Journal of Chromatography A, 1218 (2011) 4340–4348

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Enhanced monitoring of biopharmaceutical product purity using liquid chromatography–mass spectrometry

Kristoffer Laursen[a,b,*], Ulla Justesen[b], Morten A. Rasmussen[a]

[a] Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark
[b] Novo Nordisk A/S, 2880 Bagsværd, Denmark

ARTICLE INFO

Article history:
Received 8 December 2010
Received in revised form 15 February 2011
Accepted 26 April 2011
Available online 6 May 2011

Keywords:
LC–MS
Impurity detection
Principal component analysis (PCA)
Multivariate statistical process control (MSPC)
Multiple testing
Signal preprocessing

ABSTRACT

LC–MS is a widely used technique for impurity detection and identification. It is very informative and generates huge amounts of data. However, the relevant chemical information may not be directly accessible from the raw data map, particularly in reference to applications where unknown impurities are to be detected. This study demonstrates that multivariate statistical process control (MSPC) based on principal component analysis (PCA) in conjunction with multiple testing is very powerful for comprehensive monitoring and detection of an unknown and co-eluting impurity measured with liquid chromatography–mass spectrometry (LC–MS). It is demonstrated how a spiked impurity present at low concentrations (0.05% (w/w)) is detected and further how the contribution plot provides clear diagnostics of the unknown impurity. This tool makes a fully automatic monitoring of LC–MS data possible, where only relevant areas in the LC–MS data are highlighted for further interpretation.

1. Introduction

Analytical monitoring of impurity profiles in biopharmaceutical products (drug substances and drug products) is important for tracking the product quality. Impurities may potentially have adverse effects and must be identified, qualified, and reported according to the respective thresholds [1,2]. Increasing demands for higher biopharmaceutical product quality has been facilitated by developments in analytical instrumentation and computer systems. This trend leads to new and better tools for monitoring, detection, and identification of new impurities in a timely fashion.

Analytical separation techniques based on high performance liquid chromatography (HPLC) with UV detection are commonly used for determination of impurities in biopharmaceutical products. The separation and subsequent detection of compounds in a sample delivers a chromatogram, which ideally allows separation of peaks which can be attributed to individual chemical compounds. For high-purity drugs, the target compound is present in excess compared to a potential impurity. Hence, detecting the occurrence of an unknown impurity co-eluting with the target compound is a particular problematic challenge. Therefore, purity analysis of a biopharmaceutical product often entails purity examination of the target peak. Peak-purity examination should prevent co-eluting impurities to escape detection in the conventional HPLC analysis [3].

HPLC with diode array detection (HPLC-DAD) is a commonly used method to conduct peak-purity examination. However, many impurities are structurally related to the drug substance, and their structure contains very similar chromophores, making purity assessment based solely on HPLC-DAD data difficult and unreliable. Coupling a mass spectrometer to a liquid chromatograph (LC–MS) brings more selective signals to the table. LC–MS is probably the most powerful technique currently available for pharmaceutical analysis [4]. The technique is still under fast development, particularly in the mass spectrometry area, with vastly improved sensitivity and resolution. However, such state-of-the-art high-resolution instruments are considered rather costly for routine analysis in a pharmaceutical manufacturing environment. Moreover, these high-resolution LC–MS instruments may not contribute with additional required information compared to conventional low cost LC–MS instruments. Since a mass spectrometer (MS) separates compounds by their respective mass-to-charge ratios (m/z), any difference in the m/z values between the impurities and the drug substance will allow an unambiguous detection regardless of similarities in their UV spectra. Therefore an impurity co-eluting with the target peak will be separated in MS as long as their m/z values are different and ionization of the impurity is not suppressed by the target compound. The LC–MS technique is very informative and generates huge amounts of so-called three-way data, where

* Corresponding author at: Novo Nordisk A/S, 2880 Bagsværd, Denmark. Tel.: +45 30795458.
E-mail address: krfl@novonordisk.com (K. Laursen).

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.04.080

each sample is characterized by the intensity as a function of retention time and m/z. However, the relevant information from the chemical point of view is not directly accessible from the raw data map, which makes manual interpretation tedious and often generates a bottleneck in the analysis process [5]. Furthermore manual inspection of LC–MS data is prone to subjective decision-making likely to cause additional errors. Several advanced techniques for the assessment of LC–MS peak purity and co-elution problems have been reported during the last decades [6–9]. However, to comply with increased focus on process analytical technology (PAT) and quality by design (QbD) there is a need for an automatic tool that routinely monitors, detects, and extracts relevant signals from the LC–MS data where further interpretation and identification should be focused. Furthermore, such a tool should detect relevant variation in the LC–MS map quantitatively and in a statistically reliable way. This is a relatively unexplored area in LC–MS data analysis. A powerful tool has recently been demonstrated on chromatographic purity analysis by Laursen et al. [10]. That study demonstrates that multivariate statistical process control (MSPC) based on principal component analysis (PCA) [11,12] applied on chromatographic data is suitable for monitoring subtle changes in the chromatographic pattern. Unknown impurities co-eluting with the target compound were detected in the sum of squared residuals (Q) statistics, and contribution plots provided clear diagnostics of cause of the subtly deviating chromatograms [10]. However, this approach might suffer from lack of sensitivity when applied to LC–MS data. The huge amount of data points combined with the discrete nature of LC–MS signals (i.e. sharp signals in MS direction) makes detection of unknown impurities a case of needle-in-the-haystack expedition. If a new LC–MS sample containing an unknown impurity is fitted to a PCA model based on normal operation condition (NOC) LC–MS samples, the resulting residuals would ideally hold information about the unknown impurity. However, a few discrete residuals related to an unknown impurity would simply be masked when calculating the sum of squared residuals (Q). This makes Q a non-sensitive measure for monitoring and detection of unknown impurities present in low concentrations. Therefore, a more discriminative and sensitive measure is needed targeted towards the nature of LC–MS data. Ralston et al. [13] proposed a statistical enhancement to the typical application of multivariate statistical techniques. The statistical enhancement uses confidence limits on the residuals of each variable for fault detection rather than just confidence limits on the overall Q residual. The method detected faults earlier than the basic Q residual contribution method typically used, but the enhancement proved primarily as a graphical support tool and not as a single value measure for control chart monitoring.

In this study, the approach reported by Laursen et al. [10] is developed to adapt the nature of LC–MS data and to enhance monitoring and detection of unknown impurities in an industrial insulin intermediate (DesB30). In-process samples are spiked with the structurally related human insulin drug product co-eluting with DesB30. MSPC based on PCA is combined with variable wise (multiple) testing. This would enhance detection of discrete residuals from unknown impurities, as residuals of each variable are tested against corresponding model residuals.

