solução estoque $0,01 \text{ mol L}^{-1}$ de cada aminoácido.

3.8. Planejamento Fatorial

O planejamento fatorial é um método do tipo simultâneo, cujas variáveis de interesse que apresentam influências significativas na resposta são avaliadas ao mesmo tempo. Para realizar um planejamento fatorial, escolhem-se as variáveis a serem estudadas (fatores) e efetuam-se experimentos em diferentes valores destes fatores, os níveis. A seguir são realizados experimentos para todas as combinações possíveis dos níveis selecionados [52].

De um modo geral, o planejamento experimental pode ser representado por b^a , onde a é o número de fatores e b o número de níveis escolhidos. Em geral, os planejamentos fatoriais com dois níveis são os mais comuns, por envolverem a realização de poucos experimentos. Nos planejamentos experimentais em que as variáveis são exploradas em 2 níveis, é comum codificá-los usando os sinais (+) e (-). A distribuição destes sinais aos níveis superiores ou inferiores é feita de forma arbitrária e não interfere na realização dos experimentos ou interpretação dos resultados.

O desenvolvimento do conjunto de calibração multivariada foi preparado de acordo com um planejamento experimental 2^5 . Com este planejamento foram produzidas 32 misturas com diferentes proporções para os cinco aminoácidos, conforme mostrado na Tabela 1.

Tabela 1. Planejamento fatorial 2^5 realizado para calibração multivariada.

Misturas	Fatores				
	Phe	His	Pro	Trp	Ile
1	+	+	+	+	+
2	+	+	+	+	-
3	+	+	+	-	+
4	+	+	+	-	-
5	+	+	-	+	+
6	+	+	-	+	-
7	+	+	-	-	+
8	+	+	-	-	-
9	+	-	+	+	+
10	+	-	+	+	-
11	+	-	+	-	+
12	+	-	+	-	-
13	+	-	-	+	+
14	+	-	-	+	-
15	+	-	-	-	+
16	+	-	-	-	-
17	-	+	+	+	+
18	-	+	+	+	-
19	-	+	+	-	+
20	-	+	+	-	-
21	-	+	-	+	+
22	-	+	-	+	-
23	-	+	-	-	+
24	-	+	-	-	-
25	-	-	+	+	+
26	-	-	+	+	-
27	-	-	+	-	+
28	-	-	+	-	-
29	-	-	-	+	+
30	-	-	-	+	-
31	-	-	-	-	+
32	-	-	-	-	-

3.9. Análise das Componentes Principais (PCA)

As medidas espectrofotométricas das 32 misturas sintéticas foram submetidas à análise de PCA.

Para a elaboração do modelo multivariado, os dados espectrais foram centrados na média (translação do sistema de origem até o centro do conjunto de dados) visando facilitar a visualização bem como reduzir a dimensão do modelo construído. Esta operação foi realizada na matriz de dados das medidas de absorbância ($\mathbf{X}$). Esse tipo de processamento consiste, basicamente, na subtração do valor de cada elemento da coluna (X_{ij}) pelo valor médio dos elementos dessa coluna (X_{mj}), obtendo-se como resultado uma matriz onde toda a coluna tem média zero.

Após a etapa de pré-processamento a matriz de dados foi submetida à análise de PCA.

Estes procedimentos foram executados utilizando-se uma rotina do pacote quimiométrico PLS_ToolBox versão 2.0 no ambiente MatLab versão 7.0.

3.10. Utilização do PLS para Calibração Multivariada

Os dados para calibração multivariada correspondem à média de três varreduras de cada uma das 32 misturas de acordo com o planejamento experimental utilizado.

O tratamento dos dados foi realizado através do método dos mínimos quadrados parciais (PLS) contido no pacote quimiométrico “PLS-ToolBox”.

Os dados foram arranjados em duas matrizes, uma com valores das concentrações dos 5 aminoácidos, num total de 32 padrões e a segunda com os valores de absorbância no intervalo 290 a 590 nm em intervalos de 1 nm, obtendo-se 301 valores por espectro.

Como resultado, obteve-se uma matriz de concentração (32 linhas por 5 colunas), chamada de matriz $\mathbf{Y}$, e uma matriz de resposta (32 linhas e 301 colunas), chamada de matriz $\mathbf{X}$.

3.11. Pré-Processamento dos Dados

Para verificar uma melhora na eficiência dos modelos multivariados envolvendo espectroscopia UV-VIS, alguns procedimentos de pré-processamento dos dados foram utilizados:

- i) Primeira Derivada: processada com o software Origin Pro 6.1; principalmente para melhorar a separação de sinais não totalmente sobrepostos;
- ii) Alisamento: processado com o software Origin Pro 6.1; para diminuição de ruídos espectrais.

3.12. Análise em Amostras Reais

Duas amostras de suplementos alimentares protéicos, uma em pó e a outra líquida, foram submetidas a extrações. O equivalente a 3,00 g da amostra sólida e 3,00 mL da amostra líquida foram extraídos com 50,00 mL de solução fenol (1% v/v) por 30 minutos com agitação magnética. Os extratos obtidos foram filtrados em membrana de poro de 45 µm; 5,00 mL do extrato filtrado foram transferidos para um balão de 50,00 mL e completou-se o volume com água deionizada. Após diluição, 5,00 mL desta solução foi diluída para 50,00 mL, completando-se o volume com água deionizada. Todos os ensaios foram realizados em triplicata.

RESULTADOS E DISCUSSÃO

4. RESULTADOS E DISCUSSÃO

4.1. Influência do Tempo de Leitura

A estabilidade do produto da reação entre NQS e aminoácidos foi avaliada em função do tempo, utilizando-se as soluções de concentração $5,00 \times 10^{-4} \text{ mol L}^{-1}$ de cada aminoácido. Os espectros foram registrados nos tempos 0, 3, 5, 8 e 10 minutos após a coleta da amostra na cubeta de quartzo e podem ser visualizados nas Figuras 14(a) a 18(a).

A variação do valor de absorbância em 480 nm em função do tempo pode ser verificada nas Figuras 14(b) a 18(b) e na Tabela 2. Neste estudo, modificações no sinal espectral da ordem de 0,008 a 0,042 nm podem ser observadas na região de máxima absorção (480 nm).

Pela análise dos gráficos e da Tabela 2, verifica-se que o composto é estável no intervalo estudado, sendo assim, as leituras espectrais foram realizadas de maneira discreta, sem perda de informação.

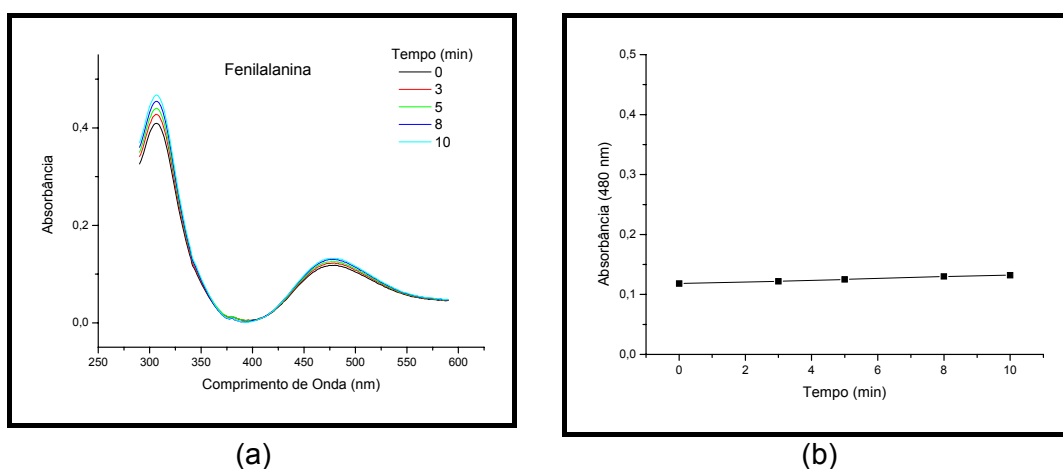


Figura 14. (a) Espectros obtidos para a solução de Fenilalanina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Fenilalanina ($5,00 \times 10^{-4} \text{ mol L}^{-1}$) no intervalo de 0 a 10 min.

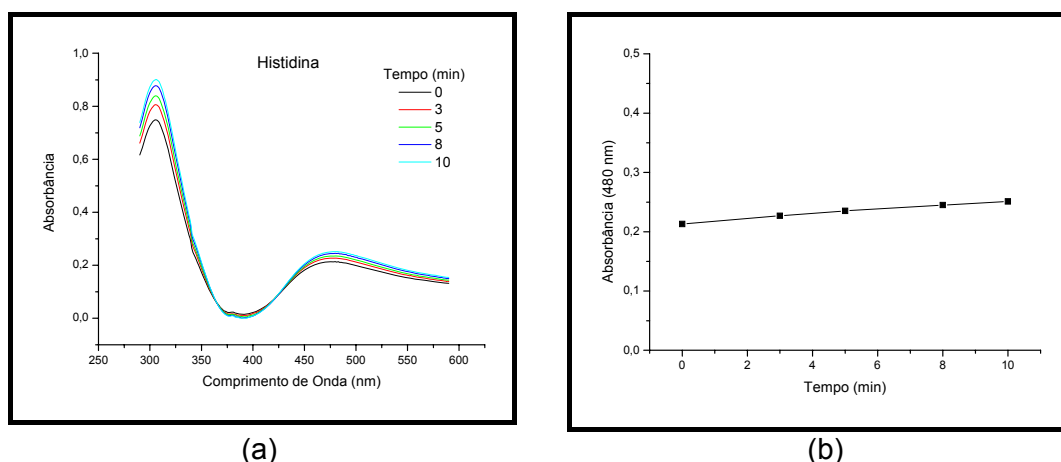


Figura 15. (a) Espectros obtidos para a solução de Histidina ($5,00 \times 10^{-4}$ mol L $^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Histidina ($5,00 \times 10^{-4}$ mol L $^{-1}$) no intervalo de 0 a 10 min.

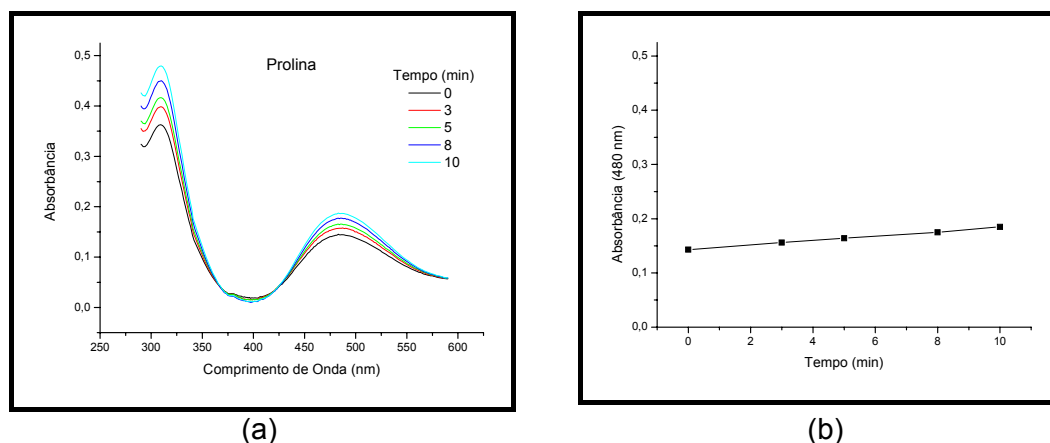


Figura 16. (a) Espectros obtidos para a solução de Prolina ($5,00 \times 10^{-4}$ mol L $^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Prolina ($5,00 \times 10^{-4}$ mol L $^{-1}$) no intervalo de 0 a 10 min.

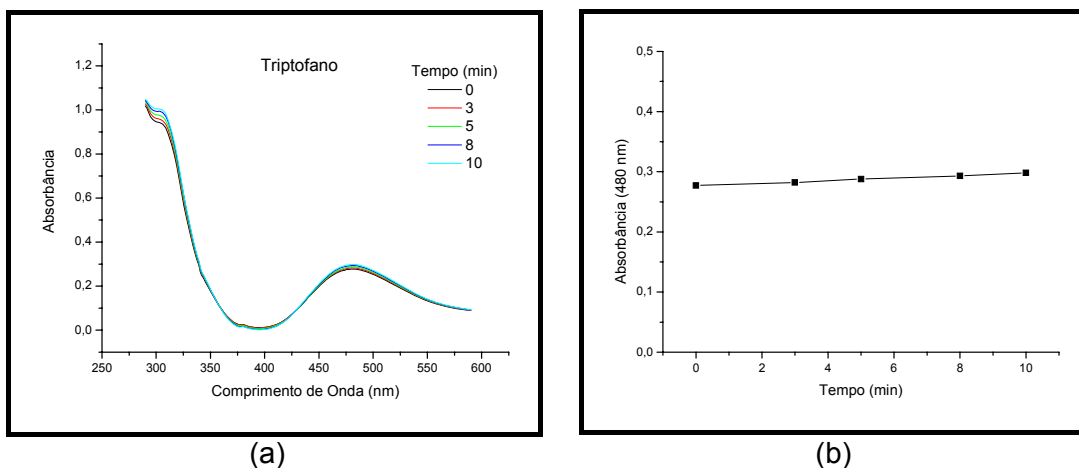


Figura 17. (a) Espectros obtidos para a solução de Prolina ($5,00 \times 10^{-4}$ mol L $^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Tryptofano ($5,00 \times 10^{-4}$ mol L $^{-1}$) no intervalo de 0 a 10 min.

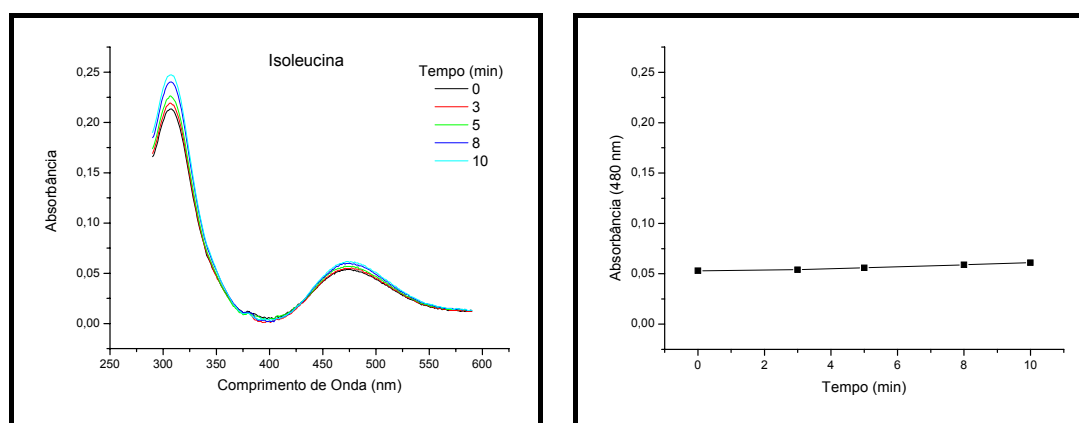


Figura 18. (a) Espectros obtidos para a solução de Isoleucina ($5,00 \times 10^{-4}$ mol L $^{-1}$) em diferentes tempos de leitura; (b) Variação do valor absorbância em 480 nm para a solução de Isoleucina ($5,00 \times 10^{-4}$ mol L $^{-1}$) no intervalo de 0 a 10 min.

Tabela 2. Variação da absorbância em função do tempo para os cinco aminoácidos.

