

## Um novo método para a rápida verificação da adulteração de whey protein baseado em espectroscopia ATR-FTIR e análise multivariada

### A New Method for Rapid Verification of Whey Protein Adulteration Based on ATR-FTIR Spectroscopy and Multivariate Analysis

Matthews Silva Martins<sup>1</sup>, Marcia Helena Cassago Nascimento<sup>1</sup>, Wanderson Romão<sup>2</sup>, Fabiano Kenji Haraguchi<sup>3</sup>, Paulo Roberto Filgueiras<sup>2</sup> & Valerio Garrone Barauna<sup>4</sup>

<sup>1</sup> Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brasil; <sup>2</sup> Universidade Estadual de Campinas (Unicamp), Campinas, SP, Brasil;

<sup>3</sup> Universidade Federal de Ouro Preto (UFOP), Ouro Preto, MG, Brasil; <sup>4</sup> Faculdade de Medicina da Universidade de São Paulo (FM-USP), São Paulo, Brasil. Martins, M. S.\*

matthewsilvam@gmail.com  
<https://orcid.org/0000-0002-0050-4490>

Nascimento, M. H. C.  
marcia.cassago@gmail.com  
<https://orcid.org/0000-0001-5252-586X>

Romão, A.  
wandersonromao@gmail.com  
<https://orcid.org/0000-0002-2254-668X>

Haraguchi, F. K.  
fabianokenji@gmail.com  
<https://orcid.org/0000-0002-1019-8888>

Filgueiras, P. R.  
filgueiras.pr@gmail.com  
<https://orcid.org/0000-0003-2617-1601>

Barauna, V. G.  
barauna2@gmail.com  
<https://orcid.org/0000-0003-2832-0922>

\* Autor correspondente: Av. Marechal Campos, nº 1468, Prédio do Básico 1, CCS/UFES, Campus de Maruípe, Vitória/ES, CEP 29.043-900

**RESUMO:** Os métodos atualmente empregados para detectar a adulteração de proteína do soro do leite bovino (*whey protein*) são demorados, dependem de reagentes e exigem mão de obra especializada. Neste trabalho, propomos a utilização da espectroscopia de infravermelho com reflectância total atenuada (ATR-FTIR), aliada à análise multivariada de dados, como um método de triagem para detectar adulteração no *whey protein*. Na análise dos espectros, as bandas de Amida I (~1550 cm<sup>-1</sup>) e Amida II (~1650 cm<sup>-1</sup>), características de proteínas, apresentaram intensidades de absorbância decrescentes à medida que o adulterante foi adicionado. Foram desenvolvidos modelos de predição utilizando o método de regressão por mínimos quadrados parciais (PLS). Três modelos foram construídos com uma marca diferente de *whey protein* cada, além de um quarto modelo com a combinação de todas as marcas. Os modelos apresentaram altos coeficientes de determinação (>0,94) e baixos erros médios (< 0,71 g/30g) para o conjunto validação externa. Em seguida, testes de predição foram conduzidos com novas amostras, adulteradas de forma aleatória, não utilizadas na construção e validação, para verificar a aplicabilidade dos modelos. Estes testes apresentaram boas predições acerca do conteúdo proteico das amostras. Já o modelo construído com a combinação de todas as marcas demonstrou ser representativo, uma vez que foi capaz de prever amostras independente da marca, sabor e tipo de *whey protein* adulterado, com um erro quadrático médio de 0,46 g/30g e variações de predição de 1,47 a 15,24%. Concluímos, portanto, que a espectroscopia ATR-FTIR, associada à análise multivariada de dados, pode ser empregada de forma auxiliar aos métodos tradicionais na triagem de identificação de *whey protein* adulterado, de maneira muito mais rápida (< 3 min por amostra), com bom custo-benefício e sem a necessidade de reagentes.

**Palavras-chave:** Adulteração de Alimentos, Espectroscopia de Infravermelho com Transformada de Fourier, Proteínas do Soro do Leite, Quimiometria.

#### ABSTRACT

Current methods for detecting whey protein adulteration are time-consuming, reliant on reagents, and require specialized manpower. In this study, attenuated total reflection Fourier-transform infrared spectroscopy (ATR-FTIR), in conjunction with multivariate data analysis, is proposed as a screening method for determining whey protein adulteration. Regarding the spectral analysis, the absorbance intensity of Amida I (~1550 cm<sup>-1</sup>) e Amida II (~1650 cm<sup>-1</sup>) bands, related to protein concentration, decreased as the adulterant was added. Four prediction models were constructed using the partial least squares regression (PLS) method, one for each brand of whey protein concentrate, and one with all brands combined. The models showed high coefficients of determination (>0.94) and low errors (<0.71 g/30g) for the external validation set. Then, blind prediction tests were conducted with new samples that were randomly adulterated to assess the applicability of the models. Accurate predictions of the protein content in samples were achieved. Furthermore, the model constructed by combining all brands has proven to be representative and capable of predicting samples irrespective of the brand, flavor, and type of whey protein with a mean squared error of 0.46 g/30g and prediction variations ranging from 1.47 to 15.24%. In conclusion, ATR-FTIR spectroscopy combined with multivariate data analysis can be employed as a supplementary method to traditional approaches in screening for the identification of adulterated whey protein, in a much faster (< 3 min per sample), cost-effective manner, and without the need for reagents.

**Keywords:** Chemometrics, Food Contamination, Fourier Transform Infrared Spectroscopy, Whey proteins.

## 1 INTRODUÇÃO

Nas últimas décadas, o uso de suplementos alimentares aumentou consideravelmente entre praticantes de atividade física, e evidências apontam que sua taxa de utilização é ainda maior entre atletas (Maughan et al., 2011). O aumento da performance esportiva, compensação de uma dieta inadequada e a urgência de atingir altas demandas dietéticas são as principais justificativas utilizadas para suportar o seu uso indiscriminado, resultando em um mercado global avaliado em mais de \$100 bilhões (Hannon et al., 2020; Huang et al., 2006). Contudo, ainda não há um consenso se os benefícios relatados superam os possíveis riscos à saúde relacionados a uma grande quantidade de produtos contaminados e adulterados encontrados no mercado (Garrido et al., 2016; Petróczy et al., 2007).