2. Theory and methods

The general workflow of MSPC based on PCA follows a previously described trajectory [10,14]. The trajectory is divided in three phases; the initial phase, the training phase and the application phase (ITA). In this modified version, the training phase involving PCA modeling is extended with multiple testing as shown in Fig. 1.

In the initial phase, appropriate historical LC–MS experiments are collected and prepared for PCA modeling. In the training phase a PCA model based on NOC LC–MS samples is developed (describing common cause variation) and multiple testing is applied on the residuals. Finally, in the application phase new samples are fitted to the model and the most significant variable is monitored in control charts developed in the training phase. Deviating samples are diagnosed using multiple testing contribution plots to determine causes of the deviating behavior.

2.1. Signal preprocessing

Once the LC–MS data has been collected, preprocessing methods are required to correct, refine and filter the data. The quality of signal preprocessing is crucial in order to extract relevant (chemical) information. The signal preprocessing was divided into the following steps: baseline correction, normalization, alignment, data reduction, and scaling. The preprocessing steps are described in the following subsections. The practical implications of these preprocessing steps are visualized in the result section.

2.1.1. Baseline correction

Baseline correction is commonly employed to eliminate interferences due to baseline drift. A variety of techniques for baseline correction of LC–MS data are applicable and is reviewed by Listgarten and Emili [15] among others. In this study an efficient and rather simple method for baseline correction is applied. The method works by fitting a global polynomial (of a user-defined order) to each extracted ion chromatogram and, through an iterative routine, down-weighting points belonging to the signal. A baseline is then constructed and subtracted from the original extracted ion chromatogram. Upon selecting the polynomial order and fraction of data points to use for determining the baseline, the algorithm provides an objective and automatic preprocessing. The baseline correction is similar to a previously described method by Gan et al. [16].

2.1.2. Normalization

MS signals are frequently corrupted by either systematic or sporadic changes in abundance measurements. Normalization will correct for bias due to errors in sample amount, possibly sample carry-over and drifts in ionization and detector efficiencies. Normalization procedures enable a more accurate matching and quantification between multiple samples. Different procedures for normalization can be applied. Normalization values can be calculated on the basis of a global distribution for all detected features (like sum, average or median of all intensities per run), or calculated from a specific sub-set of features, for instance from a spiked protein that is used as internal standard [15,17]. In this application the target peak purity might vary but the overall signal intensity should ideally be the same for each sample. Therefore the sum of all intensities is used as normalization value for each sample.

2.1.3. Alignment

As with every laboratory experiment, chromatographic separation is stable and reproducible only to a certain extent. The retention time often shows large shifts, and distortions can be observed when different runs are compared. Even the m/z dimension might show (typically much smaller) deviations. Pressure fluctuations or changes in column temperature or mobile phase may result in shifted peaks.

Alignment of shifted peaks can be performed in various ways. Very reproducible LC–MS data often need only a movement of the extracted ion chromatograms a certain integer sideways for proper alignment. This is characterized by a systematic shift and can easily be handled by the recently published *i*coshift algorithm [18].

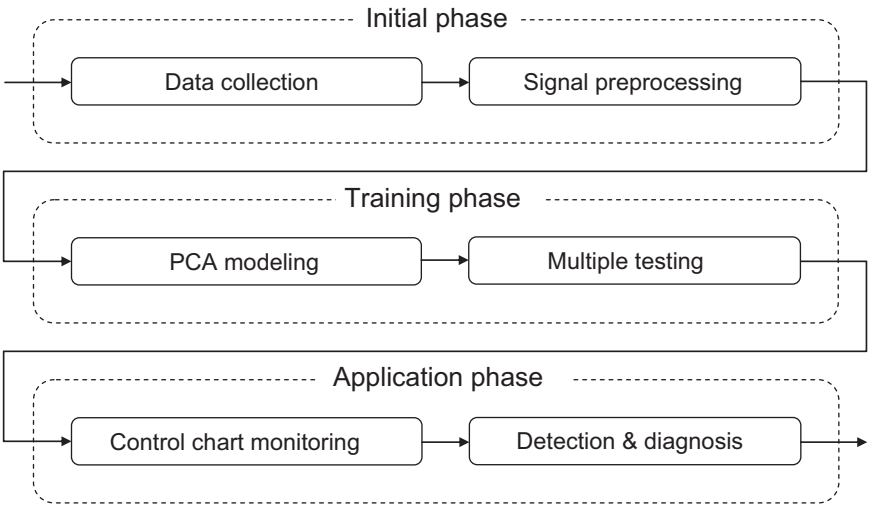

Fig. 1. The three phases according to ITA trajectory (initial, training and application phase).

The *i*coshift algorithm is based on correlation shifting of intervals and employs a fast Fourier transform engine that aligns all spectra simultaneously. The algorithm is demonstrated to be faster than similar methods found in the literature making full-resolution alignment of large datasets feasible [18]. Yet, if peaks shift independently from one another in the same extracted ion chromatogram, more complex shift correction is needed to correct for this non-systematic shift [19,20].

2.1.4. Data reduction

The LC–MS map of a sample is characterized by a collection of intensity measurements as a function of retention time and m/z value. To make the measurements more comparable, and to reduce the huge amount of data points per sample, all intensities within a user-specified bin level are summed. This technique puts all the intensities on a (time, m/z) grid. The bin size is selected based on experience.

2.1.5. Scaling and centering

Scaling is crucial for the performance of the subsequent multivariate statistical analysis. A fold difference in concentration for the target compound and an impurity is not proportional to the chemical relevance of these compounds [21]. Therefore scaling is applied to increase the model sensitivity on detecting small unknown impurities. Furthermore, scaling is crucial in order to bring the distribution of data points close to a normal distribution. This is especially important when multiple testing (like Student's t-test) is used for difference analysis [22]. In many cases, a logarithmic transformation is used for stabilization of the variance. Furthermore, using log-transformed intensities, the disparity in fold differences in between various signals is adjusted. As the final preprocessing step the samples are mean centered (the average unfolded chromatographic pattern is subtracted) to remove a common offset. This brings each variable to vary around zero. This procedure is standard in multivariate modeling that focuses on variability in data.

2.2. MSPC based on PCA modeling combined with multiple testing

PCA and variable wise (multiple) testing offers two different dimensions to statistical data analysis. Multiple testing aims at separating the variable space into variables with a significant- or non-significant change, where PCA separates data into a systematic part (D) and a non-systematic part (Q). In Fig. 2 this is schematized.