Aminoácido	Absorbância (480 nm)		
	0 min	10 min	Δt
Fenilalanina	0,118	0,132	0,014
Histidina	0,213	0,251	0,038
Prolina	0,143	0,185	0,042
Triptofano	0,277	0,298	0,021
Isoleucina	0,053	0,061	0,008

4.2. Curvas Analíticas dos Aminoácidos

Para o desenvolvimento de modelos de calibração, realizou-se a leitura espectrofotométrica para cada aminoácido como pode ser visto nas figuras 19(a) a 23(a).

A partir dos espectros de absorção obtidos construíram-se as curvas analíticas, com base nas regiões de máxima absorção na região do visível como pode ser observado nas Figuras 19(b) a 23(b). As curvas analíticas apresentaram coeficientes de regressão de 0,9980 para fenilalanina; 0,9940 para histidina; 0,9980 para triptofano; 1,0000 para prolina e 0,9940 para isoleucina.

Os dois níveis de concentração utilizados para realizar o planejamento experimental foram selecionados a partir da curva analítica de calibração tomando-se como nível inferior (-) a menor concentração e como nível superior (+) a maior concentração (Tabela 3).

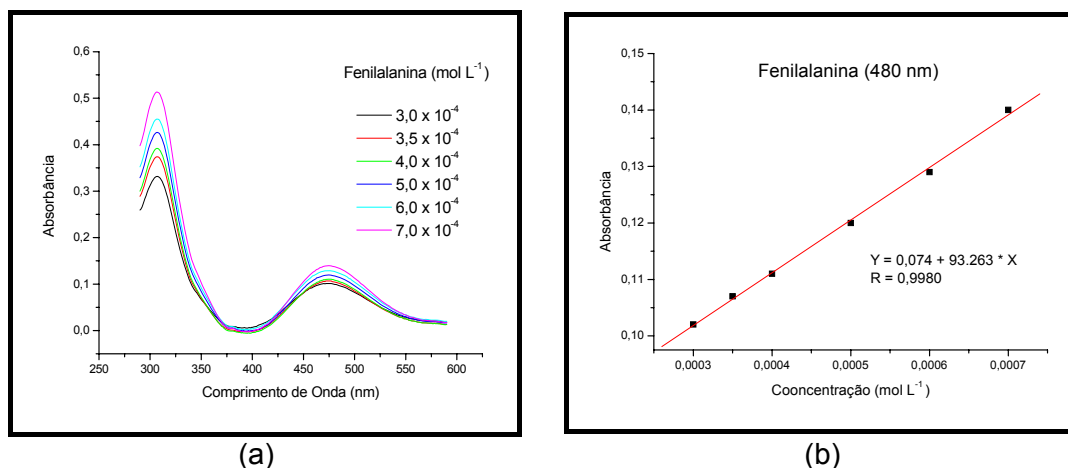


Figura 19. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Fenilalanina.

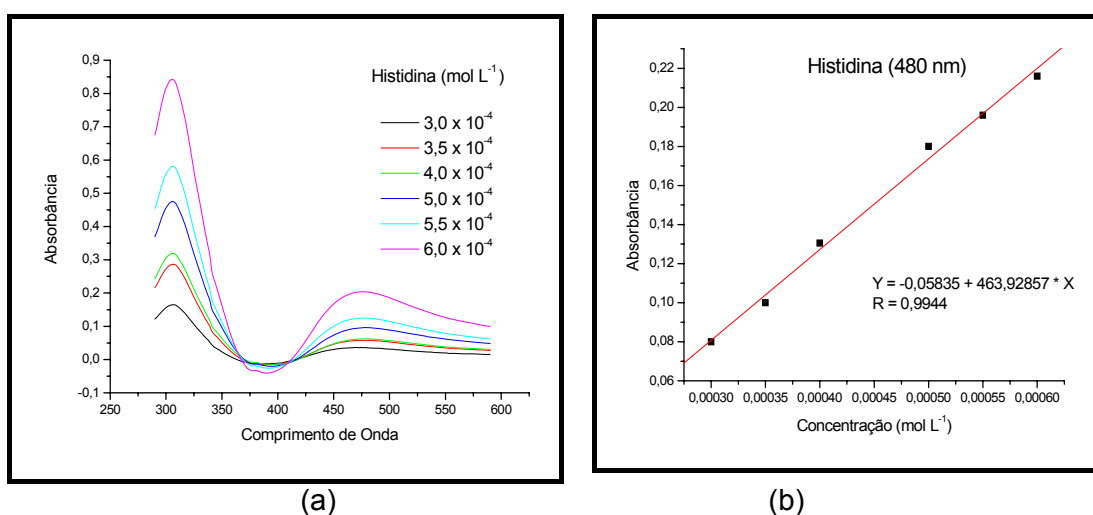


Figura 20. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Histidina.

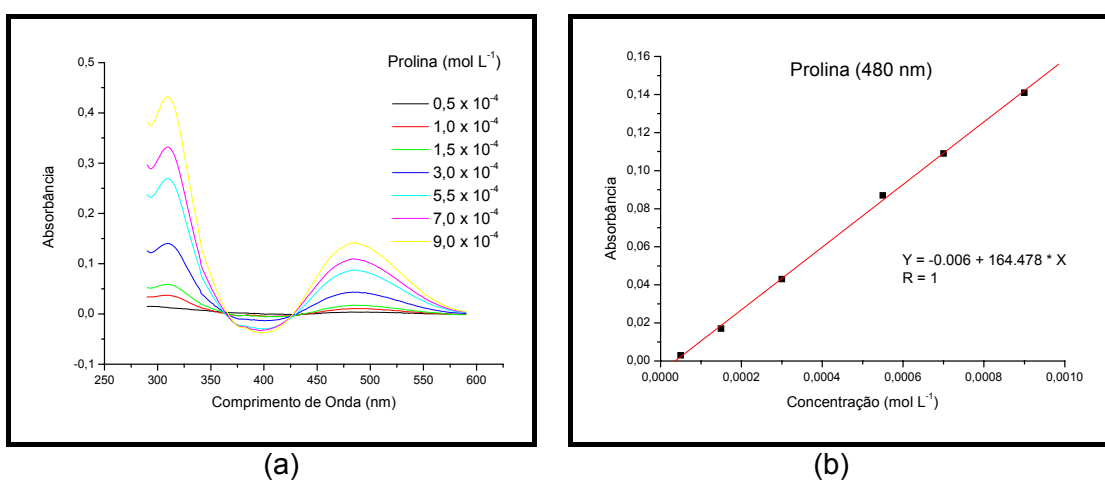


Figura 21. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Prolina.

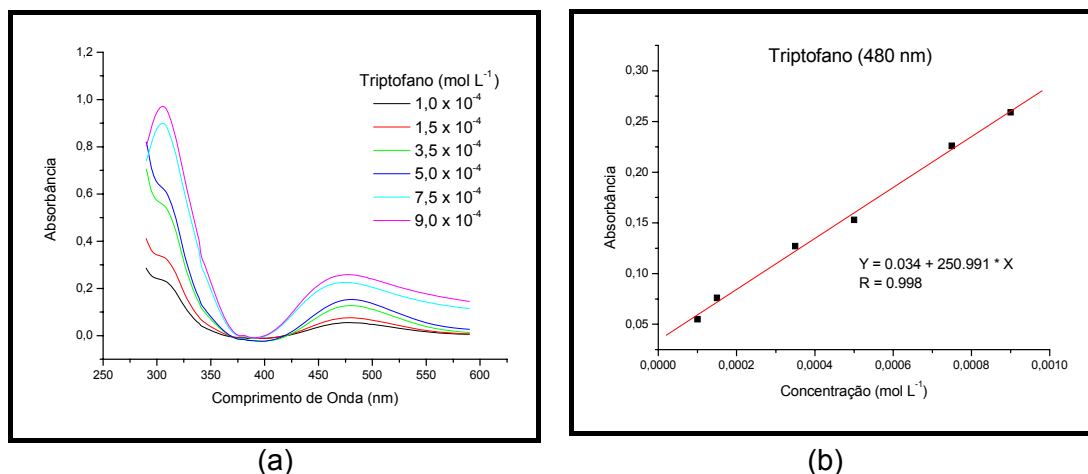


Figura 22. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Triptofano.

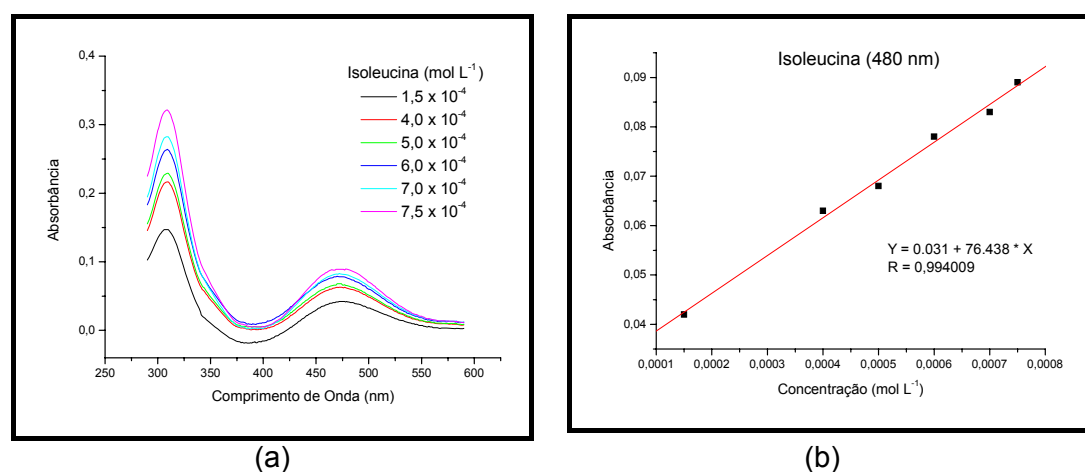


Figura 23. (a) Espectros de absorção UV-VIS; (b) Curva analítica, absorbância em função das concentrações de Isoleucina.

De acordo com o planejamento fatorial 2^5 , 32 misturas sintéticas foram preparadas. Os níveis de concentração que foram estudados e o planejamento fatorial com os respectivos valores de concentração podem ser visualizados nas Tabelas 3 e 4, respectivamente.

Tabela 3. Valores de concentrações obtidas a partir das curvas analíticas de calibração para os níveis (-) e (+).

Nível	Concentrações ($\times 10^{-4}$ mol L $^{-1}$)				
	Fenilalanina	Histidina	Prolina	Triptofano	Isoleucina
(-)	3,00	3,00	0,50	1,00	1,50
(+)	7,00	6,00	9,00	9,00	7,50

Tabela 4. Composição das misturas utilizadas nas etapas de calibração e validação para o desenvolvimento de modelos multivariados.

Misturas	Concentrações ($\times 10^{-4}$ mol L ⁻¹)				
	Phe	His	Pro	Trp	Ile
1	7,00	6,00	9,00	9,00	7,50
2	7,00	6,00	9,00	9,00	1,50
3	7,00	6,00	9,00	1,00	7,50
4	7,00	6,00	9,00	1,00	1,50
5	7,00	6,00	0,50	9,00	7,50
6	7,00	6,00	0,50	9,00	1,50
7	7,00	6,00	0,50	1,00	7,50
8	7,00	6,00	0,50	1,00	1,50
9	7,00	3,00	9,00	9,00	7,50
10	7,00	3,00	9,00	9,00	1,50
11	7,00	3,00	9,00	1,00	7,50
12	7,00	3,00	9,00	1,00	1,50
13	7,00	3,00	0,50	9,00	7,50
14	7,00	3,00	0,50	9,00	1,50
15	7,00	3,00	0,50	1,00	7,50
16	7,00	3,00	0,50	1,00	1,50
17	3,00	6,00	9,00	9,00	7,50
18	3,00	6,00	9,00	9,00	1,50
19	3,00	6,00	9,00	1,00	7,50
20	3,00	6,00	9,00	1,00	1,50
21	3,00	6,00	0,50	9,00	7,50
22	3,00	6,00	0,50	9,00	1,50
23	3,00	6,00	0,50	1,00	7,50
24	3,00	6,00	0,50	1,00	1,50
25	3,00	3,00	9,00	9,00	7,50
26	3,00	3,00	9,00	9,00	1,50
27	3,00	3,00	9,00	1,00	7,50
28	3,00	3,00	9,00	1,00	1,50
29	3,00	3,00	0,50	9,00	7,50
30	3,00	3,00	0,50	9,00	1,50
31	3,00	3,00	0,50	1,00	7,50
32	3,00	3,00	0,50	1,00	1,50

4.3. Análise Multivariada

Os aminoácidos fenilalanina, histidina, prolina, triptofano e com exceção da isoleucina, possuem estruturas semelhantes como pode ser observado na Figura 24.

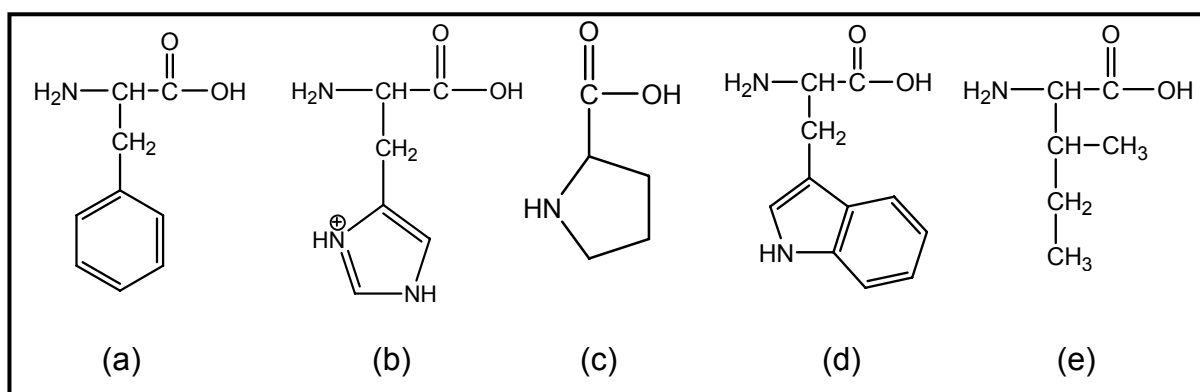


Figura 24. Estruturas químicas (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano e (e) Isoleucina.

A semelhança estrutural desses aminoácidos faz com que os espectros individuais e em misturas sejam muito semelhantes. Estas características impossibilitam a diferenciação e/ou classificação dos diferentes aminoácidos apenas pela simples análise visual dos espectros. A semelhança entre os espectros individuais e em misturas sintéticas pode ser observada nas Figuras 25 e 26, respectivamente.

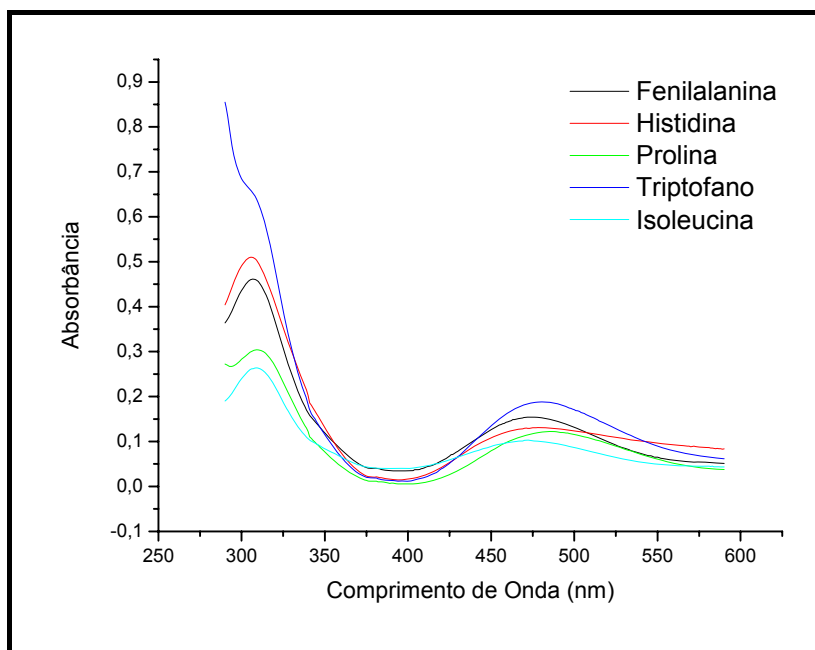


Figura 25. Espectros de absorção UV-VIS de Fenilalanina, Histidina, Prolina, Triptofano e Isoleucina obtidos em pH 10. Concentração dos aminoácidos: $5,00 \times 10^{-4} \text{ mol L}^{-1}$.