Casos de adulteração de suplementos alimentares já são bem descritos na literatura, e levantam questionamentos em relação à segurança desses produtos (Garrido et al., 2016; Molin et al., 2019; Silva & Souza, 2016). Denúncias e análises recentes apontam para suplementos com mais carboidratos e/ou menos proteínas do que o informado nos rótulos. No Brasil, as leis de rotulagem configuram-se como o principal guia para a delimitação e controle destes casos (Almeida et al., 2016; Maughan et al., 2011). Segundo a resolução da Diretoria Colegiada da Agência Nacional de Vigilância Sanitária (ANVISA) Nº 429 as quantidades de qualquer macronutriente do produto comercializado não devem diferir em mais de 20% do valor declarado no rótulo (Agência Nacional de Vigilância Sanitária [ANVISA], 2020). Já entre os principais produtos observados com estas inconformidades, destacam-se suplementos proteicos como a proteína do soro do leite, o *whey protein* (WP) (Justa Neves & Caldas, 2015).

O WP é um suplemento proteico em pó fabricado a partir do soro do leite bovino, e obtido durante a coagulação da caseína. Suas proteínas são consideradas de alto valor biológico, uma vez que apresentam altos teores de aminoácidos essenciais (Kilara & Vaghela, 2018). O WP do tipo concentrado, sua formulação mais popular, possui uma concentração proteica máxima de 80%, um baixo nível de gordura e em geral uma maior quantidade de compostos

bioativos e de carboidratos na forma de lactose (Gangurde et al., 2011). Alguns estudos demonstraram que o WP é o suplemento proteico mais consumido entre praticantes de atividade física, devido ao seu papel na síntese proteica e na construção de massa muscular, entretanto, sua adulteração também é muito comum (Børsheim et al., 2002; Pasiakos et al., 2015; Yoshizawa, 2004).

A adição de adulterantes no WP normalmente resulta em uma mistura com menor teor proteico do que o indicado no rótulo. O teor de proteína, por sua vez, é considerado a principal medida de qualidade para suplementos proteicos (Ingle et al., 2016). Os métodos tradicionais utilizados para a dosagem de proteína são baseados no método de Kjeldahl ou Dumas, baseados na quantificação do nitrogênio total das amostras (AOAC International, 2012; Instituto Adolfo Lutz, 2005). Entretanto, estes são métodos laboratoriais extensivos, que envolvem múltiplas etapas, requerem uma mão de obra especializada e gastos com reagentes, além do elevado tempo das análises. Portanto, estes métodos são limitados em analisar grandes quantidades de amostras em uma realidade crescente de denúncias (Ezhilan et al., 2017; Lukacs et al., 2018; Saxton & McDougal, 2021).

Diante desse cenário, surge a oportunidade para técnicas inovadoras que busquem solucionar as limitações das técnicas atuais. A Espectroscopia de Infravermelho por Transformada de Fourier com Reflectância Total Atenuada (ATR-FTIR) permite a captação da energia vibracional de moléculas orgânicas por meio da incidência de radiação na região do infravermelho médio. Por meio de medições dessa interação são geradas informações que possibilitam a identificação e a quantificação de grupos funcionais característicos (Andrade et al., 2019; Pirutin et al., 2023). Suas principais vantagens incluem uma análise realizada em poucos minutos, com mínima preparação amostral e que não destrói as amostras (Kazarian & Chan, 2006).

Existem hoje poucos trabalhos que utilizaram essa técnica para avaliar a adulteração de WP (Andrade et al., 2019; Braga et al., 2021). Em uma pesquisa anterior realizada pelo presente grupo, o potencial do infravermelho em detectar esse tipo de adulteração foi testado e validado (Martins et al., 2022). Porém, foram construídos apenas modelos individuais para cada marca e com uma região mais ampla do espectro

(1800–800  $\text{cm}^{-1}$ ), conhecida como *fingerprint*. Dessa forma, sua otimização pode ser alcançada através da delimitação da região espectral relacionada a proteínas (1716–1471  $\text{cm}^{-1}$ ) e de um modelo geral combinando diferentes marcas, sabores e tipos. Tais melhorias tem como objetivo a comparação entre os modelos e avaliar a aplicabilidade do método proposto independente das características do WP.

Mediante o exposto, o objetivo do presente estudo é otimizar a predição de modelos previamente construídos para a identificação rápida da adulteração de WP utilizando a espectroscopia ATR-FTIR e a análise multivariada de dados. Além disso, propõe-se a construção de um modelo representativo com a combinação de diferentes marcas e características. Por meio de tais melhorias, espera-se que essa técnica possa ser aplicada na triagem de identificação da adulteração de WP de forma rápida e eficiente, independente das características do WP analisado.

## 2 MATERIAIS E MÉTODOS

### 2.1 PREPARAÇÃO DAS AMOSTRAS

Três marcas de *whey protein* (WP) do tipo concentrado com o mesmo sabor baunilha foram adquiridas e, representadas no estudo por A, B e C. O adulterante utilizado para promover a diluição proteica foi a farinha de trigo tipo I (FT), e duas marcas diferentes representadas por X e Y foram adquiridas. Cada marca de WP foi adulterada como uma mistura binária com cada marca de FT, totalizando seis combinações de amostras (Tabela 1). Todos os produtos utilizados no estudo foram adquiridos em mercados locais.

Tabela 1  
Combinações de amostras, combinações individuais e conjunto geral

| Combinações de amostras | Combinações individuais | Conjunto geral |
|-------------------------|-------------------------|----------------|
| WP A + FT X             | WP A + FT               | WP + FT        |
| WP A + FT Y             |                         |                |
| WP B + FT X             | WP B + FT               |                |
| WP B + FT Y             |                         |                |
| WP C + FT X             | WP C + FT               |                |
| WP C + FT Y             |                         |                |

WP = whey protein; FT = farinha de trigo; A, B e C representam as marcas de WP e X e Y representam as marcas de FT.

Para cada mistura binária (WP + FT), as amostras foram adulteradas com a adição de FT de 10 em 10 wt% (0 a 100 wt%). A massa total de cada amostra foi de 5 g. As amostras foram pesadas em balança analítica (Shimadzu, modelo Unibloc, 220 g – 0,1 mg), misturadas manualmente e depois homogeneizadas em vórtex (Benchmark Scientific, BenchMixer BV1000) por 30 s. A seguir, três alíquotas de aproximadamente 200 mg, cada, foram analisadas no espectrômetro.

