Chemometric classification of low-resolution spectra of organic iron derivatives

Ernest K. PUJSZO, Jacek A. SZYMURA, Jan LAMKIEWICZ - Faculty of Chemical Technology and Engineering, University of Technology and Life Sciences, Bydgoszcz; Irena SZYMURA - University of Warmia and Mazury, Olsztyn

Please cited as: CHEMIK 2011, **65,** 4, 267-272

Introduction

Mass spectrometry (MS) is an instrumental technique which enables identification of unknown substance or determination its structure from mass spectrum. Identification of any chemical compound, for which mass spectrum can be found in popular data bases (e.g. NIST [1] or Wiley [2]), is an easy and routine task. It is usually realized by means of library search techniques [3, 4] depending on computer-aided comparison of investigated spectrum with those gathered in available MS data bases. When spectrum of unknown was not included into available collections of reference spectra its analysis can be performed by means of chemometric methods applying either isotopic profiles [5, 6] (in case of organometallics) or spectral classifiers [7, 8] (in case of typical organic compounds, i.e., containing C, H, O, N atoms in molecules). The latter represents mathematical algorithms based on multivariate methods, such as principal component analysis (PCA) [9, 10]. The primary goal of spectra classification is to find correlation between the properties of chemical compounds and their mass spectra. The results of classification reflect structural features that are determined by fragmentation ions appearing in a mass spectrum as characteristic bands.

The main idea of PCA is to reduce the dimensionality of a data set in which there are a large number of interrelated variables, while preserving as much as possible of the variance present in the data set. In relation to mass spectrometry, the data set consists of mass spectra of different compounds in which the intensities of individual m/z ratios are the original variables. PCA also serves as a very good visualization tool, which causes the relationship and clusters become more apparent.

Regarding to the classification of mass spectra, up to now PCA has been used exclusively for typical organic compounds but never for organometallics. It is well known that spectra of the latter substances in contrast to those of the former contain a numerous signals originating from rich isotopic profiles of metals. An exception from this rule is isotopic profile of iron, which indicates a striking similarity to that of carbon. In mass spectra of both elements there is a predominant signal of the main isotope (^{56}Fe or ^{12}C, respectively) exhibiting intensity over 90%. In view of all above facts the authors of present work decided to check, whether classification methods elaborated and used for low-resolution mass spectra of typical organic compounds can also be effective for such spectra of organometallics. For this purpose a set

