

## PLS based quantitative determination of insulin aspart in solution using Raman spectroscopy

Hrisanta Godzo<sup>1\*</sup>, Olga Gigopulu<sup>1</sup>, Nikola Geskovski<sup>2</sup>,  
Jelena Acevska<sup>1</sup>, Natalija Nakov<sup>1</sup>, Jasmina Tonic-Ribarska<sup>1</sup>,  
Ana Poceva Panovska<sup>1</sup>, Katerina Brezovska<sup>1</sup>

<sup>1</sup> Ss. Cyril and Methodius University - Faculty of Pharmacy, Institute of Applied Chemistry and Pharmaceutical Analysis, Majka Tereza 47, 1000 Skopje, Republic of North Macedonia

<sup>2</sup> Ss. Cyril and Methodius University - Faculty of Pharmacy, Institute of Pharmaceutical Technology, Majka Tereza 47, 1000 Skopje, Republic of North Macedonia

\*Corresponding author: Hrisanta Godzo, e-mail: hrisanta@ff.ukim.edu.mk

Received: 18 March 2024; Revised in revised form: 17 April 2024; Accepted: 18 April 2024

---

### Abstract

The complex structure of medicines containing polypeptide active substances requires the implementation of challenging analytical approaches, based on physicochemical methods, and, where necessary, biological assays for quality control, as well as for the detection of substandard and falsified products. Vibrational spectroscopic techniques, including Raman spectroscopy, are fast, powerful, and non-destructive techniques which, when combined with multivariate chemometric modelling, can provide specific identification, quantitative determination, and insight into the secondary structure of proteins and peptides. The aim of this study was to investigate the possibility of using Raman spectroscopy as a screening method for quantification of pharmaceutical products containing active substances with polypeptide structures. For that purpose, a model based on partial least square (PLS) analysis for quantitative determination of insulin aspart in solution was developed using Raman spectroscopy. The proposed model enables the establishment of a rapid approach for screening of the quality of formulations containing active substances with polypeptide structure, providing the selection of suspected samples that should be further analysed using routine techniques, which are time-consuming and costly.

**Key words:** Raman spectroscopy, polypeptides, insulin, partial least square (PLS) analysis

---

<https://doi.org/10.5937/arhfarm74-49901>

## Introduction

Active substances with protein and peptide structures comprise a very important group of compounds used in the development and production of pharmaceutical dosage forms. However, only a fraction of the dosage forms with market authorization contains this type of active substance. Compared to therapeutic proteins, peptides are composed of fewer than 50 amino acids (corresponding to a molecular weight of 500 Da to 5000 Da), implying that they are smaller biopharmaceuticals with simpler structures (1).

Peptides are a unique class of pharmaceuticals combining the attributes of biopharmaceuticals with the features of low molecular weight active substances. In terms of the complexity of their characterization, routine techniques for analyzing traditional low molecular weight active substances are insufficient for the clarification and characterization of the structure and behavior of peptide medicines (2). Pharmaceutical products containing peptide active substances can be characterized using a number of methods, including chromatographic (reversed-phase, normal-phase, ion-exchange, size-exclusion, hydrophilic interaction chromatography), electrophoretic (capillary electrophoresis, isoelectric focusing), spectroscopic (nuclear magnetic spectroscopy, UV-spectrophotometry, mass spectrometry, infrared spectroscopy, circular dichroism), and biological assays (animal-based, cell culture-based and biochemical), etc. (3, 4).

The development of novel medicinal products, including high-cost protein and peptide-based therapies, triggered the increase of the number of falsified products available on the illegal market, as well as in the legal medicines supply chain on a global level (5). The sophisticated nature of the falsification methods for biopharmaceutical products requires advanced analytical instruments for their identification. Considering that the impact on public health is significant because these medications are used to treat serious conditions such as diabetes, cancer, and infectious diseases, the market surveillance of these products requires the implementation of screening methods based on faster, simpler, and cheaper techniques (5, 6). Biological methods, as well as a majority of physicochemical techniques (chromatographic, electrophoretic) applied in quality control of these medicinal products are destructive, time-consuming, and require sample preparation. Vibrational spectroscopic techniques, including Raman spectroscopy and Fourier transform infrared spectroscopy (FTIR), are fast, powerful, and non-destructive techniques which, when combined with multivariate chemometric modelling, can provide specific identification, quantitative determination, and insight into the secondary structure of proteins and peptides (7–11).