Adicionalmente, para utilização no teste de aplicabilidade dos modelos, foram adquiridas amostras de WP do tipo concentrado sabor chocolate e WP do tipo vegano (à base de proteínas vegetais) sabor bolo de banana, de diferentes marcas comerciais, representadas no estudo respectivamente por D e E.

### 2.2 ESPECTROSCOPIA ATR-FTIR

Os espectros foram adquiridos em um espectrômetro FTIR Alpha II (Bruker Corporation) com acessório ATR com cristal de diamante e software OPUS 6.5. Os dados foram obtidos a partir de 32 *scans* no modo de absorção, na faixa espectral de 4000–400  $\text{cm}^{-1}$  com resolução de 4  $\text{cm}^{-1}$ , resultando em um tempo de medição de 40 s por amostra. A alíquota da amostra (~200 mg) foi posicionada no cristal do ATR e uma prensa de amostragem foi acionada para garantir a homogeneidade e o contato do material com o cristal. O *background* foi coletado a cada mudança de amostra e as medidas foram feitas em triplicata, resultando em 33 espectros por combinação. O cristal e a prensa de amostragem foram limpos com álcool 70% v/v e água Mili Q entre as medições das amostras.

### 2.3 PRÉ-PROCESSAMENTOS

Os espectros derivados das misturas binárias (WP + FT) de mesma marca foram unidos por média simples, reduzindo os dados a três combinações individuais, uma para cada marca de WP. Em seguida, os espectros das três combinações individuais foram concatenados para formar um único conjunto geral das três marcas de WP (Tabela 1). As análises

foram conduzidas tanto nas três combinações individuais de espectros quanto no conjunto geral, e os resultados foram comparados.

Todas as análises foram conduzidas no software Matlab R2023b. Todos os espectros foram pré-processados para correção dos desvios na linha de base por meio do algoritmo airPLS (Zhang et al., 2010) e padronização normal de sinal (*standard normal variate* – SNV) para corrigir os efeitos da dispersão da luz e diferenças no tamanho das partículas. Em seguida, os espectros foram truncados na região de absorção atribuída a Amidas I e II, que estão associadas a ligações peptídicas de proteínas (1716-1471  $\text{cm}^{-1}$ ), em conformidade com a propriedade de interesse (Wang et al., 2015).

## 2.4 DETERMINAÇÃO DE PROTEÍNA DAS AMOSTRAS

A princípio, o teor proteico de cada amostra foi atribuído a partir das informações contidas nos rótulos dos produtos. Para as misturas binárias (WP + FT) somou-se as quantidades de proteína, conforme descrição nos rótulos, contidas nas massas utilizadas de WP e FT em casa mistura. Em seguida, esses valores foram expressos em 30 g, um valor de porção comumente encontrado nos rótulos de WP (Tabelas S1-S3).

A seguir, o teor proteico foi determinado pelo método Kjeldahl para confirmar os teores descritos nos rótulos e observar se havia discrepâncias entre os valores informados. O ensaio de determinação de proteínas pelo método Kjeldahl foi realizada segundo métodos estabelecidos pelo Instituto Adolf Lutz para análise de alimentos (Instituto Adolfo Lutz, 2005). Este método baseia-se em três etapas: digestão, destilação e titulação, levando algumas horas para ser concluído. A amostra é digerida com ácido sulfúrico concentrado, transformando o nitrogênio orgânico em íons amônio. Em seguida, a solução obtida é alcalinizada com hidróxido de sódio concentrado, produzindo amônia, que é recolhida em ácido bórico na etapa de destilação. Por fim, a solução de ácido bórico é titulada com ácido padronizado (Instituto Adolfo Lutz, 2005).

Foi utilizado um destilador de nitrogênio da marca CIENLAB (modelo CE-600). O teor de nitrogênio foi convertido

em teor de proteína por meio do fator de conversão de 6,38, utilizado para produtos lácteos e o teor de proteína foi determinado conforme a Equação 1. O teor determinado de proteína de cada WP (Tabela S4) foi utilizado para corrigir os valores informados nos rótulos conforme a necessidade.

$$\text{Proteína (\%)} = \frac{Fc \times 0,0014 \times 100 \times V \times 6,38}{P}$$

Equação 1:

Onde: *Fc* é fator de correção da solução de ácido clorídrico 0,1 M; *V* é volume da solução de ácido clorídrico gasto na titulação; e *P* é massa da amostra em gramas.

## 2.5 CALIBRAÇÃO MULTIVARIADA

Quatro modelos de calibração multivariada foram construídos, um para cada combinação individual (WP + FT), gerando três modelos, e um para uma combinação de todas as marcas de WP e FT. Todos os modelos foram construídos pelo método PLS. Para facilitar a compreensão dos resultados, aqui descreveremos como modelos individuais (relativos às combinações individuais WP + FT de cada marca) e modelo geral (relativo à combinação de misturas de todas as marcas analisadas).

O PLS permite a maximização da covariância entre dois blocos de dados, *X* (espectros das amostras) e *y* (quantidade de proteína de cada amostra, em g). A partir dessa maximização, um modelo matemático é encontrado para descrever a relação entre os dois grupos de variáveis investigadas.

A partir do algoritmo mínimos quadrados parciais iterativos não lineares (NIPLAS-PLS1), mais comumente utilizado para estimar um valor de *y* univariado, a matriz *X* e o vetor *y* com a propriedade de interesse são decompostos iterativamente a cada variável latente (VL). Como resultado dessa decomposição, a soma de VL descreve a direção de máxima covariância dos dados originais direcionando-os à propriedade de interesse (Geladi & Kowalski, 1986). O modelo PLS é representado pela Equação 2, e pode ser aplicado para a predição de novas amostras através da combinação linear da nova matriz *X* com o coeficiente de regressão *b*PLS.

Equação 2:

$$y = X \times b_{PLS}$$

Os modelos construídos foram avaliados a partir de métricas de exatidão baseados na raiz quadrada do erro médio (RMSE) (Equação 3-4) e de linearidade pelo coeficiente de determinação ( $R^2$ ) (Equação 5) para os conjuntos calibração e validação. De maneira geral, uma acurácia maximizada com boas previsões requer um erro de previsão minimizado, uma proximidade entre os valores de RMSEC e RMSEP e um  $R^2$  o mais próximo possível de 1 (Bassbasi et al., 2014).

