Analysis of infrared spectra of blood serum of patients with multiple myeloma

D A Chernyshev¹, E S Mikhailets¹, E A Telnaya¹, L V Plotnikova¹, A D Garifullin², A Y Kuvshinov³, S V Voloshin²,³,⁴ and A M Polyanichko¹,⁵

¹St. Petersburg State University, Department of Physics, St. Petersburg, 198504 St. Ulyanovskaya, 3, Russia
²Russian Scientific Research Institute of Hematology and Transfusiology, 16, 2nd Sovetskaya str., Saint Petersburg, 191024, Russia
³Military Medical Academy. CM. Kirova, St. Petersburg, 194044, ul. Academician Lebedev, 6, Russia
⁴Federal State Budgetary Educational Establishment of the North-Western State Medical University named after I.I. Mechnikov, St. Petersburg, 191015, ul. Kirochnaya, 4, Russia
⁵Institute of Cytology RAS, 4 Tikhoretsky Ave., St. Petersburg, Russia, 194064, Russia

st063380@student.spbu.ru, a.polyanichko@spbu.ru

Abstract. Multiple myeloma (MM) is a serious disease that is difficult to diagnose especially at early stage. Infrared spectroscopy is a promising approach for diagnosing MM. The principal component analysis (PCA) allows us to reduce the dimension of the data and keep only the important variables. In this study, we apply principal components analysis to infrared (IR) spectra of blood serum from healthy donors and multiple myeloma patients. As a result of the analysis by PCA, it was possible to visualize the separation of patient’s and donor’s samples into two clusters. The result indicates that this method is potentially applicable for diagnosis of multiple myeloma.

1. Introduction

Multiple myeloma is a serious oncohematological disease that is difficult to diagnose especially at early stage. MM accounts for approximately 10% of all hematological cancers [1], and 1% of all oncological diseases. MM is characterized by the formation in the bone marrow of clusters of malignant plasma cells that produce and secrete monoclonal immunoglobulins (M-proteins) or their fragments (k/λ light chains). In addition, the level of albumin in the blood is usually reduced. As a result of changes in the ratio between albumin and immunoglobulins, the ratio between different types of secondary structures in the serum may also change [1, 2]. There is also a rare type of myeloma characterized by strongly reduced production of M-protein or k/λ light chains [3]. Infrared spectroscopy is proved to be very promising approach for early cancer diagnostics. By considering the spectra of blood serum, a complex picture of maxima, located on different wavenumbers, can be observed. The maxima that are localized in the range of 1700-1350 cm⁻¹ correspond to the
superposition of vibrations sensitive to the secondary structures of proteins. Since blood contains many proteins, a detailed analysis of these bands might be helpful for medical diagnostics. Development of MM is accompanied by the increasing amount of immunoglobulins. These changes can be observed in IR spectra [4].

The aim of this study was to analyze the IR spectra using the principal component analysis. Blood samples from patients with MM and from healthy donors were analyzed. PCA has proven to be a good method for solving chemometrics problems [5]. A similar approach was successfully applied earlier in diagnostic of breast cancer [6].

2. Materials and methods

2.1. Serum samples.
In this study we analyzed infrared spectra of serum samples from patients with multiple myeloma, which are under the supervision of the hematology clinic of the Russian Research Institute of Hematology and Transfusionology (St. Petersburg, Russia), where blood serum samples from MM patients and healthy donors were obtained, as well as their primary analysis was performed. In this study samples from 25 healthy donors and 13 patients with multiple myeloma were analyzed. There were 6 male and 7 female patients. All patients were in the age of 50-70 years.

Serum samples were obtained using standard clinical protocol [7]. S-Monovette (Sarstedt, Germany) clotting activator was used to obtain serum samples. Collected blood samples were left in tubes for 20 - 30 min at room temperature (18 - 24 °C), after which it was centrifuged for 15 min at 3000 rpm (1000 g) using a Heraeus Labofuge 200 centrifuge (Thermo Scientific, USA). Before physical and chemical tests, the samples were frozen and stored at a temperature of –30 °C.

