CHEMOMETRICS IN MEDICINE AND PHARMACY

JÁN MOCÁK

Department of Chemistry, Faculty of Natural Sciences, University of SS. Cyril and Methodius in Trnava, Nam. J. Herdu 2, Trnava, SK-917 01, Slovakia (jan.mocak@ucm.sk)

Abstract: This minireview summarizes the basic ways of application of chemometrics in medicine and pharmacy. It brings a collection of applications of chemometric used for the solution of diverse practical problems, e.g. exploitation of biologically active species, effective use of biomarkers, advancement of clinical diagnosis, monitoring of the patient's state and prediction of its perspectives, drug design or classification of toxic chemical substances. The aim of this contribution is a brief presentation of versatile potentialities of contemporary chemometrical techniques and relevant software. They are exemplified by typical cases from literature as well as by own research results of the Chemometrics group at Department of Chemistry, the University of Ss. Cyril & Methodius in Trnava.

Key words: chemometrics, biological activity, diagnosis prediction, drug design, QSPR, proteo-chemometrics

1. Introduction

Chemometrics put to use statistics, applied mathematics and computer science to extract and manage chemical information, which brings benefits to many experimental life sciences at and even behind the borders of chemistry - in medicine, pharmacy, biology, etc. It is a data-driven interdisciplinary science suitable for solving diverse applications. An important advantage of chemometrics and its methods is versatility, facilitated by its high level of abstraction, characteristic for the scientific disciplines extensively utilizing mathematics. When processing various data of everyday life, assembled in medicine, pharmacy, food control, environmental monitoring, etc., very similar algorithms and analogous ways of the data processing and evaluation are implemented for rather different investigated objects explored by means of miscellaneous techniques. By chemometrics, both descriptive as well as predictive issues of life sciences are solved. In descriptive applications, properties of the investigated systems are modeled in order to learn the underlying relationships and the system structure, which leads to the model identification, composition and understanding. In predictive applications, numerous system properties are utilized in the elaborated model with the intent of predicting the target properties, wanted features or behaviour of interest.

In this study an overview of the main directions of chemometric applications in medicine and pharmacy is propounded. As a part of it, the results of our own research activities oriented to medicinal and pharmaceutical chemistry are added and its impact upon several practical problems is discussed.

2. Materials and implemented software

Synthetic work is described in detail in the cited literature. The synthesis of the series of compounds, which exhibit considerable antimycobacterial activity, was

DOI 10.2478/v10296-012-0002-3

VERSITA

performed at the collaborating Faculty of Pharmacy, Charles University, Hradec Králové, Czech Republic. The series of various types of phenylcarbamic acid derivatives were prepared at Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Comenius University, Bratislava. All prepared compounds were characterized by their IR and NMR spectra and elemental analysis. Molecular properties and the quantum chemistry descriptors of the investigated compounds were calculated by means of the contemporary commercial software packages - ACD/Labs, Chemdraw, HyperChem, Dragon or directly on the VCCL web site (Virtual Computational Chemistry Laboratory). Determination of tumour markers in blood and pleural exudate was carried out in the Institute of Tuberculosis and Respiratory Diseases in Poprad. Determinations of cardiovascular markers, accomplishment of basic human biochemical tests, monitoring of the biochemical responses to the statine treatment and determination of glycated haemoglobin were performed at the collaborating Analytical - Diagnostic Laboratory, in Prešov. The children's hypertension data were assembled at the 2nd Department of Paediatrics, Faculty of Medicine, Comenius University, Bratislava. For the chemometric data processing and evaluation software packages Statistica, SPSS, Statgraphics and SAS were employed.

3. Domains of Chemometrics applications

3.1 Chemometric aids in medicine and pharmacy

Present automated laboratory instruments in biological/medical research produce a vast volume of measurement data, which are difficult to absorb and interpret. Therefore a challenging task of powerful mathematical and statistical methods of chemometrics is to reduce them and reveal all useful information. An important role of chemometrics in medicine was known a long time ago, shortly after it was widespread among life sciences (BOYD, 1986). There exist several areas of medicine as well as pharmacy where the aid of chemometrics is essential and indispensable: (a) Quality control of laboratory results, laboratory measurement standardization. (b) Decision about diagnosis, confirming or excluding a diagnosis. (c) Optimal clinical interpretation of laboratory tests in disease monitoring. (d) Detection of significant change in patient's condition during medical treatment and clinical care, prediction of the future medical state of the patient. (e) Drug synthesis, development and design, tools for drug discovery, structure-activity relationships, drug mechanism. (f) Chemometrics in metabolomics, proteochemometrics - a new bioinformatics technology. (g) Identification of dopes and toxic substances; expert systems for toxicity classification. (h) Prediction of bio-properties of various remedies, herbal medicine, and functional foods.

We have searched the application possibilities in many of the above mentioned issues and tried to provide several typical case studies (MOCAK *et al.*, 2011). Nevertheless, a more systematic approach is described in the following sub-chapters.

3.2 Quality control of laboratory tests, measurement standardization

A common way of evaluating the biochemical state of "normality" or "abnormality" is based on the statistically derived reference intervals, which serve

as the standards by which the laboratory test results are judged. Careful design of experimental protocol is the key in carrying out any evaluation of clinical diagnostic value (BOYD, 1997). Reference interval development has classically relied on concepts elaborated by the International Federation of Clinical Chemistry Expert Panel on Reference Values during the 1980s. Recent approaches to defining the reference values and decision limits are discussed in the papers (CERIOTTI *et al*., 2008; BOYD, 2010).