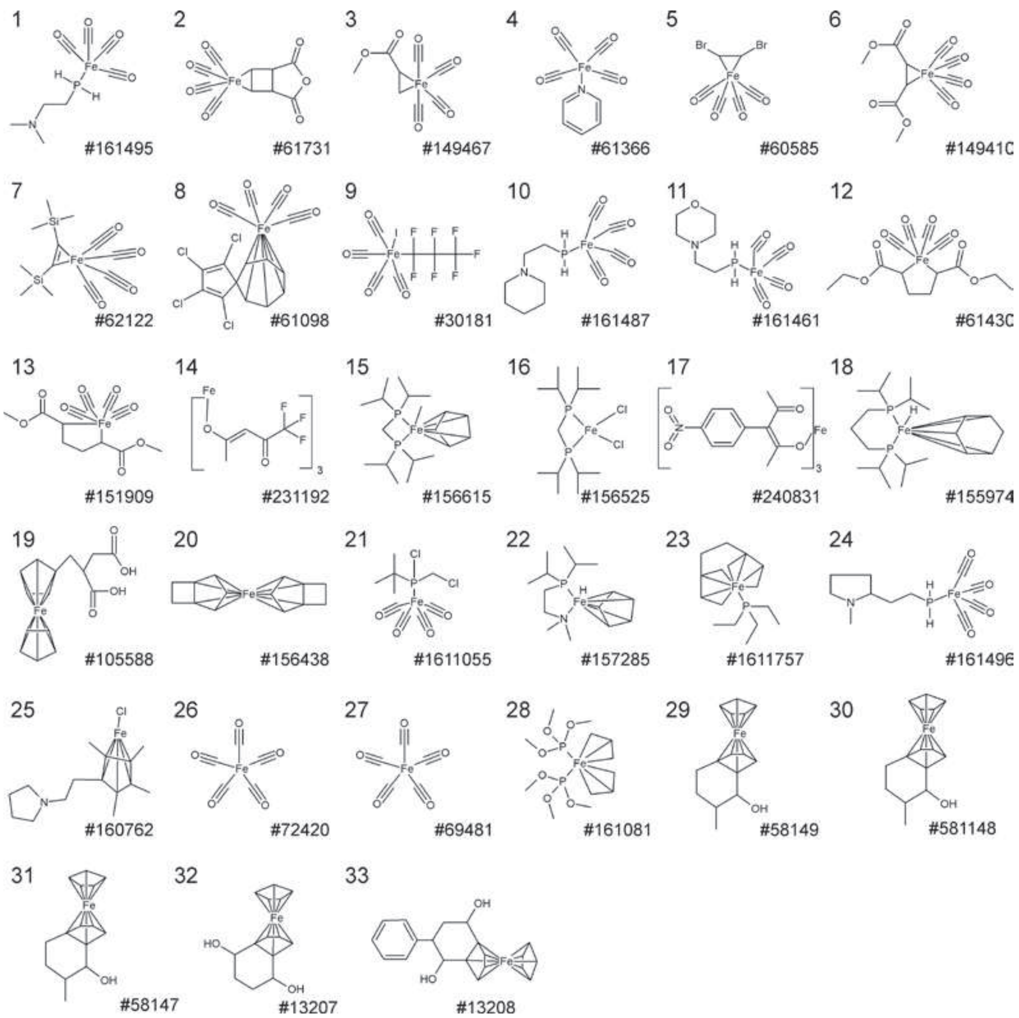

Fig. 1. Structural formulas, NIST [1] numbers (#) and serial numbers of iron organic derivatives mass spectra analyzed in this work

of mass spectra representing iron organic (carbonyl and non-carbonyl) derivatives and two computer programs MassFeatGen [13] and R [14] being chemometric tools for classification analysis were used.

Mass spectra and methods

For present study 33 low-resolution (obtained by Electron Impact) spectra of iron organic derivatives were taken from NIST 2008 database [1]. 17 of them containing at least 4 carbonyls in molecule were treated as the research set, while the rest 16 characterizing by lack of such substituent were served as the reference set. Figure 1 summarizes their serial numbers, structural formulas and numbers from NIST database.

At the beginning of chemometric analysis all selected spectra were written to the text files, which contained spectral data (m/z positions of signals and their intensities) as well as basic data identifying each spectrum, such as name of the compound, its mass, summary formula, etc. From them a one file compatible with JCAMP-DX format [15] had been created, which was next introduced to MassFeatGen program [13]. Among many spectral features included in this program two of them, namely: modulo (MD) and autocorrelation (AC), were chosen by the authors for present study. In parameters of both features necessary modifications were done, since their values had been set originally for spectra of typical organic compounds. For studying of iron organometallics spectra the value of difference between two m/z signals in spectrum was changed from 14 (referring to loss of CH_2 fragment) to 28 (expressing loss of one CO). It had been assumed that such modification made possible a detection of carbonyl group in the structures of those compounds represented by spectra investigated. The results obtained from MassFeatGen had been finally subjected to PCA, which was performed by means of computer program R [14].

Results and discussion

Figure 2 presents the results of PCA into the spaces PC_1 versus PC_2, obtained for 33 mass spectra of iron organic derivatives. Their division on carbonyl containing compounds and those without carbonyls is manifested by the solid line separating respective areas I and II in the plot. One can notice that spectrum of the non-carbonyl compound with serial number 25 is the only one exception from this rule, because of its location in the inappropriate class I.