Equação 3:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{ncal} (y_i - \hat{y}_i)^2}{ncal - VL - 1}}$$

Equação 4:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{npred} (y_i - \hat{y}_i)^2}{npred}}$$

Equação 5:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}$$

Onde:  $y_i$  é o valor de referência para a propriedade de interesse;  $\hat{y}_i$  é o valor de predição do modelo;  $\bar{y}_i$  é o valor da média de referência;  $ncal$  é o número de amostras no conjunto calibração;  $npred$  é o número de amostras no conjunto validação externa; e  $VL$  é o número de variáveis latentes utilizado na construção do modelo.

Para a construção dos quatro modelos PLS, cada conjunto de dados foi dividido a partir do algoritmo *Kennard-Stone* em 70% calibração e 30% validação externa, e estes geraram as métricas de predição desejadas, uma vez que não participaram da construção do modelo. Além disso, todas as triplicatas foram consideradas e permaneceram juntas dentro do mesmo conjunto. Na validação cruzada, o método

empregado foi o k-fold ( $k = 5$ ), para otimização do número de VL a partir do RMSECV.

Os limites de detecção (LOD) e limite de quantificação (LOQ) dos modelos também foram estimados. Esses parâmetros representam as menores concentrações que podem ser detectadas e quantificadas, respectivamente. O cálculo foi realizado por meio da abordagem padronizada pela União Internacional de Química Pura e Aplicada (IUPAC) para modelos de calibração multivariada, que fornece um intervalo de valores ao invés de um valor único (Allegrini & Olivieri, 2014).

## 2.6 TESTES DE PREDIÇÃO ÀS CEGAS

Para demonstrar a aplicação dos modelos construídos e comparar a predição entre eles, realizamos dois testes de predição às cegas. No primeiro, um analista independente preparou nove amostras adulteradas (três com cada marca de WP: A, B e C) com níveis aleatórios de substituição por FT em 30 g. As mesmas condições de aquisição e pré-processamento foram empregadas em todos os espectros, e os quatro modelos foram aplicados para determinar a quantidade em g de proteína de cada amostra.

A partir da predição de cada modelo, a comparação entre eles foi realizada através do cálculo do erro quadrático médio (soma do quadrado da diferença do valor real pelo valor previsto de cada amostra). Dessa forma, o modelo que apresentasse o menor erro quadrático médio (eqM), seria aquele com a melhor performance, uma vez que suas predições se aproximaram mais do resultado real. As predições também foram submetidas a um teste t pareado para verificar diferenças estatísticas nas acurácias.

No segundo teste de predição às cegas aplicou-se somente o modelo geral. O PLS é considerado um modelo de regressão com boa robustez, gerando predições satisfatórias para amostras novas, ou seja, que não participaram da calibração do modelo (Cassel et al., 1999). Portanto, uma vez que o modelo geral fosse representativo o suficiente, esperava-se que ele pudesse realizar boas predições de outras marcas, sabores e tipos de WP não utilizados durante a calibração do modelo.

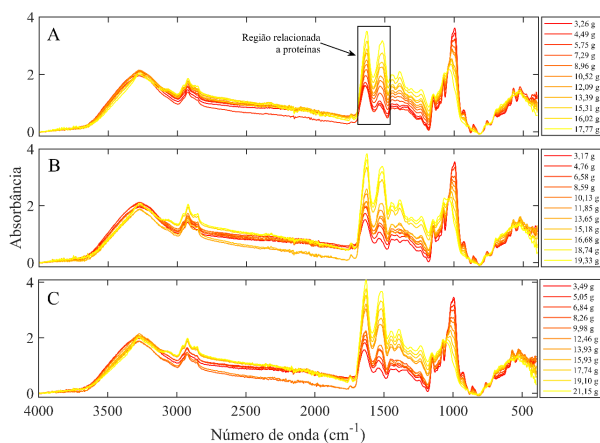
Para testar essa hipótese, quatro amostras de WP da marca D e quatro amostras de WP da marca E foram adulteradas com quantidades (wt%) aleatórias de FT. Os teores de proteína de ambas as marcas também foram determinados pelo método Kjeldahl (Tabela S3). As predições do modelo geral foram verificadas pela métrica de erro quadrático médio.

## 3 RESULTADOS E DISCUSSÃO

### 3.1 ANÁLISE DOS ESPECTROS ATR-FTIR

Na Fig. 1 são representadas as médias dos espectros na região de absorção de 4000-400  $\text{cm}^{-1}$ , com os respectivos teores de proteína de cada amostra em 30 g. A região destacada na Fig. 1A possui duas bandas atribuídas a vibrações das ligações peptídicas (CO-NH) de proteínas, e assim esta região foi truncada: Amida I ( $\nu_{\text{C=O}}$ ,  $\nu_{\text{C-N}}$ ) em 1650  $\text{cm}^{-1}$  e Amida II ( $\delta_{\text{N-H}}$ ,  $\nu_{\text{C-N}}$ ) em 1550  $\text{cm}^{-1}$  (Wang et al., 2015). A partir dessas bandas é possível estimar o conteúdo proteico das amostras, uma vez que, segundo a Lei de Lambert-Beer, a absorbância em bandas espectrais é diretamente proporcional à concentração do grupo funcional correspondente (Wang et al., 2018).

Figura 1  
Média dos espectros ATR-FTIR de cada combinação na região de 4000-400  $\text{cm}^{-1}$

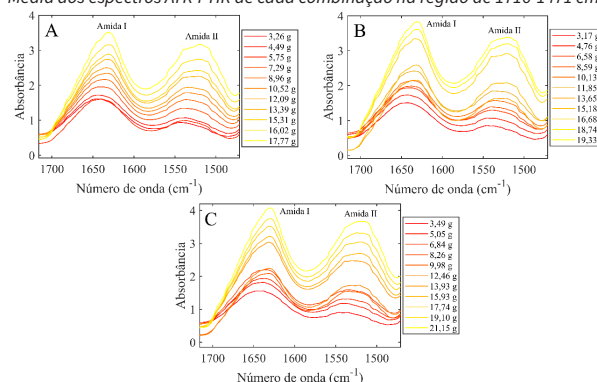


Nota: (A) WP A + FT; (B) WP B + FT; (C) WP C + FT. Variação da escala de cor de acordo com valor do teor proteico.