Experiments where a high number of variables are evaluated on possibly several outcomes involve testing of numerous hypotheses where handling of error rates is of crucial importance. This discipline is referred to as multiple testing. Multiple testing is widely used for biomarker discovery in proteomics, and has been applied in several difference analyses of LC–MS data intensities [15,23,24]. Both Wiener et al. [23] and Listgarten et al. [24] evaluate the intensity differences between samples from two classes using t-tests on every combination of time and mass to charge ratio, to find regions of interest for further interpretation. However if multiple testing is applied directly to preprocessed LC–MS data it would result in detection of all intensity differences (i.e. both known according to normal operating conditions and unknown features). Multiple testing applied to PCA residuals would only result in detection of unknown features, as the known features are described by the model and expressed in the D-statistics.

With PCA the variation from many correlated (time, m/z) bins in a data matrix $\mathbf{X}$ (with M rows of samples and N columns of bins), can be decomposed into R ($R \leq N$) linear principal components $\mathbf{TP}^{\mathrm{T}}$ and a residual part $\mathbf{E}$ ($M \times N$):

$$\mathbf{X} = t_1 {p_1}^T + t_2 {p_2}^T + \cdots + t_R {p_R}^T + \mathbf{E} = \mathbf{TP}^{\mathrm{T}} + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E} \quad (1)$$

where $\mathbf{T}$ ($M \times R$) is the score matrix and $\mathbf{P}$ ($N \times R$) is the loading matrix, with R components. $\hat{\mathbf{X}}$ is the matrix of predicted values. The correct number of significant principal components can be determined by using cross-validation to eliminate less important directions in the data matrix [25]. In this way the dimensionality of the data matrix is reduced while capturing the underlying relationship between the variables. In standard PCA, each sample is a vector of values. If one sample is a matrix of values (e.g. in the case of LC–MS data), the sample matrix can be unfolded into a vector. This allows standard application of PCA, but throws away some of the information conveyed by storage in a matrix. Using the information contained in all the measured signals simultaneously, MSPC charts are much more powerful in detecting faulty conditions than conventional single variable SPC charts [26]. Once the MSPC chart signals an alarm, the model can be scrutinized to understand the cause of the alarm; hereafter a possible corrective action can be taken. Faults can be due to deviation from *common-cause variation* (detected in Q) and in the *magnitude* of the common cause variation (detected in D). Fault detected in the D chart could for example be caused by an increased amount of already modeled compounds in the sample, and is described by the scores in

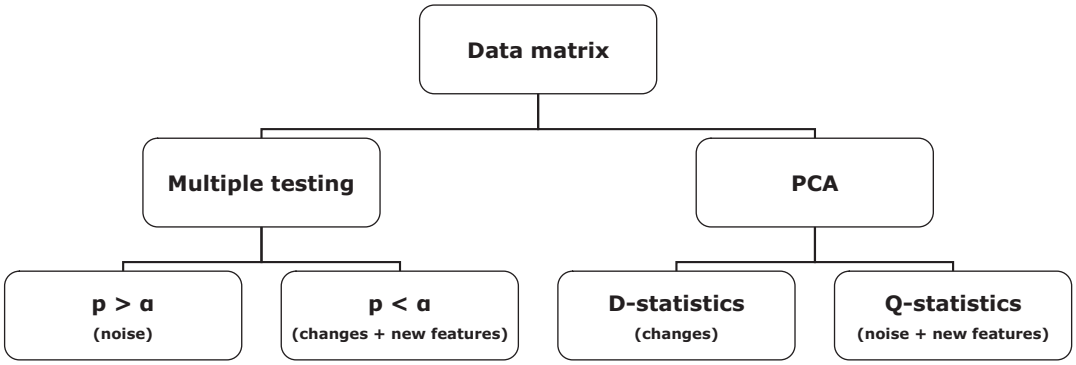

Fig. 2. Schematic overview of two different data analytical approaches for extraction of information from multivariate data. p refers to test probability, α is significance level.

Hotelling's T^2. Hotelling [27] introduced the T^2 for principal components, also referred to as D:

$$T^2 = \sum_{r=1}^{R} \frac{t_r^2}{\sigma_{t_r}^2} \tag{2}$$

where t_r is the rth principal component score, $\sigma_{t_r}^2$ is the variance of the rth component and R denotes the number of principal components retained in the PCA model. Assuming normality for the individual scores, the D-statistic can be expected to approximately follow a weighted F distribution and the upper control limit for the D-statistic can be calculated according to Jackson [28].

If a new sample, containing an unknown impurity, is predicted by the model (based on pure samples), the sample is expected to break the correlation. Indications of the unknown impurity would then be represented in the residuals and monitored in Q:

$$Q = \sum_{n=1}^{N} (x_n - \hat{x}_n)^2 = \sum_{n=1}^{N} (e_n)^2 \tag{3}$$

where x_n and $\hat{x}_n$ are a measurement of the nth variable and its predicted (reconstructed) value, respectively which result in the residual e_n. N denotes the number of variables. Most commonly, a normal distribution to approximate a weighted Chi-square distribution is used from which the upper control limit for the Q-statistic can be calculated according to Jackson and Mudholkar [29].

However, as claimed earlier, a few discrete residuals related to an unknown impurity would simply drown when calculating Q. In order to detect the needle in the haystack we device multiple testing based on a simple t-test for each bin (n) as:

$$t_n = \frac{e_{new,n} - \bar{e}_{ref,n}}{s_n \cdot \sqrt{1 + M^{-1}}} \tag{4}$$

where

$$s_n^2 = \frac{1}{M-1} \sum_{i=1}^{M} (e_{i,n} - \bar{e}_{ref,n})^2 \tag{5}$$

and

$$\bar{e}_{ref,n} = \frac{1}{M} \sum_{i=1}^{M} e_{i,n} \tag{6}$$

where $e_{new,n}$ is the residual from the new sample for bin n, $\bar{e}_{ref,n}$ is the mean of the residuals from the reference samples for bin n. M is the number of reference samples. s_n is the standard deviation of residuals from reference samples for bin n.

The critical value of t is dependent on sample size. In order to correct for this ambiguity t is transformed to a z-value through a p-value:

$$P(T_{df} \leq t_n) = \Phi(z_n) \tag{7}$$

where T_{df} is the t-distribution with df degrees of freedom, $df = M - 1$. Φ is the cumulative distribution function of the standard Gaussian distribution. This z-value is used as diagnostic measure for the corresponding (time, m/z) bin. The z-value and p-value reflects the same statistics (Eq. (7)) and hence the behavior of the system. When dealing with signals of interest in the area of $p < 0.01$, changes are more easily captured by exploring the corresponding z-values e.g. over production time.

