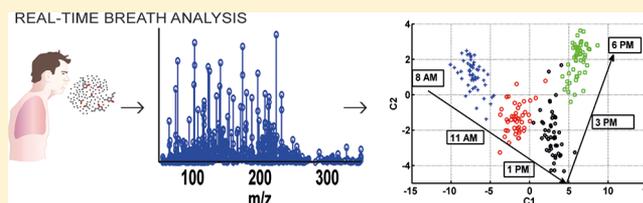


## Monitoring Diurnal Changes in Exhaled Human Breath

Pablo Martinez-Lozano Sinues,<sup>†</sup> Malcolm Kohler,<sup>‡,§</sup> and Renato Zenobi<sup>\*,†</sup><sup>†</sup>Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland<sup>‡</sup>Pulmonary Division, University Hospital Zurich, Zurich, Switzerland<sup>§</sup>Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland

## Supporting Information

**ABSTRACT:** The development of noninvasive analytical techniques is of interest to the field of chronobiology, in order to reveal the human metabolome that seems to show temporal patterns and to predict internal body time. We report on the real-time mass spectrometric analysis of human breath as a potential method to be used in this field. The breath of 12 subjects was analyzed during 9 days by secondary electrospray ionization-mass spectrometry (SESI-MS). The samples were collected during four time slots: morning (8:00–11:00), before lunch (11:00–13:00), after lunch (13:00–15:00), and late afternoon (15:00–18:00). A total of 203 mass spectra were statistically analyzed. Univariate analysis revealed a number of features with a marked temporal behavior. Principal component analysis/canonical analysis showed a clear temporal evolution of the breath patterns. A blind cross-validation yielded 84% of correct classifications of the time slot at which the breath samples were collected. We conclude that this approach seems to have potential for the investigation of biological clocks, including the description of internal body time, which may have important implications for the timing of pharmacotherapy.



Biological clocks have been studied for centuries and nowadays, chronobiology is an active discipline that seeks to understand the effect of time in living systems.<sup>1</sup> For example, the circadian clock is known to play a central role in physiology and its disruption is linked with a number of health disorders like obesity and diabetes.<sup>2</sup> In addition, as our understanding of this process grows, pharmaceutical science is now recognizing that biological clocks play a major role in disease activity, and therefore the time at which a drug should be administered may be important (i.e., chronotherapy).<sup>3</sup> For this reason, important efforts are devoted to reveal the human metabolome that shows temporal patterns<sup>4</sup> and to predict internal body time.<sup>5</sup> Therefore, there is an obvious interest in developing new analytical tools to investigate biological clocks. The present work explores the potential of breath analysis to characterize human biological clocks.

Just like other biofluids (e.g., urine, saliva), exhaled breath carries relevant biochemical information. Given that exhaled breath can be analyzed noninvasively, it is a most appropriate approach to investigate temporal fluctuations of the metabolome. Proton transfer reaction-mass spectrometry (PTR-MS) and selected ion flow tube-mass spectrometry (SIFT-MS) are two real-time techniques available for the analysis of breath. For example, ammonia, acetone, isoprene, and ethanol are among the compounds whose fluctuations over time have been investigated with these techniques.<sup>6</sup> In addition, we have shown that breath can be analyzed in real time by injecting exhaled breath into electrospray plumes of pure solvent, which causes ionization of relevant exhaled compounds.<sup>7</sup> Following this approach, we have detected a number of compounds never observed before in breath by real-time mass spectrometry. In

the present study, we show that this method allows for the investigation of diurnal fluctuations in the composition of exhaled breath in humans.

## EXPERIMENTAL SECTION

**Subjects and Sampling.** A total of 12 nonsmoking subjects (5 females/7 males) participated in this study. The sampling took place twice in the morning (prior to lunch) and twice in the afternoon (after lunch). The samples were allocated into four “time-slot” categories: morning (8:00–11:00), before lunch (11:00–13:00), after lunch (13:00–15:00), and late afternoon (15:00–18:00). Measurements were collected during nine consecutive working days. Note that not all the subjects could join the experiments all the days during the four sampling times, ranging the number of samples per subject from 10 to 26. During each individual measurement, each subject breathed three consecutive times into the system.