Os três conjuntos espectrais apresentaram o mesmo comportamento (Fig. 2). É evidente que amostras com

maior concentração de proteína apresentaram maiores intensidades de absorbância nas duas bandas indicadas. À medida que o adulterante foi adicionado, a intensidade das bandas referentes a proteínas diminuiu.

Figura 2  
Média dos espectros ATR-FTIR de cada combinação na região de 1716-1471  $\text{cm}^{-1}$



Nota: (A) WP A + FT; (B) WP B + FT; (C) WP C + FT. Variação da escala de cor de acordo com valor do teor proteico.

O mesmo padrão foi observado quando todos os espectros das combinações de diferentes marcas foram compostos em um conjunto de dados único (Figura S1). Em todos os casos, as intensidades de absorbância variaram proporcionalmente em função da quantidade de adulterante FT adicionado às amostras. Esses dados demonstram ser possível a caracterização da variação do teor proteico das amostras de WP a partir dos espectros. Portanto, se existe correlação entre o espectro e a propriedade de interesse que se deseja estimar, é possível encontrar uma expressão matemática que melhor descreva essa relação e permita estimar a mesma propriedade para espectros de novas amostras desconhecidas (Forina et al., 2007).

### 3.2 MODELOS DE CALIBRAÇÃO PLS

Os principais parâmetros dos quatro modelos são apresentados na Tabela 2, assim como o número de variáveis latentes escolhidos. Os coeficientes de determinação foram acima de 0,99 para o conjunto calibração e acima de 0,94 para o conjunto predição, indicando alta linearidade. Os erros médios (RMSE) são expressos na mesma unidade que os teores proteicos, variando de 0,31 a 0,66 g/30 g para o

conjunto calibração e de 0,53 a 0,71 g/30g para o conjunto validação externa, o que mostra que os modelos foram capazes de realizar boas predições. O erro máximo observado para o conjunto validação externa (0,71 g/30g) representa apenas 2,37% do valor total da porção de 30 g, o que reforça a potencialidade das predições dos modelos.

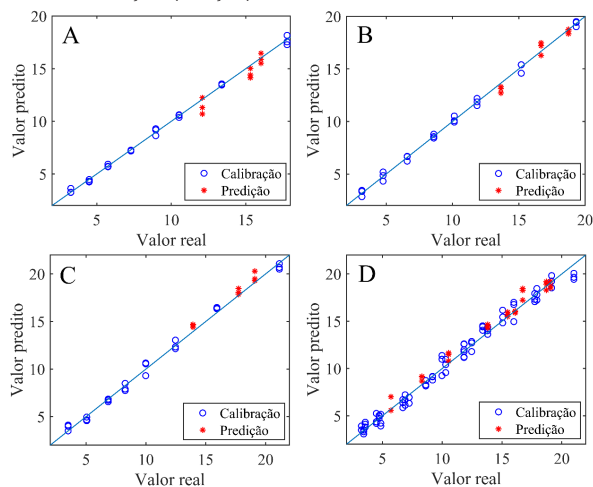
Tabela 2  
Figuras de mérito dos modelos PLS construídos a partir da região espectral truncada (1716-1471  $\text{cm}^{-1}$ )

| Modelos   | Calibração     |               | Predição       |               |    |
|-----------|----------------|---------------|----------------|---------------|----|
|           | R <sup>2</sup> | RMSEc (g/30g) | R <sup>2</sup> | RMSEp (g/30g) | VL |
| WP A + FT | 0,99           | 0,62          | 0,94           | 0,64          | 4  |
| WP B + FT | 0,99           | 0,31          | 0,96           | 0,53          | 4  |
| WP C + FT | 0,99           | 0,48          | 0,98           | 0,62          | 3  |
| WP + FT   | 0,99           | 0,66          | 0,98           | 0,71          | 4  |

Nota: R<sup>2</sup> = coeficiente de determinação; RMSEc = raiz quadrada do erro médio do conjunto calibração; RMSEp = raiz quadrada do erro médio do conjunto predição; VL = número de variáveis latentes.

As curvas de calibração dos quatro modelos (Fig. 3) também apontam para modelos lineares, uma vez as amostras do conjunto calibração e predição ficaram bem alinhadas às retas de regressão. No modelo geral (Fig. 3D) amostras do conjunto calibração e teste que ficaram sobrepostas são de marcas diferentes, e, portanto, tem espectros diferentes apesar da mesma concentração proteica.

Figura 3  
Curvas de calibração e predição para cada modelo



Nota: (A) WP A + FT; (B) WP B + FT; (C) WP C + FT; (D) WP + FT.

Os limites de detecção (LOD) e limites de quantificação (LOQ) são apresentados no formato de intervalos mínimo e máximo, na mesma unidade que os teores de proteína

(g/30 g) (Tabela 3). Abaixo do valor mínimo, as detecções e quantificações não são possíveis. Entre o valor mínimo e o máximo elas são possíveis com uma menor confiabilidade. Acima do valor máximo, as predições possuem uma maior confiabilidade no resultado (Allegrini & Olivieri, 2014). Em relação à quantificação, considerando todos os modelos avaliados, é possível obter melhores predições para valores acima de 6,70 g/30 g de proteína nas amostras, o que equivale a aproximadamente adulterações na faixa de 0-80%.

Tabela 3  
Limite de detecção (LOD) e limite de quantificação (LOQ) para cada modelo

| Modelos   | LOD (g/30g) |      | LOQ (g/30g) |      |
|-----------|-------------|------|-------------|------|
|           | min         | max  | min         | max  |
| WP A + FT | 0,65        | 1,40 | 1,96        | 4,23 |
| WP B + FT | 0,75        | 2,11 | 2,27        | 6,39 |
| WP C + FT | 0,92        | 2,21 | 2,79        | 6,70 |
| WP + FT   | 0,61        | 1,95 | 1,84        | 5,92 |

### 3.3 PREDIÇÕES ÀS CEGAS E COMPARAÇÃO DOS MODELOS

Os resultados das predições dos teores de proteína das nove amostras aleatórias podem ser observados na Tabela 4 como média e desvio padrão para cada modelo. O valor médio das predições foi utilizado para o cálculo do erro quadrático médio (eqM), por meio do qual estimou-se as performances de cada modelo (Tabela 5).