2.2. IR-spectroscopy.
The absorption spectra in the mid-infrared region (4000 — 400 cm⁻¹) were obtained using Tensor 27 FTIR spectrometer (Bruker). The samples were studied in D2O solutions using demountable BaF2 cells with an optical pathlength of 50 μm. The primary processing and analysis of the spectra was carried out using the software supplied with the device. The spectra of each sample were recorded with a resolution of 2 cm⁻¹ and averaged by 128 accumulations.

2.3. Principal component analysis.
In this study, the obtained spectra were analyzed by principal component analysis. The PCA is a method for data processing that converts complex large dimensions data into a lower dimensional space that can well represent the original data. PCA transforms the original data into a number of uncorrelated variables. New variables can be viewed as points projected from original data onto a lower-dimensional space. They are aimed to maximize variance in the original dataset. New variables carry information about original dataset, and the PC1 will carry the largest contribution to the variance, the PC2 will carry smaller contribution, etc. [6, 10].

In this study, the PCA was implemented using the singular value decomposition. The original X matrix was constructed from intensities of the serum samples. The spectra were placed in rows of the matrix, columns were corresponded to wavenumbers. Then the entire matrix was normalized. After normalization X matrix was decomposed by the singular value method to be represented as X=USVᵀ, where U is the matrix consisting of eigenvectors of the matrix XXᵀ, S is a diagonal matrix with elements equal to the root of the positive eigenvalues of XXᵀ or XᵀX, V - is the matrix consisting of eigenvectors of the matrix XᵀX. Columns U constitute the orthogonal basis of X matrix columns. Columns V constitute the orthogonal basis of X matrix rows. The PC are represented as columns of the matrix U, where the first column corresponds to PC1, the second corresponds to PC2, etc [10]. Plots will be built using principal components further.
3. Results and discussion

Typical mid-FTIR spectra of serum samples obtained from multiple myeloma patients and healthy donors are shown in figure 1. This region contains amide absorption bands, which correspond to the vibrations of C = O (Amide I), side chains and N-D bonds (Amide II) [9].

![Absorption spectra of blood serum from healthy donors and from patients with secretory and non-secretory multiple myeloma at 1750-1300 cm⁻¹. Here are represented Amide I band (1), side chains vibrations band (2) and Amide II band (3) respectively.](image)

Figure 1. Absorption spectra of blood serum from healthy donors and from patients with secretory and non-secretory multiple myeloma at 1750-1300 cm⁻¹. Here are represented Amide I band (1), side chains vibrations band (2) and Amide II band (3) respectively.

Figure 1 shows us spectra obtained from donor and MM patients. There are some differences between spectra, for example, the shape of the Amide I band and position of maximum of this band differs between donors, secretory and non-secretory patients. These differences can be explained by the fact that samples have differences in the secondary structure of proteins.

For further analysis of spectra, the principal component analysis was chosen. Spectra were analyzed in the range 1900-1300 cm⁻¹ and 1900-1500 cm⁻¹. Graphs were plotted using pairs of U-matrix columns, which contribute major variance of original data. In the spectral range of 1900-1300 cm⁻¹ PC1 explained 66% of the total data variance, PC2 explained 32%, PC3 explained 1.5%. The graph for PC1-PC2 is shown in figure 2.
Figure 2. Graph in coordinates PC1-PC2 for the range 1900-1300 cm\(^{-1}\). Samples with myeloma are highlighted in red, donor samples are in green.

This range of wavenumbers was chosen because it contains Amide vibrations and side chains vibrations. It can be noticed that in figure 2 the PCA plot in PC1-PC2 coordinates visualizes the separation of serum into two groups. The donor’s group mostly corresponds to negative values of PC2, while MM patients correspond to positive. PC1, which contributes the most to total variance, does not contribute significantly to dividing samples into two clusters. Patients are located at a positive area of PC1, while donors are located on both the positive area and the negative area.