The aim of this study was to investigate the possibility of using Raman spectroscopy as a screening method for quantification of pharmaceutical products containing active substances with polypeptide structures. For that purpose, samples containing insulin aspart as a characteristic example of therapeutic peptides were analyzed using Raman spectroscopy and a model for quantitative determination was developed based on partial least square (PLS) analysis.

## **Materials and methods**

### **Reagents and standard substances**

Insulin Aspart Secondary Reference Material (SRM), an aqueous solution containing: 602 nmol/mL insulin aspart, equivalent to 100 IU/mL (3.5 mg/mL), and Insulin Aspart standard substance (with P=95.8% on dried substance), were provided by the manufacturer with a suitable certificate of analysis. Highly purified water was obtained using a TKA-LAB Reinstwasser system (Niederelbert, Germany).

### **Preparation of samples for analysis**

Insulin Aspart SRM was used for preparation of both the calibration set of samples and prediction set of samples, with concentrations ranging from 40 IU/mL to 155.8 IU/mL. Samples with concentrations lower than 100% of the working concentration (100 IU/mL equivalent to 3.5 mg/mL) were prepared by dilution of Insulin Aspart SRM solution (3.5 mg/mL) with highly purified water (TKA-LAB Reinstwasser system, Niederelbert, Germany). Samples with a concentration higher than 100% of the working concentration were prepared by spiking Insulin Aspart SRM solution with suitable amounts of Insulin Aspart standard substance.

### **Raman spectroscopy**

Raman spectra were collected using a fiber optic probe from an ATR 3000 DH Portable Raman Spectrometer (Optosky, China) in the 2600 to 190  $\text{cm}^{-1}$  spectral region. The power of the laser (1064 nm) was set to 400 mW and the spectral resolution was 5  $\text{cm}^{-1}$ .

The integration and excitation time for the collection of the spectra for the PLS model was set to 90 seconds (180 seconds per measurement). The fiber optic probe was directed towards the sample, transferred in a transparent glass vial. For each solution, six independent spectra were recorded, using a new portion of the solution for each measurement.

GRAMS Spectroscopy Software was used for spectral smoothing and blank subtraction (Raman spectrum of purified water was used as blank). SIMCA® 14 software (Umetrics, Sweden) was used for pre-processing the spectra with the Standard normal variate (SNV) technique.

### **Statistical analysis**

SIMCA® 14 software (Umetrics, Sweden) was used for developing the chemometric model for quantification of insulin aspart based on partial least squares analysis (PLS).