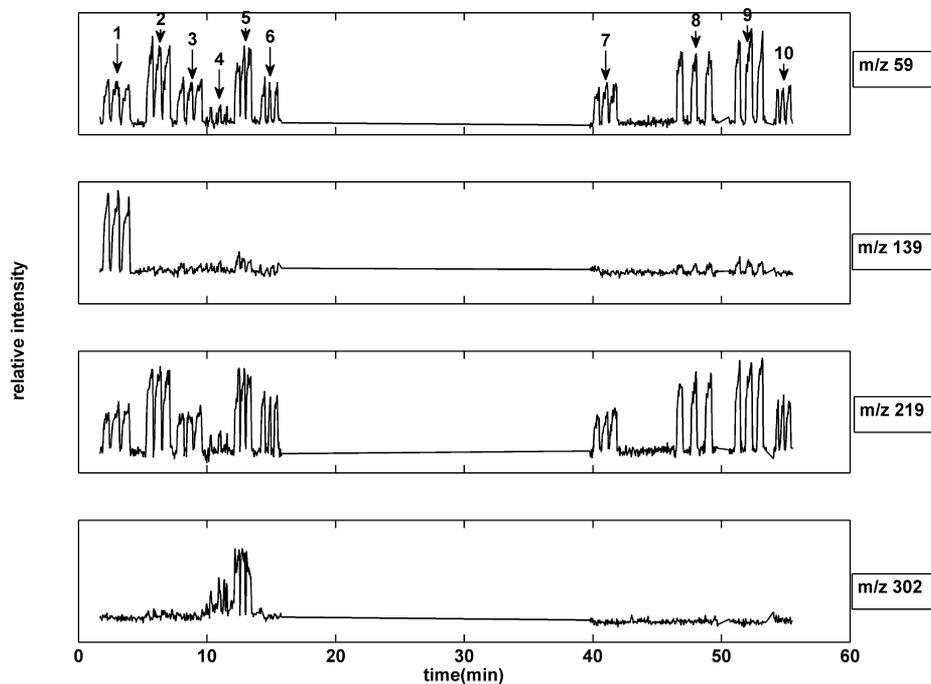
To minimize confounding effects, the participants refrained from eating, drinking, and brushing their teeth at least 30 min prior to the measurements. The local ethical committee approved the study (EK 2012-N-25) and all subjects gave written informed consent to participate.

**Real-Time Mass Spectrometric Measurements.** A quadrupole time-of-flight (QTOF; Waters, Ultima) was slightly modified to allow for the admission of exhaled breath. The participants breathed through a disposable mouthpiece into a

Received: October 6, 2012

Accepted: November 30, 2012

Published: November 30, 2012



**Figure 1.** Typical series of real-time breath analyses recorded from 11:00 a.m. to 11:55 a.m. Relative intensity vs time of four compounds ( $m/z$  59, 139, 219, and 302) present in the breath of ten subjects. Each subject provided three consecutive breath samples, which have a comparable height within each subject, but are different between the individuals.

Teflon tube (50 cm long, 3 mm i.d.), which was connected to the curtain gas channel of the mass spectrometer. The sampling Teflon tube was coated with heating tape maintained at a temperature of 90 °C. Thus, the exhaled breath samples were delivered coaxially to the sampling orifice through the commercial curtain gas nozzle, much in the same way as the early measurements of Fenn and colleagues.<sup>8</sup> To make sure that all the subjects breathed at the same flow rate, they were asked to breathe keeping the pressure through the sampling tube at 20 mbar (monitored by an electronic manometer visible to the subjects). This pressure corresponded to a flow rate of 3.8 L/min. As the breath sample issued the nozzle, it collided with an electrospray cloud formed by a lab-built nano electrospray source (PicoTip emitter i.d. 20  $\mu$ m). Previous mechanistic studies suggest that water-based electrosprays are more efficient in ionizing compounds containing amine moieties,<sup>9</sup> and for this reason we electrosprayed water (0.2% formic acid) infused at  $\sim$ 100 nL/min (2 kV,  $\sim$ 400 nA). The electrospray tip was located 6 mm from the sampling cone and 1 mm off the symmetry axis.