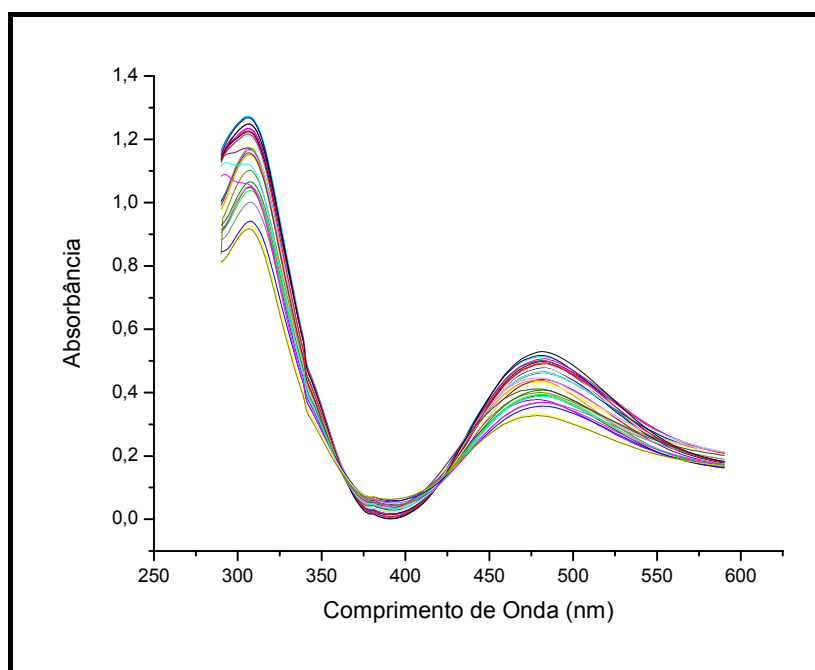


Figura 26. Espectros de Absorção UV-VIS das misturas sintéticas de Fenilalanina, Histidina, Prolina, Triptofano e Isoleucina obtidos em pH 10,00 e utilizados no modelo multivariado.

Visando contornar os problemas de interferência, os espectros de absorção UV-VIS foram processados em primeira derivada. Uma vez que a derivação aumenta o sinal do ruído, um procedimento de alisamento foi realizado simultaneamente.

Observando os espectros originais (Figura 26) e os espectros das misturas na forma da primeira derivada (Figura 27) verifica-se um aumento de informação espectral por surgimento de novos picos, que podem ser utilizados no processamento quimiométrico.

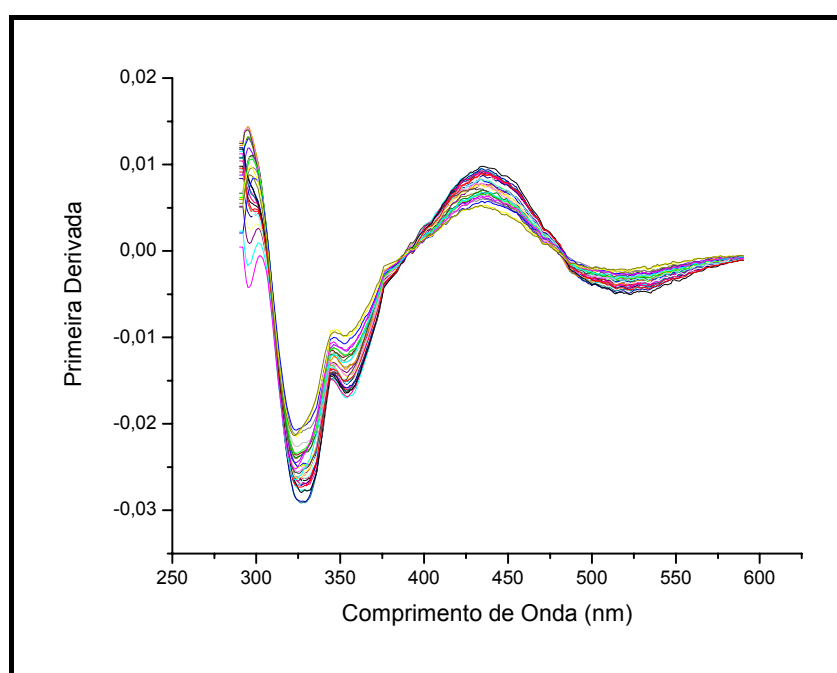


Figura 27. Espectros de absorção UV-VIS das misturas de Fenilalanina, Histidina, Prolina, Triptofano e Isoleucina pré-processados em primeira derivada, obtidos em pH 10 e utilizados no modelo multivariado.

4.3.1. Análise das Componentes Principais (PCA)

O conjunto de dados obtidos com as medidas espectrofotométricas foi submetido ao tratamento quimiométrico por análise das componentes principais (PCA), uma ferramenta relevante para visualização que permite a representação gráfica dos resultados do planejamento fatorial em grupos distintos.

Os valores das absorbâncias das 32 misturas foram dispostos em forma matricial e trabalhados no ambiente MatLab. Os dados foram inicialmente centrados na média; esse pré-processamento consistiu em calcular o valor médio de absorbância em cada comprimento de onda, subtraindo-o da absorbância correspondente em cada amostra e dividir o valor obtido pelo desvio padrão dos valores de absorbância de cada coluna da matriz de espectros.

Durante a análise dos dados, dois itens foram levados em consideração, o número de componentes principais que são necessários para descrever o sistema e o quanto da variância dos dados cada uma destas componentes é capaz de descrever. A variância descrita pelas componentes principais indica a quantidade de informação que aquele fator contém. Um dos objetivos de uma análise de componentes principais é justamente descrever o conjunto de dados com o menor número de fatores (ou componentes principais) possível.

Uma forma de encontrar as condições ótimas que obedeçam a estes critérios é selecionar um conjunto de amostras e de variáveis que melhor descreva o conjunto de dados e as amostras com um comportamento anômalo muitas vezes podem estar relacionadas a algum fenômeno, e contribuem para a sua explicação. Estas amostras não devem ser simplesmente excluídas do conjunto de dados e ter seu comportamento ignorado.

Construíram-se dois modelos com a PCA; no primeiro, com os dados sem pré-processamento e no segundo utilizaram-se os dados da primeira derivada das absorbâncias em função do comprimento de onda obtidos no programa Origin 6.1 utilizando polinômio de segunda ordem e alisamento com 5 pontos.

4.3.1.1. Análise das Componentes Principais sem pré-processamento

O estudo realizado apresentou o percentual de contribuição de cada uma das componentes principais para o total de 32 ensaios realizados.

Pelos valores de porcentagem de variância descrita pelas componentes principais, foram selecionados sete fatores para a descrição do sistema. Na Figura 28 e na Tabela 5 podem ser observadas as variâncias descritas pelas componentes principais. A componente principal 1 (PC1) descreveu 93,56% da variância total dos dados e a componente principal 2 (PC2), 3,50%.

A utilização das duas primeiras componentes, cuja soma de contribuição é superior a 95%, foi suficiente para representar o sistema estudado, ou seja, pode ser descrito como um gráfico bidimensional com apenas dois eixos, componente principal 1 e 2.

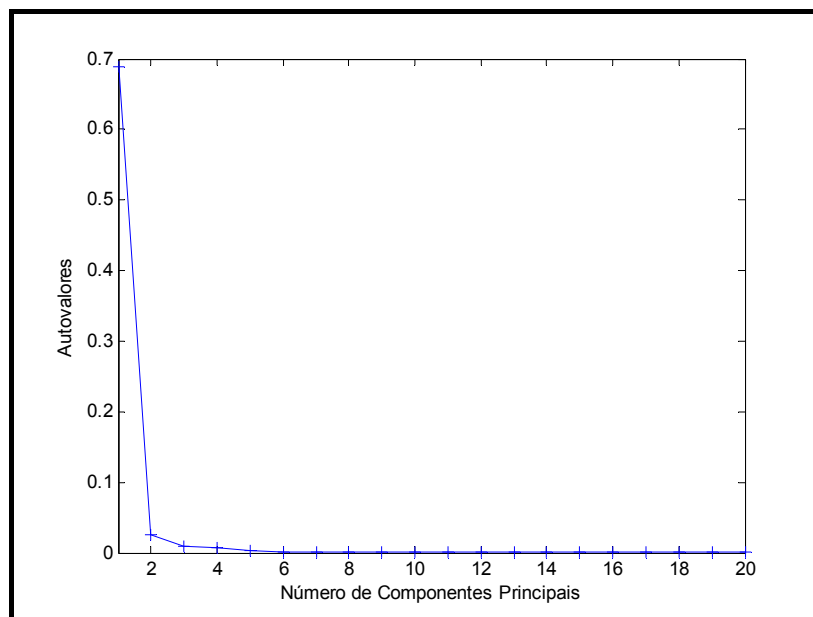


Figura 28. Autovalores em função do número de componentes principais aplicando-se PCA.

Tabela 5. Cálculo da variância e percentual de contribuição das componentes utilizadas na PCA.

Componente Principal	Variância (%)	Variância Acumulada (%)
1	93,56	93,56
2	3,50	97,06
3	1,36	98,42
4	1,02	99,43
5	0,38	99,81

A análise de componentes principais permitiu a visualização mediante representação gráfica dos resultados do planejamento fatorial em grupos distintos, como pode ser observado na Figura 29. Cada grupo é caracterizado de acordo com os níveis de concentração das soluções de aminoácidos (Tabela 4).

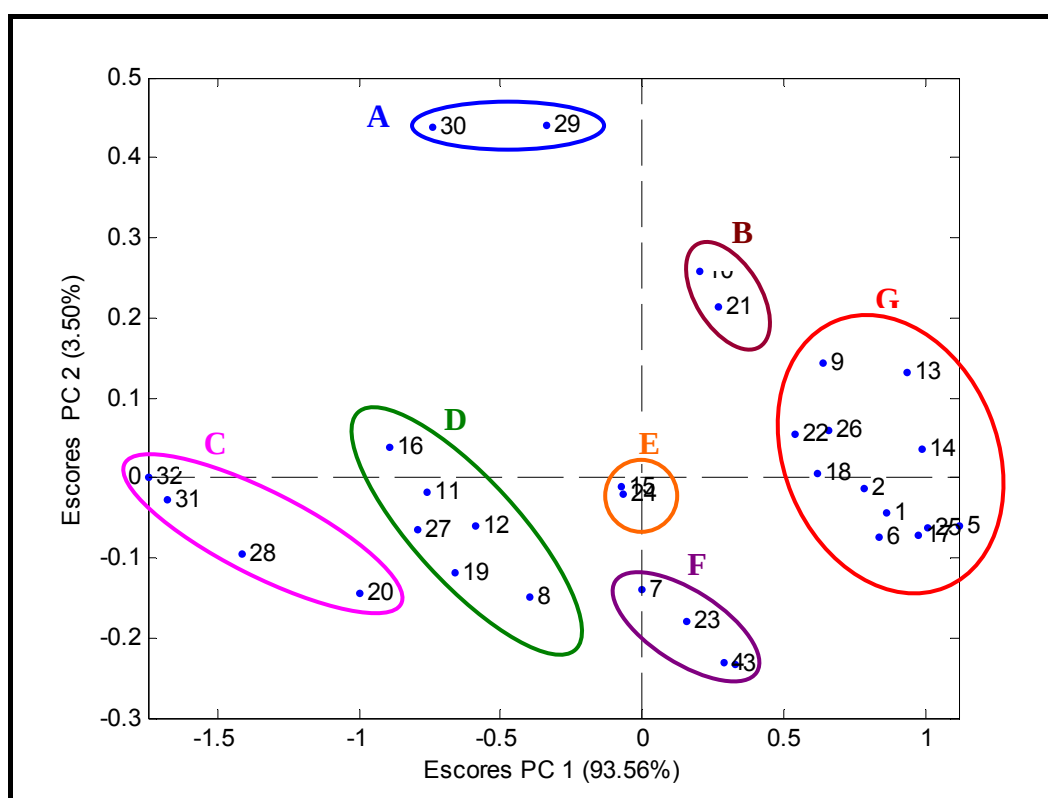


Figura 29. Análise das componentes principais (PC1 versus PC2) para as 32 misturas contendo aminoácidos em solução com os dados centrados na média.

Pela análise de escores para a primeira e a segunda componentes principais pode-se observar que a PC1, que explica 93,56% da variância, separou as soluções das misturas que contêm histidina, triptofano e fenilalanina com níveis máximo e mínimo de concentração. As soluções de fenilalanina, histidina e triptofano com nível máximo de concentração agrupam-se à direita do eixo da PC1 e as com nível mínimo de concentração, à esquerda do eixo da PC1.

A PC2, que explica 3,50% da variância, separou as soluções das misturas que apresentam isoleucina com níveis máximo e mínimo de concentração. As soluções com o nível máximo de concentração agrupam-se abaixo do eixo da PC2 e as com o nível mínimo de concentração, acima do eixo da PC2.

De acordo com os agrupamentos, foi possível separar o conjunto de dados das misturas, em sete grupos distintos e foram identificados na Figura 29 pelas letras A, B, C, D, E, F e G.

O grupo A, representado pelas misturas 29 e 30, é caracterizado por soluções de fenilalanina, histidina e prolina com nível mínimo de concentração e triptofano com nível máximo de concentração.

O grupo B, representado pelas misturas 10 e 21, é caracterizado por soluções de triptofano com nível máximo de concentração. Pode-se observar que a mistura 21 encontra-se deslocada para baixo no grupo, o que pode ser explicado pela presença de isoleucina com nível máximo de concentração.

O grupo C, representado pelas misturas 20, 28, 31 e 32, é caracterizado por soluções de fenilalanina, histidina e triptofano com nível mínimo de concentração. A mistura 20 que contém histidina com nível máximo de concentração explica o seu deslocamento para direita dentro do grupo e está de acordo com a análise da PC1 que separa misturas de histidina com nível máximo de concentração e localizam-se à direita do eixo da PC1.

O grupo D, representado pelas misturas 8, 11, 12, 16, 19 e 27, é caracterizado por soluções de triptofano com nível mínimo de concentração e fenilalanina com nível máximo de concentração. Observa-se que a mistura 8 encontra-se deslocada para a direita no grupo o que pode ser explicado pela presença de histidina com nível máximo de concentração uma vez que

a PC1 separou as misturas deste aminoácido sendo que as que apresentam nível máximo de concentração encontram-se à direita do eixo da PC1.

O grupo E, representado pelas misturas 15 e 24, é caracterizado por soluções de triptofano e prolina com nível mínimo de concentração.

O grupo F, representado pelas misturas 3, 4, 7 e 23, é caracterizado por soluções de histidina, isoleucina e fenilalanina com níveis máximo de concentração. Estas misturas encontram-se à direita do eixo da PC1 porque apresentam predominantemente nível máximo de concentração.

O grupo G, representado pelas misturas 1, 2, 5, 6, 9, 13 14, 17, 18, 22, 25 e 26, é caracterizado por soluções de triptofano com nível máximo de concentração.