The obvious way of assessment of the diagnostic value of biochemical markers is by means of sensitivity, specificity, efficiency and predictive values, which have long been used as the indices of test accuracy (BALLA *et al*., 2004a). However, newer methods such as receiver operating characteristic curve (ROC) analysis or logistic regression analysis are more robust indicators that overcome many limitations of the traditional indices. Their successful implementation is discussed in further text.

An important part of quality control is statistical comparison of agreement of laboratory tests. Selection of an optimal measurement method is usually based on comparison of a newly developed procedure with the traditional one. In such a case both compared methods represent random variables - as neither of them is error-free. Consequently, the use of ordinary standard least squares regression method is not appropriate since the basic assumption about the error-free independent variable is violated, which may cause critical errors when using standard ways of regression (MOCAK *et al.*, 2003).

Advanced regression modelling in such a case requires the use of advanced techniques like Deming, orthogonal or Passing-Bablok regressions (ISLAMČEVIČ, 2004; MRAZOVA *et al.*, 2008a). First two techniques consider the specific variance value of each variable (their ratio is assumed unity in orthogonal regression), the third one is a non-metric, robust, rank dependent technique, nowadays frequently used in medicinal chemistry. The application of the mentioned advanced regression techniques was applied to ensure correct measurements of glycated haemoglobin in blood. It is an important task since glycohaemoglobin HbA_{1C} is the best marker of the glycemic state in patients with diabetes and enables its long-term assessment (MRAZOVA *et al*., 2008b).

For measurement of glycated haemoglobin in blood two automatic analyzers were used: Advia 1200 (Chemistry Systems Bayer) based on the NGSP (U.S. National Glycohaemoglobin Standardization Program) calibration assay and Hitachi 912 (Roche Diagnostic Systems) calibrated according to the IFCC (the International Federation of Clinical Chemistry Working Group) reference method. A thorough chemometrical data treatment by Deming, orthogonal and Passing-Bablok regressions revealed (MRAZOVA *et al.*, 2010a) that the NGSP reference system gives systematically higher results of glycated haemoglobin than the IFCC reference system. If the recalculation of the IFCC into NGSP values is made via Master Equation, the regression intercept is adjusted to zero as expected, but the resulting straight-line deviates by the slope. In conclusion, further harmonisation of the two reference systems is demanded and more experimental data from the accredited laboratories are needed (MRAZOVA *et al.*, 2010b).

Similar advanced regression approach was successfully applied to determination of low-density lipoprotein cholesterol (LDLc), which serves as a marker

for hypertension and progression of cardiovascular diseases (ISLAMČEVIČ, 2004). Theoretical background of all used regression methods is there described.

3.3 Medical diagnosis confirmation or prediction, assessment of effectiveness of laboratory tests

Advanced implementation of chemometric and statistical algorithms facilitates clarifying of many important practical applications in medicine. It was found that an appropriate exploitation of multivariate statistics, e.g. the outputs of principal component analysis, techniques of discriminant analysis, logistic regression, together with the standard statistical tools like correlation analysis, analysis of variance (ANOVA) as well as the ROC curves (MOCAK *et al.*, 2005; BALLA *et al.*, 2004a; BALLA and MOCAK, 2008), may enable a better medical diagnosis confirmation and/or prediction.

The area under the ROC curve is the best global indicator of the test accuracy (BOYD, 1997). Logistic regression analysis allows the diagnostic information from several tests to be evaluated simultaneously. In addition, it provides a straightforward method for calculating likelihood ratios. Likelihood ratios are useful for interpreting the test results in the individual patient because they provide a convenient means to directly determine predictive value without having to calculate sensitivity and specificity for a given decision limit.

It is common that the physician used to evaluate all performed laboratory tests sequentially and then it depends on his/her knowledge how these results are composed into final decision about diagnosis. On the contrary, our new approach is based on a *simultaneous use* of several selected laboratory tests (BALLA *et al.*, 2004b). Utilization of biochemical markers as the chemometrical descriptors and their optimal combination into a newly calculated multicomponent marker should provide an optimally weighted collective diagnostic information. This effective new approach was applied in the treatment of cardiovascular diseases (BALLA *et al.*, 2004b; MOCAK *et al*., 2005). In this case, with regard to diagnosis the most effective were the multicomponent markers composed of eight individual cardiovascular markers directly measured in the clinical laboratory. As manifested by the ROC curve areas, the best combination was provided by logit - the output of logistic regression, the second best was the first discriminant function (as the most informative output of discriminant analysis), linearly combined from all eight individual markers.

According to the ROC curve area, the composed multicomponent markers exhibited the best diagnostic effectiveness also in diagnosing lung malignity, where CEA and CYFRA tumour markers in the blood as well as in pleural exudate were originally measured (MOCAK and BALLA, 2003; KAVKOVA *et al.*, 2007; MRAZOVA *et al*., 2008c; BALLA and MOCAK, 2008). In this case the differentiation from benign pulmonary diseases is highly demanded (KAVKOVA *et al.*, 2007). Again, the best muticomponent was logit, followed by the first principal component (the most informative output of principal component analysis) and first discriminant function; the ROC curve area for CEA in exudate, found as the best individual marker, was significantly smaller (MRAZOVA *et al*., 2009; MOCAK,

2010a). We have confirmed that the calculation of the composed multicomponent marker from the originally measured individual markers is simple and can be easily implemented into the hospital information system as a new effective diagnostic tool.