**Statistical Analysis.** The three consecutive exhaled breath mass spectra produced by each subject were averaged by using Water's MassLynx software. The averaged raw mass spectra were then saved as .txt files. These .txt files were post processed by using the commercial software MATLAB (R2011b, Mathworks Inc.). First, each mass spectrum was interpolated to 10 000 mass-to-charge ( $m/z$ ) values (56–400 Da in steps of 0.0187 Da). The 203 mass spectra were normalized by standardizing the area under the curve to the total median. After individually normalizing every signal, they were further scaled to adjust the overall maximum intensity to an overall maximum intensity of 100. Finally, we assembled a 203  $\times$  10 000 matrix with each of these normalized mass spectra. Each of the 203 mass spectra was categorized either as morning ( $n =$

55), before lunch ( $n = 46$ ), after lunch ( $n = 48$ ), or late afternoon ( $n = 54$ ).

First, univariate analysis was employed to extract the most significant features discriminating these four groups and therefore remove the noisiest part of the spectra from further modeling.<sup>10</sup> A Kruskal–Wallis test (nonparametric one-way analysis of variance) was performed for the task. A total of 1761  $m/z$  values yielded  $p$ -values  $< 10^{-4}$  (Bonferroni corrected at 95% confidence interval). Thus, the original 203  $\times$  10 000 matrix was reduced to 203  $\times$  1761 (included as Supporting Information). Further statistical analysis was performed as described previously.<sup>11</sup> The 203  $\times$  1761 matrix was autoscaled.<sup>12</sup> This resulting matrix was subjected to principal component analysis (PCA) and the numbers of principal components explaining most of the variance were identified and used for further analysis. The resulting PCA score submatrix was then subjected to one-way multivariate analysis of variance (MANOVA) which, as performed by MATLAB, provides in addition canonical analysis (CA). The resulting canonical variables are linear combinations of the original variables, chosen to maximize the separation between groups. The interpretation of the resulting discriminant functions to ascertain the contribution of the different  $m/z$  values to the separation between groups was performed by standardizing the coefficients as described previously.<sup>13</sup>

Finally, the breath samples' categories were predicted using a  $k$ -fold cross validation ( $k = 5\%$ ). The data set was randomly split (i.e., Monte Carlo repetitions) into training and test sets 1000 times. During each of these 1000 iterations, the training set was subjected to Kruskal–Wallis/PCA/CA analysis. Then the test set, which was completely disjoint of the model set, was projected into the Kruskal–Wallis/PCA/CA subspace generated by the training set, and finally a  $k$ -nearest neighbor classifier ( $k = 1$ , Euclidean distance) assigned the sample class.