## Results and discussion

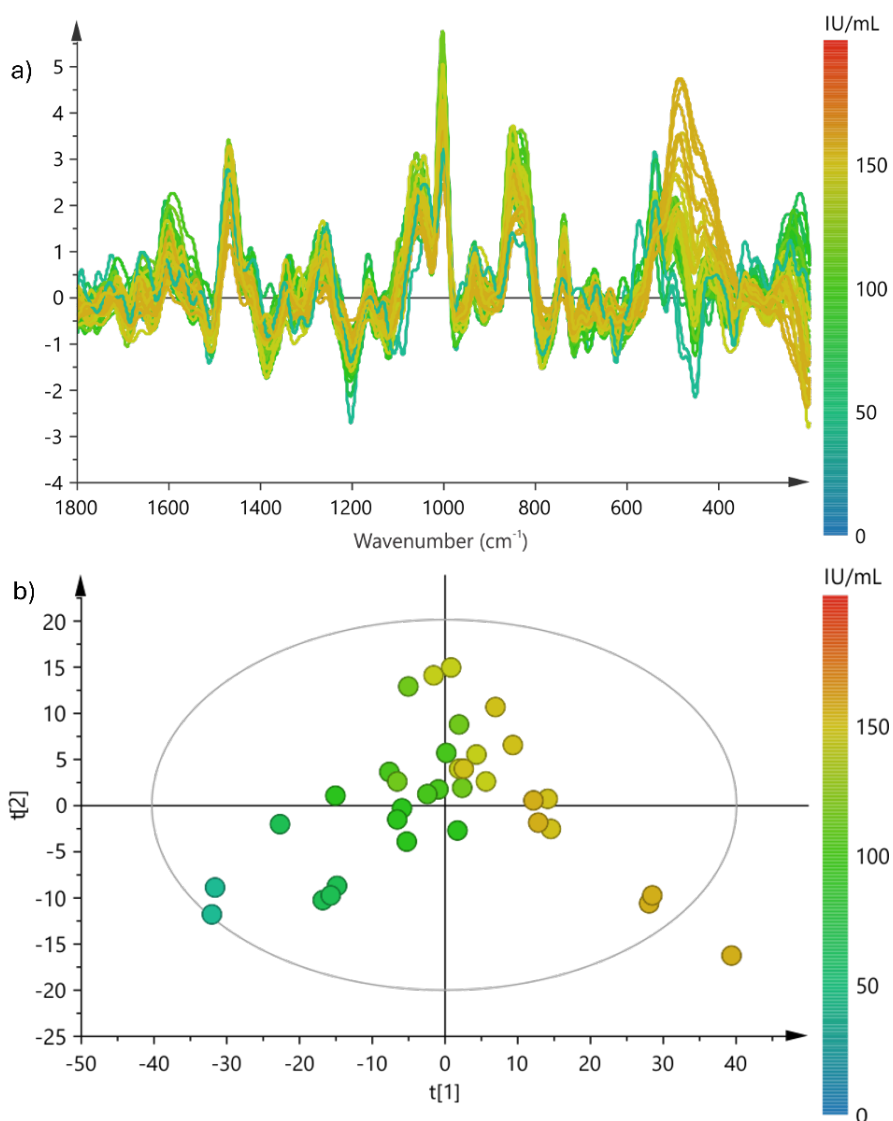
### Chemometric approach for quantification of insulin aspart

A calibration model for quantitative determination was developed using a PLS regression model based on the Raman spectra of insulin aspart in a solution. The concentration of insulin aspart (IU/mL) was used as a dependent variable (Y), while the corresponding Raman spectra (1346 X variables included) were designated as independent variables (X).

The calibration model was developed using the thirty-four different Raman spectra acquired from insulin aspart solutions with varying concentrations, whereas the prediction set was obtained using the remaining eight spectra. The overlaid spectra from the calibration and prediction set, coloured according to the insulin aspart concentration (40 IU/mL and 155.8 IU/mL) in the samples (Figure 1a) and the score scatter plot of the PLS model (Figure 1b) are presented in Figure 1. The statistical model was established within the spectral region of 1800–200  $\text{cm}^{-1}$ . The spectral region higher than 1800  $\text{cm}^{-1}$  was excluded from the model because it did not contain any characteristic bands originating from the analyte, while it additionally exhibited significant background scattering. The best-fit values for R<sup>2</sup><sub>Y</sub> (goodness of fit for Y), R<sup>2</sup><sub>X</sub> (goodness of fit for X) and Q<sup>2</sup> (the proportion of the total variation in the dependent variable that is predictable by the model) were obtained using smoothed and blank subtracted (GRAMS Spectroscopy Software), as well as SNV pre-processed spectra (SIMCA® 14 software), within the abovementioned spectral region in the model.

