

# Statistical and Chemometric Tools Applied to Pharmaceutical Analysis

Jean-Luc Veuthey and Serge Rudaz\*

**Abstract:** The Laboratory of Pharmaceutical Analytical Chemistry (LCAP) is involved in numerous projects and therefore has to use different analytical instrumentations. Separation techniques, such as chromatography or electrophoresis, remain the techniques of choice but need to be adapted in each specific challenge. To better develop, optimize and validate analytical methodologies, the use of the multivariate approach is of utmost interest. Statistical experimental design is a powerful tool to quantify the effect of one or more variables on a set of measured responses. It provides a methodological framework for changing operating parameters simultaneously by the help of experimental designs. These approaches involve the smallest possible number of useful experiments and provide maximum information. The relevant results obtained in pharmaceutical analysis compared to a univariate approach come from their ability to detect and quantify variable interactions and therefore considerably increase the system knowledge. On the other hand, with the increase of modern computerized analytical techniques, the information could be present as hidden structures in the huge amount of data. As a result, chemometric methods have been applied to extract information. For this purpose, principal component analysis (PCA) combined with other statistical techniques are now necessary to efficiently reduce and interpret complex data structure.

**Keywords:** Chemometrics · Experimental designs · Pharmaceutical analysis · Statistical analysis

## Introduction

Chemometrics was born in the eighties and can be defined as “the science of relating measurements made on a chemical system or process to the state of the system *via* application of mathematical or statistical methods and design or selection of optimal measurement procedures or experiments to provide maximum information content of collected data” [1]. Chemometric tools for developing analytical methods, determining optimized or robust conditions as well as for treating data with pattern rec-

ognition techniques are therefore applied in many projects in the Laboratory of Pharmaceutical Analytical Chemistry (LCAP). Numerous works have demonstrated the potential of experimental design in the development and optimization of several analytical techniques such as the enantiomeric separation of pharmaceutical compounds [2–6] by capillary electrophoresis (CE) or liquid chromatography (LC) with several detection modes (UV, fluorescence, LIF, mass spectrometry). After development and optimization phase steps, the method validation needs to be performed. Validation is “...the confirmation that the particular requirements for a specific intended use are fulfilled...” and among the numerous investigated criteria, robustness remains a fundamental aspect of method validation to identify factors which could influence the results. Robustness testing is widely applied in the pharmaceutical analytical field because of strict regulations set by authorities who require extensively validated methods. Factors are examined by a multivariate approach with experimental designs selected as a function of their number. Robustness tests are generally performed by screening designs such as fractional factorial (FF) or Plackett-Burman (PB) to simultaneously investigate a relatively large number of factors in a relatively small number of experiments [7][8].

On the other hand, with the development of new powerful and rapid analytical procedures, we are faced with a huge number of data which can be also very complex. Therefore, new approaches for data mining are mandatory. Hence, handling, interpretation and prediction procedures are necessary to extract the most valuable chemical information from the recorded chemical data. Such techniques, issued from the multivariate analysis, involve Principal Component Analysis (PCA), Factor Analysis (FA), Cluster Analysis (CA), K Nearest Neighbors (KNN), N-way tools and so on, which were applied in several research projects at the LCAP.

## Chemometrics in Method Development

### *Development of a Bioreactor Based on Trypsin Immobilized on Monolithic Support for the On-line Digestion and Identification of Proteins*

The preparation and characterization of a new trypsin-based bioreactor was investigated for on-line protein digestion and peptide analysis [9]. Following the previous successful development of bioreactors based on monolithic silica support, this material was tested for tryptic digestion of proteins. Trypsin was immobilized on an

\*Correspondence: Dr S. Rudaz  
School of Pharmaceutical Sciences,  
Ecole de Pharmacie Genève-Lausanne  
Laboratory of Pharmaceutical Analytical Chemistry  
University of Geneva  
20 Bd d'Yvoy  
CH-1211 Geneva 4  
Tel.: +41 22 379 65 72  
Fax: +41 22 379 68 08  
E-Mail: serge.rudaz@pharm.unige.ch