2.2.1. Multiple testing

Handling of issues related to multiple testing is becoming more important as number of features detectable from modern analytical instruments is rapidly increasing. For example within the field of proteomics from different platforms such as micro arrays, LC–MS, GC–MS, and NMR often numbers in thousands to tens of thousands or even more is common [30]. Performing numerous univariate significance tests on such highly multivariate data will lead to a high false positive rate (FPR). The conservative Bonferroni factor is a way of controlling the error rate across all tests, known as the family wise error rate (FWER) [31]. The Bonferroni factor is simply a proportionality correction of the p-value threshold (α) with the inverse of the number of test. The Bonferroni correction is a crude up front correction where *all* null hypotheses are assumed true i.e. no difference what so ever. But data is seldom collected under the assumption that there is no relation with a specified outcome. In 1995 Benjamini and Hochberg [31] developed control of false discovery rate (FDR) as an alternative to Bonferroni factor in multiple testing. Estimation of the FDR, contrary to FWER, does not assume that all null hypotheses are true but estimates the proportion of null cases and non-null cases from data. This procedure is shown more powerful in detecting true non-null cases than procedures controlling the FWER [32]. Where the FPR predicts how many of the truly null hypotheses are rejected, the FDR predicts how many of the rejected hypotheses are in fact likely to be truly null. In proteomics the aim is to discover biomarkers in order to develop biological understanding. Here a list of significant biomarkers supported by a FDR is relevant for reporting of results including statistical inference. In MSPC the primary scope is to deem a sample pure or impure and secondly if impure to investigate the impurity contribution. Both cases are dealing with issues related to multiple testing, but as the scope is different, the estimation and extraction of a relevant statistics is likewise. In the following subsection we derive a single measure statistics, and estimate its distribution under normal operator conditions.

2.2.2. Single measure statistic for control chart

In Laursen et al. [10] the Q value was used for a new sample as a measure for detecting subtle differences in the chromatographic pattern. The methodology devised here produces not one but N significance tests where N is the number of bins. These are expressed as a list of z-values; $z_1, z_2, \ldots, z_K$. The largest values of $z_1, z_2, \ldots, z_K$ reflect the bins where the new sample is most deviating. Impurities are in excess and hence only large positive z-values are of interest. The present method proposes use of the *maximum* z-value across all K bins as a measure in control chart monitoring.

2.2.3. Distribution of the maximum z-value across N bins

Under normality assumptions for residuals within each bin, with equal variance for calibration and new samples, the derived t-test statistics is T distributed with $M-1$ degrees of freedom (M number of calibration samples). The corresponding z-values are normally distributed with mean zero and variance one. Assuming independence between the K z-values it is easy to compute the distribution of the maximum z-value:

$$P(z_{\max} \le z) = P(z_1 \le z) \cdot P(z_2 \le z), \ldots$$

$$, P(z_K \le z) = \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt \right)^K \quad (8)$$

In standard two-sided SPC charts an observation more than three standard deviations (3σ) from normal operating conditions is often used as the critical limit. This correspond to a coverage probability of 0.9973 ($1 - 2\Phi(-3) = 0.9973$). As only maximum positive z-values are of interest here, the one-sided control chart threshold should reflect the same coverage probability. In accordance it is possible to calculate the corresponding threshold for the maximum z-values ($z_{0.9973}$) such that $P(z_{max,K} \le z_{0.9973}) = 2\Phi(-3)$. This threshold only depends on number of bins (N). For $N = 1000$, $z_{0.9973} = 4.55$, and for $N = 500$, $z_{0.9973} = 4.40$. Independence between bins might be an overly optimistic assumption, especially when chemical compounds give signal in more than one bin. In order not to rely on assumptions concerning independence we use a heuristic iterative approach on the calibration samples to estimate the critical threshold. The critical 3σ limit is calculated by iteratively testing one reference sample against the remaining reference samples, creating a distribution of $z_{\max}$ values ($z_{\max,1}, z_{\max,2}, \ldots, z_{\max,25}$). From this a 3σ limit is calculated as:

$$Limit_{3\sigma} = \bar{z}_{\max} + 3sz_{\max} \quad (9)$$

where

$$\bar{z}_{\max} = \frac{1}{M} \sum_{i=1}^{M} z_{\max,i} \quad (10)$$

and

$$sz^2_{\max} = \frac{1}{M} \sum_{i=1}^{M} (z_{\max,i} - \bar{z}_{\max})^2 \quad (11)$$

M is the number of reference samples.

3. Experimental

Thirty in-process samples of the insulin intermediate DesB30 were collected for routine quality control testing. All samples were collected under NOC, i.e. the process has been running consistently and only high quality products have been obtained. The 30 samples represent a substantial time period representing possible changes in production. One sample was spiked with human insulin drug product in five various levels from 0.01% to 0.15%. Human insulin is co-eluting with the structurally related target compound DesB30-insulin, but has a different molecular weight and thus different m/z values. Samples were injected into a gradient (0.05% TFA/10% acetonitrile and 0.05% TFA/70% acetonitrile) LC–MS system consisting of an Alliance reverse phase HPLC system (Waters, MA, USA), a Kinetex C18 column (150 mm × 3 mm, 2.6 μm) (Phenomenex, CA, USA), and a MicroTOF-Q II mass spectrometer (Bruker Daltonics, Bremen, D) operated with electrospray (ESI) in the positive ion mode. ESI provides maximum intentsity of the MH^{4+} ions, why this charge state was used in the calculations. All 30 NOC samples were measured in one replicate, whereas the five spiked samples were measured in five replicates each. The LC–MS data was collected and exported as text files using a software tool called DataAnalysis (Bruker Daltonics) and imported to Matlab version 7 (Mathworks, MA, USA) for further analysis. All software was written in Matlab using tools from PLS_Toolbox (Eigenvector Research, WA, USA) and Statistics Toolbox (Mathworks).