3.4 Monitoring of the health state of the patients

Further task of chemometrics is its assistance in detecting the changes in patient's condition concerning progress of the disease and reaction to the medical treatment. Another target of chemometrics is determination of patient prognosis - prediction of the future medical state of a patient on the basis of present laboratory test results and clinical care. Several typical examples are shortly discussed in the next paragraphs.

A complex monitoring of the health state of the patients during a long term treatment by some drugs involves simultaneous evaluation of numerous biochemical tests. A good example of chemometrical processing of the laboratory test results was described in (DURCEKOVA *et al.*, 2009a) where the state of the probands before and one year after statin administration was studied. Statins are effective inhibitors of 3-hydroxy-3-methylglutaryl coenzyme A reductase and represent a major advance in prevention and treatment of cardiovascular diseases. Their benefit is caused by the lipids suppression. On the other hand, it is important to find whether the hypolipidemic effect of statins is not accompanied by the unfavourable impact on some human body organ. Before, during and after the statin administration the level of many biochemical tests is therefore determined and statistical evaluation of the changes is performed.

The results obtained in (DURCEKOVA *et al.*, 2009b; 2011) unambiguously confirmed the positive changes in the lipid metabolism of the patients with cardiovascular risk. The multivariate data processing techniques allow clear visualisation of the statin administration effect. ROC curves, box-plots and discriminant analysis exhibit a distinct discrimination of the patients' samples into two categories - before and one year after the statin treatment. Analysis of variance reveals that four biochemical tests are capable to differentiate the statin treatment also with regard to the patient gender: total cholesterol, low-density lipoprotein cholesterol, triacylglycerols and aterogenity index. A substantial outcome of the achieved results is a very high diagnostic effectiveness of three composed multicomponent variables, calculated by linear combination of individual laboratory tests. Changes in the kidney and hepatic test values caused by the statin uptake were found not statistically significant (DURCEKOVA *et al.*, 2011).

Hypertension is not only a serious disease of adults, but as demonstrated in recent years, it is now increasingly encountered in children. The progress of this latent disease leads to damage and subsequent failure of vital organs. Diagnosis of hypertension in children is very difficult. The problems with the blood pressure measurement require a long-term monitoring of the state of the patient via biochemical tests leading the physician to diagnose. Chemometrical approach evaluates the impact of all risk factors; the applied multidimensional classification techniques (artificial neural networks, Kth nearest neighbour and general discriminant analysis) enable quantification of the hypertension risk in children (MRAZOVA *et al.*, 2012).

Another interesting study involves a combination of HPLC and chemometrics data evaluation concerning determination of morphine and its two main metabolites, morphine-3-glucuronide (M3G) and morphine-6-glucuronide (M6G), in the serum of oncological patients (NETRIOVA *et al.*, 2006). The results are important for appropriate analgesic treatment of the patients under palliative care. In the mentioned work it was clearly manifested that a higher M3G/M6G ratio is reflected in a stronger patient's dissatisfaction with the morphine treatment. It was found that the morphine and morphine-3-glucuronide concentrations depend on the patient's gender, with the men values higher than the corresponding women ones. Since the morphine-6-glucuronide serum concentration is about the same for men and women, the M3G/M6G ratio is in general higher for men than for women. In addition to the specific role of the patient gender it was also found a remarkable influence of the tumour location on the serum concentration of morphine and its glucuronide metabolites, and consequently the M3G/M6G concentration ratio.

3.5 Drug design, structure-activity and structure-property relationships

Multivariate chemometric models can be used also for prediction of bio-properties from chemical data obtained either from various instrumental measurements or by calculation of the descriptors derived from the molecular structure. Drug development (DREWS, 2000) is a tedious and resource demanding process consisting of several steps. Chemometrics plays important role especially in the starting step – in the search of the compounds with a strong activity against the originators of diseases. In the last decades the QSAR (Quantitative Structure – Activity Relationships) became an important tool in drug discovery and in toxicology. The goal of QSAR is investigation how to predict the biological activity of a set of compounds in demand and elucidation of the specific action of the administrated drug; it deals with generalization of relations between the chemical and biological properties of the examined set of chemically resembling derivatives. There exist numerous publications devoted to this attractive topic, two recent books were published by (PUZYN *et al.*, 2010; MERZ *et al.*, 2010). Some examples of QSAR studies are discussed in the next paragraph. In the second revised edition of the book (TODESCHINI and CONSONNI, 2009a; 2009b) the entire relevant literature over the previous six years was surveyed. Volume I contains an alphabetical listing of more than 3300 descriptors and related terms for chemoinformatic analysis of chemical compound properties, while the second volume lists over 6000 references selected from 450 journals.