epoxy-modified silica monolithic support with a single reaction step and the resulting bioreactor was coupled through a switching valve to an analytical column for identification of test proteins. The influence of various parameters on enzymatic activity such as flow rate, contact time, bioreactor temperature, buffer type and pH, were investigated by an experimental design. Preliminary experiments were carried out to identify factors acting on the hydrolysis rate of the trypsin-based bioreactor. In this study, six factors mainly related to trypsin activity were examined by applying a Plackett-Burman (PB) design. As the latter saturated design assumes that interaction effects are negligible, only main effects of factors were evaluated. From these results, only one parameter (sample concentration) did not influence the analytical response and the other experimental factors were statistically significant. The flow-rate appeared as the most important factor to control the enzymatic activity of the trypsin bioreactor. Thus, the optimum conditions were retained for the on-line digestion of some proteins on the monolithic bioreactor, coupled to a HPLC/UV/ESI-MS/MS system for the on-line peptide separation. This work was realized in collaboration with the Department of Pharmaceutical Chemistry from the University of Pavia (Italy).

## Chemometrics in Method Optimization

### Chiral Capillary Electrophoresis Coupled with Mass Spectrometry: Optimization of the Partial Filling Technique

The simultaneous evaluation of the experimental factors involved in the enantioseparation of methadone by capillary electrophoresis coupled with mass spectrometry (CE-ESI-MS) with the partial-filling technique was carried out by means of a sequential chemometrics strategy [10]. This work was carried out in collaboration with the Laboratory of Chemometrics from the University Claude Bernard I (Lyon, France). Two different experimental designs were chosen. First, a full factorial design (FFD) was selected to examine the effects and significance of three relevant factors, namely the chiral selector concentration, the separation zone length in the capillary and the nebulization gas pressure, as well as their interactions. A complete investigation was independently conducted for three different cyclodextrins (CD) used as chiral selector and thus three different FFD were performed. For each CD, the following responses were measured: enantiomeric resolution of methadone, migration time, and efficiency for both enantiomers, and apparent selectivity, which gives important in-

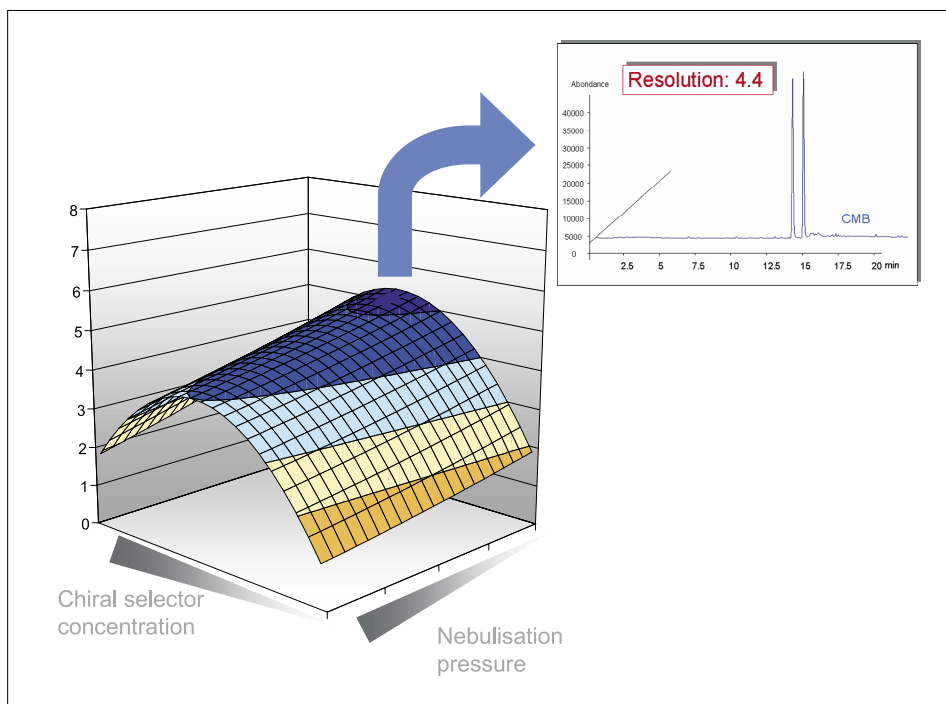


Fig. 1. Response surface plot for the enantiomeric resolution of methadone using a carboxymethyl cyclodextrin (CMB) as chiral selector, as a function of nebulization pressure and CD concentration with a 90% capillary separation zone length

formation about the enantiodiscrimination. Relevant results were obtained by the FFD, but an important curvature was observed for numerous analytical responses. Therefore, the quality of the mathematical model was further improved by means of a central composite design (CCD) allowing the setup of a robust experimental domain. From the drawn response surfaces, an optimum set of experimental factors allowed maximum resolution to be obtained (Fig. 1).