Such behavior can be explained basing on chemical properties of its molecule $C_{15}H_{24}ClFeN$. As a consequence of phenomena occurring in ionization chamber of mass spectrometer, this molecule formed $C_5H_{10}N^+$ fragmentation ion having mass 84, which is the multiple of the value 28 corresponding to the mass of CO. High intensity of m/z=84 is responsible for its erroneous inclusion to the class I. According to Biemann nomenclature described in work of McLafferty and Tureček [16], this ion originating during fragmentation represents in mass spectrometry so called cleavage of B type. In case of the compound 25, such process starts with electron knock-out from the free electron pair located on nitrogen atom of its molecular ion and leads finally to the formation of $C_5H_{10}N^+$ ion stabilized by resonance (Fig. 3).

It is worth to notice that for present studies also the replicas of the spectra for the same compound recorded by various investigators (e.g., 29 and 27) and those recorded by the same author (e.g., 29, 30 and 31) were considered. The visual evaluation of the data in Figure 2 indicates that the positions representing those spectra are located both in close vicinity (26 and 27) as well as in certain distance to each other (29 and 30). The latter observation is a consequence of significant differences between the intensities of the corresponding m/z signals of various mass spectra of the same compound. Such differences do not disturb classification done in this work, however they inform that correctness of the spectrum of the iron organic derivative 30 should be verified. For such purpose the methods of mass spectra analysis applying isotopic profiles [5, 6] could be very suitable.

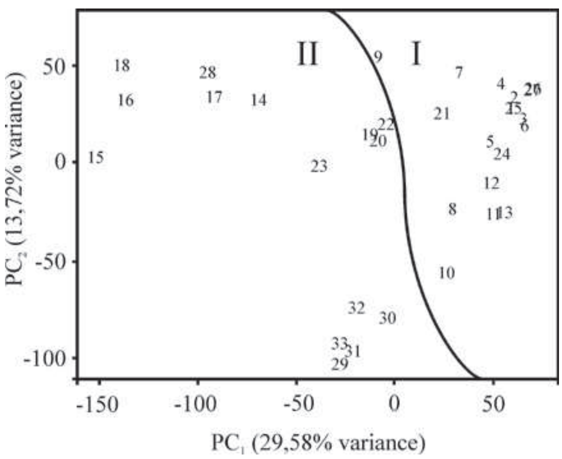

Fig. 2. PCA plot of 33 mass spectra of iron organic derivatives analyzed in this work, showing their division on two classes: I[st] – gathering compounds containing 4 or more carbonyls in molecules and II[nd] – representing non-carbonyl compounds

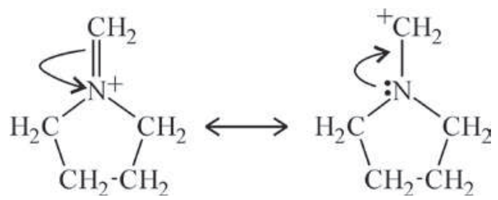

Fig. 3. Resonance structure of even-electron ion $C_5H_{10}N^+$ formed in fragmentation of molecular ion of the compound 25 (Fig. 1)

Conclusions

The classification performed in this work by chemometric methods routinely used for mass spectra of typical organic compounds enabled division of all investigation spectra of iron organic derivatives on two classes representing: carbonyl compounds containing at least 4 CO groups (I[st]) and non-carbonyl compounds (II[nd]). Therefore it can be concluded that the correlation between the compound structure and its mass spectrum was found for iron organometallics. It was also shown that spectrum quality can have the influence on the correctness of spectra classification performed by means of chemometric multivariate methods.

Acknowledgements

The authors thank professor Kurt Varmuza of Vienna University of Technology for providing the MassFeatGen program.