With regard to high antimycobacterial activity, important mainly in investigation of new strongly active antituberculotic drugs, we have explored QSAR of two sets of compounds: (1) *N*-benzylsalicylamide derivatives (NEMECEK *et al.*, 2006; 2009) and 3-phenyl-2*H*-1,3-benzoxazin-2,4(3*H*)-dione (NEMECEK *et al.*, 2012). Crucial for a successful prediction of biological activity of the studied set of compounds is reduction of the descriptor number and finding an optimal set of descriptors, which is due to their mutual dependence not a simple task. For this purpose several tools are customarily used: (a) the correlation analysis applied to all descriptors, (b) sensitivity

analysis, or (c) in the recent time most popular genetic algorithms. After selection the descriptors with the strongest correlation to biological activity (the activity against the investigated strains of mycobacteria), the structures of new derivatives with the strongest effect were found with the predicted lowest minimal inhibition concentration (MIC). Since a part of the mentioned studies was performed exclusively with the calculated molecular descriptors and the simulated ^{1}H and ^{13}C NMR signals, the MIC prediction was possible without the need of a previous synthesis. In this way, the collection of the best structures represents effective and economical way of the first step in drug design.

Thank to the development of chemoinformatics, public availability of bioactivity databases (like PubChem, ChEM-BL, DrugBank, etc.) and web-based tools for drug discovery and toxicology assessment, several open access web sites are available for effective QSAR calculations (http://www.vvclab.org, http://www.ochem.eu, http://www.opentox.org, http://www.collaborativedrug.com, http://qsardb.org, http://qsardb.jrc.it, http://chembench.mml.unc.edu, http://pharmaexpert.ru, etc.).

Molecular structure is often related to further properties in addition to biological ones and the area of interest in QSAR modelling has been expanded from *activity* towards *properties.* QSPR (Quantitative Structure – Property Relationships) connects the knowledge from biology, chemistry, physics, mathematics, and computer sciences with the aim of exploring properties of substances and processes. The novel characteristics of organic, metal organic and coordination compounds can be calculated by QSPR and applied in practice. QSPR is applicable to a vast number of different properties. The calculation procedures utilizable in these cases are similar to those known in QSAR. The QSPR methodology is based at first on finding the correlations between a desired property already experimentally measured for a series of molecules and then using these correlations to predict the same property for additional molecules, including the compounds that have not yet been synthesized. Even for these calculations the open access web systems are available (e.g. http://www.qspr-thesaurus.eu).

When chromatographic properties are concerned, instead of QSPR the specific term QSRR (Quantitative Structure – Retention Relationships) is used. At present, the QSRR studies are more and more widespread since they contribute to a deeper understanding the relations between the structure of the chromatographically separated species and the stationary and mobile phases (KALISZAN, 1997; KALISZAN 2007; HÉBERGER, 2007).

In our QSRR (Quantitative Structure – Retention Relationships) studies the HPLC retention factor was used as the target variable. Retention factor is in general very closely related to lipophilicity, which seems to be the descriptor most influencing biological activity. A very successful prediction of the PBOD retention factors were published in (NEMECEK *et al.*, 2010a; 2010b). Successful QSRR model was recently elucidated also for esters of alkoxyphenylcarbamic acid (DURCEKOVA *et al.*, 2010a; 2012), which exhibit a strong anaesthetical activity. Details about the factors most influencing the surface anaesthetical activity and infiltration anaesthetical activity are described in (DURCEKOVA *et al.*, 2010b). For prediction of anaesthetical activity the artificial neural network models were employed; further properties

of the investigated set of compounds were well characterized by principal component analysis, cluster analysis and discrimination analysis.

A unique use of the whole chromatogram instead of individual peaks for fingerprinting of various phytochemicals was described by (KOMSTA, 2012). The chromatographic fingerprinting technique is essentially a kind of high-throughput and integral tools to explore complex and multivariate data. It was used e.g. to investigate the complexity of herbal medicines where confirmation or identification the most important chemical components is needed (LIANG *et al.*, 2006; NI *et al.*, 2007); further seven similar papers are cited in the review of (KOMSTA, 2012).

A special approach to calculate new precise values of many important physico-chemical properties for a homologous series of organic compounds is based on mathematical recurrence relations, which can be easily computed even in a common spreadsheet, e.g. Microsoft Excel (MOCAK and RABAROVA, 2010b).

3.6 Chemometrics in metabolomics

Metabolomics concerns the systematic study of the unique chemical fingerprints that specific cellular processes leave behind - the study of their small-molecule metabolite profiles. METLIN metabolomics web database contained in 2001 over 40,000 metabolites. It is connected mainly to the development of gas chromatography - mass spectrometry techniques; another crucial instrumental technique in metabolomics is high-resolution NMR spectroscopy. Recent very sensitive ^{1}H NMR metabolomic studies became to be widespread in Analytical Chemistry, e.g. in papers (SANDS, *et al.*, 2009; ALVES *et al.*, 2009; ROBINETTE *et al.*, 2009; PEARCE *et al.*, 2009; MAHER *et al.*, 2008; CLOAREC *et al.*, 2005).

GC/MS and NMR techniques together with multivariate chemometrics are capable to characterize the metabolic profiles of biofluids, understand the mechanism of patogenesis and uncover potential biomarkers of disease progression. Chemometrics constitutes an integral part of the metabolomic technology by providing organized multivariate statistical approaches for modelling and interpreting changes in the metabolic profiles (BRUCE *et al.*, 2008; MADSEN *et al.*, 2010; ELIASSON *et al.*, 2011).

Combinations of „omics“ investigations (transcriptomic, proteomic, metabolomic) are increasingly applied to get comprehensive understanding of biological systems. They allow a more holistic approach to understanding the complexity of human biochemistry and biomolecular medicine.