A partir dos grupos formados pela análise das componentes principais, construiu-se um gráfico dos espectros de absorção UV-Vis com uma mistura de cada grupo.

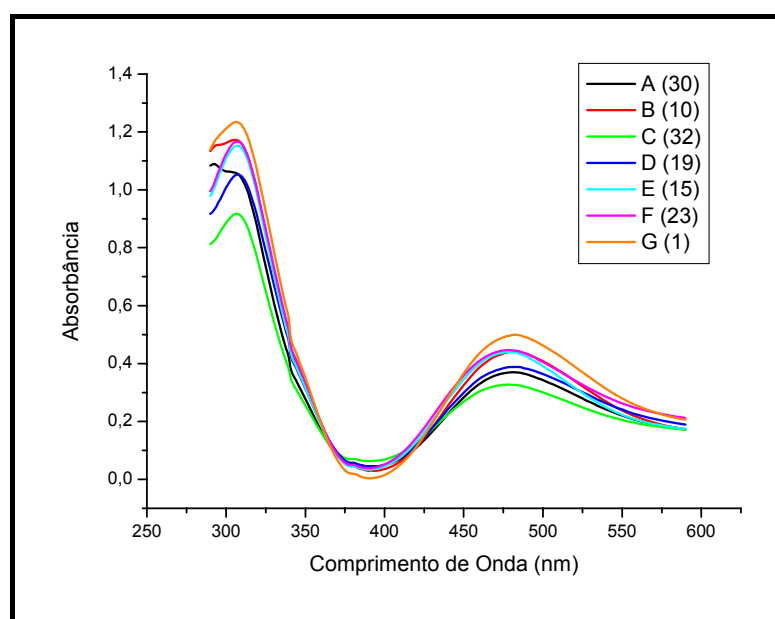


Figura 30. Espectros de Absorção UV-Vis dos sete grupos da PCA

Pela análise da Figura 30, pode-se observar diferentes valores de absorbância para as misturas estudadas (1, 10, 15, 19, 23, 30, 32), escolhidas aleatoriamente em cada grupo. As maiores intensidades observadas correspondem às amostras que apresentam soluções de aminoácidos com nível máximo de concentração. As misturas 1 e 32

apresentam maior e menor valor de absorbância, respectivamente. Estes valores estão de acordo com a Lei de Beer que diz que a absorbância é proporcional à concentração de uma espécie absorvente. Na amostra 1 (grupo G) observa-se maior uma intensidade de absorção uma vez que apresenta soluções com nível máximo de concentração. Na mistura 32 (grupo C) observa-se uma menor intensidade de absorção uma vez que apresenta soluções com nível mínimo de concentração.

As soluções dos grupos A, B, D, E e F, apresentam diferentes proporções de soluções com nível máximo e mínimo de concentração o que resultou em diferentes valores de absorbância.

4.3.1.2. Análise das Componentes Principais com Dados Derivados.

Este modelo foi obtido pela aplicação da primeira derivada aos dados espectrais; geralmente, neste procedimento realiza-se uma espécie de deconvolução parcial de sinais não totalmente sobrepostos, o que possibilita que pequenas diferenças sejam evidenciadas.

O conjunto de dados espectrais foi submetido à transformação de dados pela primeira derivada no programa Origin 6.1, concomitantemente ao procedimento de pré-processamento de dados centrados na média, utilizando-se o programa PLS_Toolbox em ambiente MatLab para Windows.

Os valores das derivadas das absorbâncias em função do comprimento de onda das 32 misturas foram dispostos em forma matricial e os dados foram submetidos à análise de componentes principais.

Pelos valores da porcentagem da variância descrita pelas componentes principais, foram selecionados 7 fatores para a descrição do sistema. Na Figura 31 e na Tabela 6 podem ser observadas as variâncias descritas pelas componentes principais. A análise de escores da PC1 versus PC2 (Figura 32) apresenta uma separação das amostras pelos seus constituintes.

A componente principal 1 descreveu (PC1) 55,19% da variância total dos dados e a componente principal 2 (PC2), 30,28%.

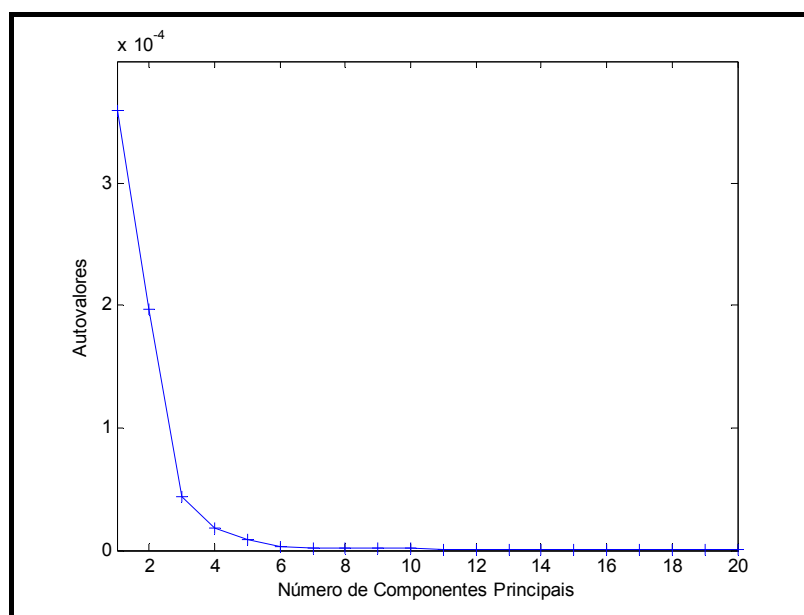


Figura 31. Autovalores em função do número de componentes principais aplicando-se PCA.

Tabela 6. Cálculo de variância e percentual de contribuição das componentes utilizadas no cálculo da PCA.

Componente Principal	Variância (%)	Variância Acumulada (%)
1	55,19	55,19
2	30,28	85,48
3	6,73	92,20
4	2,79	94,99
5	1,43	96,42
6	0,49	96,91
7	0,35	97,26

A análise de componentes principais possibilitou a visualização mediante representação gráfica dos resultados do planejamento fatorial em grupos distintos, de acordo com os valores máximos e mínimos de concentração das soluções de aminoácidos.

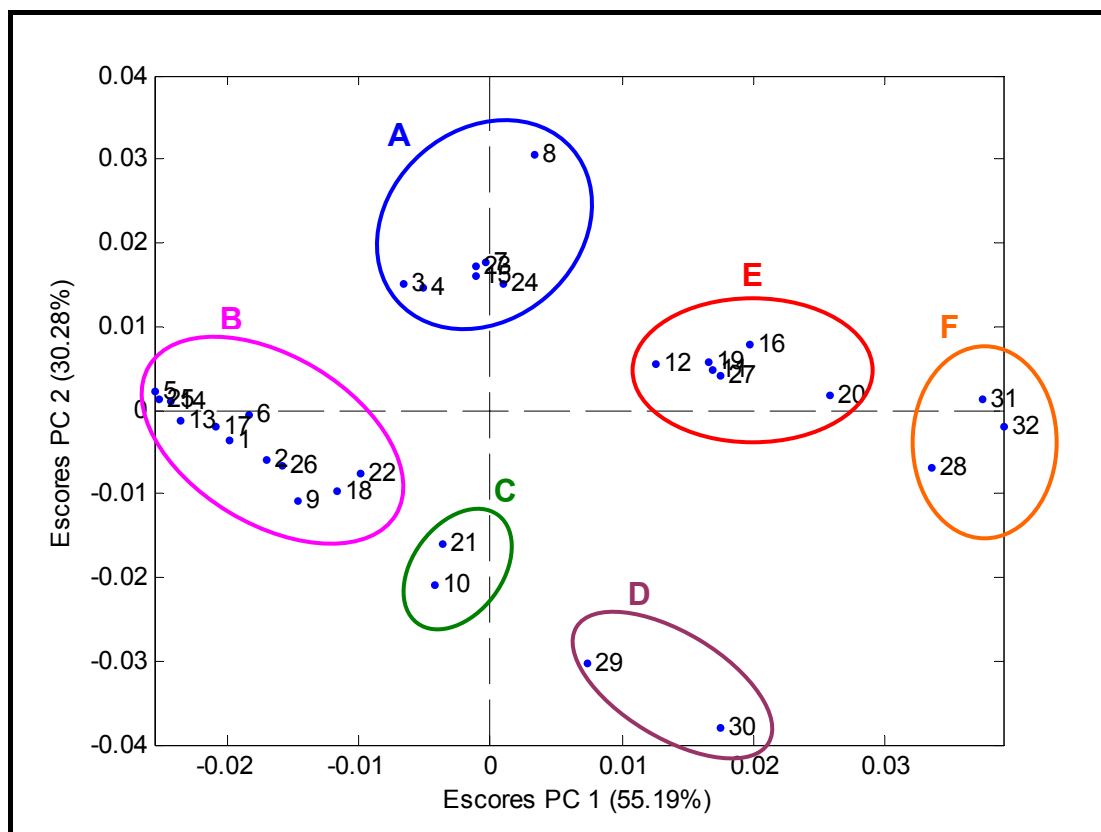


Figura 32. Análise das componentes principais (PC1 versus PC2) para as 32 misturas contendo aminoácidos em solução com dados da primeira derivada concomitantemente com pré-processamento de centrar na média.

Pela análise de escores para a primeira e a segunda componentes principais pode-se observar que a PC1, que explica 55,19% da variância, é responsável por separar as soluções das misturas que contêm histidina e fenilalanina com níveis máximo e mínimo de concentração. As soluções de fenilalanina e histidina com nível máximo de concentração agrupam-se à esquerda do eixo da PC1 e as com nível mínimo, à direita do eixo da PC1.

A PC2, que explica 30,28% da variância, é responsável por separar as soluções das misturas que apresentam triptofano e isoleucina com níveis máximo e mínimo de concentração. As soluções com níveis máximo de concentração de triptofano e isoleucina agrupam-se abaixo e acima do eixo da PC2, respectivamente e com níveis mínimo de concentração de triptofano e fenilalanina encontram-se acima e abaixo do eixo da PC2, respectivamente.

De acordo com os agrupamentos, foi possível separar o conjunto de dados das misturas, em seis grupos distintos e foram identificados na Figura 32, pelas letras A, B, C, D, E e F.

O grupo A, representado pelas misturas 3, 4, 7, 8, 15, 23, e 24, é caracterizado por soluções de fenilalanina e histidina com nível máximo de concentração e triptofano e prolina com nível mínimo de concentração. Pode-se observar que a mistura 8 encontra-se deslocada para cima dentro do grupo, o que pode ser explicado pela presença de isoleucina com nível máximo de concentração.

O grupo B, representado pelas misturas 1, 2, 5, 6, 9, 13, 14, 17, 18, 22, 25 e 26, é caracterizado por soluções de fenilalanina e triptofano com nível máximo de concentração.

O grupo C, representado pelas misturas 10 e 21, é caracterizado por soluções de triptofano com nível máximo de concentração.

O grupo D, representado pelas misturas 29 e 30, é caracterizado por soluções de triptofano com nível máximo de concentração e fenilalanina, histidina e prolina com nível mínimo de concentração. Pode-se observar que a mistura 30 encontra-se deslocada para baixo dentro do grupo, o que pode ser explicado pela presença de isoleucina com nível mínimo de concentração o que está de acordo com a separação obtida com análise da PC1 (misturas de isoleucina com nível mínimo de concentração agrupam-se abaixo do eixo da PC1).

O grupo E, representado pelas misturas 11, 12, 16, 19, 20 e 27, é caracterizado por soluções de prolina e triptofano com nível máximo e mínimo de concentração, respectivamente.

O grupo F, representado pelas misturas 28, 31 e 32, é caracterizado por soluções de fenilalanina, histidina e triptofano com nível mínimo de concentração.

4.3.2. Método dos Mínimos Quadrados Parciais (PLS)

Os valores de absorbâncias das 32 misturas assim como os valores de concentrações foram dispostos em forma matricial.

O método dos mínimos quadrados parciais (PLS) foi construído com os dados correspondentes aos valores de absorbância (variáveis independentes, x) e os dados de concentração (variáveis dependentes, y). A determinação foi específica para cada aminoácido. Cada amostra do conjunto é representada por 301 valores de absorbância, registrados em igual número de valores de comprimento de onda.

Na primeira etapa do procedimento multivariado, grande parte da variância dos dados pode ser representada por um pequeno número de novas variáveis (VLs), as quais surgem da combinação linear das variáveis originais.

O melhor modelo é sempre escolhido de modo a otimizar o ajuste de calibração e sua capacidade preditiva simultaneamente, para que não ocorra nem o sobreajuste nem a falta de ajuste do modelo selecionado.

Para definir o melhor número de variáveis latentes (VLs), utilizou-se o método de validação interna cruzada, particularmente a rotina denominada “Leave one out”. Neste procedimento, um espectro é retirado do conjunto de calibração, sendo utilizado como procedimento de previsão. Esta operação é repetida tantas vezes quanto necessário de maneira a garantir que todos os elementos do conjunto participem como elemento de previsão. A concentração da amostra que foi retirada é prevista e por comparação com o valor esperado, a raiz quadrada da média dos erros quadráticos gerada por validação cruzada é estimada para cada fator estudado. Este procedimento é repetido para todas as amostras e um erro médio é calculado, dando-se a escolha ao número de fatores que apresentar o menor RMSECV.

O RMSECV (soma dos quadrados dos erros de previsão da validação cruzada) é uma medida da capacidade do modelo em prever novas amostras. Ele está relacionado ao valor de PRESS (soma dos quadrados dos erros de previsão) para o número de variáveis latentes incluídas no modelo. Na equação relativa ao cálculo do PRESS (equação 20), o que

muda são os valores de $\hat{y}_i$, que agora se referem às previsões para as amostras não incluídas na formulação do modelo.

$$PRESS = \sum_i (\hat{y}_i - y_i)^2 \quad \text{Equação 20}$$

em que $\hat{y}_i$ é o valor previsto para a amostra i utilizando o modelo; y_i é o valor medido para a amostra i e n é o número de amostras do conjunto de calibração [23].

O RMSECV é definido pela equação 21:

$$RMSECV_k = \sqrt{\frac{PRESS_k}{n}} \quad \text{Equação 21}$$

onde $PRESS_k$ é a soma dos quadrados dos erros de previsão para o modelo que inclui k fatores [41].

Para os modelos construídos, o número de variáveis latentes foi determinado com base nos gráficos de RMSECV versus variável latente de cada aminoácido estudado. O número de variáveis latentes empregado na construção de cada um dos modelos foi o que produziu o menor erro de previsão para seus respectivos conjuntos de calibração.

A presença de amostras anômalas foi verificada pela utilização dos gráficos de Resíduos Estudentizados versus “Leverage”. A “leverage” define a influência que uma determinada amostra terá sobre o modelo. O resíduo estudentizado é uma indicação da falta de ajuste do valor y de uma amostra. Estudentizar alguma coisa significa tornar a média zero e a variância unitária. A presença de amostras anômalas é caracterizada pela presença simultânea de altos leverage e altos resíduos “Student”.

Construíram-se três modelos com o PLS; no primeiro, os dados da matriz **X** não foram submetidos a um pré-processamento; no segundo, os dados da matriz **X** foram alisados com 5 pontos e utilizou-se polinômio de segunda ordem e derivada de primeira ordem e no terceiro, utilizou-se os dados da primeira derivada das absorbâncias em função do comprimento de onda obtidos no programa Origin 6.1 utilizando polinômio de segunda ordem e alisamento com 5 pontos. O segundo e o terceiro modelos apresentaram

resultados muito semelhantes, sendo assim, neste trabalho, optou-se por detalhar o segundo modelo.