### Modeling of the Separation Buffer Viscosity in Affinity Capillary Electrophoresis

In order to determine the effect of main experimental parameters on enantioselectivity in chiral capillary electrophoresis, the knowledge of analyte-chiral selector complexation constants and complex mobility is of prime importance [11]. Due to its ease of implementation, affinity capillary electrophoresis (ACE) has often been used to determine complexation constants of chiral analytes into CD and to rationalize chiral CE method development. In ACE, obtained experimental effective mobilities should be corrected for viscosity, which depend on variations of chiral selector concentration, ionic strength and temperature. These corrections are crucial for an accurate determination of binding parameters. In order to perform viscosity corrections due to the modification of electrolyte composition, viscosity modeling was undertaken by a CCD, which allows the modeling of quadratic effects of variables. The selected factors were ammonium formate concen-

tration, chiral selector concentration and capillary temperature. Three levels were set for each parameter, leading therefore to use a face-centered CCD. Multiple regressions yielded a mathematical relationship between responses and independent variables. Enough data were generated to fit a second-order polynomial equation. The effects of these factors, as well as their interactions were simultaneously studied by means of 14 randomized experiments. In addition, four additional trials were performed under conditions corresponding to the center of the experimental domain. The detection time of the neutral marker (mesityl oxide) was measured as the analytical response. After data transformation using the Hagen-Poiseuille law, the estimated viscosity value was retained for further data modeling. Finally, estimations of complex mobilities and complexation constants shed more light on enantioselectivity achieved for tramadol, a centrally acting analgesic agent used in the treatment of chronic with pain the chiral selector.

## Chemometrics in Method Validation

### Validation of a Heart-cutting Liquid Chromatography-Mass Spectrometry Method for the Determination of Process-related Substances in Cetirizine Tablets

Because liquid chromatography coupled to mass spectrometry (LC-MS) is particularly advantageous for the analysis of samples containing trace impurities or by-products, a sensitive and simple LC-MS

method was developed for the simultaneous determination of seven cetirizine process-related substances (PRS) in tablets [12]. In order to reduce MS source contamination and ensure the absence of interferences on the PRS specific channels, a heart-cutting approach was employed. The analytical method was fully validated following the approach proposed by the Société française des Sciences et Techniques Pharmaceutiques. Validation of the method established that performance characteristics (figures of merits, quality parameters) were adequate for the intended use. It entailed the evaluation of a number of parameters such as selectivity, concentration–detector response relationship, trueness, precision (repeatability and intermediate fidelity), accuracy, sensitivity, determination limits, and robustness. Robustness tests are performed to evaluate the influence of small changes in the operating conditions on responses and to simulate variations expected during the method transfer. In this study, robustness testing was performed by applying a PB design in which six factors (mainly related to analyte ionization) were examined. The influence of drying gas flow rate and temperature, nebulizing gas pressure, capillary voltage, mobile phase pH, and electrospray voltage was investigated for the two most relevant responses (retention time and peak area) with the seven tested analytes. In addition to regular experiments, supplementary trials were conducted to check method performance as well as to estimate the experimental error. The normalized coefficients were calculated with a multilinear regression approach. None of the tested factors had a relevant effect on PRS areas. Similar results were obtained for retention times. From these observations, it was concluded that the method was robust towards all the considered factors.

### Robustness of Chiral Capillary Electrophoresis for the Determination of Drugs in Biological Matrices

Capillary electrophoresis (CE) was used for the simultaneous chiral determination of Venlafaxine (Vx), an antidepressant drug and its main active metabolite, O-desmethyl venlafaxine (ODV) in biological samples [13]. After optimization and validation of the method, robustness was carried out with the help of a  $2^3$  full factorial design with additional points at the central values to evaluate the standard deviation of the method. The robustness was undertaken in function of the buffer concentration, the chiral selector concentration and the buffer pH (2.4–2.6). Vx and ODV concentrations as well as the enantiomeric ratio were investigated as analytical responses. As presented in Fig. 2, no factors influenced significantly Vx and ODV resolution ( $p = 0.05$ ) because all factors and their confidence intervals include