Tabela 4  
Predições às cegas para nove amostras de três marcas diferentes realizadas por cada modelo. Valores expressos em uma porção de 30 g

| Marcas | Valor real (g) | WP A + FT (g) | WP B + FT (g) | WP C + FT (g) | WP + FT (g)  |
|--------|----------------|---------------|---------------|---------------|--------------|
| A      | 17,34          | 16,27 ± 0,45  | 19,53 ± 0,63  | 18,50 ± 0,25  | 18,00 ± 0,33 |
| A      | 16,62          | 15,64 ± 0,36  | 18,53 ± 0,15  | 17,00 ± 0,52  | 16,77 ± 0,21 |
| A      | 7,92           | 7,03 ± 0,31   | 9,17 ± 0,55   | 9,03 ± 0,31   | 9,14 ± 0,26  |
| B      | 17,36          | 15,11 ± 0,28  | 17,05 ± 0,72  | 16,13 ± 0,49  | 17,28 ± 0,73 |
| B      | 15,50          | 14,33 ± 0,84  | 15,90 ± 0,02  | 14,69 ± 0,43  | 15,43 ± 0,34 |
| B      | 8,83           | 7,66 ± 0,51   | 8,04 ± 0,78   | 7,10 ± 0,53   | 8,79 ± 0,51  |
| C      | 18,12          | 16,26 ± 0,10  | 19,42 ± 0,08  | 18,52 ± 0,10  | 17,71 ± 0,10 |
| C      | 16,86          | 15,05 ± 0,37  | 17,76 ± 0,09  | 16,85 ± 0,54  | 16,09 ± 0,16 |
| C      | 11,64          | 12,25 ± 0,51  | 13,19 ± 0,07  | 12,56 ± 0,25  | 12,85 ± 0,05 |

Tabela 5  
 Erro quadrático médio (eqM) de cada modelo baseado nas predições às cegas de nove amostras

| Modelo    | Erro quadrático médio (g/30g) |
|-----------|-------------------------------|
| WP A + FT | 1,98                          |
| WP B + FT | 1,75                          |
| WP C + FT | 0,99                          |
| WP + FT   | 0,46                          |

O modelo geral (WP + FT) foi o que apresentou o menor eqM (0,46 g/30g) com boas predições para amostras de todas as marcas. O resultado é esperado, uma vez que este modelo foi construído com as três marcas com as quais as amostras às cegas foram preparadas. De maneira geral, os modelos dos conjuntos individuais tiveram melhores predições nas marcas de amostras com as quais foram calibrados, o que também era esperado.

As predições dos modelos também foram submetidas a um teste t pareado para verificar se havia diferença estatística significativa. A acurácia do modelo geral foi significativamente maior que a do modelo WP A + FT ( $p < 0,01$ ). Entretanto, não foram observadas diferenças estatísticas entre o modelo geral e o modelo WP B + FT ( $p = 0,97$ ) e entre o modelo geral e o modelo WP C + FT ( $p = 0,27$ ). Apesar das acurácias para predições com as marcas A, B e C serem estatisticamente iguais para estes modelos, o uso do modelo geral justifica-se pela sua capacidade de generalização para WP com outras características do que os utilizados na calibração do modelo.

Na Tabela 6 estão descritas as predições das marcas D e E para o modelo geral (WP + FT). Ressaltando que a marca D é de um sabor diferente do utilizado na construção do modelo (sabor chocolate) e a marca E é de tipo e sabor diferentes (proteína vegetal sabor bolo de banana).

As variações das predições foram calculadas a partir da divisão da diferença entre o valor real e previsto pelo valor real, e variaram na faixa de 1,47 a 15,24%. Tais resultados reforçam a hipótese de que o modelo construído com o conjunto geral é capaz de realizar boas predições em wheys de marcas, sabores e tipos diferentes dos que foi calibrado. Além disso, a adição de mais amostras diferentes como estas no modelo geral pode contribuir com a melhora das predições do modelo, devido ao aumento da representatividade, e com a diminuição dos erros médios e da variação preditiva.

Tabela 6  
 Predições às cegas para oito amostras de duas marcas diferentes para o modelo construído com o conjunto geral

| Marcas | Valor real (g) | Valor predito (g) | Varição da predição (%) |
|--------|----------------|-------------------|-------------------------|
| D      | 16,02          | 15,22 ± 0,64      | 4,99                    |
| D      | 15,42          | 15,01 ± 0,48      | 2,66                    |
| D      | 20,22          | 17,50 ± 0,13      | 13,45                   |
| D      | 18,54          | 16,97 ± 0,27      | 8,47                    |
| E      | 16,02          | 14,76 ± 0,75      | 7,87                    |
| E      | 12,24          | 12,06 ± 0,43      | 1,47                    |
| E      | 14,76          | 12,51 ± 0,63      | 15,24                   |
| E      | 10,38          | 9,30 ± 0,09       | 10,40                   |

Nota: A variação da predição foi calculada a partir da divisão da diferença entre o valor real e previsto pelo valor real. Valores expressos em uma porção de 30 g.

Destaca-se ainda que o tempo de análise para uma amostra foi medido para fins de comparação com o método padrão. Entre coleta de espectro, limpeza do equipamento e medição do *background* foram gastos aproximadamente 2,5 minutos por amostra. Quando comparado com as várias horas necessárias para percorrer todas as etapas do método Kjeldahl, fica evidente que a análise por espectroscopia ATR-FTIR permitiria um aumento considerável do número de amostras analisadas em um espaço de tempo relativamente curto e com mínima preparação amostral.

## 4 CONCLUSÃO

O presente estudo investigou a utilização da espectroscopia ATR-FTIR, aliada a técnicas de análise multivariada, como um método na identificação rápida da adulteração de WP. Modelos previamente desenvolvidos para marcas individuais (A, B e C) foram otimizados e um modelo geral com a combinação de todas as marcas foi construído. Em geral, os modelos apresentaram alta linearidade e baixos erros médios. No teste de predição às cegas inicial, o modelo geral apresentou o menor erro quadrático médio. No teste de predição às cegas com amostras de outras marcas, sabores e tipos de WP (D, E) o modelo geral também apresentou boas predições, corroborando com a hipótese de que esse modelo é representativo e apresenta robustez para lidar com amostras diferentes de WP. Portanto, os dados sugerem que a espectroscopia ATR-FTIR pode ser utilizada de forma rápida para determinar a adulteração de WP e análise da

conformidade da rotulagem desses produtos, independente da marca, sabor e tipo de WP avaliado em menos de 3 min por amostra. Essa ferramenta de triagem permitiria um aumento expressivo do número de amostras avaliadas pelas agências regulamentadoras. Como perspectivas para a continuidade do trabalho, pretende-se empregar o modelo geral construído para avaliar amostras apreendidas pelos órgãos regulamentadores, além da incorporação de mais amostras para a constante otimização do modelo. Espera-se que esse método possa ser empregado efetivamente nas análises de forma auxiliar, em conjunto com os métodos tradicionais, para reforçar a fiscalização e garantir aos consumidores produtos mais seguros.