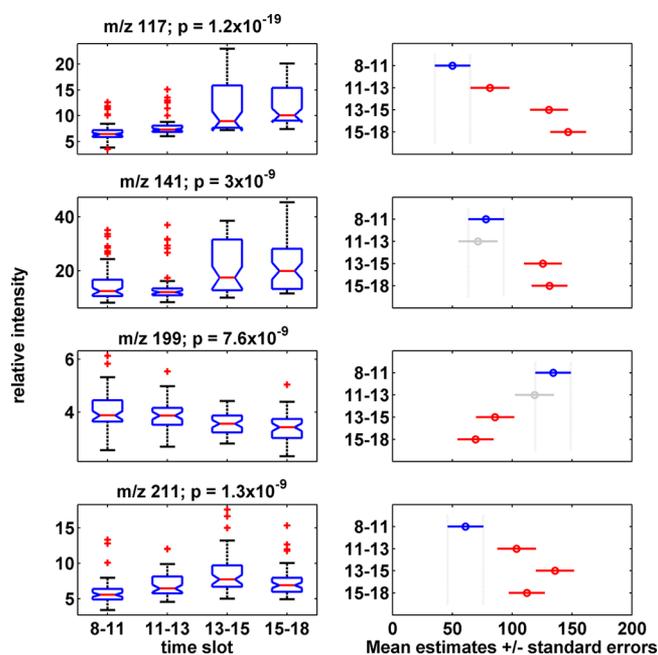
## RESULTS AND DISCUSSION

Figure 1 shows the intensity as a function of time of some compounds detected in breath. It illustrates a typical series of measurements between 11 a.m. to 12 a.m. for ten subjects (labels indicated at the top of the figure). For each compound and for a given individual, the signal increased above the background level during each breath stroke. The three individual consecutive measurements had comparable heights, but there were differences between individuals. For example, the compound at  $m/z$  59 was found to be present at highest concentrations in subjects 2 and 9 and lowest in subject 4. This ion has been previously characterized and assigned to protonated acetone.<sup>7b</sup> This metabolite has been extensively studied by PTR-MS and SIFT-MS. For example, it has been found to show substantial intra- and intersubject variability, with typical concentrations in the 100–1000 parts-per-billion range.<sup>14</sup> In this regard, quantification in SIFT-MS and PTR-MS is more readily achievable than in SESI-MS, mostly due to the different response of the mass spectrometers to which SESI is interfaced. For this reason, calibrations with standards are required in SESI-MS for quantification. For example, concentrations of free fatty acids in the 100 ppt range were found in breath following such a procedure.<sup>7d</sup> Similar quantitative measurements have concluded that limits of detection below ppt are achievable by combining SESI with modern mass spectrometers.<sup>15</sup>

Some other compounds could only be detected in a few subjects. For example,  $m/z$  139 was prominent for subject 1 and almost absent or completely absent for the rest of the individuals. The existence of individual metabolic phenotypes has been shown through the analysis of urine and plasma via nuclear magnetic resonance (NMR).<sup>11</sup> The data shown in Figure 1, previous work,<sup>7d</sup> and further (unpublished) results supports this hypothesis through the analysis of exhaled breath.

With the aim of identifying individual peaks which showed significant differences between the time slots during which the breath samples were analyzed, we conducted a Kruskal–Wallis test. Figure 2 displays four box plots for some of the representative peaks showing different intensities depending on the time of the day at which the breath samples were measured. The  $m/z$  value and the corresponding  $p$  value are quoted on top of each box plot. For example,  $m/z$  117 was found to increase during the course of the day. A multi-comparison test (Bonferroni; 95% confidence interval) revealed that “morning” and “before lunch” measurements had significantly different mean ranks compared to the other three groups. One peak at  $m/z$  141 showed comparable average values for both morning measurements, and higher (also comparable) values for the afternoon measurements. The multicomparison confirmed significant differences between the two morning measurements vs the two afternoon experiments. Some other peaks showed a different behavior. For example, the peak at  $m/z$  211 increased steadily over the morning, reached a maximum after lunch, and declined during the late afternoon. The multicomparison indicated that after lunch levels of this compound were significantly higher than early and before lunch measurements. In contrast, one peak at  $m/z$  199 declined progressively from morning to afternoon. The Bonferroni posthoc test yielded that both morning values were significantly higher than the afternoon mean values.

