*Anal. Chem.* **2006,** *78,* 4281−4290

*Articles*

# Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in $^1$H NMR Metabonomics

**Frank Dieterle, Alfred Ross, Götz Schlotterbeck, and Hans Senn***

*F. Hoffman-La Roche Ltd, Pharmaceuticals Division, PRBD-E, Building 065/512, 4070 Basel, Switzerland*



**For the analysis of the spectra of complex biofluids, preprocessing methods play a crucial role in rendering the subsequent data analyses more robust and accurate. Normalization is a preprocessing method, which accounts for different dilutions of samples by scaling the spectra to the same virtual overall concentration. In the field of $^1$H NMR metabonomics integral normalization, which scales spectra to the same total integral, is the de facto standard. In this work, it is shown that integral normalization is a suboptimal method for normalizing spectra from metabonomic studies. Especially strong metabonomic changes, evident as massive amounts of single metabolites in samples, significantly hamper the integral normalization resulting in incorrectly scaled spectra. The probabilistic quotient normalization is introduced in this work. This method is based on the calculation of a most probable dilution factor by looking at the distribution of the quotients of the amplitudes of a test spectrum by those of a reference spectrum. Simulated spectra, spectra of urine samples from a metabonomic study with cyclosporin-A as the active compound, and spectra of more than 4000 samples of control animals demonstrate that the probabilistic quotient normalization is by far more robust and more accurate than the widespread integral normalization and vector length normalization.**

## INTRODUCTION

Among the postgenomic technologies, metabonomics has recently gained increasing attention as it allows the generation and extraction of comprehensive biochemical information, which is of diagnostic and prognostic value. Metabonomics, which is also called metabolomics[1] or metabolic profiling,[2] involves the determination of changes in the concentration levels of small endogenous metabolites in biological samples resulting from physiological stimuli or genetic modification.[3] The power of metabonomics can be traced back to fast and straightforward measurements without extensive sample preparation steps, high reproducibility, especially in the case of NMR, and direct reflection of biological events. Until now, mainly $^1$H NMR spectroscopy has been used for measuring biofluids of mammals. The spectroscopic data are subsequently analyzed by multivariate data analysis methods to extract information about changes in the distinct metabolites, which directly represent the functional endpoints of biological action or drug effects.

Thereby data preprocessing is a crucial intermediate step transforming raw spectroscopic data for dedicated data analysis methods. Data preprocessing transforms the data in such a way that subsequent analyses are easier, more robust, and more accurate. For NMR spectra of metabonomic studies, preprocessing methods are typically used to reduce variances and influences, which might interfere with data analysis. These variances can be caused by typical NMR effects such as incorrect phasing of the spectra or varying baselines, by biology-related effects such as varying overall concentrations of biofluids, and by combined effects such as shifting peak locations. In addition to typical data processing techniques for NMR spectra such as baseline correction and phasing, two preprocessing steps are crucial for analyzing a set of complex biofluids.

The first preprocessing step accounts for the effect of shifting peaks. The matrix of biofluids and in particular urine is highly variable because of changes in the pH, salt concentration, overall dilution of samples, relative concentrations of specific ions, relative concentration of specific metabolites, and many more. All these parameters can influence the chemical shifts of peaks; not all peaks are affected, and different peaks are affected to a different extent even when they belong to the same metabolite. Although it has been demonstrated in a recent publication[4] that peak shifts can be beneficial for separation of different groups of samples under certain circumstances, this effect is unwanted for most applications. Two approaches have been proposed to handle variations

* To whom correspondence should be addressed. E-mail: hans.senn@roche.com. Tel: +41(0)61 6882028. Fax: +41 (0)61 6887408.

(1) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155−171.
(2) Trethewey, R N.; Krotzky, A. J.; Willmitzer, L. *Curr. Opin. Plant Biol.* **1999**, *2*, 83−85.
(3) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181−1189.
(4) Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *Anal. Chem.* **2005**, *77*, 517−526.