## AGRADECIMENTOS

Agradecemos ao suporte financeiro da FAPES, CAPES e CNPq. Agradecemos também ao Laboratório de Análises de Alimentos do Curso de Nutrição da Universidade Federal do Espírito Santo por ceder o espaço e os equipamentos para a realização da dosagem de proteínas.

## REFERÊNCIAS

- Agência Nacional de Vigilância Sanitária [ANVISA]. Resolução da Diretoria Colegiada – RDC No 429, de 8 de outubro de 2020.
- Allegrini, F., & Olivieri, A. C. (2014). IUPAC-Consistent Approach to the Limit of Detection in Partial Least-Squares Calibration. *Analytical Chemistry*, *86*(15), 7858-7866. <https://doi.org/10.1021/ac501786u>
- Almeida, C. C., Alvares, T. S., Costa, M. P., & Conte-Junior, C. A. (2016). Protein and Amino Acid Profiles of Different Whey Protein Supplements. *Journal of Dietary Supplements*, *13*(3), 313-323. <https://doi.org/10.3109/19390211.2015.1036187>
- Andrade, J., Pereira, C. G., Almeida Junior, J. C. de, Viana, C. C. R., Neves, L. N. de O., Silva, P. H. F. da, Bell, M. J. V., & Anjos, V. de C. dos. (2019). FTIR-ATR determination of protein content to evaluate whey protein concentrate adulteration. *LWT*, *99*, 166-172. <https://doi.org/10.1016/j.lwt.2018.09.079>
- AOAC International. (2012). *Official Methods of Analysis* (19th ed). AOAC International.
- Bassbasi, M., Platikanov, S., Tauler, R., & Oussama, A. (2014). FTIR-ATR determination of solid non fat (SNF) in raw milk using PLS and SVM chemometric methods. *Food Chemistry*, *146*, 250-254. <https://doi.org/10.1016/j.foodchem.2013.09.044>
- Børsheim, E., Tipton, K. D., Wolf, S. E., & Wolfe, R. R. (2002). Essential amino acids and muscle protein recovery from resistance exercise. *American Journal of Physiology-Endocrinology and Metabolism*, *283*(4), E648-E657. <https://doi.org/10.1152/ajpendo.00466.2001>
- Braga, S. C. G. N., Braga, F. L., Boschetti, A. de F., Gerardth, L. F. F., da Rocha, M. A. C., & Cecatto, L. (2021). Whey protein supplement adulteration with rice flour quantification: A simple method using ATR-FT-MIR and iPLS. *Scientia Agropecuaria*, *12*(3), 379-383. <https://doi.org/10.17268/sci.agropecu.2021.041>
- Cassel, C., Hackl, P., & Westlund, A. H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, *26*(4), 435-446. <https://doi.org/10.1080/02664769922322>
- Ezhilan, M., Gumpu, M. B., Ramachandra, B. L., Nesakumar, N., Babu, K. J., Krishnan, U. M., & Rayappan, J. B. B. (2017). Design and development of electrochemical biosensor for the simultaneous detection of melamine and urea in adulterated milk samples. *Sensors and Actuators B: Chemical*, *238*, 1283-1292. <https://doi.org/10.1016/j.snb.2016.09.100>
- Forina, M., Lanteri, S., & Casale, M. (2007). Multivariate calibration. *Journal of Chromatography A*, *1158*(1-2), 61-93. <https://doi.org/10.1016/j.chroma.2007.03.082>
- Gangurde, H., Patil, P., Chordiya, M., & Baste, N. (2011). Whey protein. *Scholars' Research Journal*, *1*(2), 69. <https://doi.org/10.4103/2249-5975.99663>
- Garrido, B. C., Souza, G. H. M. F., Lourenço, D. C., & Fasciotti, M. (2016). Proteomics in quality control: Whey protein-based supplements. *Journal of Proteomics*, *147*, 48-55. <https://doi.org/10.1016/j.jprot.2016.03.044>
- Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, *185*, 1-17. [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9)
- Hannon, B. A., Fairfield, W. D., Adams, B., Kyle, T., Crow, M., & Thomas, D. M. (2020). Use and abuse of dietary supplements in persons with diabetes. *Nutrition & Diabetes*, *10*(1), 14. <https://doi.org/10.1038/s41387-020-0117-6>
- Huang, S.-H. (Susan), Johnson, K., & Pipe, A. L. (2006). The Use of Dietary Supplements and Medications by Canadian Athletes at the Atlanta and Sydney Olympic Games. *Clinical Journal of Sport Medicine*, *16*(1), 27-33. <https://doi.org/10.1097/01.jsm.0000194766.35443.9c>
- Ingle, P. D., Christian, R., Purohit, P., Zarraga, V., Handley, E., Freel, K., & Abdo, S. (2016). Determination of Protein Content by NIR Spectroscopy in Protein Powder Mix Products. *Journal of AOAC International*, *99*(2), 360-363. <https://doi.org/10.5740/jaoacint.15-0115>
- Instituto Adolfo Lutz. (2005). *Métodos físico-químicos para análise de alimentos* (4th ed.). Ministério da Saúde.