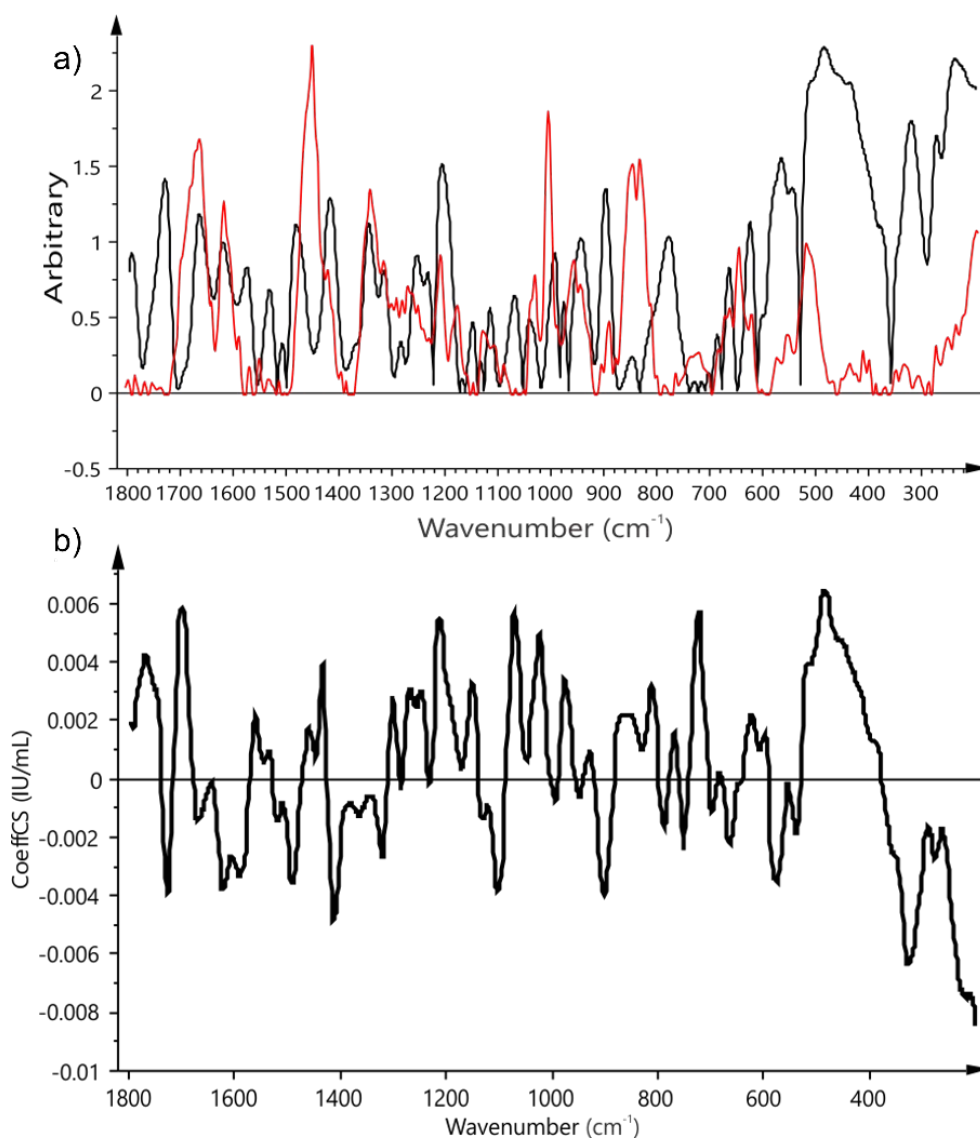
Four main components were employed for building the PLS quantification model (R<sup>2</sup><sub>X</sub>=0.495, R<sup>2</sup><sub>Y</sub>=0.969, Q<sup>2</sup>=0.767). The root mean square error of estimation (RMSEE) was 6.75 IU/mL. The model variable importance in projection (VIP) plot (Figure 2a, black line) was utilized for assigning the most prominent spectral bands with the highest importance in the regression model. A VIP score larger than 1 indicates that the corresponding predictor variable has a relatively high importance in the model (12). The bands at 1728, 1662, 1481, 1417, 1342, 1204, 897, 622, 563, 485, 320, 270 and 235  $\text{cm}^{-1}$ , assigned with the highest VIP factors, are the most relevant features of the spectral data which were used for quantification purposes. In the Raman spectrum of solid insulin aspart (Figure 2a, red line), several characteristic bands, deriving from both the polypeptide backbone and the amino acid side chain vibrations, were observed. Bands at 1728 and 1662  $\text{cm}^{-1}$  correspond to Amide I band (which arises from the C=O stretching vibration of the peptide bond), while the band at 1448  $\text{cm}^{-1}$  can be assigned to Amide II (linked to N-H bending and C-N stretching vibrations). The Amide III band (associated with a combination of N-H bending, C-N stretching, and N-H in-plane bending vibrations) was found at 1337  $\text{cm}^{-1}$ , whereas the disulfide bond band, which is of great importance for stabilizing the tertiary structure of insulin, appears at 563  $\text{cm}^{-1}$ . A strong band observed at 1003  $\text{cm}^{-1}$  can be assigned to the C-C stretching vibrations of the phenyl ring of the Phenylalanine (Phe) residue (9, 11, 13–15). The bands assigned with the highest VIP factors are complementary to the