10.1021/ac051632c CCC: $33.50 
Published on Web 05/28/2006

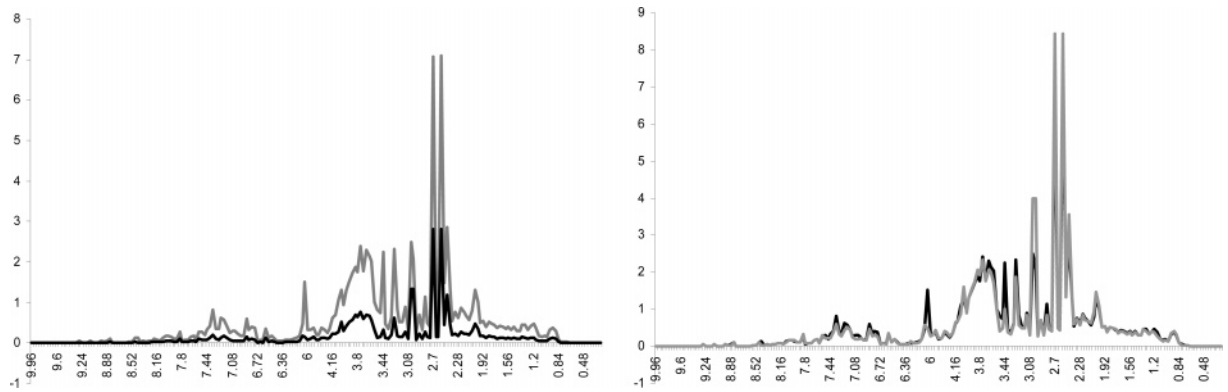

**Figure 1.** Two binned NMR urine spectra before normalization (left side) and after normalization (right side). The black spectrum is diluted by a factor of 3 compared to the gray spectrum. After normalization, it is obvious that only very few differences of the relative concentrations of metabolites between both spectra exist.

in chemical shifts: the first approach is based on a reduction of the spectral resolution by equidistant or nonequidistant binning methods, and the second approach is based on an alignment of the peaks with the help of genetic algorithms, beam searches, or sophisticated correlations.[5−8]

The second crucial preprocessing step for NMR spectra of metabonomic studies is the so-called normalization. The normalization step tries to account for variations of the overall concentrations of samples caused by different dilutions. Especially for urine, unspecific variations of the overall concentrations of samples are very distinctive. Dependent on their water uptake behavior, the urine of animals and humans typically spans a range of half an order of magnitude, but dilution resulting from food deprivation or drug effects can exceed a factor of 10 compared to that of normal urines.[9] In addition, sample preparation steps, such as the sample-specific addition of buffer, can further change the dilution of urine. Dilution is defined as a process, which influences the concentrations of all metabolites and thus all intensities of peaks of the corresponding spectrum by the same factor, which can also be referred to as unspecific changes of metabolites. In contrast, metabonomic responses and fluxes mainly influence a few metabolites in body fluids and, consequently, only a few peaks of the corresponding spectrum. These specific changes are visible as relative changes of concentrations of few metabolites related to the concentrations of all other metabolites, which represent the overall concentration of the sample. Usually these specific relative changes are of interest in metabonomic studies in contrast to the overall concentration of the sample. Therefore, a normalization step, which compensates for the differences of the overall concentration, is crucial, as variations of the overall concentrations obscure specific changes of metabolites. An example of two spectra of urine samples before normalization and after normalization is shown in Figure 1. It is obvious that before normalization the difference of the overall concentrations obscures the fact that only very small differences of relative concentrations between these two samples exist. These small differences are only visible after an adequate normalization of the spectra.

In additon to changes of the overall concentration of samples caused by biological factors, normalization can also be necessary for technical reasons. If spectra are recorded using a different number of scans or if spectra are recorded with different devices, the absolute signal intensities are different, rendering a joint analysis of the spectra without prior normalization impossible.

In contrast to the binning[8] and peak alignment steps,[5,6,7] no investigations concerning the normalization step of a set of NMR spectra of complex biofluids have been published. For the $^1$H NMR spectra of complex biofluids, the so-called integral normalization is the de facto standard of normalizing NMR spectra.[10−15] The integral normalization assumes that the integrals of spectra are mainly a function of the overall concentrations of samples. A linear concentration series of urine will result in a linear series of integrals of the corresponding spectra. Influences of the changes of individual concentrations of single analytes are supposed to be small compared to the changes from the variation of the overall concentrations of urine. In addition, the influence of specific down-regulation of metabolites should be balanced to a certain extent by the up-regulation of other metabolites. The integral normalization procedure divides each signal or bin of a spectrum by the integral of the complete spectrum or part of it. Thus, all spectra are normalized to the same integral.

In this publication, it is shown that integral normalization has severe drawbacks concerning the robustness and the accuracy,


(5) Stoyanova, R.; Nicholls, A. W.; Nicholson, J. K.; Lindon, J. C. Brown, T. R. *J. Magn. Reson.* **2004**, *170*, 329−335.