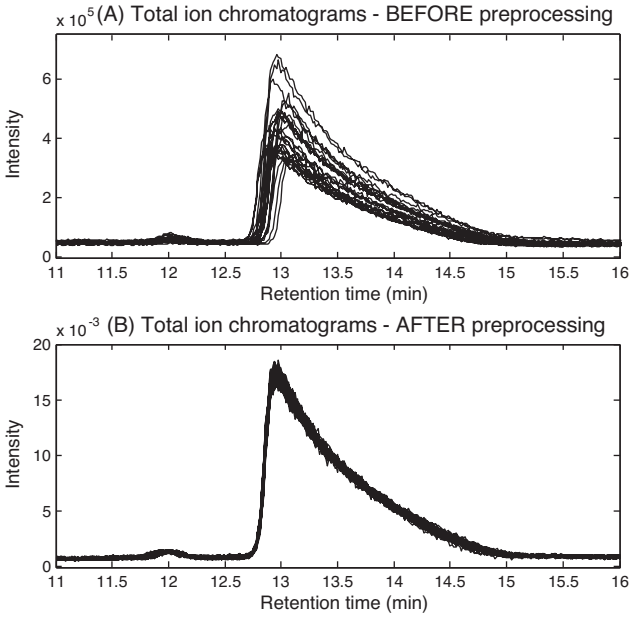

Fig. 3. TIC profiles of all samples before (A) and after (B) preprocessing (baseline correction, normalization and time alignment).

4. Results and discussion

4.1. Initial phase

The 55 LC–MS samples (30 NOC samples and 5 × 5 spiked samples) were collected and organized as an $M \times N \times O$ dataset X, with M samples, N elution times, and O m/z values. A relevant LC–MS window was chosen around the target peak, resulting in a 55 (samples) × 300 (retention times) × 200 (m/z values) dataset. For baseline correction of the data, a second order polynomial was fitted to each extracted ion chromatogram from each sample, based on 50% of all data points. The settings were chosen upon initial investigation of different alternatives. Once the samples were normalized by the sum of all intensities for each sample, time alignment using *i*coshift was sufficient for proper alignment of the LC–MS data. The corrected time axis was calculated from the total ion chromatogram (TIC) profiles, and then applied to the extracted ion chromatograms (EIC) of the corresponding LC–MS sample. The profile which showed the highest correlation with the remaining TIC profiles was selected as the target. For illustrative purpose, the corrected versus the original TIC profiles for all samples is presented in Fig. 3.

To make the measurements more comparable, and to reduce the huge amount of data points per sample, all intensities within a user-specified bin were summed. In this study, intensities within a bin size of 0.5 min and 2 m/z were summed. The bin size was chosen so that single peaks were approximately represented within in a bin. The binning reduced the number of data points from 60.000 to 1000 bin values per sample. Finally, a logarithmic transformation was used to adjust the variation in fold differences between the target peak and minor surrounding peaks, and to reduce the heteroscedasticity of the noise [33]. In Fig. 4, the effect of data reduction transformation of a NOC sample is illustrated, showing that smaller features around the target compound are enlarged due to binning and scaling.

4.2. Training phase

The essence of the training phase is to model the common cause variation present in the LC–MS samples obtained under normal operating conditions. The number of samples needed to construct a representative NOC model and control charts depends on the application. In this case study, a calibration set consisting of the first 25 chronologically ordered LC–MS NOC samples was used to develop a two component PCA model describing nearly 82% of the variation. The optimal number of PCA components to include in the model was selected based on the variance captured and on the results of leave-one-out cross-validation (data not shown). Variance captured flattens out somewhat after two components, and root mean squared error of cross-validation (RMSECV) has a clear local minimum at two components, indicating that after this point, the components just reflect noise. Furthermore, the inspection of loadings confirmed that the first two components reflect real chemical variation (Fig. 5).

The model was validated using an independent validation set consisting of the last five LC–MS samples. By inspection of the D- and Q-statistics (Fig. 6) it was confirmed that two components describe the common-cause variation. All 30 NOC samples were within the 95% quantile in both the D-statistic chart and the Q-statistic chart.

Both D- and Q-statistics are monitored during the training phase. Nevertheless, as this study focuses on purity analysis; we are primarily interested in the residuals. We use the residuals to identify new, unanticipated peaks, which are not part of the normal chromatographic pattern and thus, the model. On the other hand, when developing the model in the training phase, both the

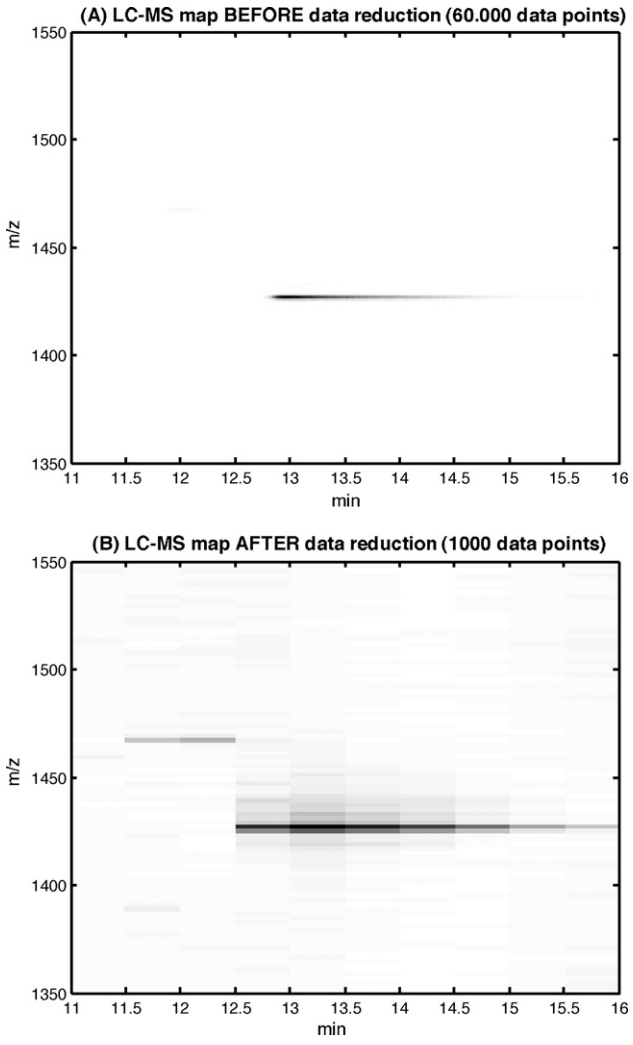

Fig. 4. LC–MS maps before (A) and after (B) data reduction of 60.000 data points to 1000 bin values, using a bin size of 0.5 min and 2 m/z.