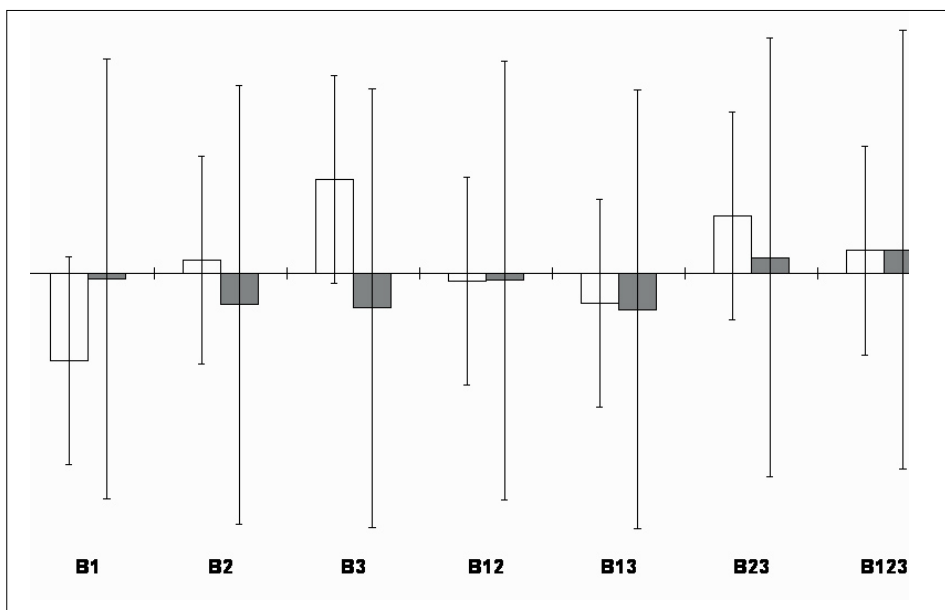


Fig. 2. Effects of main factors and interactions calculated for the resolution of Vx (white) and ODV (grey) obtained by FFD (Tested factors; B1: buffer concentration, B2: chiral selector concentration, B3: buffer pH, B12-B13-B23-B123: interactions between factors, see [12])

the value of zero and were therefore not statistically significant. Furthermore, all interactions between factors were negligible, and thus small variations in BGE preparation did not affect the stereoselective separation of both analytes. The method was therefore considered as robust.

### Robustness in Phytochemistry: Coca Leaf Extraction

An accelerated solvent extraction of cocaine and benzoylecgonine from coca leaves was developed [14]. Several parameters such as nature of extracting solvent, pressure, temperature, extraction time, addition of alkaline substances and sample granulometry were investigated to find the best extraction conditions. Experimental design was used to optimize critical parameters (pressure, temperature, and extraction time) and to assess the robustness of the extraction method. A CCD was used to optimize these three factors, since it offers the possibility to evaluate the robustness of the method by drawing response surfaces. In Fig. 3, influence of pressure and temperature on cocaine recovery is illustrated and demonstrated that pressure had no effect on cocaine extraction efficiency. Therefore, around the central conditions extraction, the method was perfectly robust.

### Chemometrics in Data Treatment

#### Characterization of Chromatographic Supports for the Analysis of Basic Compounds

Today reversed-phase liquid chromatography is the most employed chromato-

graphic technique for the analysis of basic compounds in the pharmaceutical domain. Unfortunately, strong ionic interactions of basic compounds with residual silanol groups on the chromatographic support result in asymmetrical peaks and irreproducible retention. The continuous increase in reversed-phase liquid chromatography techniques has furthered the need for a new generation of 'base-deactivated' stationary phases. Today, more than 100 base-deactivated columns are commercially available. Thus, selecting the appropriate stationary phase for a specific separation remains a challenging task in the development of LC methods.