(6) Forshed, J.; Schuppe-Koistinen, I.; Jacobsson, S. P. *Anal. Chim. Acta* **2003**, *487*, 189−199.

(7) Lee, G.; Woodruff, D. L. *Anal. Chim. Acta* **2004**, *513*, 413−416.

(8) Lefebre, B.; Golotvin, S.; Schoenbachler, L.; Beger, R.; Price, P.; Megyesi, J.; Safirstein, R. Intelligent Bucketing for Metabonomics. Part 1; Poster; Advanced Chemistry Development, Inc.: Toronto, Canda; http://www.acdlabs.com/download/publ/2004/enc04/intelbucket.pdf.

(9) Ciba-Geigy, A. G. *Wissenschaftliche Tabellen Geigy 1, Körperflüssigkeiten*, 8th ed.; CIBA-GEIGY Limited: Basel, Switzerland, 1977; pp 51−53.

(10) Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *NMR Biomed.* **2005**, *18*, 143−162.

(11) Keun, H. C.; Ebbels, T. M. D.; Bollard, B. E.; Beckonert, O.; Antti, H.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Chem. Res. Toxicol.* **2004**, *17*, 580−587.

(12) Brindle, J. T.; Antti, H.; Holmes, E.; Tranter, G.; Nicholson, J. K.; Bethell, W. L.; Clarke, S.; Schofield, P. M.; McKilligin, E.; Mosedale, D. E.; Grainger, D. J. *Nat. Med.* **2002**, *8*, 1439−1444.

(13) Water, N. J.; Waterfield, C. J.; Duncan Farrant, R.; Holmes, E.; Nicholson, J. K. *Chem. Res. Toxicol.* **2005**, *18*, 639−654.

(14) Mortishire-Smith, R. J.; Skiles, G. L.; Lawrence, J. W.; Spence, S.; Nicholls, A. W.; Johnson, B. A.; Nicholson, J. K. *Chem. Res. Toxicol.* **2004**, *17*, 165−173.

(15) Ringeissen, S.; Connor, S. C. Browin, H. R.; Seatman, B. C.; Hodson, M. P.; Kenny, S. P.; Haworth, R. I.; McGill, P.; Price, M. A.; Aylott, M. C.; Nunez, D. J.; Haelden, J. N.; Waterfield, C. J. *Biomarkers* **2003**, *8*, 240−271.

as the assumption of the total integral being only a function of dilution is not universally valid. It is demonstrated that the integral normalization completely fails if the samples contain extreme amounts of single endogenous or drug metabolites. The limitations of the integral normalization were demonstrated also in a recent publication.[16] It was shown that integral normalization can negatively influence multivariate data analysis methods and correlation methods, and the need for improved normalization methods was raised.

Here, a new normalization method is proposed called probabilistic quotient normalization. This method scales spectra on the basis of the most probable dilutions. This most probable dilution of a spectrum is estimated by analyzing the distribution of the quotients of the amplitudes of the spectrum to be normalized with those of a reference spectrum. Using simulated spectra, spectra from a metabonomic study, and more than 4000 spectra of control samples, it is demonstrated that the new quotient normalization is more exact and by far more robust than the wide-spread integral normalization and vector length normalization, which is hardly used in metabonomics[17] but is often used in other fields of pattern recognition.[18,19] Especially in the case of spectrum-fitting methods and metabolite quantifications, the improvements of the quotient normalization are crucial.

## MATHEMATICAL METHODS

**Common Normalization Methods.** All normalization procedures scale complete spectra in a way that these spectra represent the same overall concentration. In general, the scaling procedures of most normalization methods (also in other scientific fields) are variations of the general equation

$$I(i) = \frac{I^{\mathrm{old}}(i)}{\sum_k \left( \int_{j_k^{\mathrm{l}}}^{j_k^{\mathrm{u}}} (I(x))^n \, \mathrm{d}x \right)^{1/n}} \qquad (1)$$

where $I^{\mathrm{old}}(i)$ and $I(i)$ are the intensities of the variable $i$ (spectral feature, wavelength, bin, chemical shift) before and after normalization, respectively, $k$ is an index of the spectral regions used for normalization, $j_k^{\mathrm{l}}$ and $j_k^{\mathrm{u}}$ are the lower and upper borders, respectively, of the spectral region $k$, for which the power $n$ of the intensities $I(x)$ are integrated.