Proteochemometrics is a bioinformatic approach to protein engineering, novel drug design, target analysis and functional genomics analysis (LAPINSH *et al.*, 2001; WIKBERG 2004). Proteochemometrics combines chemical organic synthesis and design with molecular biological molecular pharmacological analysis and with bioinformatics/mathematical modeling. Using proteochemometrics the detailed maps can be built, down to the physico-chemical and structural level, exhibiting the interactions of organic molecules with macromolecules in a cell and these maps may be directly used for the design of new molecules with desired properties. Many further useful details can be found on the web site of Uppsala University (http://www.proteochemometrics.org).

3.7 Toxicity assessment and prediction

In the last two decades a great deal of effort has been put into the study of the relationships between a compound's structure and its toxicity (CRONIN *et al.*, 2003; GALLEGOS, 2006). Attempts have been made to classify chemicals according to the mechanism of their toxicity and to screen them for their environmental risk assessment. Most of the QSAR applications to toxicities have been developed for congeneric sets, but non-congeneric sets of compounds have also been documented (CRONIN and DEARDEN, 1995a; 1995b; 1995c; 1995d). It is necessary to note that many QSAR toxicity applications have been called QSTR (Quantitative structure – Toxicity Relationships). Numerous calculated descriptors and recommended software packages are discussed e.g. in (KATRITZKY and TATHAM, 2001).

The European policy for evaluation of chemicals has lead to a controversy with regard to the need of additional animal safety testing. In order to save time and resources, alternative *in vitro* or *in silico* tests for the assessment of toxic effects of chemicals were suggested. In toxicology, besides in vitro models the computer-assisted prediction models are used as predictive tools especially to elucidate acute, chronic and/or organ toxicities (SIMON-HETTICH and STEGER-HARTMANN, 2006). Computer-assisted prediction tools play an important role mainly in the assessment of chemicals for which no data is available or for which toxicological testing is impractical due to the lack of availability of sufficient compounds for testing (e.g. impurities in pharmaceuticals). EU and OECD have produced several documents concerning toxicity prediction, e.g. (WORTH *et al.*, 2005; VRAČKO *et al.*, 2006).

There exist many applications concerning toxic substances where chemometrics plays an important explaining role, e.g. in expert systems for toxicity classification, in identification of drugs and toxic substances in complex mixtures and in forensic chemistry (GARKANI-NEJAD, 2004). In analytical procedures dealing with the analytes containing toxic components very important are also the rules of good laboratory practice and exact evaluation of the measurement results according to metrological principles. Two examples of quality assurance application are in the papers of BOROŠOVÁ *et al.* (2001; 2002).

3.8 Further chemometrics applications

Another area demanding prediction of bio-properties from chemical data is the study of bioactive compounds in food. The presence of specific molecules which have a positive pharmacological effect when consumed in sufficient quantities represents the basic benefit of functional food (MCGOVERIN *et al.*, 2010). Functional foods are foods or dietary ingredients that provide a health benefit beyond basic nutrition. However, the legal framework within the European Union now demands that any claim about the nutritional or physiological effects of a product must be scientifically demonstrated. An example of chemometric study of metabolomics in functional food is described in (LUNDSTEDT *et al.*, 2010) where a combined intake of soybean and grapefruit was analyzed with respect to both pharmacological and physiological effects. Resulting multivariate models showed that this diet induced

decrease of lactate, cholesterols and triglycerides; further biotransformation details are there also discussed.

Food commodities are the objects of so many chemometrics application that they cannot be included here and will be described and discussed in another review. Another special attention should be paid to the details of data processing like data pre-treatment, normalization, denoising, baseline removal, warping, then to the proper choice and combination of chemometrics tools and implementation of pertinent software. Despite of their importance they are outside the scope of this paper. Nevertheless, at least one recent excellent book should be mentioned in this connection (BRERETON, 2007) where not only older as well as newest techniques of chemometrics are clarified but Chapter 10 is directly devoted to biological and medical applications of chemometrics.

4. Conclusions

Chemometrical processing of the experimentally measured data and/or the use of diverse calculated molecular descriptors represents at present a substantial aid to medicine and pharmacy. Chemometric methods and specialized software packages enable (a) quality control of laboratory tests and standardization of measurements in laboratory medicine, (b) assessment of effectiveness of laboratory tests used for confirmation as well as prediction of clinical diagnosis, (c) long-term monitoring of the disease, success or failure in medical treatment and assessment of effectiveness of the applied drugs, (d) targeted design of new potentially high effective drugs using structure – activity relationships, (e) prediction of desired property of the investigated group of substances by QSPR studies, (f) deeper insight into chemical processes involving metabolites, which can be utilized in deeper knowledge of the disease but also in more effective drug design, (g) elucidation of toxicity of chemical substances related to their structure and molecular properties, (h) prediction of bio-properties from chemical data in food. Progressive chemometric tools exhibit a notable role in advancement of theoretical cognition in all mentioned fields.

Acknowledgement: This study is supported by the grant VEGA 1/0233/12. Collaboration of all co-authors of the cited own publications is highly acknowledged.

References

ALVES, A.C., RANTALAINEN, M., HOLMES, E., NICHOLSON, J.K., EBBELS, T.M.: Analytic properties of statistical total correlation spectroscopy based information recovery in 1H NMR metabolic data sets. Anal. Chem., 81, 2009, 2075-2084.