4.3.2.1. PLS com Espectros sem pré-processamento

Para a construção do modelo PLS é essencial encontrar o número ideal de variáveis latentes que devem ser incluídos no modelo. Nas Figuras 33 (a) – 33 (e) encontram-se os gráficos de RMSECV versus variável latente para cada aminoácido. Pode-se observar que o RMSEC atinge um valor mínimo em um determinado número de variáveis latentes. Na Figura 33 (a), o valor mínimo para o RMSECV é atingido com 7 variáveis latentes enquanto que na 33 (d) o valor mínimo é atingido com 12 variáveis latentes.

Deve-se escolher o número de variáveis latentes que corresponda ao valor mínimo para RMSECV, pois a partir desse valor não se tem mais uma variância significativa, ou seja, a maior parte da informação estatística foi “capturada”. De acordo com este princípio, 12 variáveis latentes foram selecionadas para descrever o modelo uma vez que este foi construído simultaneamente para os cinco aminoácidos em estudo. Se mais variáveis latentes fossem incorporadas ao modelo, ruídos poderiam estar sendo incluídos. Porém no caso de se usar um número de variáveis latentes, menor do que 12 informações importantes deixariam de ser utilizadas sendo assim, deve-se escolher um número de variáveis latentes que forneça uma boa previsão de concentração para todos os aminoácidos.

Nas figuras 34 (a) – 34 (e) encontram-se os dois critérios de maior importância para a verificação de anomalias no conjunto de calibração, os resíduos de “Student” e “leverage”. Por definição, os valores máximos permitidos para estes parâmetros são $\pm 2,5$ e $1,12 (3VL/n)$ respectivamente. Utilizando-se estes argumentos, observa-se que nenhuma anomalia existe no conjunto utilizado para a elaboração do modelo, sendo assim, nenhuma amostra foi retirada para a construção do modelo.

Os resultados obtidos com o modelo PLS construído com 12 variáveis latentes e os valores para os erros de previsão, podem ser verificados nas Tabelas 7 a 11. Como os dados não foram submetidos a nenhum

pré-processamento, o ruído adicionado aos dados foi mantido durante a análise. Observa-se que não existe boa concordância entre os dados obtidos pela análise espectrofotométrica e os previstos pelo modelo para todos os aminoácidos, fazendo-se necessário submeter os dados a etapas de pré-processamento.

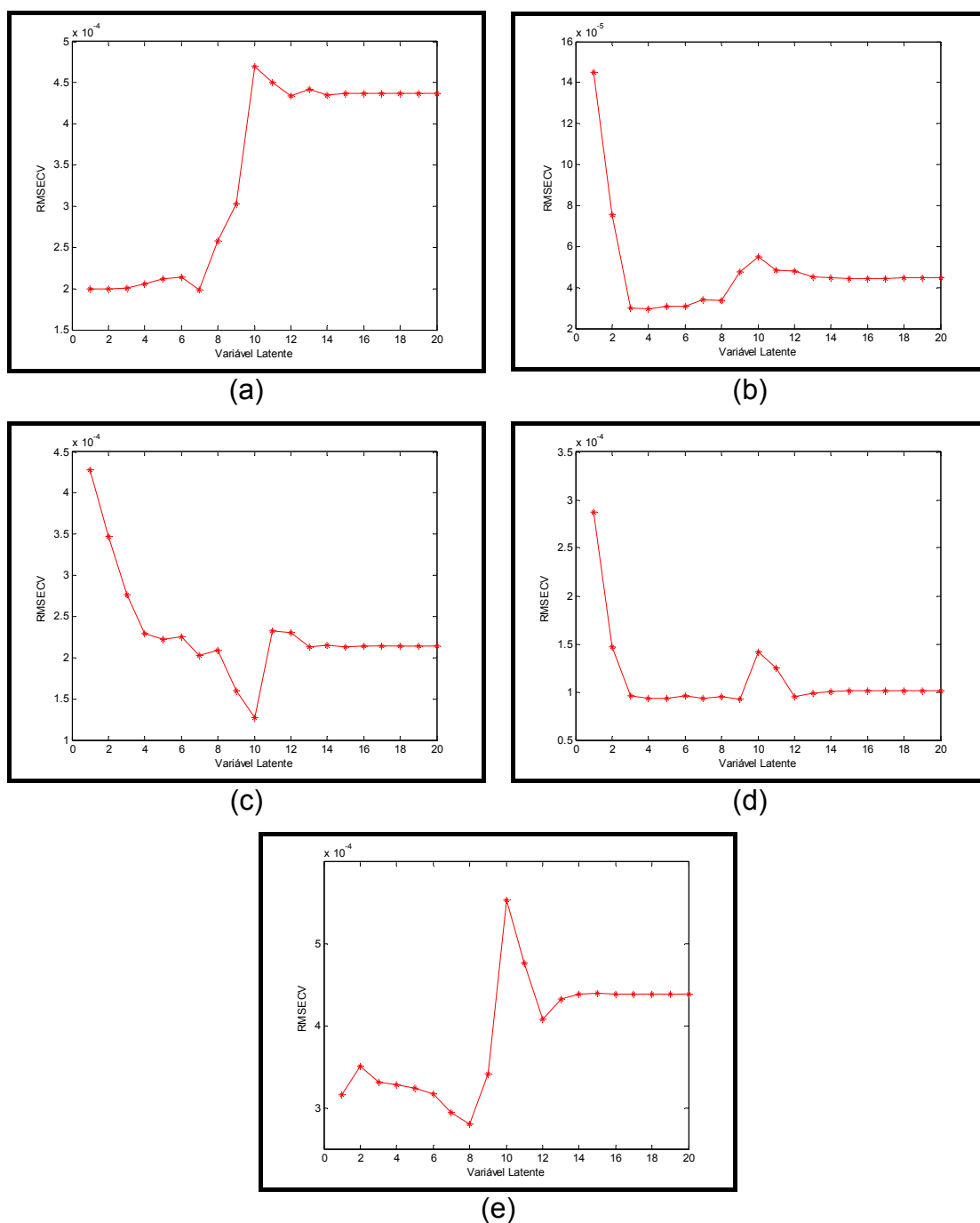


Figura 33. RMSECV versus número de variáveis latentes com dados sem pré-processamento (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano, (e) Isoleucina.

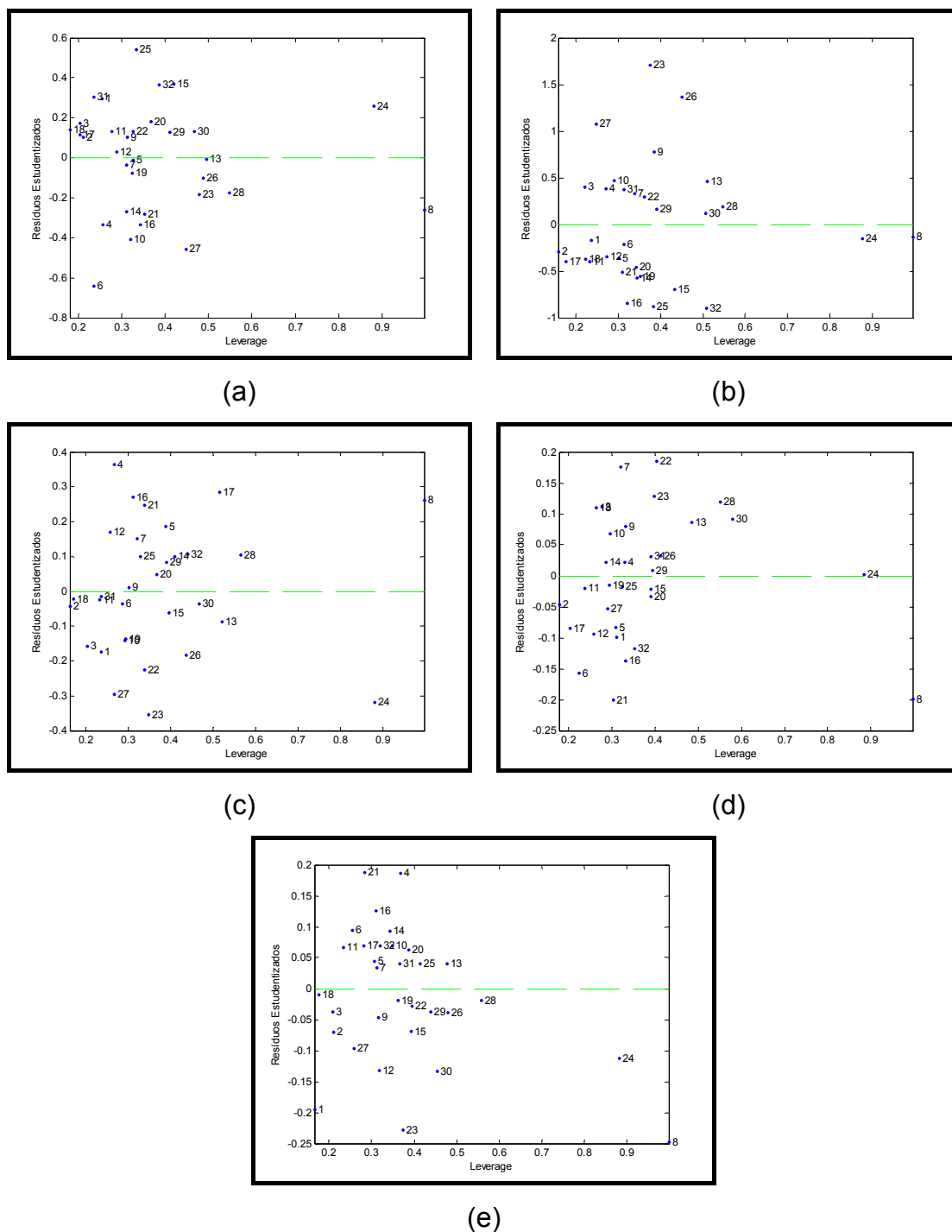


Figura 34. Resíduos Estudantizados versus “Leverage” para Modelo Desenvolvido com 12 VLS e dados sem pré-processamento (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano, (e) Isoleucina.

Tabela 7. Resultados da previsão das concentrações das misturas sintéticas de Fenilalanina para validação, utilizando 12 VLS, sem pré-processamento.

Amostra de teste	Concentração Real ($\times 10^{-4}$ mol L $^{-1}$)	Concentração Prevista ($\times 10^{-4}$ mol L $^{-1}$)	Erro de previsão (%)
1	7,00	7,48	6,86
2	7,00	7,17	2,43
3	7,00	7,29	4,14
4	7,00	6,45	7,86
5	7,00	6,98	0,28
6	7,00	5,95	15,00
7	7,00	6,94	0,86
8	7,00	6,97	0,43
9	7,00	7,16	2,28
10	7,00	6,37	9,00
11	7,00	7,21	3,00
12	7,00	7,04	0,57
13	7,00	6,99	0,14
14	7,00	6,58	6,00
15	7,00	7,53	7,57
16	7,00	6,49	7,28
17	3,00	3,19	6,33
18	3,00	3,24	8,00
19	3,00	2,88	4,00
20	3,00	3,27	9,00
21	3,00	2,58	14,00
22	3,00	3,20	6,67
23	3,00	2,75	8,33
24	3,00	3,17	5,67
25	3,00	3,83	27,67
26	3,00	2,86	4,67
27	3,00	2,36	21,33
28	3,00	2,78	7,33
29	3,00	3,18	6,00
30	3,00	3,18	6,00
31	3,00	3,5	16,67
32	3,00	3,53	17,67

Tabela 8. Resultados da previsão das concentrações das misturas sintéticas de Histidina para validação, utilizando 12 VLS, sem pré-processamento.

Amostra de teste	Concentração Real ($\times 10^{-4}$ mol L⁻¹)	Concentração Prevista ($\times 10^{-4}$ mol L⁻¹)	Erro de previsão (%)
1	6,00	5,98	0,33
2	6,00	5,95	0,83
3	6,00	6,06	1,00
4	6,00	6,06	1,00
5	6,00	5,95	0,83
6	6,00	5,97	0,50
7	6,00	6,04	0,67
8	6,00	6,00	0,00
9	3,00	3,10	3,33
10	3,00	3,07	2,33
11	3,00	2,94	2,00
12	3,00	2,95	1,67
13	3,00	3,05	1,67
14	3,00	2,92	2,67
15	3,00	2,91	3,00
16	3,00	2,88	4,00
17	6,00	5,94	1,00
18	6,00	5,94	1,00
19	6,00	5,92	1,33
20	6,00	5,94	1,00
21	6,00	5,93	1,20
22	6,00	6,04	0,67
23	6,00	6,23	3,83
24	6,00	5,99	0,16
25	3,00	2,88	4,00
26	3,00	3,17	5,67
27	3,00	3,16	5,33
28	3,00	3,02	0,67
29	3,00	3,02	0,67
30	3,00	3,01	0,33
31	3,00	3,05	1,67
32	3,00	2,89	3,67

Tabela 9. Resultados da previsão das concentrações das misturas sintéticas de Prolina para validação, utilizando 12 VLS, sem pré-processamento.

Amostra de teste	Concentração Real (x 10⁻⁴ mol L⁻¹)	Concentração Prevista (x 10⁻⁴ mol L⁻¹)	Erro de previsão (%)
1	9,00	8,68	3,56
2	9,00	8,91	1,00
3	9,00	8,70	3,33
4	9,00	9,66	7,33
5	0,50	0,81	62,00
6	0,50	0,44	12,00
7	0,50	0,76	52,00
8	0,50	0,53	6,00
9	9,00	9,02	0,22
10	9,00	8,75	2,78
11	9,00	8,95	0,56
12	9,00	9,31	3,44
13	0,50	0,37	26,00
14	0,50	0,66	32,00
15	0,50	0,40	20,00
16	0,50	0,97	94,00
17	9,00	9,42	4,67
18	9,00	8,96	0,44
19	9,00	8,76	0,00
20	9,00	9,08	2,67
21	0,50	0,92	0,89
22	0,50	0,11	84,00
23	0,50	0,10	78,00
24	0,50	0,27	80,00
25	9,00	9,17	46,00
26	9,00	8,71	1,89
27	9,00	8,46	3,22
28	9,00	9,14	6,00
29	0,50	0,64	1,56
30	0,50	0,45	28,00
31	0,50	0,47	10,00
32	0,50	0,67	6,00

Tabela 10. Resultados da previsão das concentrações das misturas sintéticas de Triptofano para validação, utilizando 12 VLS, sem pré-processamento.

Amostra de teste	Concentração Real ($\times 10^{-4}$ mol L⁻¹)	Concentração Prevista ($\times 10^{-4}$ mol L⁻¹)	Erro de previsão (%)
1	9,00	8,79	2,33
2	9,00	8,89	1,22
3	1,00	1,24	24,00
4	1,00	1,04	4,00
5	9,00	8,82	2,00
6	9,00	8,64	4,00
7	1,00	1,37	37,00
8	1,00	0,98	2,00
9	9,00	9,17	1,89
10	9,00	9,15	1,67
11	1,00	0,95	5,00
12	1,00	0,79	21,00
13	9,00	9,16	1,78
14	9,00	9,05	0,55
15	1,00	0,96	4,00
16	1,00	0,71	29,00
17	9,00	8,80	2,22
18	9,00	9,24	2,67
19	1,00	0,97	3,00
20	1,00	0,93	7,00
21	9,00	8,57	4,78
22	9,00	9,37	4,11
23	1,00	1,26	26,00
24	1,00	1,00	0,00
25	9,00	8,96	0,44
26	9,00	9,06	0,66
27	1,00	0,88	12,00
28	1,00	1,21	21,00
29	9,00	9,02	0,22
30	9,00	9,15	1,67
31	1,00	1,06	6,00
32	1,00	0,75	25,00

Tabela 11. Resultados da previsão das concentrações das misturas sintéticas de Isoleucina para validação, utilizando 12 VLS, sem pré-processamento.