characteristic spectral bands of insulin, thus confirming the model's capability in explaining the correlation among the variations in the Y (insulin concentration) and X (Raman spectra) data matrices. The coefficient plot (Figure 2b) confirms the previously mentioned findings, as the abovementioned spectral regions are assigned with the largest regression coefficients.



**Figure 1.** (a) Overlaid spectra from the calibration and prediction sets used in the statistical model for insulin aspart sample solutions, (b) score scatter plot of the PLS model. The spectra and score points are colored according to insulin aspart concentration (IU/mL)

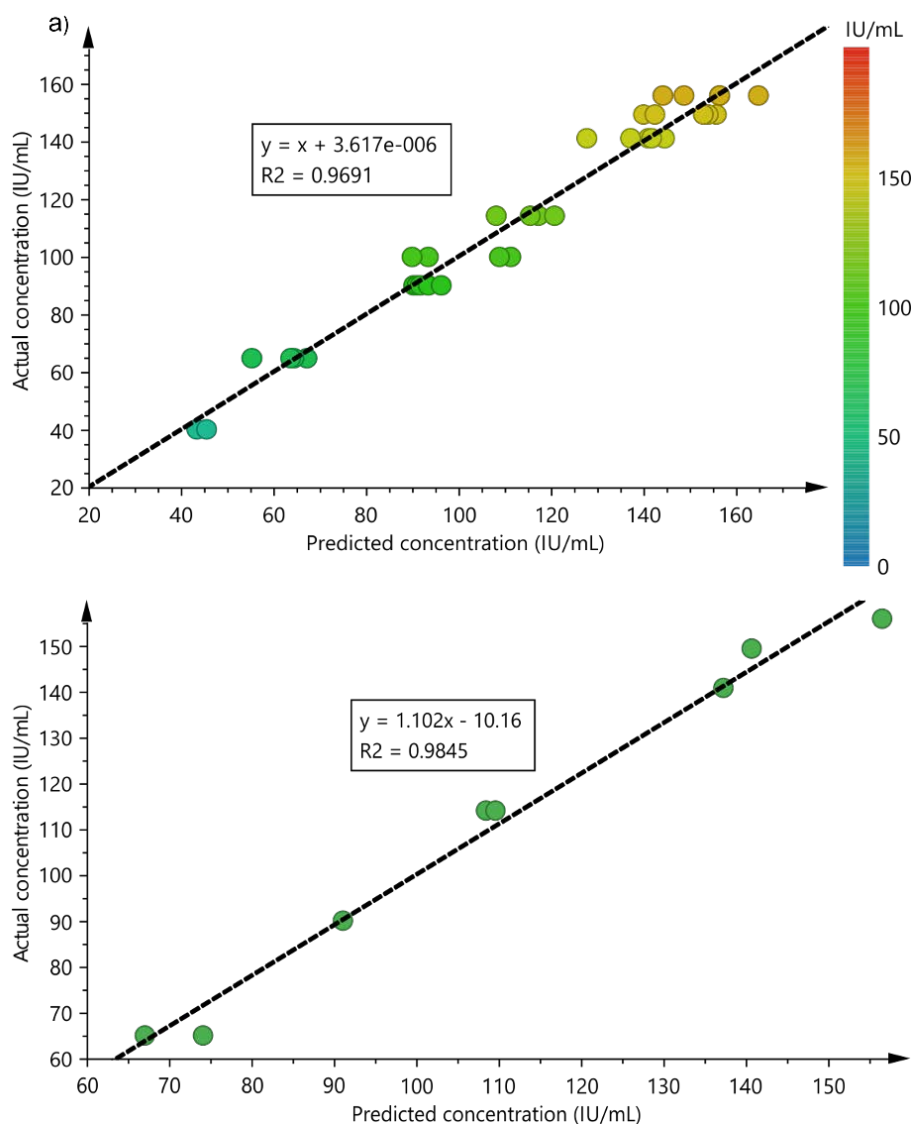
**Slika 1.** Preklapljeni spektri iz kalibracionog i prediktivnog skupa korišćeni u statističkom modelu za uzorke rastvora insulin asparta, (b) raspored tačaka rezultata modela PLS. Spektri i tačke rezultata obojeni su prema koncentraciji insulin asparta (IU/mL)