In order to characterize and evaluate the relative performances of stationary phases, a chromatographic test was developed in the LCAP [15]. A set of 14 test compounds was used to evaluate some base-deactivated chromatographic supports at two pH values (pH 3 and 7) with three isocratic mobile phases. In order to evaluate column and batch variability, five columns for each of the selected supports were tested: three from the same batch and two from different batches. Different chromatographic parameters measured in these conditions allowed column evaluation studies. Therefore, a total of 1890 measured chromatographic parameters were evaluated for each of the studied supports. For a simplified graphical representation of this complex data system, PCA was chosen. The PCA application demonstrated a good classification of the chromatographic supports and allowed the evaluation of their performances in the analysis of basic compounds. The new generation of chromatographic supports

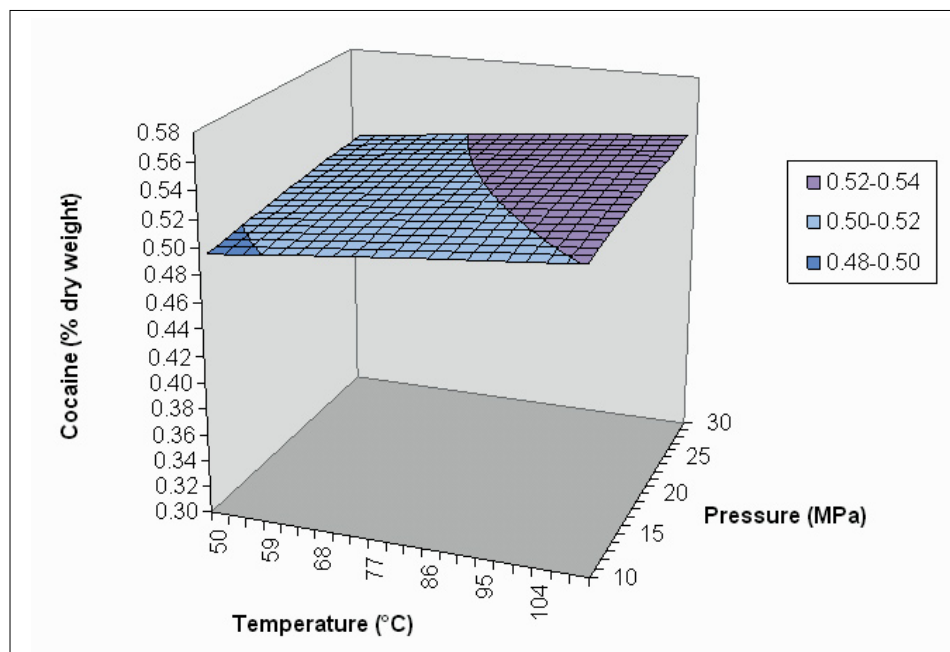


Fig. 3. Response surface plot for the cocaine extraction percentage by varying temperature and pressure

showed better batch and column reproducibility in comparison to conventional supports. This study showed a strong similarity between electrostatically shielded supports. In addition, results demonstrated that some of the test compounds often gave similar information, suggesting a simplified procedure. This simplification allowed a faster and more accessible addition of new supports in the database for further comparison and evaluation. The number of compounds was reduced (7 instead of 14) and selected according to their physico-chemical properties. Only inter-batch and not intra-batch variability was tested (three instead of five

columns per chromatographic support), and two mobile phases were retained for a rapid column evaluation (pH 3 and pH 7). Observations suggested that column evaluation was still ensured when data acquisition was drastically reduced. PCA classification of chromatographic supports with the reduced test was almost identical to that obtained with the complete test. Furthermore, this simplified procedure entailed a large reduction of the number of measured chromatographic parameters per support (252 instead 1890). Score plots obtained for the two mobile phases were in fact almost identical to those previously obtained

with the whole set of data and loading plots suggested that contributions to principal axes did not change. In order to assess the reliability of the obtained classification, hierarchical cluster analysis (HCA) was further performed. Clustering techniques are widely used with different objectives such as visual clustering of variables on the PC axes of projection. As presented in Fig. 4, the six first PC were selected (95% variance explained) and Euclidian distance determination was performed to obtain tree diagrams. The results reported showed the effectiveness of PCA combined with HCA in furnishing interesting information for the description of data samples.

**Data Analysis of Cannabis Samples after Solid-phase Microextraction**

A headspace-solid phase microextraction (SPME) method combined with gas chromatography–mass spectrometry was developed for the direct analysis of cannabis samples. SPME, in headspace mode, was used for the extraction of cannabinoids directly from the plants [16]. The developed method was applied to the analysis of two series of cannabis samples from Switzerland, harvested in Geneva, Neuchâtel, and Zürich. Each sample was analyzed in triplicate. With measured cannabinoid concentrations, a PCA was applied and demonstrated sample discrimination.