Amostra de teste	Concentração Real ($\times 10^{-4}$ mol L⁻¹)	Concentração Prevista ($\times 10^{-4}$ mol L⁻¹)	Erro de previsão (%)
1	7,50	6,52	13,06
2	1,50	1,16	22,60
3	7,50	7,31	2,53
4	1,50	2,31	54,00
5	7,50	7,70	2,67
6	1,50	1,95	30,00
7	7,50	7,65	2,00
8	1,50	1,44	4,00
9	7,50	7,29	2,80
10	1,50	1,80	20,00
11	7,50	7,82	4,27
12	1,50	0,90	40,00
13	7,50	7,66	2,13
14	1,50	1,91	27,33
15	7,50	7,20	4,00
16	1,50	2,07	38,00
17	7,50	7,82	4,27
18	1,50	1,45	3,33
19	7,50	7,42	1,09
20	1,50	1,77	18,00
21	7,50	8,37	11,60
22	1,50	1,38	8,00
23	7,50	6,51	13,20
24	1,50	1,29	14,00
25	7,50	7,67	2,27
26	1,50	1,34	10,67
27	7,50	7,04	6,13
28	1,50	1,43	4,67
29	7,50	7,35	2,00
30	1,50	0,96	36,00
31	7,50	7,68	2,40
32	1,50	1,81	20,67

4.3.2.2. PLS com Dados Pré-Processados.

Neste modelo, os dados da matriz **X** foram submetidos ao pré-processamento do tipo Savgol (Savitzky-Golay), também conhecido como Mínimos Quadrados ou DISPO (*Digital Smoothing Polynomial*, em português, Polinômio de Suavização Digital) que utiliza Erros Quadrados Mínimos para calcular um polinômio de tamanho e grau determinados que melhor ajuste o conjunto de dados. A convolução do polinômio com o intervalo de dados originais da janela de interesse gera um novo vetor com os dados filtrados. A idéia principal deste filtro é de aproximar a função base dentro da janela de suavização não por uma constante, a qual estima a média, mas por um polinômio de ordem maior. Utilizações recursivas desse filtro sobre um dado conjunto resultarão em dados filtrados com formato semelhante ao polinômio utilizado. Os dados foram alisados com 5 pontos, utilizou-se um polinômio de segunda ordem e derivada de primeira ordem.

Nas Figuras 35 (a) a 35 (e) podem ser observados os gráficos de RMSECV versus variável latente para o modelo construído. Na Figura 35 (b) o mínimo para o RMSECV é atingido com 3 variáveis latentes e na 35 (d), com 9 variáveis latentes. Deve-se escolher o número de variáveis latentes que corresponda ao valor mínimo para RMSECV, o qual leva à obtenção dos menores erros possíveis nas previsões. Deve-se selecionar um número de variáveis latentes que forneça uma boa previsão de concentração para todos os aminoácidos em estudo, desta forma, foram selecionadas 9 variáveis latentes uma vez que o modelo foi construído simultaneamente para os cinco aminoácidos em estudo.

Nas Figuras 36 (a) a 36 (e) encontram-se os gráficos de Resíduos Estudentizados versus “Leverage”. Observa-se que todas as amostras estão dentro da faixa permitida para os valores de resíduo de Student ($\pm 2,5$). Com relação a leverage o valor máximo permitido para este parâmetro é 0,85 ($3VL/n$), as amostras 8 e 24 apresentam valores superiores ao permitido e poderiam ser fontes de erros, mas apresentam valores de resíduos estudentizados dentro da faixa permitida, sendo assim, não foram retiradas na construção do modelo.

A capacidade de previsão dos modelos desenvolvido com a primeira derivada pode ser avaliada a partir dos resultados apresentados nas Tabelas 12 a 16. Observa-se este modelo permite previsões coerentes com os valores conhecidos, com erros relativos abaixo de 6,0%.

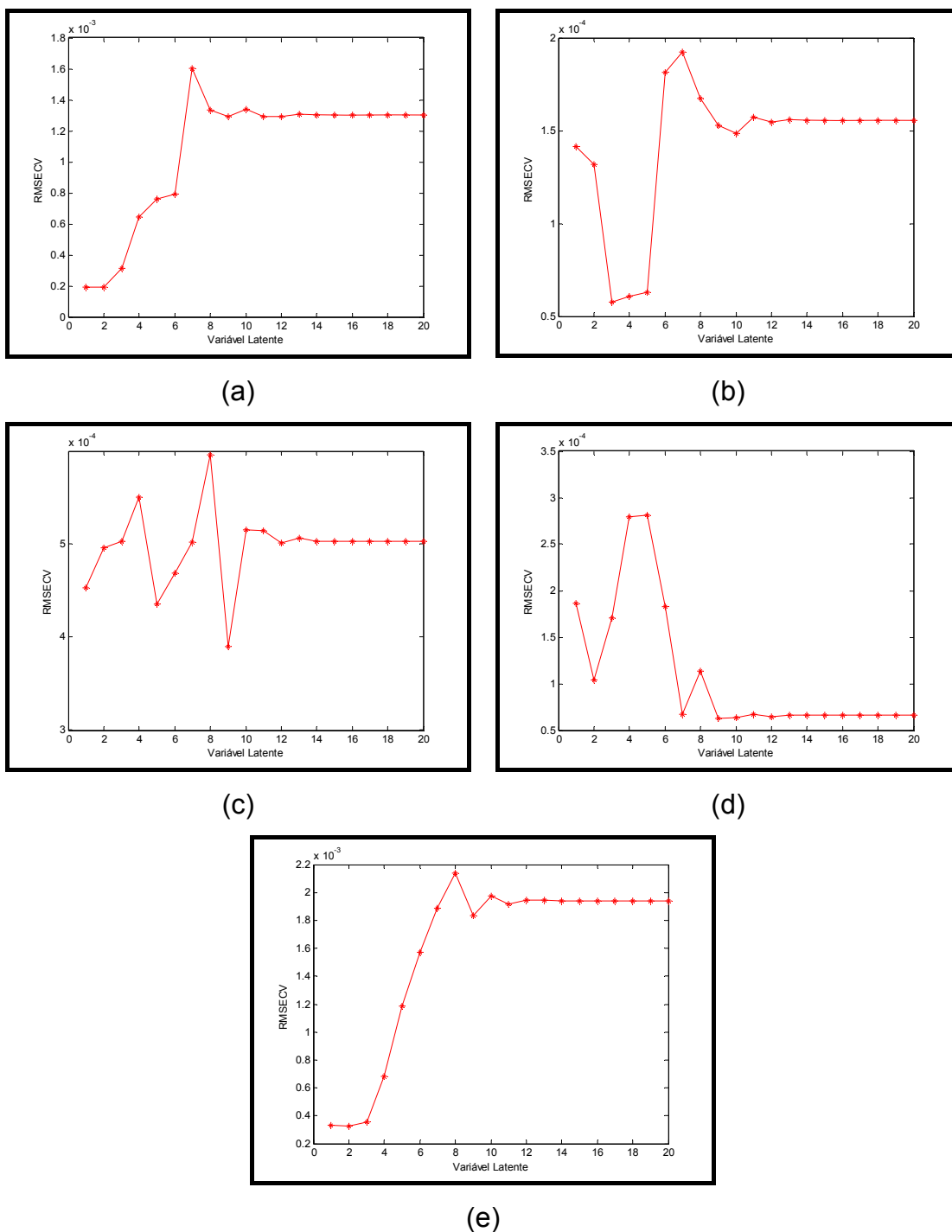


Figura 35. RMSECV versus número de variáveis latentes com dados pré-processados (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano, (e) Isoleucina.

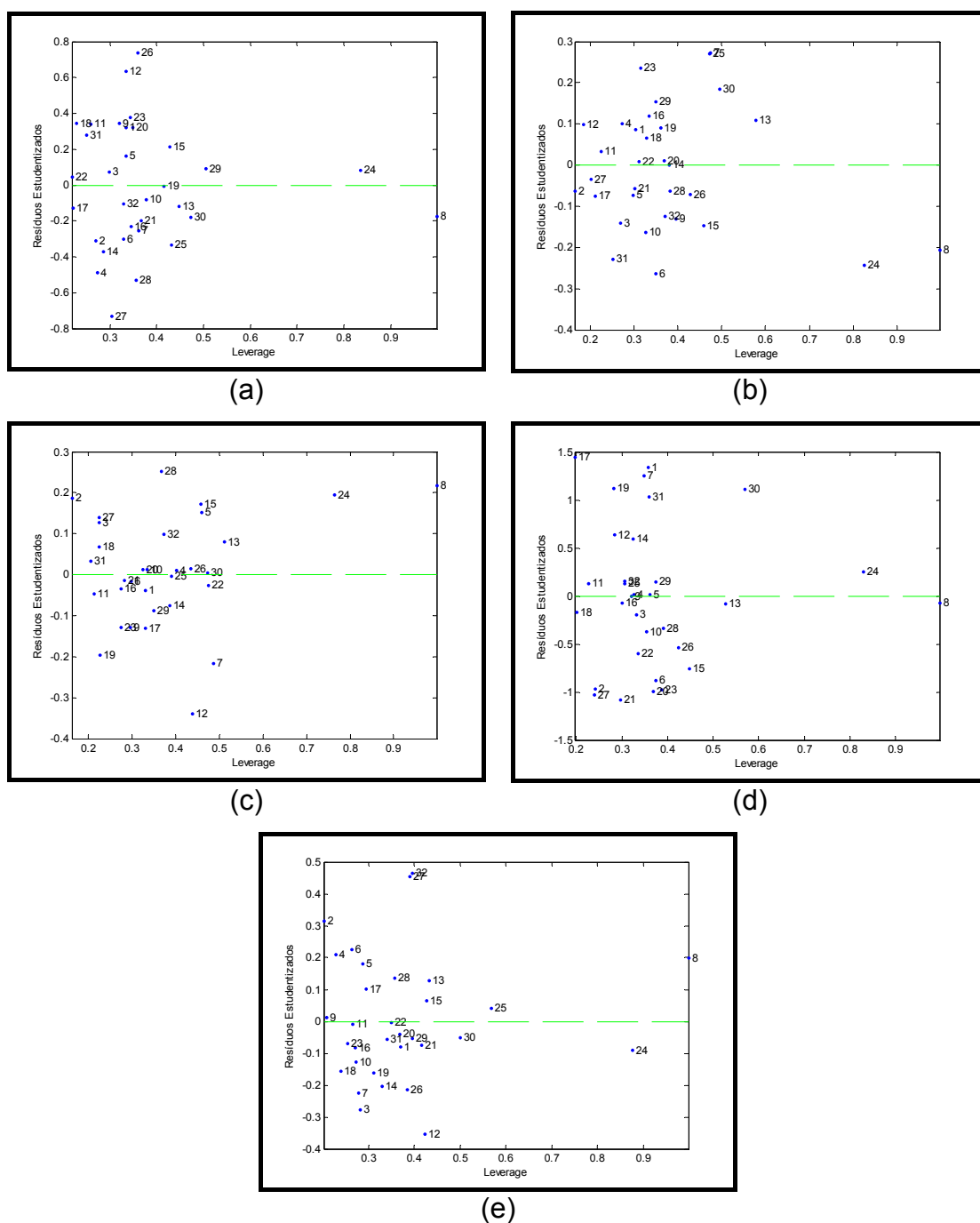


Figura 36. Resíduos Estudatizados versus “Leverage” para modelo desenvolvido com 9 VLS com dados pré-processados (a) Fenilalanina, (b) Histidina, (c) Prolina, (d) Triptofano, (e) Isoleucina.

Tabela 12. Resultados da previsão das concentrações das misturas sintéticas de Fenilalanina para validação, utilizando 9 VLS, pré-processados com primeira derivada.

Amostra de teste	Concentração Real ($\times 10^{-4}$ mol L⁻¹)	Concentração Prevista ($\times 10^{-4}$ mol L⁻¹)	Erro de previsão (%)
1	7,00	6,99	0,14
2	7,00	7,02	0,28
3	7,00	6,99	0,14
4	7,00	7,01	0,14
5	7,00	7,02	0,28
6	7,00	7,01	0,14
7	7,00	6,99	0,14
8	7,00	7,00	0,00
9	7,00	7,00	0,00
10	7,00	7,01	0,14
11	7,00	6,99	0,14
12	7,00	6,98	0,28
13	7,00	7,00	0,00
14	7,00	6,99	0,14
15	7,00	7,00	0,00
16	7,00	6,99	0,14
17	3,00	2,99	0,33
18	3,00	3,00	0,00
19	3,00	2,97	1,00
20	3,00	3,00	0,00
21	3,00	2,99	0,33
22	3,00	3,00	0,00
23	3,00	2,99	0,33
24	3,00	3,02	0,67
25	3,00	2,99	0,33
26	3,00	3,02	0,67
27	3,00	3,02	0,67
28	3,00	3,01	0,33
29	3,00	3,02	0,67
30	3,00	2,97	1,00
31	3,00	3,00	0,00
32	3,00	3,02	0,67

Tabela 13. Resultados da previsão das concentrações das misturas sintéticas de Histidina para validação, utilizando 9 VLS, pré-processados com primeira derivada.

Amostra de teste	Concentração Real (mol L⁻¹)	Concentração Prevista (mol L⁻¹)	Erro de previsão (%)
1	6,00	6,00	0,00
2	6,00	6,01	0,16
3	6,00	5,98	0,33
4	6,00	6,01	0,16
5	6,00	6,01	0,16
6	6,00	6,01	0,16
7	6,00	5,99	0,17
8	6,00	6,00	0,00
9	3,00	2,99	0,33
10	3,00	3,01	0,33
11	3,00	3,00	0,00
12	3,00	3,00	0,00
13	3,00	3,00	0,00
14	3,00	2,99	0,33
15	3,00	2,98	0,67
16	3,00	3,00	0,00
17	6,00	5,99	0,17
18	6,00	5,99	0,17
19	6,00	6,00	0,00
20	6,00	6,00	0,00
21	6,00	6,00	0,00
22	6,00	6,00	0,00
23	6,00	6,01	0,16
24	6,00	6,00	0,00
25	3,00	3,01	0,33
26	3,00	3,00	0,00
27	3,00	3,02	0,67
28	3,00	3,00	0,00
29	3,00	3,00	0,00
30	3,00	3,00	0,00
31	3,00	2,99	0,33
32	3,00	3,00	0,00

Tabela 14. Resultados da previsão das concentrações das misturas sintéticas de Prolina para validação, utilizando 9 VLS, pré-processados com primeira derivada.

Amostra de teste	Concentração Real ($\times 10^{-4}$ mol L⁻¹)	Concentração Prevista ($\times 10^{-4}$ mol L⁻¹)	Erro de previsão (%)
1	9,00	9,01	0,11
2	9,00	9,01	0,11
3	9,00	9,01	0,11
4	9,00	8,99	0,11
5	0,50	0,53	6,00
6	0,50	0,49	2,00
7	0,50	0,51	2,00
8	0,50	0,50	0,00
9	9,00	9,00	0,00
10	9,00	8,99	0,11
11	9,00	8,98	0,22
12	9,00	9,02	0,22
13	0,50	0,52	4,00
14	0,50	0,51	2,00
15	0,50	0,50	0,00
16	0,50	0,50	0,00
17	9,00	9,01	0,11
18	9,00	8,97	0,33
19	9,00	8,98	0,22
20	9,00	8,97	0,33
21	0,50	0,49	2,00
22	0,50	0,50	0,00
23	0,50	0,51	2,00
24	0,50	0,49	2,00
25	9,00	9,03	0,33
26	9,00	9,00	0,00
27	9,00	9,00	0,00
28	9,00	9,03	0,33
29	0,50	0,50	0,00
30	0,50	0,51	2,00
31	0,50	0,50	0,00
32	0,50	0,52	4,00

Tabela 15. Resultados da previsão das concentrações das misturas sintéticas de Triptofano para validação, utilizando 9 VLS, pré-processados com primeira derivada.