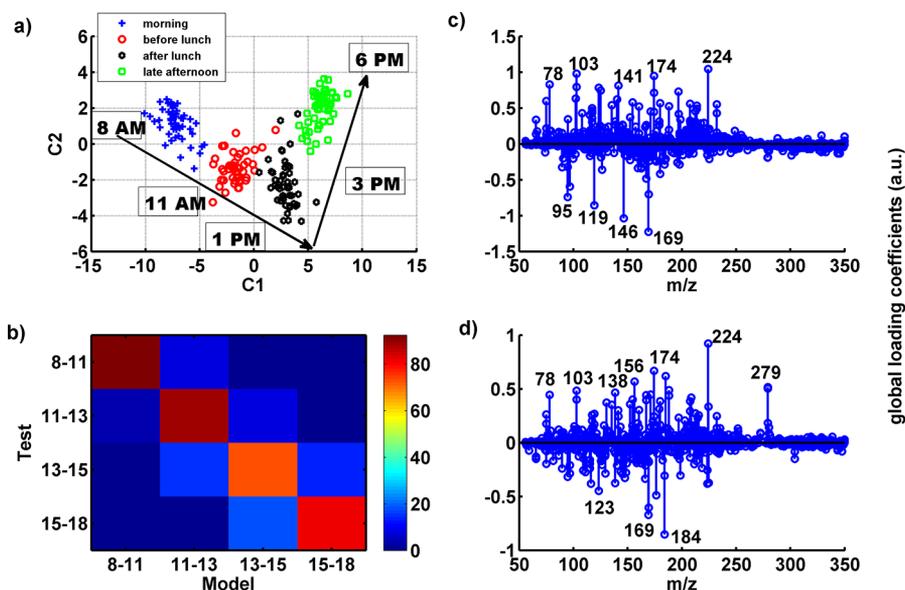
This illustrates that a common core of compounds present in the breath of a group of healthy volunteers seem to have a



**Figure 2.** Box plot of some exemplarily features found to show temporal oscillations. The  $p$ -values shown on the top of the figures were computed upon a Kruskal–Wallis test. The figures on the right show the results of a multicomparison test (Bonferroni corrected). They display each group’s mean represented by “O” and an interval around the symbol (95% confidence level). Two means are significantly different if their intervals are disjoint, and are not significantly different if their intervals overlap.

similar temporal behavior during the course of the day. This notion was strengthened through further multivariate analysis. As described above, the 1761  $m/z$  variables resulting from the Kruskal–Wallis test were retained for PCA. The resulting score submatrix (110 principal components) explaining 99.5% of the variance was subjected to CA, which yielded the existence of three dimensions (all  $p$ -values  $< 0.005$ ) which maximize the separation of the four categories analyzed. Figure 3a shows the projection of the original mass spectra onto the first two canonical dimensions. It becomes clear that there exists a temporal evolution during the course of the day common to all subjects investigated and repeated during 9 nonconsecutive days. Parts c–d of Figure 3 show the loading plots, illustrating the relative contributions of each peak in the spectrum to the separation shown in Figure 3a.

The PCA/CA model was further challenged through a  $k$ -fold cross validation ( $k = 5\%$ ) with the aim of discarding potential overtraining. Thus, 10 “unknown” breath mass spectra were projected onto the PCA/CA subspace generated by the remaining 193 mass spectra, and classified accordingly (knn classifier). This process was repeated 1000 times and the test samples were each time shuffled (i.e., Monte Carlo repetitions) and the most significant features reselected in each iteration. The overall correct classification score was 84%. The confusion matrix shown in Figure 3b summarizes the results of this blind classification exercise. The reddish diagonal indicates that most of the times the test sample presented to the model was correctly classified. Clearly, each time slot of the day investigated had a common characteristic breath signature in all subjects that could be recognized. Interestingly, when this was not the case, the sample was most of the times assigned to an adjacent class. For example, 92.5% of the morning samples



**Figure 3.** (a) Projection of the 203 breath mass spectra onto a PCA/CA 2-dimensional subspace. The temporal evolution during the course of the day becomes evident. (b) Validation of the model: confusion matrix summarizing the classification results of a  $k$ -fold cross validation ( $k = 5\%$ ) repeated 1000 times. The average recognition score of the time slot at which the sample was exhaled was 84%. Note that in the cases in which the breath sample was misclassified, it was confused with an adjacent time slot. (c and d) Global loading coefficients representing the relative contribution of each of the mass spectral peaks to the separation shown in part a. Positive values in part c (dimension C1) denote increased levels in measurements taken after 1 p.m., and negative values decreased levels. Positive values in part d (dimension C2) represent relative increased levels in morning and late afternoon vs. before and after lunch measurements, and negative values relatively decreased levels.