- Justa Neves, D. B. da, & Caldas, E. D. (2015). Dietary supplements: International legal framework and adulteration profiles, and characteristics of products on the Brazilian clandestine market. *Regulatory Toxicology and Pharmacology*, 73(1), 93-104. <https://doi.org/10.1016/j.yrtph.2015.06.013>
- Kazarian, S. G., & Chan, K. L. A. (2006). Applications of ATR-FTIR spectroscopic imaging to biomedical samples. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1758(7), 858-867. <https://doi.org/10.1016/j.bbmem.2006.02.011>
- Kilara, A., & Vaghela, M. N. (2018). Whey proteins. In *Proteins in Food Processing* (pp. 93-126). Elsevier. <https://doi.org/10.1016/B978-0-08-100722-8.00005-X>
- Lukacs, M., Bazar, G., Pollner, B., Henn, R., Kirchler, C. G., Huck, C. W., & Kovacs, Z. (2018). Near infrared spectroscopy as an alternative quick method for simultaneous detection of multiple adulterants in whey protein-based sports supplement. *Food Control*, 94, 331-340. <https://doi.org/10.1016/j.foodcont.2018.07.004>
- Martins, M. S., Nascimento, M. H., Barbosa, L. L., Campos, L. C. G., Singh, M. N., Martin, F. L., Romão, W., Filgueiras, P. R., & Barauna, V. G. (2022). Detection and quantification using ATR-FTIR spectroscopy of whey protein concentrate adulteration with wheat flour. *LWT*, 172, 114161. <https://doi.org/10.1016/j.lwt.2022.114161>
- Maughan, R. J., Greenhaff, P. L., & Hespel, P. (2011). Dietary supplements for athletes: Emerging trends and recurring themes. *Journal of Sports Sciences*, 29(sup1), S57-S66. <https://doi.org/10.1080/02640414.2011.587446>
- Molin, T. R. D., Leal, G. C., Muratt, D. T., Marcon, G. Z., Carvalho, L. M. de, & Viana, C. (2019). Regulatory framework for dietary supplements and the public health challenge. *Revista de Saúde Pública*, 53, 90. <https://doi.org/10.11606/s1518-8787.2019053001263>
- Pasiakos, S. M., McLellan, T. M., & Lieberman, H. R. (2015). The Effects of Protein Supplements on Muscle Mass, Strength, and Aerobic and Anaerobic Power in Healthy Adults: A Systematic Review. *Sports Medicine*, 45(1), 111-131. <https://doi.org/10.1007/s40279-014-0242-2>
- Petróczi, A., Naughton, D. P., Mazanov, J., Holloway, A., & Bingham, J. (2007). Performance enhancement with supplements: incongruence between rationale and practice. *Journal of the International Society of Sports Nutrition*, 4(1). <https://doi.org/10.1186/1550-2783-4-19>
- Pirutin, S. K., Jia, S., Yusipovich, A. I., Shank, M. A., Parshina, E. Yu., & Rubin, A. B. (2023). Vibrational Spectroscopy as a Tool for Bioanalytical and Biomonitoring Studies. *International Journal of Molecular Sciences*, 24(8), 6947. <https://doi.org/10.3390/ijms24086947>
- Saxton, R., & McDougal, O. M. (2021). Whey Protein Powder Analysis by Mid-Infrared Spectroscopy. *Foods*, 10(5), 1033. <https://doi.org/10.3390/foods10051033>
- Silva, L. V., & Souza, S. V. C. de. (2016). Qualidade de suplementos proteicos: avaliação da composição e rotulagem. *Revista Do Instituto Adolfo Lutz*, 75, 01-17. <https://doi.org/10.53393/rial.2016.v75.33516>
- Wang, T., Tan, S. Y., Mutilangi, W., Aykas, D. P., & Rodriguez-Saona, L. E. (2015). Authentication of Whey Protein Powders by Portable Mid-Infrared Spectrometers Combined with Pattern Recognition Analysis. *Journal of Food Science*, 80(10), C2111-C2116. <https://doi.org/10.1111/1750-3841.13006>
- Wang, X., Esquerre, C., Downey, G., Henihan, L., O'Callaghan, D., & O'Donnell, C. (2018). Feasibility of Discriminating Dried Dairy Ingredients and Preheat Treatments Using Mid-Infrared and Raman Spectroscopy. *Food Analytical Methods*, 11(5), 1380-1389. <https://doi.org/10.1007/s12161-017-1114-9>
- Yoshizawa, F. (2004). Regulation of protein synthesis by branched-chain amino acids in vivo. *Biochemical and Biophysical Research Communications*, 313(2), 417-422. <https://doi.org/10.1016/j.bbrc.2003.07.013>
- Zhang, Z.-M., Chen, S., & Liang, Y.-Z. (2010). Baseline correction using adaptive iteratively reweighted penalized least squares. *The Analyst*, 135(5), 1138. <https://doi.org/10.1039/b922045c>

## MATERIAIS SUPLEMENTARES

Tabela S1

Quantidade de proteína (g) em uma porção de 30 g para cada nível de adulteração no conjunto individual WP A + FT

| Adulterante (%) | TOTAL (g) |
|-----------------|-----------|
| 100             | 3,26      |
| 90              | 4,49      |
| 80              | 5,75      |
| 70              | 7,29      |
| 60              | 8,96      |
| 50              | 10,52     |
| 40              | 12,09     |
| 30              | 13,39     |
| 20              | 15,31     |
| 10              | 16,02     |
| 0               | 17,77     |

Tabela S2

Quantidade de proteína (g) em uma porção de 30 g para cada nível de adulteração no conjunto individual WP B + FT

| Adulterante (%) | TOTAL (g) |
|-----------------|-----------|
| 100             | 3,17      |
| 90              | 4,76      |
| 80              | 6,58      |
| 70              | 8,59      |
| 60              | 10,13     |
| 50              | 11,85     |
| 40              | 13,65     |
| 30              | 15,18     |
| 20              | 16,68     |
| 10              | 18,74     |
| 0               | 19,33     |

Tabela S3

Quantidade de proteína (g) em uma porção de 30 g para cada nível de adulteração no conjunto individual WP C + FT

| Adulterante (%) | TOTAL (g) |
|-----------------|-----------|
| 100             | 3,49      |
| 90              | 5,05      |
| 80              | 6,84      |
| 70              | 8,26      |
| 60              | 9,98      |
| 50              | 12,46     |
| 40              | 13,93     |
| 30              | 15,93     |
| 20              | 17,74     |
| 10              | 19,10     |
| 0               | 21,15     |

Tabela S4

Teor de proteína das marcas de whey protein determinado pelo método de Kjeldahl em 100g

| Marca | Proteína (%) |
|-------|--------------|
| A     | 59,24        |
| B     | 64,43        |
| C     | 70,49        |
| D     | 71,08        |
| E     | 57,14        |

Figura S1

Média dos espectros ATR-FTIR do conjunto geral na região de 1716-1471  $\text{cm}^{-1}$ .

Cada cor representa uma concentração proteica diferente