Each sample (individual) was described with its content in ten detected cannabinoids (variables), which implied a complex description based on the ten dimensions given by the detected cannabinoids. As previously mentioned, PCA provides the possibility to present in a reduced number of axes the information originally contained in the data. Around 50% of the total variance was explained by the two first principal components (PC1 and PC2). As shown on the score plot in Fig. 5, samples were clustered in two main groups with some outliers. The former included, on one hand, samples from Zürich (except ZH 6), and on the other hand, samples from Geneva (GE 2 and GE 4 to GE 13) and Neuchâtel (NE). This statistical method of data treatment allowed an important discrimination of cannabis samples depending of its origin.

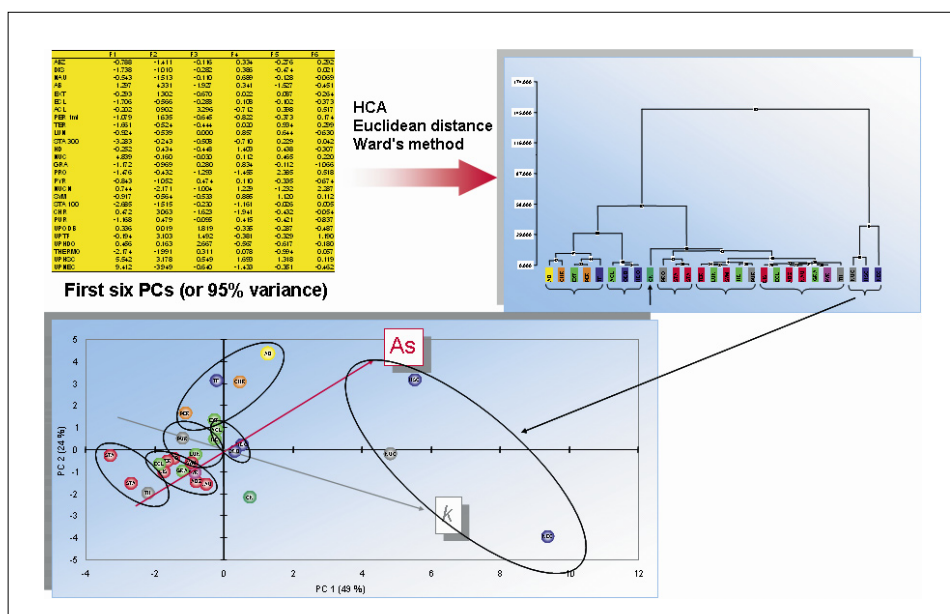


Fig. 4. Distribution of the tested analytical HPLC columns by PCA and HCA. Representation of the score plot.

**Conclusion**

The implementation of chemometric tools in various fields of pharmaceutical analysis has been achieved in the LCAP. Numerous applications demonstrated the usefulness of statistical and mathematical methodologies to improve quality of the results.

From method development to data analysis, well-established chemometric approaches significantly modified the way