For integral normalization, the power $n$ is set to 1. Virtually all publications dealing with $^1$H NMR metabonomics of urine samples use only the spectral range between 0 and 10 ppm for integration and exclude the nonquantitative spectral range of urea and water resonances. Finally, each variable is typically multiplied by 100 to achieve a total integral of 100 for each spectrum.

The creatinine normalization is a special version of the integral normalization originating from clinical chemistry. For the investigation of the urine of humans and animals in clinical chemistry, it is a common practice to normalize concentrations of analytes by the concentration of creatinine.[20−23] The assumption is a constant excretion of creatinine into urine (often referred to as creatinine clearance). Thus, creatinine is an indicator of the concentration of urine. As the concentration of creatinine can be determined by its peaks in the spectra, the creatinine normalization can also be expressed in terms of general eq 1, whereby $j_k^{\mathrm{l}}$ and $j_k^{\mathrm{u}}$ mark the borders of the $k$ parts (left and right borders of the 2 creatinine peaks at 3.05 and 4.05 ppm). In the field of metabonomics, the creatinine normalization was proposed in early publications.[24] Yet, the practical application of the creatinine normalization is faced by technical and biological difficulties. From the technical point of view, metabolites with peaks overlapping with creatinine signals can interfere with the determination of the creatinine concentration (e.g., creatine at 3.04 ppm). In addition, the chemical shift of creatinine at 4.05 ppm depends on the pH of samples rendering a sophisticated peak-picking algorithm necessary. The biological challenges for creatinine normalization are changes of the concentrations of creatinine caused by metabonomic responses, which has been shown in several studies.[25−27] In that case, the normalization by creatinine is not applicable because, at the stage of normalization, a possible increase of the creatinine level caused by metabonomic responses is usually not yet known. Thus, a creatinine-based normalization is not of general use in metabonomics and will not be investigated in this manuscript. Nevertheless, a partial integral normalization analogous to the creatinine normalization can be a very simple and powerful normalization method under certain circumstances. If it is a priori known for a specific problem or biological matrix that certain parts of the spectrum are not influenced by metabolic changes, these integrals could be used for an integral normalization. Yet, this approach is less general than other normalization approaches because it has to be ensured that no varying peaks appear within the integral regions. Therefore, this approach will not further be pursued in this work.

A common normalization technique, found in pattern recognition approaches in many scientific fields, is based on looking at the spectra as vectors. Therefore, it is assumed that the concentration of a sample determines the length of the corresponding vector, whereas the content of a sample determines the direction of the vector. Thus, the adjustment of different concentrations is performed by setting the lengths of the vectors to 1. This is equivalent to setting the power $n$ of general eq 1 to 2.

**Probabilistic Quotient Normalizaton.** The approach of probabilistic quotient normalization assumes that changes in the concentrations of single analytes only influence parts of the


(16) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, C. *Anal. Chem.* **2006**, *78*, 2262−2267.
(17) Jansen, J. J.; Hoefloot, H. C. J.; Boelens, H. F. M.; van der Greef, J.; Smilde, A. K. *Bioinformatics* **2004**, *20*, 2438−2446.
(18) Jurs, P. C.; Bakken, G. A.; McClelland, H. E. *Chem. Rev.* **2000**, *100*, 2649−2678.
(19) Vandeginiste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics*, 1st ed.; Elsevier: Amsterdam; 1998.
(20) *The Merck Veterinary Manual*; http://www.merckvetmanual.com/mvm/index.jsp?cfile=htm/bc/150204.htm.
(21) Jatlow, P.; McKee, S.; O'Malley, S. *Clin. Chem.* **2003**, *49*, 1932−1934.
(22) Sterling Reference Laboratories, http://www.sterlingreflabs.com/thccreatios.html.
(23) Fauler, G.; Leis, H. J.; Huber, E.; Schellauf, C.; Kerbl, R.; Urban, C.; Gleispach, H. *J. Mass Spectrom.* **1997**, *32*, 507−514.
(24) Holmes, E.; Foxall, P. J. D.; Nicholson, J. K.; Neild, G. H.; Brown, S. M.; Beddell, C. R.; Sweatman, B. C.; Rahr, E.; Lindon, J. C.; Spraul, M.; Neidig, P. *Anal. Biochem.* **1994**, *220*, 284−296.
(25) Shockor, J. P.; Holmes, E. *Curr. Top. Med. Chem.* **2002**, *2*, 35−51.
(26) Beckwith-Hall, B. M.; Nicholson, J. K.; Nicholls, A. W.; Foxall, P. J. D.; Lindon, J. C.; Connor, S. C.; Abdi, M.; Connelly, J.; Holmes, E. *Chem. Res. Toxicol.* **1998**, *11*, 260−272.
(27) Gray, J.; Nicholson, J. K.; Creasy, D. M.; Timbrell, J. A. *Arch. Toxicol.* **1990**, *64*, 443−450.