D- and Q-statistics are of interest. These statistics may contribute with important and complementary indications about samples to exclude from the NOC model due to deviation in *common-cause variation* (Q) and *magnitude* (D). In this case all 30 samples used

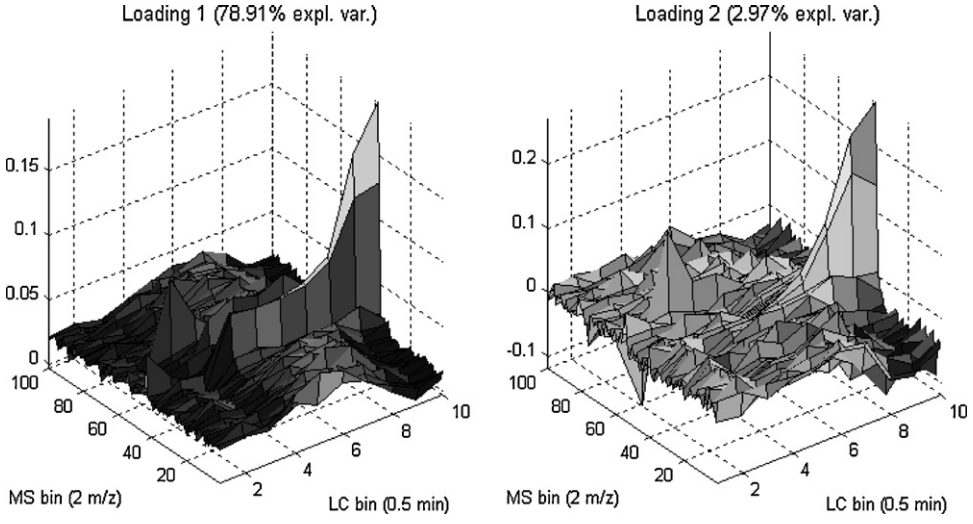

Fig. 5. 3D plot of the first two PCA loadings.

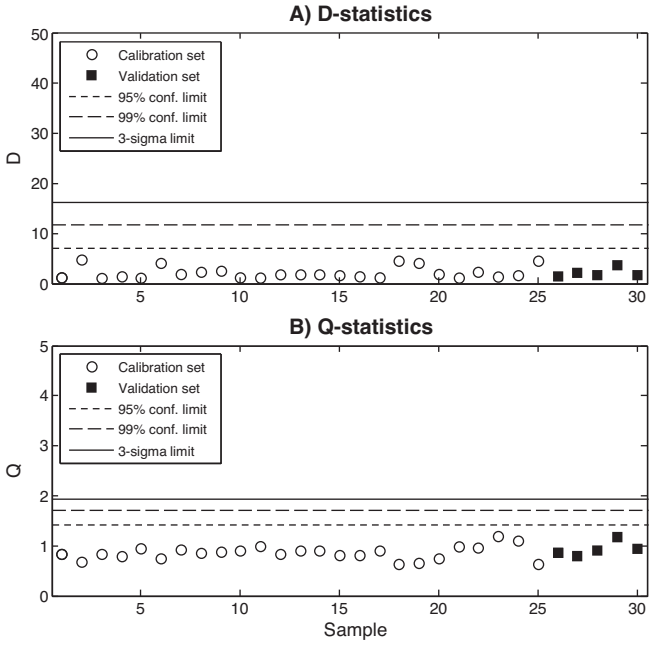

Fig. 6. Plot of (A) *D*-statistics and (B) *Q*-statistics of calibration (circle) and validation (square) sample sets.

in the training phase are within their respective 3σ limits in both *D*- and *Q*-statistics charts, and are therefore assumed to describe common-cause variation.

4.3. Application phase

To demonstrate the lack of sensitivity of ordinary *Q*-based MSPC applied to LC–MS data, a sample from the validation set was spiked with human insulin drug product in five various levels from 0.01% to 0.15%. Human insulin is co-eluting with the structurally related target compound DesB30-insulin, but has slightly different *m*/*z* values. The five spiked samples (measured in five replicates) were used to evaluate the ability of detecting an unknown impurity co-eluting with the target compound. As indicated in the *Q*-statistic chart (Fig. 7) none of the simulated chromatograms were detected as faulty by falling outside the 3σ limit.

As discussed earlier the *Q*-statistic measure suffers from lack of sensitivity due to the needle-in-the-haystack expedition. In Fig. 8 the *Q* contributions are presented for a sample spiked with 0.15% impurity. Even though the contributions provide indications of an abnormality around *m*/*z* 1450–1454 eluting at 12.5–13 min, the

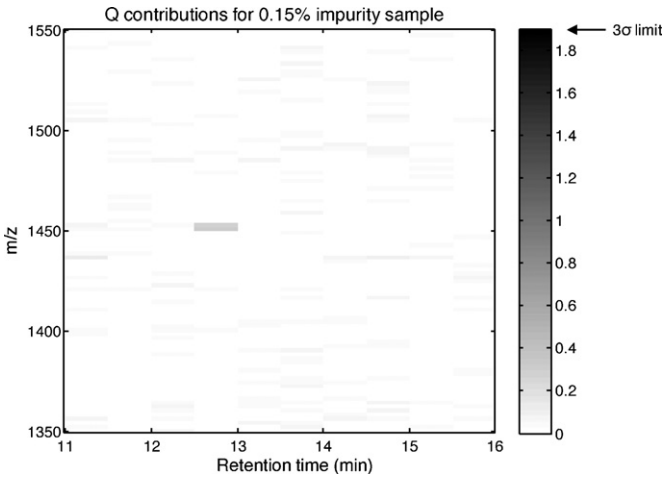

Fig. 8. Plot of *Q* contributions from PCA prediction of sample spiked with 0.15% HI.

relevant diagnostics seems to drown when calculating *Q*. As a consequence the relevant information is not detected and exploited.

Therefore, possible deviations were detected in the *individual* bins using multiple testing rather than testing the overall residual variation. The critical 3σ limit was calculated by iteratively testing one calibration sample against the remaining calibration samples. In comparison with the theoretically derived critical value (4.55), the data generated 3σ limit is slightly lower (3.75). This controversy is primarily due to the incorrect independence assumption which produces a more conservative limit, but maybe also deviation from the normality assumption in the *t*-tests. As indicated in Fig. 9, spike levels down to 0.05% HI was detected as faulty, falling outside the 3σ limit.