Literature

1. NIST/EPA/NIH Mass Spectral Library. National Institute of Standards and Technology and Advanced Chemistry Development Washington, DC, 2008.
2. Wiley Registry™ of Mass Spectral Data 9th Edition, John Wiley & Sons, Inc. 2009.
3. Stein S.E., Scott D.R.: *Optimization and testing of mass spectral library search algorithms for compound identification.* Journal of American Society for Mass Spectrometry 1994, **5**, 859-866.
4. McLafferty F.W., Stauffer D.A., Loh S.Y., Wesdemiotis C.: *Unknown identification using reference mass spectra. Quality evaluation of databases*. American Society for Mass Spectrometry 1999, **10**, 1229-1240.
5. Szymura J.A., Lamkiewicz J.: *Band composition analysis: a new procedure for deconvolution of the mass spectra of organometallic compounds*. Journal of Mass Spectrometry 2003, **38**, 817-822.
6. Meija J.: *Mathematical tools in analytical mass spectrometry*. Analytical and Bioanalytical Chemistry 2006, **385**, 486-499.
7. Varmuza K., Werther W.: *Mass spectral classifiers for supporting systematic structure elucidation*. Journal of Chemical Information and Computer Sciences 1996, **36**, 323-333.
8. Schymanski E.L., Meinert C., Meringer M., Brack W.: *The use of MS classifiers and structure generation to assist in the identification of unknowns in effect-directed analysis.* Analytica Chimica Acta 2008, **615** (No. 2), 136-147.
9. Varmuza K., Filzmoser P.: *Introduction to Multivariate Statistical Analysis in Chemometrics.* Taylor & Francis - CRC Press, Boca Raton, FL, 2009.

10. Mazerski J.: *Podstawy chemometrii*. Wydawnictwo Politechniki Gdańskiej, Gdańsk 2000.
11. Aruga R., Mirti P., A. Casoli A., Palla G.: *Classification of ancient proteinaceous painting media by the joint use of pattern recognition and factor analysis on GC/MS data*. Fresenius' Journal of Analytical Chemistry 1999, **365**, 559-566.
12. Varmuza K.: *Recognition of relationships between mass spectral data and chemical structures by multivariate data analysis*. Analytical Sciences 2001, **17** (supplement), i467-i470.
13. Demuth W., Varmuza K.: *MassFeatGen: Software for generation of mass spectral features.* Version 107-3, Laboratory for Chemometrics, Vienna University of Technology, www.lcm.tuwien.ac.at, Vienna, Austria 2005.
14. *The R project for statistical computing, version 2.9.2. The R Foundation for Statistical Computing*. www.r-project.org.
15. Lampen P., Lambert J., Lancashire R.J., McDonald R.S., McIntyre P.S. Rutledge D.N., Fröhlich T., Davies A.N.: *An extension to the JCAMP-DX standard file format, JCAMP-DX V.5.01.* Pure and Applied Chemistry 1999, **71** (No. 8), 1549-1556.
16. F.W. McLafferty, F. Tureček: *Interpretation of mass spectra, 4th ed*. University Science Books, Mill Valley, CA, 1993.

Translation into English by the Author

Ernest K. PUJSZO, M.Sc., Eng., is a graduate of the University of Technology and Agriculture (at present University of Technology and Life Sciences) in Bydgoszcz, Faculty of Technology and Chemical Engineering (2006). At present he is Ph.D. student in Faculty of Chemical Technology of Poznań University of Technology. Scientific interests: chemometrics, computer-aided methods of mass spectra interpretation, analytics in environmental protection.

Jacek A. SZYMURA, (Sc.D.), Professor of University of Technology and Life Sciences in Bydgoszcz, is a graduate of the University of Nicolaus Copernicus in Toruń, Faculty of Mathematic-Physics-Chemistry (1972). He acquired his Ph.D. and Sc.D. degrees at the Technical University of Łódź in 1979 and 1989, respectively. Currently he is a head of Department of Inorganic Chemistry in Faculty of Technology and Chemical Engineering at the University of Technology and Life Sciences in Bydgoszcz. Scientific interests: catalysis, mass spectrometry, chemometrics, computer-aided methods in chemistry.