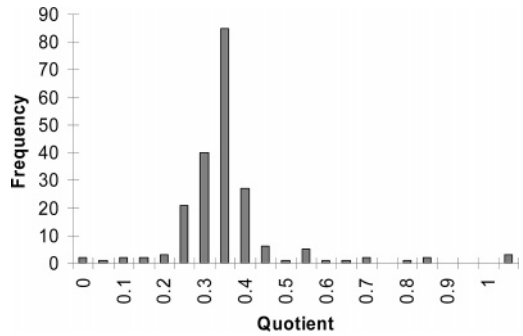


**Figure 2.** Histogram of the quotients of the two spectra shown in the left part of Figure 1. The dilution factor of 3 of the black spectrum is represented as the most probable quotient around 0.33.

spectra, whereas changes of the overall concentration of a sample influence the complete spectrum. In contrast to integral normalization, which supposes that the total integral, which covers all signals, is a function of dilution only, the quotient normalization assumes that the intensity of a majority of signals is a function of dilution only. Therefore, a most probable quotient between the signals of the corresponding spectrum and of a reference spectrum is calculated as normalization factor, which replaces the total integral as marker of the sample concentration. This most probable quotient for a specific spectrum can be derived from the distribution of signals (variables, spectral values, data points, bins) of a spectrum divided by the corresponding signal of a reference spectrum. The histogram shown in Figure 2 represents the quotients of the 205 variables of the black and gray spectrum shown in the left part of Figure 1. It can be seen that the distribution of the quotients is rather sharp. The most probable quotient is located around 0.33, which corresponds to a 3-fold dilution of the black spectrum. The most probable quotient, which is needed for the normalization, can be estimated in several ways. Using the median as an estimation of the most probable quotient has proven to be a very robust and exact method. For the calculation of the quotients, a reference spectrum is needed, in contrast to integral normalization. The reference spectrum can be a single spectrum of the study, a "golden" reference spectrum from a database, or a calculated median or mean spectrum on the basis of all spectra of the study or on the basis of a subset of the study. Later in this work it is shown that the choice of reference spectrum is uncritical: yet a median spectrum of control animals seems to be the most robust reference spectrum for studies with only few animals. It is also recommended that an integral normalization is performed prior to the quotient normalization to scale the different studies to the same absolute magnitude. The algorithm of the quotient normalization can be summarized as follows:

1. Perform an integral normalization (typically a constant integral of 100 is used).

2. Choose/calculate the reference spectrum (the best approach is the calculation of the median spectrum of control samples).

3. Calculate the quotients of all variables of interest of the test spectrum with those of the reference spectrum.

4. Calculate the median of these quotients.

5. Divide all variables of the test spectrum by this median.

The quotient normalization can be applied to raw spectra or to binned spectra, while variables not containing signals should be excluded from the calculation of the most probable quotient if feasible.

## EXPERIMENTAL SECTION

Three types of data sets with different backgrounds are examined in this report. The first type of data is based on simulations. Thereby a typical experimental urine NMR spectrum is systematically varied simulating different effects, which can influence the normalization. The simulations cover realistic changes up to rather unrealistically extreme changes. A second data set is based on NMR measurements of samples from a complete metabonomic study. Thereby, normalization procedures are challenged by samples with extreme amounts of metabolites and simultaneously by varying concentrations of urine. The third data set is a collection of 4023 NMR measurements of control animals (nondosed) originating from COMET studies.[28] These samples only show normal biological and analytical variations. This data set allows a comparison of the performance of the normalization procedures under the least demanding conditions.