**Figure 2. PLS model statistical output: (a) VIP plot overlaid with the Raman spectrum of Insulin Aspart standard substance (red), (b) coefficients plot for the first main component of the model**

**Slika 2. Statistički rezultati PLS modela: (a) VIP grafikon preklopljen sa Ramanovim spektrom standardne supstance insulin asparta (crvena), (b) grafikon koeficijenata za prvu glavnu komponentu modela**

For estimation of the predictive capability of the model, a separate predictive set was used for determination of the root mean square error of prediction (RMSEP). The prediction set, composed of eight samples containing insulin aspart with concentrations varying from 65 IU/mL to 155.8 IU/mL, resulted in an RMSEP value of 7.21 IU/mL, indicating an acceptable predictive performance of the chemometric model (Figure 3).



**Figure 3. PLS model statistical output: (a) actual vs predicted plot for insulin aspart quantification, (b) actual versus predicted plot for the prediction set of the PLS model for calculation of insulin aspart**

**Slika 3. Statistički rezultati PLS modela: (a) grafikon stvarnih naspram predvidenih vrednosti za kvantifikaciju insulin asparta, (b) grafikon stvarnih naspram predvidenih vrednosti za skup predikcije PLS modela za izračunavanje insulin asparta**

The obtained value for RMSEP can be considered satisfactory, taking into account the relatively low concentration of insulin aspart (40 IU/mL to 155.5 IU/mL, corresponding to 1.4 mg/mL to 5.5 mg/mL), compared to similar models described in the literature data, where insulin concentrations in aqueous solutions used for analysis with Raman spectroscopy are higher (around 100 mg/mL), probably due to the weak Raman

scattering properties of the peptides in aqueous solution (9, 14). The main factors influencing the accuracy descriptors that should be taken into consideration are the sample solution preparation procedure (where water was used for dilution instead of the secondary reference material matrix, due to insufficient information about the declared composition), the complexity of the matrix, the limited number of samples and concentration range, as well as the relatively low concentration of insulin in the sample solutions.

## **Conclusion**

For the first time, a PLS regression model for quantitative determination of insulin aspart in a solution using Raman spectroscopy was developed. The VIP plots of the model indicated that the band regions of highest importance (1728, 1662, 1481, 1417, 1342, 1204, 897, 622, 563, 485, 320, 270 and 235  $\text{cm}^{-1}$ ) correspond with the insulin aspart characteristic bands. These band regions were used for predicting the concentration of the sample solutions in the developed PLS model. The sample solution preparation procedure, the complexity of the matrix and the relatively low concentration of insulin in the sample solutions were identified as the main factors influencing the accuracy descriptors of the model.