Amostra de teste	Concentração Real ($\times 10^{-4}$ mol L⁻¹)	Concentração Prevista ($\times 10^{-4}$ mol L⁻¹)	Erro de previsão (%)
1	9,00	9,01	0,11
2	9,00	9,00	0,00
3	1,00	1,00	0,00
4	1,00	1,00	0,00
5	9,00	8,99	0,11
6	9,00	8,99	0,11
7	1,00	0,99	1,00
8	1,00	1,00	0,00
9	9,00	9,02	0,22
10	9,00	9,00	0,00
11	1,00	0,99	1,00
12	1,00	1,00	0,00
13	9,00	9,00	0,00
14	9,00	8,98	0,22
15	1,00	1,01	1,00
16	1,00	1,01	1,00
17	9,00	9,01	0,11
18	9,00	9,01	0,11
19	1,00	0,99	1,00
20	1,00	1,00	0,00
21	9,00	9,00	0,00
22	9,00	9,00	0,00
23	1,00	0,97	3,00
24	1,00	1,02	2,00
25	9,00	9,01	0,11
26	9,00	8,98	0,22
27	1,00	1,00	0,00
28	1,00	0,98	2,00
29	9,00	8,98	0,22
30	9,00	9,01	0,11
31	1,00	1,01	1,00
32	1,00	1,00	0,00

Tabela 16. Resultados da previsão das concentrações das misturas sintéticas de Isoleucina para validação, utilizando 9 VLS, pré-processados com primeira derivada.

Amostra de teste	Concentração Real ($\times 10^{-4}$ mol L^{-1})	Concentração Prevista ($\times 10^{-4}$ mol L^{-1})	Erro de previsão (%)
1	7,50	7,49	0,13
2	1,50	1,60	6,67
3	7,50	7,41	1,20
4	1,50	1,57	4,67
5	7,50	7,60	1,33
6	1,50	1,57	4,67
7	7,50	7,42	1,07
8	1,50	1,50	0,00
9	7,50	7,51	0,13
10	1,50	1,46	2,66
11	7,50	7,60	1,33
12	1,50	1,48	1,33
13	7,50	7,54	0,53
14	1,50	1,46	2,67
15	7,50	7,52	0,26
16	1,50	1,47	2,00
17	7,50	7,49	0,13
18	1,50	1,48	1,33
19	7,50	7,45	0,67
20	1,50	1,48	1,33
21	7,50	7,5	0,00
22	1,50	1,50	0,00
23	7,50	7,51	0,93
24	1,50	1,49	0,67
25	7,50	7,56	0,80
26	1,50	1,48	1,33
27	7,50	7,56	0,80
28	1,50	1,54	2,67
29	7,50	7,48	0,26
30	1,50	1,49	0,67
31	7,50	7,48	0,26
32	1,50	1,48	1,33

Uma comparação entre as médias dos erros de previsão para cada fator é mostrada na Tabela 17, pode-se observar que as médias dos erros de previsão para cada aminoácido estudado estão abaixo de 13,36% para os dados sem pré-processamento (modelo 1) e abaixo de 1,370% para os dados com pré-processamento (modelo 2) evidenciando os melhores resultados.

Tabela 17. Média dos erros de previsão (%) para os aminoácidos estudados.

Aminoácido	Modelo 1	Modelo 2
Fenilalanina	7,595	0,284
Histidina	1,813	0,145
Prolina	21,24	0,957
Triptofano	8,069	0,458
Isoleucina	13,36	1,370

4.4. Análise de Amostras Reais

4.4.1. Análise das Componentes Principais

Neste estudo, os dados das 32 soluções das misturas sintéticas e 4 das amostras de suplementos alimentares protéicos foram dispostos na forma matricial; os dados foram centrados na média e submetidos à análise de componentes principais.

Foram analisadas duas amostras de suplementos alimentares protéicos uma em pó e outra líquida que foram submetidas a duas diluições. Na Figura 37, as amostras 33 e 34 correspondem às amostras de aminoácidos líquidos diluídos 10 e 100 vezes respectivamente e as amostras 35 e 36 correspondem às amostras de aminoácidos em pó, diluídas 10 e 100 vezes, respectivamente.

Pela análise da Figura 38, observa-se que o modelo é explicado com cinco componentes principais que foram selecionados para a descrição do sistema.

A componente principal 1 (PC1) explicou 95,13% da variância e a componente principal 2 (PC2) explicou 3,59%.

Pela análise do gráfico de escores (Figura 39) para as primeira e segunda componentes principais pode-se observar que somente a PC2 apresenta semelhança entre as amostras e as misturas sintéticas.

A PC2 separou as soluções das misturas que contém fenilalanina e triptofano com níveis máximo e mínimo de concentração. As soluções com nível máximo de concentração dos dois aminoácidos agrupam-se abaixo do eixo da PC2 e as com nível mínimo de concentração, acima do eixo da PC2.

Como as amostras 34 e 36 encontram-se acima do eixo da PC2, sugere-se que estas apresentam baixa concentração de fenilalanina e triptofano enquanto que as amostras 33 e 35 que se encontram abaixo do eixo da PC2, apresenta considerável concentração de fenilalanina e triptofano.

A observação do conjunto de espectros (Figura 37) permite identificar algumas diferenças. As amostras 34 (L:100) 35 (D:10) e 36 (D:100) apresentam concentração abaixo dos níveis estudados neste trabalho e a

amostra 33 (L:10) apresenta espectro semelhante aos das misturas sintéticas utilizadas no planejamento, apesar disto apresenta características diferentes. Esta diferenciação pode ser observada a partir do gráfico de escores. Este parâmetro, que representa a coordenada de cada amostra no sistema de eixos formado pelas novas componentes principais (PC1 e PC2) indica que as amostras 33 a 36 são diferenciáveis do resto do conjunto uma vez que estas encontram-se afastadas de todas as misturas sintéticas utilizadas no planejamento, apesar de apresentar algumas características semelhantes identificadas pela análise da PC2. Em primeira instância, é possível assumir que discrepâncias deste tipo sejam devidas à presença de excipientes, espécies que não foram consideradas na modelagem e introduzem modificações espectrais que inviabiliza a adequada previsão das espécies de interesse. As amostras analisadas apresentam além dos aminoácidos considerados neste estudo muitos outros, o que pode estar influenciando nos resultados obtidos.

Com base nos resultados obtidos pela PCA, verifica-se que o modelo desenvolvido com o PLS não pode ser aplicado a nenhuma das amostras. Uma vez que estas apresentam características diferenciáveis do conjunto de misturas sintéticas, as previsões serão comprometidas.

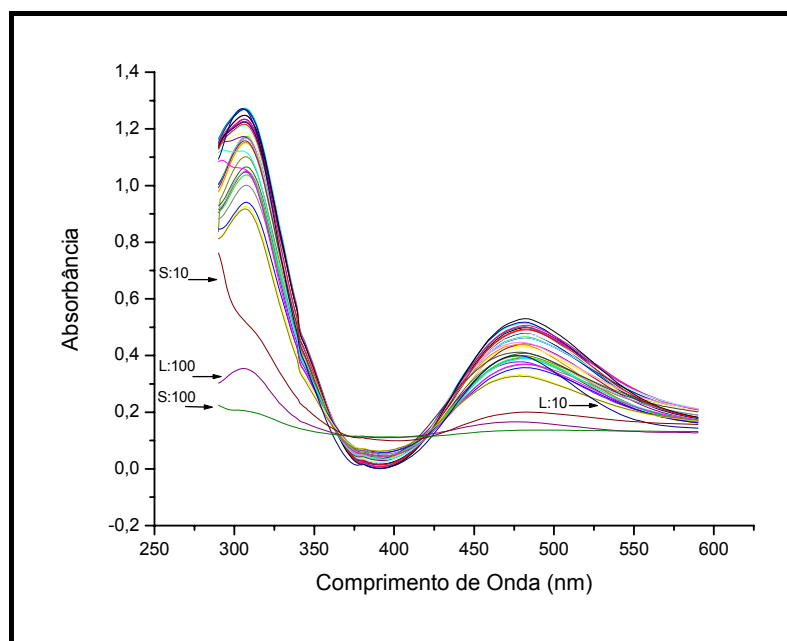


Figura 37. Espectros das 36 soluções utilizadas nos estudos por análise de componentes principais.

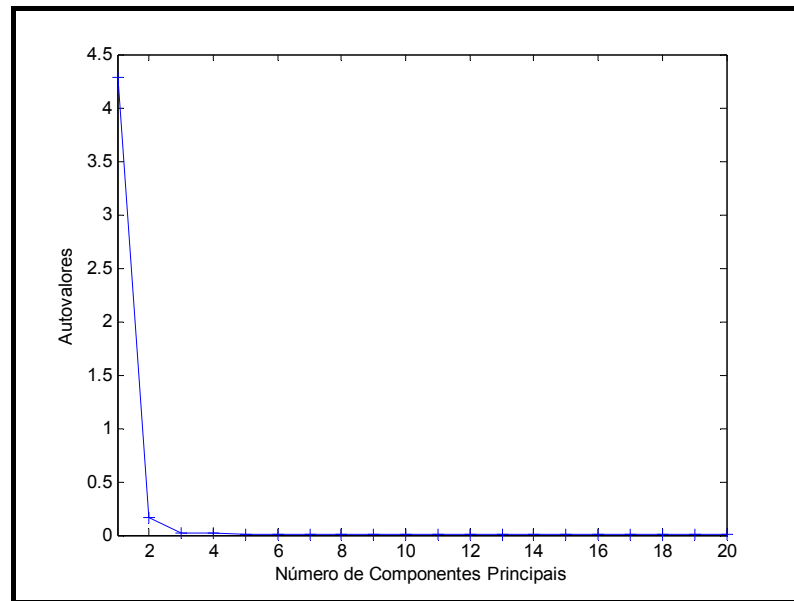


Figura 38. autovalores em função do número de componentes principais aplicando-se PCA.

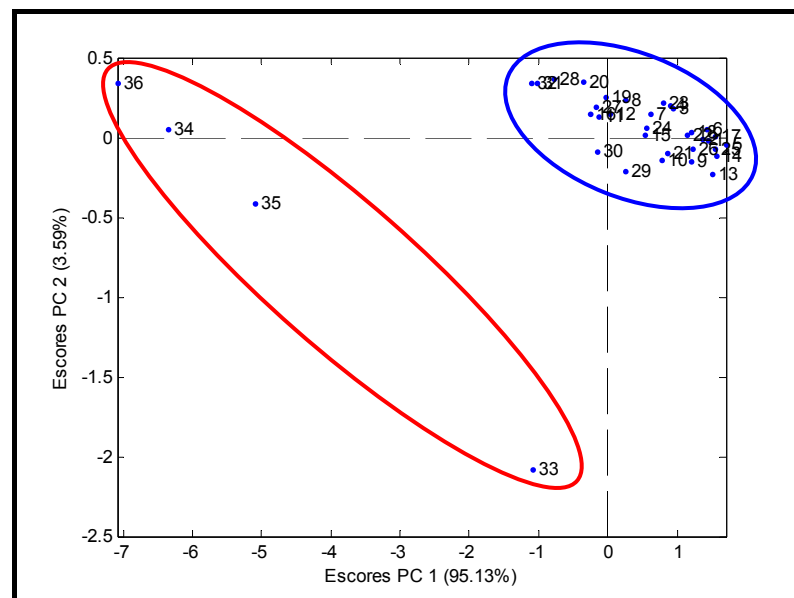


Figura 39. Escores da PC1 versus PC2.

CONCLUSÕES

5. CONCLUSÕES

Foi possível demonstrar pelo trabalho realizado que a utilização de métodos espectrofotométricos em conjunção com métodos quimiométricos pode resultar em aplicações analíticas de grande potencial, tanto para determinações qualitativas como quantitativas.

Para a determinação espectrofotométrica de aminoácidos, pela reação destes com NQS, foram empregadas duas técnicas de estatística multivariada, análise de componentes principais (PCA) e o método dos mínimos quadrados parciais (PLS). A intensa sobreposição das bandas espectrais das misturas de aminoácidos impossibilitou que as técnicas convencionais de análise espectrofotométrica sejam realizadas. A utilização de técnicas multivariadas permitiu contornar este problema, propiciando bons resultados.

Para a análise das componentes principais, foram aplicados dois modelos, um com dados originais e o outro com dados derivados.

No primeiro modelo, a PC1 separou as soluções das misturas que apresentam histidina, triptofano e fenilalanina com níveis maiores e menores de concentrações, sendo que as que apresentam maiores níveis de concentração agruparam-se à esquerda do eixo da PC1. A PC2 separou as soluções das misturas que apresentaram maiores e menores níveis de concentração; as que apresentaram níveis maiores agruparam-se abaixo do eixo da PC2.

No segundo modelo, a PC1 foi responsável por separar as soluções das misturas que apresentaram histidina e fenilalanina com níveis maiores e menores de concentrações, as que apresentaram maiores níveis de concentração agruparam-se à esquerda do eixo da PC1. A PC2 propiciou as separações das soluções das misturas que apresentam triptofano e isoleucina com níveis maiores e menores de concentrações. As soluções com os maiores níveis de concentração de triptofano e isoleucina agruparam-se abaixo e acima do eixo da PC2 respectivamente.

Nos dois modelos estudados verificou-se separação do conjunto de dados em agrupamentos que foram caracterizados pela presença de amostras com características em comum.

Para o processo de calibração fundamentado em regressão por mínimos quadrados parciais (PLS) aplicou-se os dados originais e os derivados, como realizado na PCA, com 12 e 9 VLS, respectivamente.

Os resultados obtidos com dados sem pré-processamento não foram satisfatórios, com erros de previsão acima de 90% para Prolina.

O modelo realizado com dados derivados permitiu obter ótimos resultados, com erros de previsão abaixo de 6%.

A análise de amostras reais não pôde ser determinada principalmente em razão de modificações estruturais ocasionadas pela presença de substâncias não modeladas.

RECOMENDAÇÕES FUTURAS

6. RECOMENDAÇÕES FUTURAS

A realização deste trabalho permitiu evidenciar inúmeros aspectos relevantes da calibração multivariada, aspectos estes que permitem reconhecer necessidades que podem ser tema de trabalhos futuros.

Devido a complexidade das amostras que contêm aminoácidos, recomenda-se desenvolver modelo de calibração multivariada em uma faixa de concentração levando-se em consideração o teor nominal de aminoácidos presentes nas amostras comercialmente disponíveis.

Melhores resultados provavelmente podem ser obtidos utilizando uma cubeta em fluxo uma vez que as leituras serão realizadas em tempo real, com renovação do material reacional, eliminando-se a instabilidade do produto corado.