**Sample Preparation and NMR Spectroscopy.** Before measurement, 200 $\mu$L of 0.2 M sodium phosphate buffer (pH 7.4) containing 1 mM TSP 20% $D_2O$, and 3 mM sodium azide was added to 400 $\mu$L urine samples. Then the samples were centrifuged for 15 min. The spectra were measured with a 600 MHz flow-injection NMR system (Bruker Biospin, Karlsruhe, Germany) at 300 K at the Biological Chemistry Research Section at the Imperial College (London). The spectral acquisition was based on a NOESY pulse sequence with an acquisition time of 1.36 s. For suppression of the water signal, irradiation of the water frequency during a 1 s relaxation delay and during a 100 ms mixing time was used. Data points (32 K) were collected during 64 scans with a spectral width of 20.036 ppm. A line-broadening factor of 1 Hz prior to Fourier transformation was applied. The spectra were phased, baseline corrected, and referenced to TSP using a routine in MATLAB (NMRProc 0.3, Dr. Tim Ebbels, Imperial College, London). Each NMR spectrum (0.2−10 ppm) was reduced by an equidistant binning method with a bin width of 0.04 ppm. The spectral region of 4.50−5.98 ppm was deleted to remove variability caused by the suppression of water resonances and cross-relaxation effects. The regions around the citrate resonances (2.50−2.58 and 2.66−2.74 ppm) were summed to account for the highly shifting citrate signals.

**Simulated Data Sets.** For the simulations concerning the stability of normalization methods, a "golden spectrum" was systematically varied. The golden spectrum was calculated as the median spectrum of 4023 spectra of urine of nondosed rats (described later) when dealing with the control sample data set. The spectrum, which was reduced to 201 bins (0.4−9.96 ppm), was normalized to a total integral of 100. As a marker bin for the normalization quality, the intensity of the 201st bin (0.4 ppm) was artificially set to 0.5. For the simulated spectra, this marker bin will not be touched by simulated specific metabolite changes. The marker bin will be only scaled by simulated dilution effects. This

(28) Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Antti, H.; Bollard, M. E.; Keun, H.; Beckonert, O.; Ebbels, T. M.; Reily, M. D.; Robertson, D.; Stevens, G. J.; Luke, P.; Breau, A. P.; Cantor, G. H.; Bible, R. H.; Niederhauser, U.; Senn, H.; Schlotterbeck, G.; Sidelmann, U. G.; Laursen, S. M.; Tymiak, A.; Car, B. D.; Lehman-McKeeman, L.; Cole, J. M.; Loukaci, A.; Thomas, C. *Tox. Appl. Pharmacology* **2003**, *187*, 137−146.

bin will be used as reference to judge the quality of the normalization procedures, as the intensity of this bin after an optimal normalization should be 0.5 for all normalized spectra.

The first data set simulates systematic variations of unspecific overall concentrations of samples in steps of 10% ending up in a doubled concentration. This was performed by multiplying the intensity of each bin of the golden spectrum with a factor of 1.1, 1.2, 1.3, ..., 2.

The second simulated data set simulates strong specific metabonomic changes by systematically varying one single bin. Thereby the peak at 2.7 ppm (one of the two peaks of citrate typically visible in the spectrum) was increased in steps of 10% of the total intensity integral with 10 steps in total. It is obvious that this single peak dominates the complete spectrum for the last few steps.

The third simulated data set can be seen as a combination of the modifications of the first and second simulated data sets. Therefore, unspecific overall concentrations of samples were increased in steps of 10%, and simultaneously, 10% of the integral intensity of the golden spectrum was added to the peak at 2.7 ppm for each step, simulating strong specific changes (10 steps in total).

For the fourth simulated data set, blocks of 10 peaks were systematically modified simulating specific changes of several peaks. The intensity of the first 10 bins was increased by 1% of the integral intensity of the golden spectrum for each of the 10 bins. For the second spectrum, the first 20 bins were increased. In total, 20 spectra were generated, always increasing 10 additional bins from step to step, ending up with 200% additional integral intensity for the last spectrum (compared with the golden spectrum).

It is clear that for all normalization procedures investigated in this work, the location of changed peaks or bins is irrelevant in contrast to the intensity changed and the number of bins changed.