BALLA, B., MOCAK, J., VARMUSOVA, E., KAVKOVA, D., TUDIK, I.: Assessment of effectivity of laboratory methods (in Slovak). Chem. Listy, 97, 2004a, 333-338.

BALLA, B., MOCAK, J., PIVOVARNIKOVA, H., BALLA, J.: Comparative study of cardiovascular markers data by various techniques of multivariate analysis. Chemom. Intell. Lab. Systems, 72, 2004b, 259-267.

BALLA, B., MOCAK, J.: Evaluation of laboratory tests and prospects of computer assisted clinical diagnosis. *In*: R. GASKO, B. BALLA, L. DOHNAL, V. JANOUT, E. KLIMOVA, J. MOCAK (Eds.) Statistical methods in clinical, epidemiological and laboratory practice (in Slovak). Aprilla, Kosice, 2008, Chap. 14. ISBN 978-80-89346-00-4.

BOROŠOVÁ, D., MOCAK, J., BEINROHR, E., BOBROWSKI, A.: Determination of arsenic in water – quality assurance and calculation of metrological characteristics. Centr. Europ. J. Public Health, 8, 2001, 196-198.

BOROŠOVÁ, D., MOCAK, J., BEINROHR, E., MIŠKOVIČ, P.: Validation and Quality Assurance of Arsenic Determination in Urine by GFAAS after Toluene Extraction. Polish J. Environ. Stud., 11, 2002, 617-623.

BOYD, J.C.: Perspectives on the use of chemometrics in laboratory medicine. Clin. Chem., 32, 1986, 1726-1733.

BOYD, J.C.: Mathematical tools for demonstrating the clinical usefulness of biochemical markers. Scand. J. Clin. Lab. Invest. Suppl. 227, 1997, 46-63.

BOYD, J.C.: Defining laboratory reference values and decision limits: populations, intervals, and interpretations. Review. Asian J. Androl., 12, 2010, 83-90.

BRERETON, R.: Applied chemometrics for scientists. John Wiley & Sons Ltd, Chichester, 2007, 379 pp.

BRUCE, S. J., TAVAZZI, I., PARISOD, V., REZZI, S., KOCHHAR, S., GUY, P.A.: Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. Anal. Chem., 81, 2009, 3285-3296.

CERIOTTI, F., BOYD, J.C., KLEIN, G., HENNY, J., QUERALTÓ, J. KAIRISTO, V., PANTEGHINI, M.: IFCC Committee on Reference Intervals and Decision Limits. Reference intervals for serum creatinine concentrations: assessment of available data for global application. Clin. Chem., 54, 2008, 559-566.

CLOAREC, O., DUMAS, M.E., CRAIG, A., BARTON, R.H., TRYGG, J., HUDSON, J., BLANCHER, C., GAUGUIER, D., LINDON, J.C., HOLMES, E., NICHOLSON, J.K.: Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. Anal. Chem., 77, 2005, 1282-1289.

CRONIN, M.T., JAWORSKA, J.S., WALKER, J.D., COMBER, M.H.,WATTS, C.D., WORTH, A.P.: Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. Environ. Health Perspect., 111, 2003, 1391-1401.

CRONIN, M.T.D., DEARDEN, J.C.: QSAR in Toxicology. 1. Prediction of Aquatic Toxicity. Quant. Struct.-Act. Relat., 14, 1995a, 1-7.

CRONIN, M.T.D., DEARDEN, J.C.: QSAR in Toxicology. 2. Prediction of Acute Mammalian Toxicity and Interspecies Correlations. Quant. Struct.-Act. Relat., 14, 1995b, 117-120.

CRONIN, M.T.D., DEARDEN, J.C.: QSAR in Toxicology. 3. Prediction of Chronic Toxicities. Quant. Struct.-Act. Relat., 14, 1995c, 329-334.

CRONIN, M.T.D., DEARDEN, J.C.: QSAR in Toxicology. 4. Prediction of Non-lethal Mammalian Toxicological Endpoints, and Expert Systems for Toxicity Prediction. Quant. Struct.-Act. Relat., 14, 1995d, 518-523.

DREWS, J.: Drug Discovery: A Historical Perspective. Science, 287, 2000, 1960-1964.

DURCEKOVA, T., MOCAK, J., BALLA J., GROMANOVA, G., BORONOVA, K.: Relationship between administration of statins and blood serum levels of selected biochemical parameters. Nova Biotechnol., 9, 2009a, 83-90.

DURCEKOVA, T., MOCAK, J., BALLA J., GROMANOVA, G., BORONOVA, K.: Comparison of ten biochemical laboratory tests before and after treatment by statins. Nova Biotechnologica, 9, 2009b, 133-140.

DURCEKOVA, T., BORONOVA, K., LEHOTAY, J., MOCAK J., CIZMARIK, J.: Exploitation of artificial neural networks for finding the correlations between the calculated and measured values of retention factors of basic esters of alkoxysubstituted phenylcarbamic acid (in Slovak). Czech Slovak Pharm., 59, 2010a, 205-209.

DURCEKOVA, T., MOCAK, J., LEHOTAY, J., CIZMARIK, J., BORONOVA, K.: Study of anaesthetical activity of esters of alkoxyphenylcarbamic acid using chemometrical methods. Pharmazie, 65, 2010b, 169-174.

DURCEKOVA, T., MOCAK, J., BALLA, J., BORONOVA, K.: Effect of statin therapy on biochemical laboratory tests - A chemometrics study. J. Pharm. Biomed. Analysis, 54, 2011, 141-147.