(8–11) were correctly classified, while 7.5% of them were assigned as “before lunch” (11–13). A plausible explanation for the misclassified mass spectra is that, even though the sample was collected and tagged at a given time slot, the internal body time—and therefore the breathprint—was closer to the adjacent time slot. Note also that the discretization of the samples into the four classes led to some samples collected some 10 min apart, but falling into different time-slots categories. Thus it is reasonable to obtain erroneous classifications with the adjacent categories. For example, the worst case was for the measurements taken between 1 p.m. and 3 p.m. (*after lunch*), which were correctly classified 72% of the times, and it was confused with measurements from 11 a.m. to 1 p.m. 15% of the times, and 13% of the times with breath samples recorded afterward (3 p.m. to 6 p.m.).

Biological clocks play a major role in many organisms. A number of chronobiology studies covering gene expression and metabolome profiling have sought to better understand the underlying mechanisms of biological clocks. In this regard, the estimation of internal body time is essential in chronotherapy.<sup>16</sup> A recent study has shown that by identifying temporal fluctuations via liquid-chromatography–mass spectrometry (LC-MS) in plasma metabolites, internal body time could accurately be predicted within 3 h with two samples drawn 12 h apart.<sup>5</sup> We have shown that, similarly to LC-MS identification of plasma fluctuating metabolites, real-time breath analysis can be used to track human physiological modulations ultimately reflected in exhaled breath. Moreover, it should be possible to predict the time at which a given sample was collected with a precision of at least 2–3 h.

## CONCLUSIONS

By monitoring the diurnal variations of the breath composition of 12 individuals during 9 days, we found a group of compounds whose concentrations in breath significantly differ,

depending on the time of the analysis. Statistical analysis of the breathprints reveals a clear temporal evolution common to all subjects. Furthermore, the time slot at which the breath samples were analyzed could be correctly predicted in 84% of the cases. We conclude that real-time mass spectrometric breathprinting may become a useful analytical approach, complementary to traditional techniques (i.e., NMR and LC/MS) to investigate open questions in the field of chronobiology. In particular, for the development of methods to estimate individual's internal body time to maximize the efficiency of drug delivery (i.e., chronotherapy). Further investigations will be conducted to extend these measurements during a 40 h cycle to cover the whole circadian clock.

## ASSOCIATED CONTENT

### Supporting Information

Matrix data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [zenobi@org.chem.ethz.ch](mailto:zenobi@org.chem.ethz.ch).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This research was supported by a Marie Curie European Reintegration Grant (to P.M.L.S.) within the 7th European Community Framework Programme (276860). René Dreier is gratefully acknowledged for constructing the breath sampling device. We thank Dr. Robert Dallmann for his numerous suggestions to improve this study, and the volunteers for providing their breath samples.