**Metabonomic Study Data Set.** The spectra of a real metabonomics study, in which cyclosporin-A was administered to animals, allow testing the performance of different normalization procedures with experimental data. In this study, single animals showed strong metabonomic responses including the extreme excretion of glucose caused by glucosuria. The animal study and the measurements of the samples were performed in the Consortium of Metabonomics (COMET).[28] The animal study was performed according to the COMET protocol, which can be summarized as follows:

Male 8−10 week old Sprague−Dawley rats were housed in metabolism cages. All animals had free access to water and food (Purina chow 5002 standard diet). Animals were randomly assigned to to the groups "control group", "low-dosed group", and "high-dosed group". Each group contained 10 animals; 5 animals per group were sacrificed 48 h post dose, and 5 animals per group were sacrificed 168 h post dose. At 0 h, corn oil was orally administered to the control group, 70 mg/kg cyclosporin-A in corn oil was administered to the low-dosed group, and 700 mg/kg cyclosporine A was orally administered to the high-dosed animals. Urine was collected between 24 and 16 h predose and then continuously while fractionating at 0, 8, 24, 48, 72, 96, 120, 144, and 168 h. The samples were collected over ice into 1 mL of a 1% sodium azide solution. The samples were centrifuged for 15 min and stored at −80 °C until measurement. As 12 samples were missing, the total number of samples was 213.

**Control Samples Data Set.** To validate the performance of the different normalization procedures not only under the difficult conditions of strong metabonomic responses but also in a situation, in which animals show only "normal" metabolic changes, a collection of the NMR spectra of nondosed rats was compiled; 4521 predosed and control samples were collected, measured, and preprocessed according to the COMET protocol (as described before). PCA models with 5 components were built, and spectra, which were outlying according to the augmented distance to the model DmodX+,[29] were iteratively removed from the model. Finally, a set of 4023 samples was obtained. This collection of samples, which could be considered to be the least-demanding sample set for normalization procedures, should represent normal variations of rats similar to a baseline.

## RESULTS

**Simulated Data Sets.** *Comparison of Normalization Methods.* In this section, the performance of the two common normalization procedures, namely, integral normalization and vector length normalization, is compared with quotient normalization using simulated data. For the quotient normalization, the golden spectrum was used as the reference spectrum. Systematic variations of the reference spectrum for the quotient normalization are described later. Here, the quotients of the intensity of the 0.4 ppm bin of the normalized spectra and of the golden spectra are calculated after application of the normalization procedures as a quality measure. As this reference bin was modified artificially only according to changes of the overall concentration during the different simulations, the quotient directly shows the performance of the different normalization procedures. An optimal normalization must result in a quotient of 1, which means that the dilution of the corresponding sample is totally recovered by the normalization.

In Figure 3, the results of the three normalization procedures for the four simulated data sets are shown. For data set 1, which only contains unspecific variations of the overall concentrations, all three methods show an optimal normalization. The recovery (quotient) of 1 means that the peaks and the analytes, respectively, which only vary with the total concentration (like dilution of samples), are normalized to the same relative concentration. As expected, all three procedures are capable of perfectly normalizing spectra of an ideal series of dilutions.

The second data set, which contains only specific changes of one single signal and no dilution, shows a rather different performance of the three procedures. The normalization to vector length is highly sensitive to changes of the single peak. Thereby, the increase of the length of the concentration vector due to the increase of a single bin of the spectrum is dramatic because the vector length is calculated using quadratic terms. The rescaling of the vector length is distributed over all bins equally resulting in an underestimation of the bins not changed. The deviations for the integral method from the ideal normalization are also remarkable but less than the vector length normalizaton, as the

(29) *User's Guide to SIMCA-P, SIMCA-P+*, version 11; Umetrics AB: Umeå, Sweden, 2005.