The proposed model enables the establishment of a rapid approach for screening of the quality of finished products sampled from the market, containing active substances with polypeptide structure. By providing a preliminary insight into the quality of medicines, Raman spectroscopy can enable the selection of suspected samples that should be further analysed using routine techniques which are time-consuming and costly, thus eliminating unnecessary analyses. This can be used to verify authenticity and identify potential falsified products and products with substandard quality.

## **Acknowledgements**

The authors would like to acknowledge the support of Center for Drug Quality Control, Ss. Cyril and Methodius University - Faculty of Pharmacy, Republic of North Macedonia.

## **Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## **CRediT authorship contribution statement**

**Hrisanta Godzo:** Investigation; Formal analysis; Writing-original draft. **Olga Gigopulu:** Investigation; Visualization; Statistical and data analysis. **Nikola Geskovski:** Conceptualization; Statistical and data analysis; Writing - review & editing. **Jelena Acevska:** Writing-Review & Editing. **Natalija Nakov:** Writing-Review & Editing.



**Jasmina Tonic-Ribarska:** Writing - Review & Editing. **Ana Poceva Panovska:** Writing - Review & Editing. **Katerina Brezovska:** Conceptualization; Supervision; Writing-Review & Editing.

## References

1. Wang L, Wang N, Zhang W, Cheng X, Yan Z, Shao G, et al. Therapeutic peptides: current applications and future directions. *Sig Transduct Target Ther*. 2022 Feb 14;7(1):1–27.
2. D’Addio SM, Bothe JR, Neri C, Walsh PL, Zhang J, Pierson E, et al. New and Evolving Techniques for the Characterization of Peptide Therapeutics. *JPharmSci*. 2016 Oct 1;105(10):2989–3006.
3. Vergote V, Burvenich C, Van de Wiele C, De Spiegeleer B. Quality specifications for peptide drugs: a regulatory-pharmaceutical approach. *J Pept Sci*. 2009;15(11):697–710.
4. Rastogi S, Shukla S, Kalaivani M, Singh GN. Peptide-based therapeutics: quality specifications, regulatory considerations, and prospects. *Drug Discov Today*. 2019 Jan 1;24(1):148–62.
5. Janvier S, De Spiegeleer B, Vanhee C, Deconinck E. Falsification of biotechnology drugs: current dangers and/or future disasters? *J Pharm Biomed Anal*. 2018 Nov 30;161:175–91.
6. Bakker-’t Hart IME, Ohana D, Venhuis BJ. Current challenges in the detection and analysis of falsified medicines. *J Pharm Biomed Anal*. 2021 Apr 15;197:113948.
7. Yang S, Zhang Q, Yang H, Shi H, Dong A, Wang L, et al. Progress in infrared spectroscopy as an efficient tool for predicting protein secondary structure. *Int J Biol Macromol*. 2022 May 1;206:175–87.
8. Delbeck S, Heise HM. Quality Assurance of Commercial Insulin Formulations: Novel Assay Using Infrared Spectroscopy. *J Diabetes Sci Technol*. 2021 Jul 1;15(4):865–73.
9. Ortiz C, Zhang D, Xie Y, Davisson VJ, Ben-Amotz D. Identification of insulin variants using Raman spectroscopy. *Anal Biochem*. 2004 Sep 15;332(2):245–52.
10. Jesus JIS da S de, Löbenberg R, Bou-Chacra NA. Raman Spectroscopy for Quantitative Analysis in the Pharmaceutical Industry. *J Pharm Pharm Sci*. 2020 Mar 2;23:24–46.
11. Durickovic I. Using Raman Spectroscopy for Characterization of Aqueous Media and Quantification of Species in Aqueous Solution. In: *Applications of Molecular Spectroscopy to Current Research in the Chemical and Biological Sciences*. IntechOpen. 2016. doi: 10.5772/64550.
12. Geskovski N, Stefkov G, Gigopulu O, Stefov S, Huck CW, Makreski P. Mid-infrared spectroscopy as process analytical technology tool for estimation of THC and CBD content in Cannabis flowers and extracts. *Spectrochim Acta A Mol Biomol Spectrosc*. 2021 Apr 15;251:119422.
13. Tuma R. Raman spectroscopy of proteins: from peptides to large assemblies. *J Raman Spectrosc*. 2005;36(4):307–19.
14. Ishigaki M, Morimoto K, Chatani E, Ozaki Y. Exploration of Insulin Amyloid Polymorphism Using Raman Spectroscopy and Imaging. *Biophys J*. 2020 Jun 16;118(12):2997–3007.
15. Dolui S, Roy A, Pal U, Kundu S, Pandit E, N Ratha B, et al. Raman Spectroscopic Insights of Phase-Separated Insulin Aggregates. *ACS Phys Chem Au*. 2024 Feb 15. doi: 10.1021/acspchemau.3c00065.