Then the 1900–1500 cm\(^{-1}\) range was analyzed, which does not contain the Amid III band. In this case, PC1 explained 54% of the total variance in the data, PC2 explained 44%, and PC3 explained 1.5%. Graph for this case is shown in figure 3.

Figure 3. Graph in coordinates PC1-PC2 for the range 1900-1500 cm\(^{-1}\). Samples with myeloma are highlighted in red, donor samples are in green.
Figure 3 also visualizes the division of blood serum into two clusters of points when PC1-PC2 are used as x and y axes respectively. Moreover, fewer points corresponding to donors fell into a positive area of PC2 in this case. Also, several points corresponding to MM patients turned out to be higher in PC2 axis than they were in figure 2. And PC1 still does not contribute too much to clustering.

Results obtained from the analysis suggest the applicability of PCA to IR spectra for the diagnosis of multiple myeloma. In both cases that were considered the greatest contribution to clustering was made by PC2. In addition, results that were obtained from processing of two spectral ranges by PCA allowed us to conclude that the Amide II band could be excluded from the analysis of the spectrum. However, further research on this issue and more statistical data are required.

4. Conclusion
Thus, the usage of infrared spectroscopy in conjunction with spectral analysis by PCA can help to detect multiple myeloma. This approach is quite promising for the diagnosis of multiple myeloma since spectra of MM patients and healthy people are different in both secretory and non-secretory cases that can be analyzed by PCA. However, further research in this direction is required.

Acknowledgments
The work was carried out using the equipment of the resource centers of the Research Park of St. Petersburg State University ("Centre for Optical and Laser Materials Research", "Centre for Diagnostics of Functional Materials for Medicine, Pharmacology and Nanoelectronics", "Cryogenic Department").

References
[1] Rajkumar S V and Kumar S 2016 Multiple Myeloma: Diagnosis and Treatment Mayo Clinic Proceedings 91 101–19
[2] Palumbo A and Anderson K 2011 Multiple Myeloma New England Journal of Medicine 364 1046-60
[3] Ryzhko V V, Klodzinsky A A, Varlamova E Yu, Sataeva M S, Kaplansky I B, Nakastoev I M and Paraskevova O V 2010 "Nonsecretory" multiple myeloma (a review of literature and own data) Clinical Oncohematology 3 270-7
[4] Plotnikova L V, Polyanichko A M, Uspenskaya M V, Garifullin A D, Voloshin S V 2017 Comparative analysis of FTIR spectra of serum from multiple myeloma patients and healthy donors Vestnik SPbSU. Physics and Chemistry 4 34-40
[5] Kramer R 1998 Chemometric Techniques for Quantitative Analysis (Boca Raton: CRC Press) p 1-2
[6] Sitnikova V E, Kotkova M A, Nosenko T N, Kotkova T N, Martynova D M and Uspenskaya M V 2020 Breast cancer detection by ATR-FTIR spectroscopy of blood serum and multivariate data-analysis Talanta 214 1-8
[7] Telnaya E A, Plotnikova L V, Garifullin A D, Kuvshinov A Yu, Voloshin S V and Polyanichko A M 2020 Infrared Spectroscopy of Blood Serum from Patients with Oncohematological Diseases. Biophysics 65 981-6
[8] Polyanichko A M, Andrushchenko V V, Chikhirzhina E V, Vorob'ev V I and Wieser H 2004 The effect of manganese (II) on DNA structure: electronic and vibrational circular dichroism studies Nucleic Acids Research 32 989–96
[9] Plotnikova L V, Kobeleva M O, Borisov E V, Garifullin A D, Povolotskaya A V, Voloshin S V and Polyanichko A.M. 2019 Infrared Spectroscopy of Blood Serum from Patients with Multiple Myeloma Cell and Tissue Biology 13 130-5
[10] Hsu Y L, Huang P Y and Chen D T 2014 Sparse principal component analysis in cancer research Transl. Cancer Res. 3 182-90