Jan LAMKIEWICZ, Ph.D.(Eng.), is a graduate of the University of Technology and Agriculture (at present University of Technology and Life Sciences) in Bydgoszcz, Faculty of Technology and Chemical Engineering (1999). In 2006 he acquired Ph.D. degree at the Technical University of Łódź. Currently, he is assistant professor of Department of Inorganic Chemistry in Faculty of Technology and Chemical Engineering at the University of Technology and Life Sciences in Bydgoszcz. Scientific interests: mass spectrometry, chemometrics, computer-aided methods of mass spectra interpretation.

Irena SZYMURA, Ph.D., is a graduate of the University of Nicolaus Copernicus in Toruń, Faculty of Mathematic-Physics-Chemistry (1974). She acquired her Ph.D. degree in 1982 at the University of Technology and Agriculture (at present University of Technology and Life Sciences) in Bydgoszcz. Currently she is an assistant professor of Department of Chemistry at the University of Warmia and Mazury in Olsztyn. Scientific interests: analytics in environmental protection, physical chemistry of soil.

Events IYC'2011

July 2011

6th International Symposium on Macrocyclic and Supramolecular Chemistry (6-ISMSC) - Jul 03 - Jul 07, 2011 - University of Sussex, Brighton, UK

Gold: Faraday Discussion 152 - Jul 04 - Jul 06, 2011 - Cardiff, UK

10th International Conference on Materials Chemistry (MC10) - Jul 04 - Jul 07, 2011 - University of Manchester, UK

Challenges in Renewable Energy International Symposia on Advancing the Chemical Sciences - Jul 05 - Jul 08, 2011 - MIT, Boston, USA

Giftedness, Creativity & development - Jul 06 - Jul 09, 2011 - İstanbul University, İstanbul - Türkiye

22nd International Symposium: Synthesis in Organic Chemistry - Jul 11 - Jul 14, 2011 - Churchill College, University of Cambridge, UK

Chem Ed 2011 conference for chemistry educators - Jul 24 - Jul 28, 2011 - Western Michigan University in Kalamazoo, Michigan, USA

Coherence and Control in Chemistry: Faraday Discussion 153 - Jul 25 - Jul 27, 2011 - University of Leeds, UK

Analytical Research Forum (ARF) - Jul 25 - Jul 27, 2011 - University of Manchester, UK

Challenges in Chemical Biology International Symposia on Advancing the Chemical Sciences - Jul 26 - Jul 29, 2011 - Manchester, UK

"Science is for Everybody" Australian National Chemistry Quiz - Jul 28, 2011 - Australia

Children's Chemical Experiment Show - Jul 29 - Jul 31, 2011 - The National Museum of Emerging Science and Innovation Hall, Odaiba, Tokyo

IUPAC World Chemistry Congress - Chemistry Bridging Innovation among the Americas and the World - Jul 30 - Aug 07, 2011 - San Juan, Puerto Rico

August 2011

World Chemistry Leadership Meeting (WCLM) Accelerating the Contributions of Chemistry to Sustainable Development - Aug 02, 2011 - San Juan, Puerto Rico

14th Symposium of the Natural Product Research Network for Eastern and Central Africa Natural Products Research Network for Eastern and Central Africa (NAPRECA) - Aug 08 - Aug 12, 2011 - Nairobi, Kenya

Ionic Liquids: Faraday Discussion 154 - Aug 22 - Aug 24, 2011 - Queen's University, Belfast, UK

Colloquium Spectroscopicum Internationale XXXVII in Brazil - Aug 28 - Sep 02, 2011 - Buzios - Rio de Janeiro - Brazil

ACS National Meetings & Exposition - Aug 28 - Sep 01, 2011 - Denver, Colorado