# **Kvantitativno određivanje insulin asparta u rastvoru pomoću Ramanove spektroskopije zasnovano na PLS metodi**

**Hrisanta Godzo<sup>1\*</sup>, Olga Gigopulu<sup>1</sup>, Nikola Geshkovski<sup>2</sup>,  
Jelena Acevska<sup>1</sup>, Natalija Nakov<sup>1</sup>, Jasmina Tonic-Ribarska<sup>1</sup>,  
Ana Poceva Panovska<sup>1</sup>, Katerina Brezovska<sup>1</sup>**

<sup>1</sup>Ss. Cyril and Methodius University - Faculty of Pharmacy, Institute of Applied Chemistry and Pharmaceutical Analysis, Majka Tereza 47, 1000 Skopje, Severna Makedonija

<sup>2</sup>Ss. Cyril and Methodius University - Faculty of Pharmacy, Institute of Pharmaceutical Technology, Majka Tereza 47, 1000 Skopje, Severna Makedonija

\*Autor za korespondenciju: Hrisanta Godzo, e-mail: hrisanta@ff.ukim.edu.mk

---

## **Kratak sadržaj**

Kompleksna struktura lekova koji sadrže aktivne supstance polipeptidnog porekla iziskuje primenu zahtevnih analitičkih pristupa, zasnovanih na fizičko-hemijskim metodama, a u slučajevima kada je to neophodno i bioloških testova za kontrolu kvaliteta, kao i za detekciju nekvalitetnih i falsifikovanih proizvoda. Vibracione spektroskopske tehnike, uključujući Ramanovu spektroskopiju, brze su, moćne i nedestruktivne tehnike koje, kada se kombinuju sa multivarijantnim hemometrijskim modelovanjem, mogu da obezbede preciznu identifikaciju, kvantitativno određivanje i uvid u sekundarnu strukturu proteina i peptida. Cilj ovog istraživanja bio je da se ispita mogućnost korišćenja Ramanove spektroskopije kao skrining metode za kvantifikaciju farmaceutskih proizvoda koji sadrže aktivne supstance sa polipeptidnom strukturom. U te svrhe razvijen je model zasnovan na analizi parcijalnih najmanjih kvadrata (PLS) za kvantitativno određivanje insulin asparta u rastvoru pomoću Ramanove spektroskopije. Predloženi model omogućava uspostavljanje brzog pristupa za skrining kvaliteta formulacija koje sadrže aktivne supstance sa polipeptidnom strukturom, obezbeđujući tako selekciju sumnjivih uzoraka koje bi trebalo dalje analizirati upotrebom rutinskih tehnika koje su dugotrajne i skupe.

**Ključne reči:** Ramanova spektroskopija, polipeptidi, insulin, analiza parcijalnih najmanjih kvadrata (PLS)

---