DURCEKOVA, T., BORONOVA, K., MOCAK, J., LEHOTAY, J., CIZMARIK, J.: QSRR models for potential local anaesthetic drugs using high performance liquid chromatography. J. Pharm. Biomed. Analysis, 59, 2012, 209-216.

GALLEGOS, A.: Mini-review on chemical similarity and prediction of toxicity. Curr. Comp.-Aided Drug Design, 2, 2006, 105-122.

GARKANI-NEJAD, Z., KARLOVITS, M., DEMUTH, W., STIMPFL, T., VYCUDILIK, W., JALALI-HERAVI, M., VARMUZA, K.: Prediction of gas chromatography retention indices of a diverse set of toxicologically relevant compounds. J. Chromatogr. B, 1028, 2004, 298-295.

HÉBERGER, K.: Quantitative Structure-(Chromatographic) Retention Relationships. J. Chromat. A, 2007, 1158, 273–305.

ISLAMČEVIČ, M., MOCÁK, J., BRODNJAK-VONČINA, D., BLAŽÍČEK, P.: Correct Ways of Using Regression for Methods Comparison Studies - Determination of LDL-Cholesterol. Sci. Pap. Univ. Pardubice, Ser. A, 10, 2004, 111-118.

KALISZAN, R.: Structure and Retention in Chromatography. A Chemometric Approach, Harwood Academic: Amsterdam, 1997, 210 pp.

KALISZAN, R.: Quantitative Structure-(Chromatographic) Retention Relationships. Chem. Revs., 2007, 107, 3212–3246.

KATRITZKY, A.R., TATHAM, D.B.: Theoretical Descriptors for the Correlation of Aquatic Toxicity of Environmental Pollutants by Quantitative Structure - Toxicity Relationships. J. Chem. Inf. Comput. Sci., 41, 2001, 1162-1176.

KAVKOVA, D., VARMUSOVA, E., TUDIK, I., MOCAK, J., BALLA, B.: Utilization of multicomponent tumour marker for improvement of diagnostic effectiveness in differential diagnostics of malígnant pleural exudations (in Slovak). Studia Pneumol. & Phtizeol., 67, 2007, 199-203.

KOMSTA, L.: Chemometrics in Fingerprinting by Means of Thin Layer Chromatography. Chromatogr. Res. Int., 2012, in press, doi:10.1155/2012/893246.

LIANG, Y.-Z., YI L.-Z., XU Q.-S.: Chemometics and modernization of traditional Chinese medicine. Sci. China B, 51, 2008, 718-728.

LUNDSTEDT, T., HEDENSTRÖM, M., SOERIA-ATMADJA, D., HAMMERLING, U., GABRIELSSON, J., OLSSON, J., TRYGG, J.: Dynamic modelling of time series data in nutritional metabonomics - A powerful complement to randomized clinical trials in functional food studies. Chemometrics & Intell. Lab. Syst., 104, 2010, 112-120.

MAHER, A.D., CROCKFORD, D., TOFT, H., MALMODIN, D., FABER, J.H., MCCARTHY, M.I., BARRETT, A., ALLEN, M., WALKER, M., HOLMES, E., LINDON, J.C., NICHOLSON, J.K.: Optimization of human plasma 1H NMR spectroscopic data processing for high-throughput metabolic phenotyping studies and detection of insulin resistance related to type 2 diabetes. Anal. Chem., 80, 2008, 7354-7362.

MCGOVERIN, C.M., WEERANANTANAPHAN, J., DOWNEY, G., MANLEY, M.: The application of near infrared spectroscopy to the measurement of bioactive compounds in food commodities. J. Near Infrared Spectr., 18, 2010, 87-111.

MERZ, K.M., RINGE, D., REYNOLDS, C.H.: Drug Design: Structure- and Ligand-Based Approaches. Cambridge University Press, New York, 2010.

MOCAK, J., BALLA, B.: Chemometrically supported diagnostics of civilization diseases (in Slovak). Chem. Listy, 97, 2003, 736-737.

MOCAK, J., BALLA, B., BALLA, J., PIVOVARNIKOVA, H., VARMUSOVÁ, E., KAVKOVÁ, D., TUDÍK, I.: Use of Multivariate Data Analysis for Diagnosis Confirmation or Prediction. Diagnoza & Lab. Diagn., 10, 2005, 40-44.

MOCÁK, J.: Effective Use of Combination of Lung Malignity Markers. *In*: European Bioanalysis Forum - 3rd Annual Open Symposium. Barcelona, Dec. 1-3, 2010a, p. 1-23. Open internet version: http://bcn2010.europeanbioanalysisforum.eu/slides.

MOCAK J., RABAROVA E.: Recurrent relationships in chromatography and electrochemistry. Chem. Listy, 104, 2010b, s464-s468.

MOCAK, J., NEMEČEK, P., MRÁZOVÁ, V., BORONOVÁ, K., LEHOTAY, J., ČIŽMÁRIK, J.: Application of chemometrics to exploitation of biologically active species in pharmacy and medicine. *In*: P. Nemeček, M. Ondrejovič (Eds.) Proceedings Int. Conf. Appl. Nat. Sciences 2011. Publ. UCM, Trnava, 2011, 75-79. ISBN 978-80-8105-266-8.