The detection level was tested using different selections of bin size and consequently bin number. The detection level is here defined as the lowest spike level where all five replicate samples were detected as faulty, falling outside the 3σ limit. In Fig. 10 the results of different selections of bin size and corresponding impurity detection level is presented. It appears from Fig. 10 that the lowest detection level is obtained with a bin size from 30 to 60 s and 1–2 *m*/*z* value. The number of bins in that region varies from 500 up to 2000 bins. Clearly too high complexity in terms of number of bins will result in a higher critical test limit followed by a higher level of detection. On the other hand in a coarse binning the signal disappears with higher level of detection as a consequence. Though the same consequence, the origin is different for the two cases. For high number of bins the detection limit is dependent on the false positive control in the modeling part, whereas for coarse binning effect

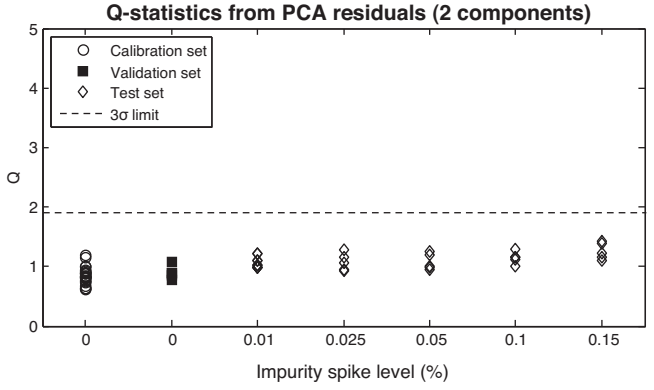

Fig. 7. Plot of *Q*-statistics of calibration- (circle), validation- (square), and test samples (diamond).

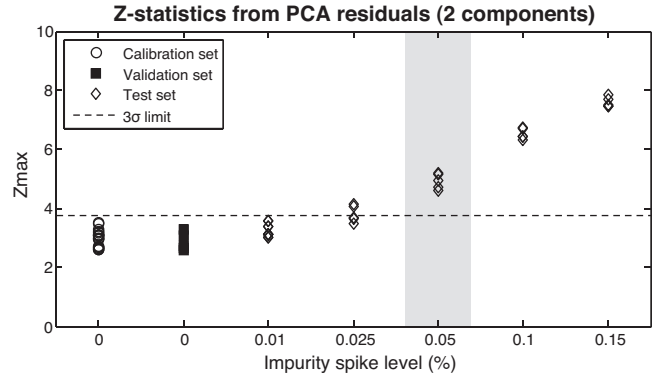

Fig. 9. Plot of *Z*-statistics of calibration- (circle), validation- (square), and test samples (diamond).

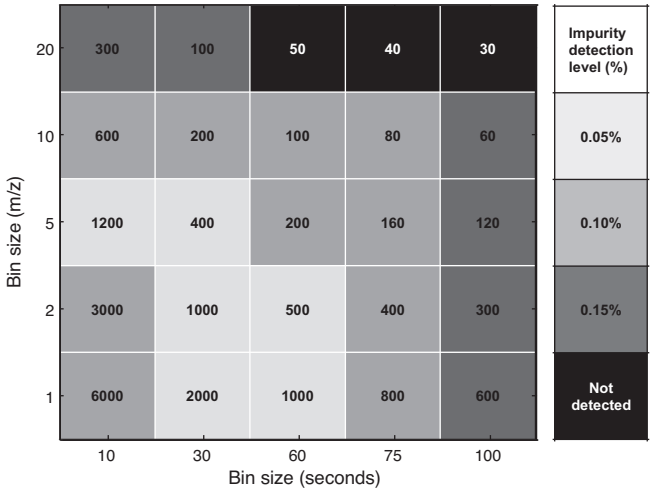

Fig. 10. Results of different selections of bin size (and number of bins) and corresponding detection level. The total bin numbers are indicated in the figure for each bin setting and colored according to the impurity detection level.

size vanishes for impure samples. Of course the impurity detection level examination presented here is optimized for this particular impurity, and is hence slightly biased downwards due to selection of bin size. For a true detection level determination an independent test set could be applied using the selected bin size. Ideally, a more objective method for selection of bin size should be considered. This is more likely an analytical discipline rather than a mathematical discipline. Future unknown impurities eluting close to the drug substance are most likely structurally related to the drug substance, and the impurities can be expected to show up in a 1000-fold difference compared to the drug substance. Hence, the examination presented in Fig. 10 may not be that misleading, and could serve as a preliminary bin-tuning procedure before setting up a reliable monitoring scheme.

To determine those variables responsible for the faulty detection the Z contribution plot is examined (Fig. 11). Clear diagnostics of the detected sample is provided, indicating that an unknown impurity is found around m/z 1450–1454 eluting around 12.5–13 min. Further inspection of the highlighted area (data not shown) revealed clear ion trace signals with a maximum intensity at m/z 1453. For more detailed diagnostics an extracted ion chromatogram (EIC) of m/z 1453 can be examined (Fig. 12).

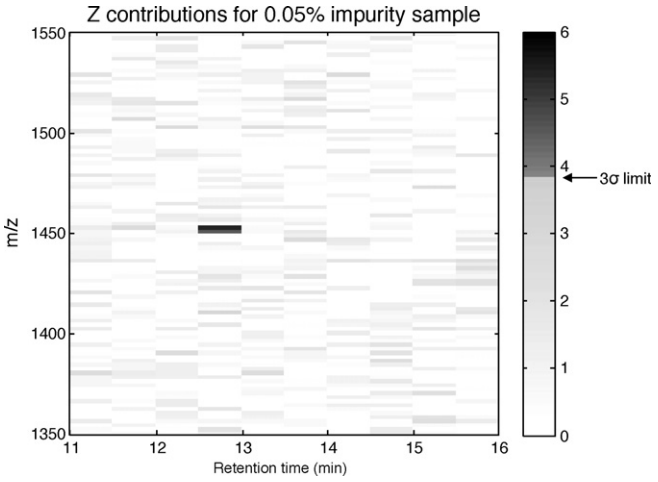

Fig. 11. Plot of Z contributions from PCA prediction of sample spiked with 0.05% HI.

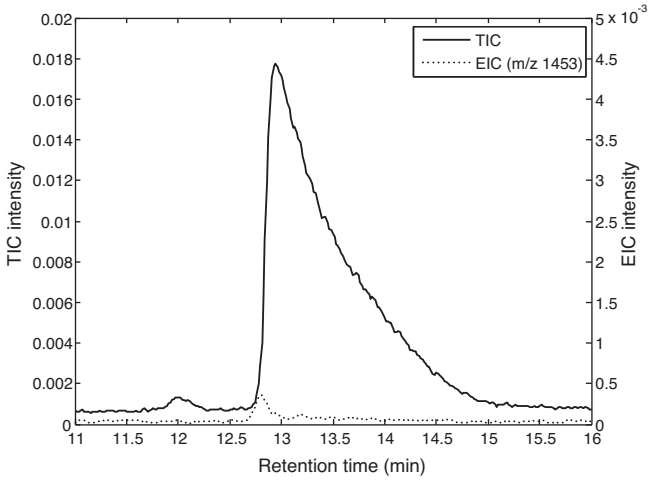

Fig. 12. Plot of TIC and EIC (m/z 1453) of sample spiked with 0.05% HI.