REFERÊNCIAS BIBLIOGRÁFICAS

7. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] CARREIRA, R. L.; BARBOSA, C. M. S.; JUNQUEIRA, R. G. et al. Use of High Performance Liquid Hydrophilic Chromatography For Determining Amino Acids In Casein Hydrolysates. **Ciência e Tecnologia de Alimentos**, v.22, n.3, p.229-232. ISSN 0101-2061. 2002.
- [2] SIMAS, R. C. Determinação de Proteína Bruta e Aminoácidos em Farelo de Soja por Espectrofotometria no Infravermelho Próximo. 2005. 119 f. Dissertação (Mestrado em Química) – UNICAMP, Campinas, 2005.
- [3] TANNUS, A. F. S. Nitrogênio Urinário e Tecidual em Ratos Desnutridos com Suplementação de Glutamina. 2001. 96 f. Dissertação (Mestrado em Biociências Aplicadas à Clínica Médica) - Universidade de São Paulo. Faculdade de Medicina de Ribeirão Preto, SP, 2001.
- [4] SCARMINIO, I. S.; ISHIKAWA, D. N.; BARRETO, W. J.; PACZKOWSKI, É. L.; ARRUDA, I. C. Calibração Multivariada para Sistemas Com Bandas Sobrepostas Através da Análise de Fatores do Tipo Q. **Química Nova**, v. 21, n. 5, p.590-596, 1998.
- [5] COSCIONE, A. R. O Uso de Calibração Multivariada para a Determinação Espectrofotométrica Simultânea de Alumínio e Ferro: Aplicação na Análise de Plantas e Solos. 2001. 129 f. Tese (Doutorado em Química) - UNICAMP, Campinas, 2001.
- [6] NOEMI, N.; BUENO, M. I. M. S. Métodos Matemáticos para Correção de Interferências Espectrais e Efeitos Interelementos na Análise Quantitativa por Fluorescência de Raios-X. **Química Nova**, v. 24, n. 4, p. 531-539, 2001.
- [7] LEHNINGER, A. L.; DAVID, L. N.; MICHAEL, M. C. **Princípios de Bioquímica**. São Paulo: Savier, 2002. 975 p.

- [8] NAVES, M. M. V.; SILVA, M. S.; CERQUEIRA, F. M.; PAES, M. C. D. Avaliação Química e Biológica da Proteína do Grão em Cultivares de Milho de Alta Qualidade Protéica. **Pesquisa Agropecuária Tropical**, v. 34, n.1, p. 1-8, 2004.
- [9] MOLINA et al. Manipulação de cereais para acúmulo de lisina em sementes. **Scientia Agricola**, v. 58, n.1, p. 205-211, 2001.
- [10] SAURINA, J. HERNÁNDEZ-CASSOU, S. **Analytica Chimica Acta**, 1993, 283, 414.
- [11] CIENFUEGOS, F.; VAITSMAN, D. **Análise Instrumental**. Rio de Janeiro: Interciência, 2000. 606 p.
- [12] REIS, B. F. Análise por Injeção em Fluxo: Vinte Anos de Desenvolvimento. **Química Nova**, v. 19, n.1, p. 51-58, 1996.
- [13] REIS, B. F.; Bergamin, H. Evolução dos Injetores Empregados em Sistema de Análise Química por Injeção em Fluxo. **Química Nova**, v. 16, n.6, p. 570-573, 1993.
- [14] REIS, B. F.; MARTELLI, P. B.; MENEGÁRIO, A. A.; GINÉ, M. F. Sistema de Análise Química por injeção em Fluxo Controlado por Microcomputador para Diluição Automática em Espectrometria de Absorção Atômica. **Química Nova**, v. 16, n.2, p. 109-112, 1992.
- [15] REIS, B. F.; BERGAMIN, H. F. Evolução dos Injetores Empregados em Sistema de Análise Química por Injeção em Fluxo. **Química Nova**, v. 16, n.6, p. 570-573, 1993.
- [16] REIS, B. F.; GINÉ, M. F.; KRONKA, E. A. A Análise Química por Injeção em Fluxo Contínuo. **Química Nova**, v. 12, n.1, p. 82-91, 1989.

- [17] COUTO, C. M. C. M.; MONTENEGRO, M. C. B. S. M. Detectores Potenciométricos para Sistemas de Análise por Injeção em Fluxo, Evolução e Aplicação. **Química Nova**, v. 23, n. 6, p. 774 -784, 2000.
- [18] FARIA, L. C.; PASQUINI, C. Um Módulo Automático de Introdução de Amostra para Sistema de Análise por Injeção em Fluxo. **Química Nova**, v. 14, n.3, p. 216-218, 1991.
- [19] MIRANDA, C. E. S.; CARRILHO, E. et al. Sistemas Interfaceados de Análise por Injeção em Fluxo e Eletroforese Capilar (FIA–CE): Desafios, Aplicações e Perspectivas. **Química Nova**, v. 25, n. 3, p. 412-419, 2002.
- [20] PALGROSSI, F. S.; PEDROTTI, J. J.; GUTZ, I. G. R. Injetor multicanal com válvulas de estrangulamento para análise em fluxo. **Química Nova**, v. 24, n. 5, p. 689-692, 2001.
- [21] VIEIRA, J. A. Estudo Comparativo de Sistemas de Análises Químicas em Fluxo Empregando FIA com Multicomutação e Amostragem Binária, FIA com Amostragem Sanduíche, Fluxo monossegmentado e Injeção Seqüencial. 2001. 69f. Tese (Doutorado em Química Analítica) – UNICAMP, Campinas, 2001.
- [22] BATALHA, M. C. Desenvolvimento de um sistema de análise por injeção em fluxo para determinação multicomponentes e simultânea de íons metálicos em amostras sintéticas, ambientais e em ligas. 2000. 109p. Dissertação (Mestrado Agroquímica) – Universidade Federal de Viçosa, Viçosa - MG.
- [23] SILVA, J. B. B.; SOUZA, I. G. DE; GERVASIO, A. P. G. Eletrodissolução anódica em sistemas de injeção em fluxo: uma alternativa rápida e eficiente para dissolução de ligas metálicas. **Química Nova**, v. 23, n. 2, p. 244-250, 2000.

- [24] CERDA, V.; ESTELA, J. M.; FORTEZA, R.; CLADERA, A.; BECERRA, E.; ALTIMIRA, P.; SITJAR, P. Flow techniques in water analysis. **Talanta**, v. 50, p. 695-705, 1999.
- [25] KRUG, F. J. A determinação de sulfato em águas naturais e digeridos de vegetais por turbidimetria empregando sistema de injeção em fluxo. 1984. 97p. Dissertação (Doutorado Solos e Nutrição de Plantas) – Universidade de São Paulo, Piracicaba - SP.
- [26] SKOOG, D. A.; WEST, D. M.; HOLLER, F.J.; CROUCH, S. R. **Fundamentos de Química Analítica**. São Paulo: Thomson, 2006. 999 p.
- [27] SKOOG, D. A.; HOLLER, F.J.; NIEMAN, T. A. **Princípios de Análise Instrumental**. 5 ed. Porto Alegre: Bookman, 2002. 836 p.
- [28] OLIVEIRA, M. C.; NOGUEIRA, R. F. P.; NETO, J. A. G. Sistema de Injeção em Fluxo Espectrofotométrico para Monitorar Peróxido de Hidrogênio em Processo de Fotodegradação por Reação Foto-Fenton. **Química Nova**, v. 24, n. 2, p. 188-190, 2001.
- [29] ROCHA, F. R. P.; TEIXEIRA, L. S. G. Estratégias para Aumento de Sensibilidade em Espectrofotometria UV-VIS. **Química Nova**, v. 27, n. 5, p. 807-812, 2004.
- [30] MENDHAM, J.; DENNEY, R. C.; BARNES, J. D.; THOMAS, M. J. K. **Análise Química Quantitativa**. 6 ed. Rio de Janeiro: LTC, 2002. 462 p.
- [31] SALDANHA, T. C. B.; ARAÚJO, M. C.; NETO, B. B. Análise Multicomponente Simultânea por Espectrofotometria de Absorção Molecular UV-VIS. **Química Nova**, v. 22, n. 6, p. 847-853, 1999.

- [32] SENA, M. M.; FREITAS, C. B.; SILVA, L. C.; PÉREZ, C. N.; PAULA, Y. O. Determinação Espectrofotométrica Simultânea de Paracetamol e Ibuprofeno em Formulações Farmacêuticas Usando Calibração Multivariada. **Química Nova**, v. xy, n. 00, p.1-x, 200_.
- [33] Karpinska, J.; Szostak, J. Determination of chlorprothizene and amitryptiline hydrochlorides by UV-derivative spectrophotometry and UV-solid-phase spectrophotometry. *Spectrochim. Acta Part A* 61 (2005) 975-981.
- [34] FERREIRA, M. M. C.; ANTUNES, A. M. MELGO, M. S. VOLPE, P. L. O. Quimiometria I: Calibração Multivariada, um Tutorial. **Química Nova**, v. 22, n. 5, p. 724-731, 1999.
- [35] VALDERRAMA, P. Avaliação de Figuras de Mérito em Calibração Multivariada na Determinação de Parâmetros de controle de Qualidade em Indústria Alcooleira por Espectroscopia no Infravermelho Próximo. 2005. 128f. Dissertação (Mestrado em Química Analítica) – UNICAMP, Campinas, 2005.
- [36] MARTENS, H.; NAES, T. **Multivariate Calibration**. 2 ed. Biddles Ltd, Great Britain, p. 419, 1994.
- [37] FREDERICO, F. R. Desenvolvimento de um Sensor Químico de Fibras Ópticas para a Determinação Simultânea de Metais Pesados Empregando Métodos de Calibração Multivariada. 2003. 108 f. Tese (Doutorado em Química Analítica) – UNICAMP, Campinas 2003.
- [38] FORTES, P. R. Calibração Multivariada e Cinética Diferencial em Sistemas de Análises em Fluxo com Detecção Espectrofotométrica. 2006. Dissertação de Mestrado em Ciências – CENA, Centro de Energia Nuclear na Agricultura da Universidade de São Paulo.

- [39] RIBEIRO, F. A. L. Aplicação de Métodos de Análise Multivariada no Estudo de Hidrocarbonetos Policíclicos Aromáticos. 2001. 174f. Dissertação (Mestrado em Química). UNICAMP, Campinas, 2001.
- [40] THOMAS, E. V. A Primer on Multivariate Calibration; **Anal. Chem.**; 66; 1994; 795A-804A.
- [41] NAGATA, N. Uso de Métodos de Calibração Multivariada na Análise de Amostras de Interesse Industrial e Espécies de Industrial e Espécies de Importância Ambiental por Fluorescência de Raios-X. 2001. 139 f. Tese (Doutorado em Química Analítica) – UNICAMP, Campinas 2003.
- [42] MARTENS, H.; NAES, T. **Multivariate Calibration**; John Wiley & Sons, New York 1989.
- [43] BORIN, A. Aplicação de Quimiometria e Espectroscopia no Infravermelho no Controle de Qualidade de Lubrificantes. 2003. 104f. Dissertação (Mestrado em Química Analítica) – UNICAMP, Campinas, 2003.
- [44] SMITH, G. L. Principal component analysis: an introduction. **Analytical Proceedings**. v. 28, n. 5, p. 150-151, 1991.
- [45] FIDÊNCIO, P. H. Análise de Solos por Espectroscopia no Infravermelho Próximo e Aplicações de Métodos Quimiométricos. 2001. 104 f. Tese (Doutorado em Química Analítica) – UNICAMP, Campinas, 2001.
- [46] GELADI, P.; KOWALSKI, B. R. Partial Least-Square Regression: A tutorial. **Analytica Chimica Acta**, Amsterdam, v. 185, p. 1-7, 1986.
- [47] Grupo PET – Engenharia Elétrica. Curso de Matlab. [on-line]. Disponível em: <<http://www.del.ufms.br/tutoriais/matlab/apresentacao.htm>>. Acesso em: 25/09/06.
- [48] MATHWORKS, Matlab, Mathworks, Natick, Ma, 1993.

[49] PORTUGAL, R. S. Instituto de Geociências – UNICAMP. [on-line]. Disponível em: <<http://www.ime.unicamp.br/~cheti/intmatlab>>. Acesso em: 25/09/06.

[50] ZUBELLI, J. P. SILVA, M. PASTORE, D, H. Tutorial para MATLAB - Curso de Métodos Matemáticos em Finanças I. [on-line]. Disponível em: <<http://w3.impa.br/~zubelli/tutorial/index.html>> Acesso em: 25/09/06.

[51] WISE, B. M. **PLS-Toolbox for use with MATLAB**. Center for Process Analytical Chemistry and Department of Chemical Engineering BF-10 University of Washington Seattle, 1991. 150 p.

[52] ELIAS, S. P.; CUELBAS, C. J.; ANDRADE, J. C.; Um Estudo Comparativo sobre a Eficiência de Estratégias Quimiométricas de Otimização. **Química Nova**, v. 16, p. 216-219, 1994.

APÊNDICE

8. APÊNDICES

APÊNDICE A

1. Pré-processamento dos dados

[X,mx,msg] = mncn(X) - Pré-processamento para centrar os dados na média;

[X,cm] = savgol(X,pontos em filtro,ordem do polinômio,ordem da derivada) - Pré-processamento dos dados com suavização e derivação (Savgol);

2. Análise das Componentes Principais

model = pca(X,ncomp,options) - Fornece o modelo (etapa de calibração);
% ncomp = número de componentes a ser calculado;

pred = pca(X,model,options) - Aplica o modelo construído em uma nova matriz de dados.

3. Modelagem com o PLS

model = pls(X,Y,ncomp,options) - Fornece o modelo de regressão (etapa de calibração);

pred = pls(X,model,options) - Realiza a previsão em uma nova matriz de dados;

valid = pls(X,Y,model,options) - Realiza a previsão com novas matrizes x e y.

4. Validação Cruzada

[press,cumpress,rmsecv,rmsec,cvpred,misclassified] = crossval(X,Y,rm,cvi,ncomp,out,pre) - Realiza a validação cruzada do tipo "leave one out" – "deixa um fora";

APÊNDICE B

1. Seqüência de comandos para construção e validação do modelo PLS

load X - Matriz de dados de absorbância;

load Y - Matriz de concentração:

[X,cm] = savgol(X,5,2,1) - Pré-processamento dos dados com suavização e derivação;

2. Edição das colunas da matriz de concentração

ya=Y(:,1);

yb=Y(:,2);

yc=Y(:,3);

yd=Y(:,4);

ye=Y(:,5);

3. Construção e validação o modelo de regressão

rm = 'sim' - PLS via algoritmo SIMPLS

cvi = {'loo'} - Validação Cruzada "leave one out"

vls=20;

vl=12;

[ba,ssqa,pa,qa,wa,ta,ua,bina] = pls(X,ya,vls);

[bb,ssqb,pb,qb,wb,tb,ub,binb] = pls(X,yb,vls);

[bc,ssqc,pc,qc,wc,tc,uc,binc] = pls(X,yc,vls);

[bd,ssqd,pd,qd,wd,td,ud,bind] = pls(X,yd,vls);

[be,ssqe,pe,qe,we,te,ue,bine] = pls(X,ye,vls);

**[pressa,cumpressa,rmsecva,rmseca,cvpreda,misclassifieda]=
rossval(X,ya,rm,cvi,20);**

**[pressb,cumpressb,rmsecvb,rmsecb,cvpredb,misclassifiedb]=
crossval(X,yb,rm,cvi,20);**

**[pressc,cumpressc,rmsecvc,rmsecc,cvpredc,misclassifiedc]=
crossval(X,yc,rm,cvi,20);**

```
[pressd,cumpressd,rmsecvd,rmsecd,cvpredd,misclassifiedd]=  
crossval(X,yd,rm,cvi,20);  
[presse,cumpresse,rmsecve,rmsece,cvprede,misclassifiede]=  
crossval(X,ye,rm,cvi,20);  
ypreda=X*ba(vl,:);  
ypredb=X*bb(vl,:);  
ypredc=X*bc(vl,:);  
ypredd=X*bd(vl,:);  
yprede=X*be(vl,:);
```

4. Cálculo do erro de previsão

```
era=((ypreda-ya)./ya)*100;  
erb=((ypredb-yb)./yb)*100;  
erc=((ypredc-yc)./yc)*100;  
erd=((ypredd-yd)./yd)*100;  
ere=((yprede-ye)./ye)*100;
```