MRÁZOVÁ, V., MOCÁK, J., BLAŽÍČEK, P., JAKUBCOVÁ, I.: Comparison of analytical methods by regression – determination of glycated haemoglobin (in Slovak). CHEMagazin, 18, 2008a, 8-10.

MRÁZOVÁ, V., MOCÁK, J., BLAŽÍČEK, P., JAKUBCOVÁ, I.: Determination of glycated haemoglobin – regression comparison of analytical methods (in Slovak). Chem. Listy, 102, 2008b, 716-717.

MRAZOVA, V., MOCAK, J., VARMUSOVA, E., KAVKOVA, D.: Computer-aided diagnosis of lung malignity using multidimensional analysis of tumour markers data. Nova Biotechnol., 8, 2008c, 65-70.

MRAZOVA, V., MOCAK, J., VARMUSOVA, E., KAVKOVA, D., BEDNAROVA, A.: Use of multi-dimensional data analysis for prediction of lung malignity. J. Pharm. Biomed. Analysis, 50, 2009, 210-215.

MRAZOVA, V., MOCAK, J., BALLA, J., BEDNAROVA, A.: Comparison of analytical methods using various types of regression. Chem. Listy, 104, 2010a, s676-s679.

MRAZOVA, V., MOCAK, J., BEDNAROVA, A, BALLA, J.: Comparison of IFCC and NGSP methods for determination of glycated haemoglobin using advanced regression techniques. Centr. Eur. J. Chem., 8, 2010b, 1216-1222.

MRÁZOVÁ V., MAKOHUSOVÁ M., MOCAK J., KOVÁCS L., BABINSKÁ K.: Assessment of biochemical tests and further factors related to hypertension in children using chemometrical techniques. Clin. Chim. Acta, 2012 (in press).

NEMECEK, P., DURCEKOVA, T., MOCAK, J., LEHOTAY, J., WAISSER, K.: Study of relationships among biological activity and physico-chemical properties of potential antituberculotics (in Slovak). Nova Biotechnol., 6, 2006, 37- 47.

NEMECEK, P., DURCEKOVA, T., MOCAK, J., WAISSER, K.: Chemometrical analysis of new QSAR parameters and their exploitation for prediction of biological activity. Chem. Pap., 63, 2009, 84-91.

NEMECEK, P., MOCAK, J., LEHOTAY, J., WAISSER, K.: QSRR study of potential antituberculotic agents. Chem. Listy, 104, 2010a, s470-s474.

NEMECEK, P., MOCAK, J., LEHOTAY, J., WAISSER, K.: Prediction of HPLC Retention Factor of Potential Antituberculotics by QSRR. J. Liq. Chrom. & Rel. Tech., 34, 2010b, 168-181.

NEMECEK, P., MOCÁK, J., LEHOTAY, J., WAISSER, K.: Prediction of antimycobacterial activity of 3-phenyl-2H-1,3-benzoxazin-2,4(3H)-dione derivatives. Chem. Pap., 2012, in press.

NETRIOVA J., JOHANEDISOVA Z., BRANDSTETEROVA E., LEHOTAY J., SERDT K., MOCAK J.: HPLC determination of morphine, morphine-3-glucuronide and morphine-6-glucuronide in human serum of oncological patients after administration of morphine drugs. Pharmazie, 61, 2006, 528-534.

NI, Y. Zhang, CHURCHILL, J., KOKOT, S.: Application of high performance liquid chromatography for the profiling of complex chemical mixtures with the aid of chemometrics. Talanta, 72, 2007, 1533-1539

PEARCE, J.T., ATHERSUCH, T.J., EBBELS, T.M., LINDON, J.C., NICHOLSON, J.K., KEUN, H.C.: Robust algorithms for automated chemical shift calibration of 1D 1H NMR spectra of blood serum. Anal. Chem., 80, 2008, 7158-7162.

PUZYN, T., LESZCZYNSKI J., CRONIN, M.T. (Eds.): Recent Advances in QSAR Studies. Methods and Applications. Springer, 2010, 414 pp.

ROBINETTE, S.L., VESELKOV, K.A., BOHUS, E., COEN, M., KEUN, H.C., EBBELS, T.M., BECKONERT, O., HOLMES, E.C., LINDON, J.C., NICHOLSON, J.K.: Cluster analysis statistical spectroscopy using nuclear magnetic resonance generated metabolic data sets from perturbed biological systems. Anal. Chem., 81, 2009, 6581-6589.

SANDS, C.J., COEN, M., MAHER, A.D., EBBELS, T.M., HOLMES, E., LINDON, J.C., NICHOLSON, J.K.: Statistical total correlation spectroscopy editing of 1H NMR spectra of biofluids: application to drug metabolite profile identification and enhanced information recovery. Anal. Chem., 81, 2009, 6458-6466.

SIMON-HETTICH, B., ROTHFUSS, A., STEGER-HARTMANN, T.: Use of computer-assisted prediction of toxic effects of chemical substances. Toxicology, 224, 2006, 156-162.

TODESCHINI, R., CONSONNI, V.: Molecular Descriptors for Chemoinformatics: Vol. I: Alphabetical Listing. 2nd Ed., Wiley-VCH, Weinheim, 2009.
TODESCHINI, R., CONSONNI, V.: Molecular Descriptors for Chemoinformatics: Vol. II: Appendices, References. 2nd Ed., Wiley-VCH, Weinheim, 2009.