From the EIC the elution profile of the unknown impurity is provided. It would be difficult or impossible to detect a co-eluting 0.05% impurity peak if measured with HPLC. However with LC–MS this challenge is possible and becomes practicable if assisted by the automated methods demonstrated in this study. However, it is important to clarify that MSPC should not be regarded as a replacement of analytical knowledge when interpreting the LC–MS data. Instead, MSPC should be seen as the means for creating robust and highly interpretable multivariate models with the aim of monitoring and detecting unknown features in large and complex LC–MS data.

5. Conclusion and perspectives

This study demonstrates that MSPC based on PCA in conjunction with multiple testing is very powerful for monitoring and detection of unknown and co-eluting impurities measured with LC–MS. A spiked impurity present at low concentrations (0.05%) was detected and comprehensible contribution plot containing clear diagnostics of the unknown impurity was provided. From examination of contribution plots for lower spike levels than 0.05% (0.025% and 0.01%) large contributions from the unknown impurity were highlighted, emphasizing the sensitivity of this method. Trading off false negative signals by lowering of the critical limit from e.g. 3σ to 2σ might enhance the detection limit further. This tool will monitor and highlight only relevant areas in the complex LC–MS data where further effort on interpretation should be applied. Furthermore the tool proved robust towards treating instrumental artifacts such as baseline- and retention time drift. Applying this procedure for the detection of new peaks makes a fully automatic monitoring of LC–MS data possible. Furthermore, if implemented and operating while the purity analyses runs, this tool may considerably reduce time needed for subsequent assessment of data, and operate according to the PAT concept aiming for real-time release. Obviously the actual root cause of the alarm is not automatically given when applying this tool. Such an analysis would need incorporation of chemical and technical process knowledge and possibly applying MS/MS fragmentation for further compound identification. Label-free LC–MS data analysis is already widespread in proteomics and may well be increasingly important in the pharmaceutical industry. However, many different types of applications can be developed with LC–MS. Due to such variety of possible applications and approaches it may also be challenging to develop and incorporate a generic solution for processing and analysis of LC–MS data in commercial software. Nevertheless, this study

point towards development and incorporation of more advanced multivariate data analysis methods in commercial software solutions.

Acknowledgements

The authors thank Professor Rasmus Bro (University of Copenhagen, Department of Food Science), and Niels Væver Hartvig (Novo Nordisk A/S, Compliance Support) for their helpful suggestions on revising this paper.

References

[1] International Conference on Harmonization (ICH) – Guidance for Industry: Q3A(R2) Impurities in New Drug Substances, 2006.
[2] International Conference on Harmonization (ICH) – Guidance for Industry: Q3B(R2) Impurities in New Drug Products, 2006.
[3] K. Wiberg, M. Andersson, A. Hagman, S.P. Jacobsson, J. Chromatogr. A 1029 (2004) 13.
[4] C.K. Lim, G. Lord, Biol. Pharm. Bull. 25 (2002) 547.
[5] M.J. Frederiksson, P. Petersson, B.-O. Alexsson, D. Bylund, J. Sep. Sci. 32 (2009) 3906.
[6] D. Lincoln, A.F. Fell, N.H. Anderson, D. England, J. Pharm. Biomed. Anal. 10 (2010) 837.
[7] J.S. Salau, M. Honing, R. Tauler, D. Barceló, J. Chromatogr. A 795 (1998) 3.
[8] D. Bylund, R. Danielsson, K.E. Markides, J. Chromatogr. A 915 (2001) 43.
[9] E. Per-Trepat, S. Lacorte, R. Tauler, J. Chromatogr. A 1096 (2005) 111.
[10] K. Laursen, S.S. Frederiksen, C. Leuenhagen, R. Bro, J. Chromatogr. A 1217 (2010) 6503.
[11] H. Hotelling, J. Educ. Psychol. 24 (1933) 417.
[12] S. Wold, K. Esbensen, P. Geladi, Chemometr. Intell. Lab. Syst. 2 (1987) 37.
[13] P. Ralston, G. DePuy, J.H. Graham, ISA Trans. 43 (2004) 639.
[14] H.J. Ramaker, E.N.M. van Sprang, S.P. Gurden, J.A. Westerhuis, A.K. Smilde, J. Process Control 12 (2002) 569.
[15] J. Listgarten, A. Emili, Mol. Cell. Proteomics 4 (2005) 419.
[16] F. Gan, G. Ruan, J. Mo, Chemometr. Intell. Lab. Syst. 82 (2006) 59.
[17] S.J. Callister, R.C. Barry, J.N. Adkins, E.T. Johnson, W.J. Qian, B.J. Webb-Robertson, R.D. Smith, M.S. Lipton, J. Proteome Res. 5 (2006) 277.
[18] F. Savorani, G. Tomasi, S.B. Engelsen, J. Magn. Reson. 202 (2010) 190.
[19] N.P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17.
[20] G. Tomasi, F. van den Berg, C. Andersson, J. Chemometr. 18 (2004) 231.
[21] R. van den Berg, H. Hoefsloot, J. Westerhuis, A. Smilde, M. van der Werf, BMC Genomics 7 (2006) 142.
[22] W. Urfer, M. Grzegorczyk, K. Jung, Proteomics 6 (2007) 48.
[23] M.C. Wiener, J.R. Sachs, E.G. Deyanova, N.A. Yates, Anal. Chem. 76 (2004) 6085.
[24] J. Listgarten, R.M. Neal, S.T. Roweis, P. Wong, A. Emili, Bioinformatics 23 (2007) E198.
[25] S. Wold, Technometrics 20 (1978) 397.
[26] A. Ferrer, Qual. Eng. 19 (2007) 311.
[27] H. Hotelling, in: C. Eisenhart, M.W. Ha0stey, W.A. Wallis (Eds.), Techniques of Statistical Analysis, McGraw-Hill, New York, 1947, p. 113.
[28] J.E. Jackson, A User's Guide to Principal Components, John Wiley and Sons, 1991.
[29] J.E. Jackson, G.S. Mudholkar, Technometrics 21 (1979) 341.
[30] J.D. Storey, R. Tibshirani, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 9440.
[31] Y. Benjamini, Y. Hochberg, J. R. Stat. Soc. Ser. B: Methodol. 57 (1995) 289.
[32] J. Trygg, E. Holmes, T. Lundstedt, J. Proteome Res. 6 (2006) 469.
[33] O.M. Kvalheim, F. Brakstad, Y. Liang, Anal. Chem. 66 (1994) 43.