## ■ REFERENCES

- (1) Piggins, H. D.; Guilding, C., *Chronobiology*; Portland Press: Colchester, UK, 2011.
- (2) (a) Levi, F.; Schibler, U. *Annu. Rev. Pharmacol. Toxicol.* **2007**, *47*, 593–628. (b) Delezie, J.; Challet, E. *Ann. N.Y. Acad. Sci.* **2011**, *1243*, 30–46.
- (3) Farrow, S. N.; Solari, R.; Willson, T. M. *Expert Opin. Drug Discovery* **2012**, *7*, 535–541.
- (4) (a) Dallmann, R.; Viola, A. U.; Tarokh, L.; Cajochen, C.; Brown, S. A. *Proc. Natl. Acad. Sci.* **2012**. (b) Eckel-Mahan, K. L.; Patel, V. R.; Mohny, R. P.; Vignola, K. S.; Baldi, P.; Sassone-Corsi, P. *Proc. Natl. Acad. Sci.* **2012**, *109*, 5541–5546.
- (5) Kasukawa, T.; Sugimoto, M.; Hida, A.; Minami, Y.; Mori, M.; Honma, S.; Honma, K.-i.; Mishima, K.; Soga, T.; Ueda, H. R. *Proc. Natl. Acad. Sci.* **2012**.
- (6) (a) Diskin, A. M.; Španěl, P.; Smith, D. *Physiol. Measurement* **2003**, *24*, 107. (b) King, J.; Kupferthaler, A.; Unterkofler, K.; Koc, H.; Teschl, S.; Teschl, G.; Miekisch, W.; Schubert, J.; Hinterhuber, H.; Amann, A. *J. Breath Res.* **2009**, *3*, 027006.
- (7) (a) Chen, H.; Wortmann, A.; Zhang, W.; Zenobi, R. *Angew. Chem., Int. Ed.* **2007**, *46*, 580–583. (b) Martínez-Lozano, P.; de la Mora, J. F. *Int. J. Mass Spectrom.* **2007**, *265*, 68–72. (c) Martínez-Lozano, P.; Fernández de la Mora, J. *Anal. Chem.* **2008**, *80*, 8210–8215. (d) Martínez-Lozano, P.; Zingaro, L.; Finiguerra, A.; Cristoni, S. *J. Breath Res.* **2011**, *5*, 016002. (e) Gamez, G.; Zhu, L.; Disko, A.; Chen, H.; Azov, V.; Chingin, K.; Kramer, G.; Zenobi, R. *Chem. Commun.* **2011**, *47*, 4884–4886.
- (8) Whitehouse, C. M.; Levin, F.; Meng, C. K.; Fenn, J. B. in *Proceedings of the 34th ASMS Conference*, Cincinnati, OH, 1986.
- (9) Martínez-Lozano Sinues, P.; Criado, E.; Vidal, G. *Int. J. Mass Spectrom.* **2012**, *313*, 21–29.
- (10) Theodoridis, S.; Pikrakis, A.; Koutroumbas, K.; Cavouras, D. *Introduction to Pattern Recognition: A Matlab Approach*; Academic Press: London, UK, 2010; p 240.
- (11) (a) Assfalg, M.; Bertini, I.; Colangiuli, D.; Luchinat, C.; Schäfer, H.; Schütz, B.; Spraul, M. *Proc. Natl. Acad. Sci.* **2008**, *105*, 1420–1424. (b) Bernini, P.; Bertini, I.; Luchinat, C.; Nepi, S.; Saccenti, E.; Schäfer, H.; Schütz, B.; Spraul, M.; Tenori, L. *J. Proteome Res.* **2009**, *8*, 4264–4271.
- (12) van den Berg, R.; Hoefsloot, H.; Westerhuis, J.; Smilde, A.; van der Werf, M. *BMC Genomics* **2006**, *7*, 142.
- (13) (a) Rencher, A. *Methods of Multivariate Analysis*; John Wiley & Sons: New York, NY, 2002. (b) Napoli, C.; Sperandio, N.; Lawlor, R. T.; Scarpa, A.; Molinari, H.; Assfalg, M. *J. Proteome Res.* **2011**, *11*, 1274–1283.
- (14) Španěl, P.; Dryahina, K.; Rejskova, A.; Chippendale, T. W. E.; Smith, D. *Physiol. Meas.* **2011**, *32*, N23–N31.
- (15) Martínez-Lozano, P.; Rus, J.; Fernández de la Mora, G.; Hernández, M.; Fernández de la Mora, J. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 287–294.
- (16) Hrushesky, W. J. *Science (N.Y.)* **1985**, *228*, 73–75.