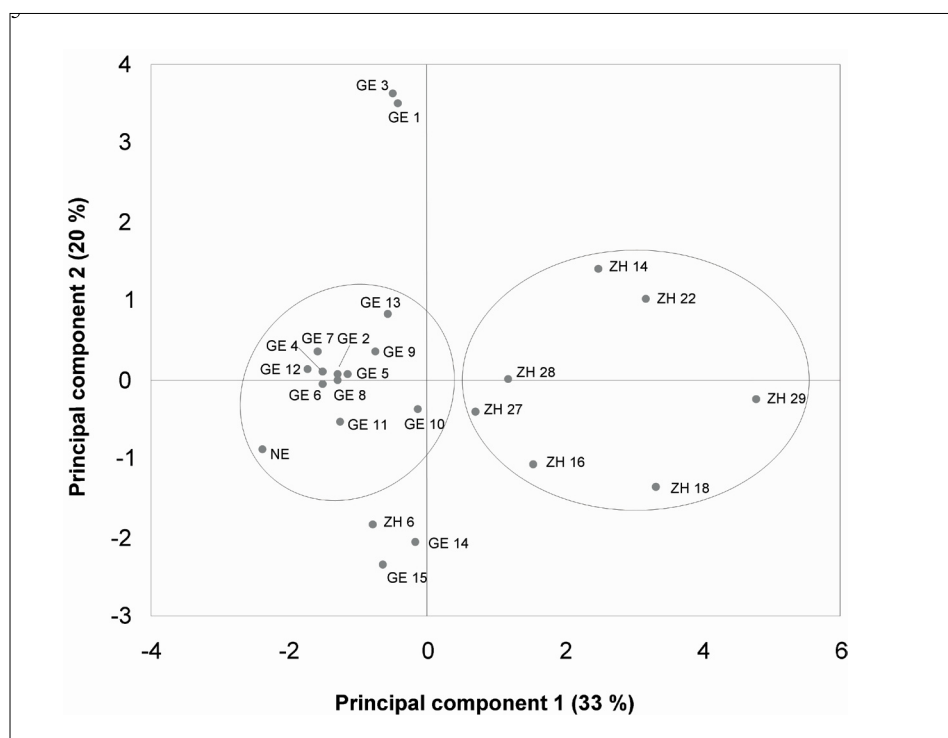


Fig. 5. Distribution of the analyzed samples by PCA. Representation of the score plot. GE = samples from Geneva; ZH = samples from Zurich; NE = sample from Neuchâtel.

to optimize the analytical process. Experimental designs are very useful for the selection of relevant operating parameters in method optimization or for robustness evaluation during validation. The large increase of experimental data, as a direct result of the number and sophistication of the analytical instrumentation, drives the necessity to develop new data treatment techniques and general strategies of extracting scientific knowledge. Therefore, statistical and chemometric tools applied to pharmaceutical and chemical analysis will continue to significantly modify the way that data are collected and information obtained.

Received: April 15, 2005

- [1] D.L. Massart, L. Buydens, *J. Pharm. Biomed. Anal.* 1988, 6, 535–545.
- [2] E. Varesio, J.Y. Gauthier, R. Longerey, P. Lanteri, J.L. Veuthey, *Electrophoresis* 1997, 18, 931–937.
- [3] E. Varesio, J.Y. Gauthier, R. Longerey, P. Lanteri, J.L. Veuthey, *Chromatographia* 1999, 50, 195–201.
- [4] K.D. Altria, B.J. Clark, S.D. Filbey, M.A. Kelly, D.R. Rudd, *Electrophoresis* 1995, 16, 2143–2148.
- [5] V.L. McGuffin, M.F.M. Tavares, *Anal. Chem.* 1997, 69, 152–164.
- [6] S. Rudaz, J.L. Veuthey, *Handbook of Experimental Pharmacology* 2003, 153, 45–75.
- [7] Y. Vander Heyden, K. De Braekeleer, Y. Zhu, E. Roets, J. Hoogmartens, J. De Beer, D.L. Massart, *J. Pharm. Biomed. Anal.* 1999, 20, 875–887.
- [8] E. Hund, Y. Vander Heyden, M. Haustein, D.L. Massart, J. Smeyers-Verbeke, *J. Chromatogr. A* 2000, 874, 167.
- [9] E. Calleri, C. Temporini, E. Perani, C. Stella, S. Rudaz, D. Lubda, G. Mellerio, J.L. Veuthey, G. Caccialanza, G. Massolini, *J. Chromatogr. A* 2004, 1045, 99.
- [10] S. Rudaz, S. Cherkaoui, J.Y. Gauthier, P. Lanteri, J.L. Veuthey, *Electrophoresis* 2001, 22, 3316.
- [11] S. Rudaz, T. le Saux, J. Prat, P. Gareil, J.L. Veuthey, *Electrophoresis* 2004, 25, 2761.
- [12] S. Rudaz, S. Souverain, C. Schelling, M. Deleers, A. Klomp, A. Norris, T.L. Vu, B. Ariano, J.L. Veuthey, *Anal. Chim. Acta* 2003, 492, 271.
- [13] S. Rudaz, C. Stella, A.E. Balant-Gorgia, S. Fanali, J.L. Veuthey, *J. Pharm. Biomed. Anal.* 2000, 23, 107.
- [14] A. Brachet, S. Rudaz, L. Mateus, P. Christen, J.L. Veuthey, *J. Sep. Sci.* 2001, 24, 865.
- [15] C. Stella, P. Seuret, S. Rudaz, P.A. Carrupt, J.Y. Gauthier, P. Lanteri, J.L. Veuthey, *J. Sep. Sci.* 2002, 25, 1351.
- [16] Y. Ilias, S. Rudaz, P. Mathieu, J.L. Veuthey, P. Christen, *Chimia* 2004, 58, 219.