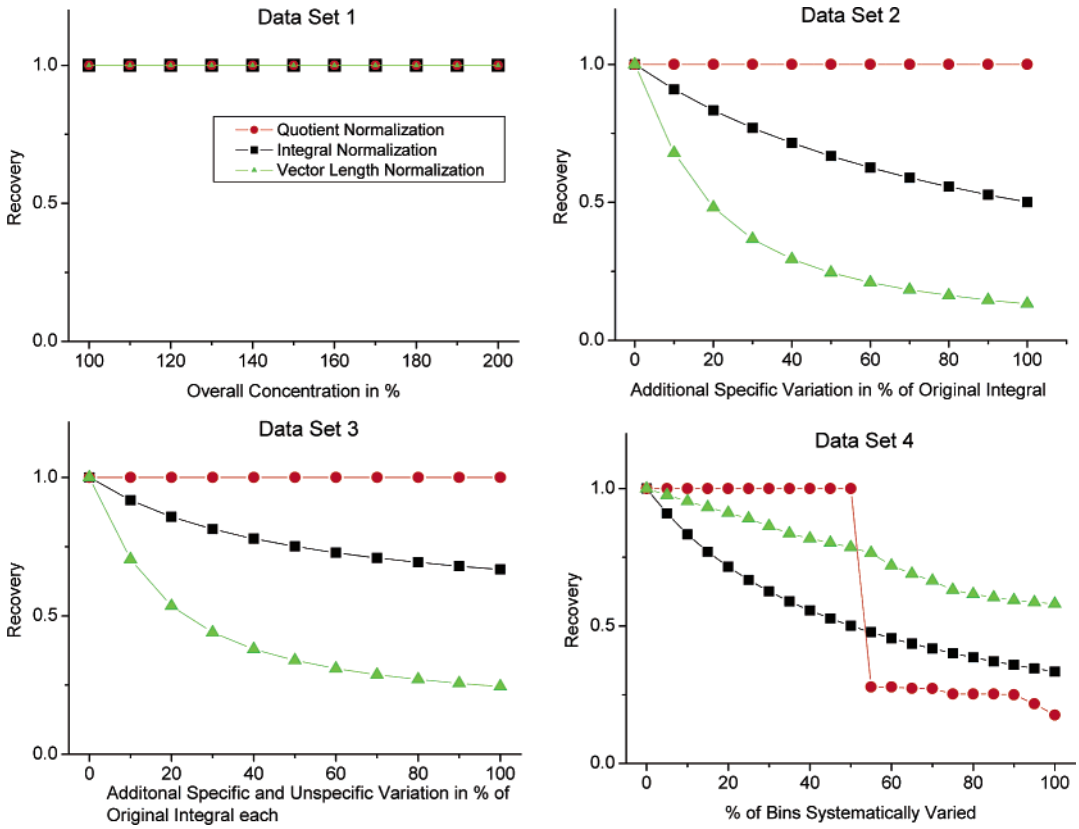


**Figure 3.** Performance of the three normalization procedures in terms of recovery of the reference signal for the systematically varied spectra of the four simulated data sets. An optimal normalization corresponds to a recovery of 1. In the first data set, only the overall concentration is systematically varied, whereas in the second data set only one specific signal is systematically varied. The third data set combines the variation of the concentration and the variation of one specific signal. The fourth data set simulates a systematical specific variation of several signals, and the number of varied signals is systematically increased.

effect of an increased intensity of one bin is distributed equally over all bins. The quotient normalization is not influenced by changes of a single bin and performs an optimal normalization for all spectra independent of the extent of changes of the single signal.

The third data set, which contains a combined variation of the first and second data set, shows rather similar results to the second data set. The variation of the single bin still highly influences the normalization to vector length and to total integral, whereas the quotient normalization shows optimal results.

The fourth data set simulates the combined changes of several signals, which corresponds to changes of metabolites with several signals or to correlated changes of several metabolites. For the first spectrum, the intensity of 10 variables (out of 201) is increased (1% of total integral per variable), for the second spectrum the intensity of 20 variables is increased the same way, and so on. In this case, the vector length normalization shows a better performance than the integral normalization but both methods show a significant deviation from optimal normalization, even for the first simulated spectrum. On the other side, the quotient normalization shows an optimal normalization until a systematical increase of intensity of 100 out of 201 variables has been reached. For spectra with systematical increases of more variables, the performance drops dramatically. Yet this case of systematical changes with an excess of more than half of the signals in the same direction is very rare. Systematic variations of 0−20% of the bins into one direction can be considered to be realistic scenarios, and systematic variations of more than 20% of the bins are extreme scenarios, especially as specific changes of variables in different directions balance each other. Therefore, specific changes of more than 50% of the signals in one direction will be never obtained for real-world data.

*Influence of Reference Spectrum for Quotient Normalization.* In contrast to vector length normalization and integral normalization, a reference spectrum is needed for quotient normalization. The influence of the reference spectrum on the performance of the quotient normalization is investigated in this section. The following reference spectra were used:

(A) The golden spectrum is the unaltered spectrum, out of which all simulated spectra were calculated.

(B) The median spectrum of all the simulated spectra of the four simulated data sets is a very extreme reference spectrum compared to real-world studies because the specific changes in the simulations were highly exaggerated.

(C) The median spectrum of all simulated spectra, which differed only by unspecific overall concentration changes or by specific changes, which are less or equal to 20% of the total integral is equivalent to calculating a median spectrum of all spectra of a real-world metabonomic study with strong metabonomic changes.

(D) The median spectrum of all simulated spectra, which differ only by unspecific overall concentration changes simulates a median spectrum of control samples, but it neglects small specific changes typically present in control samples

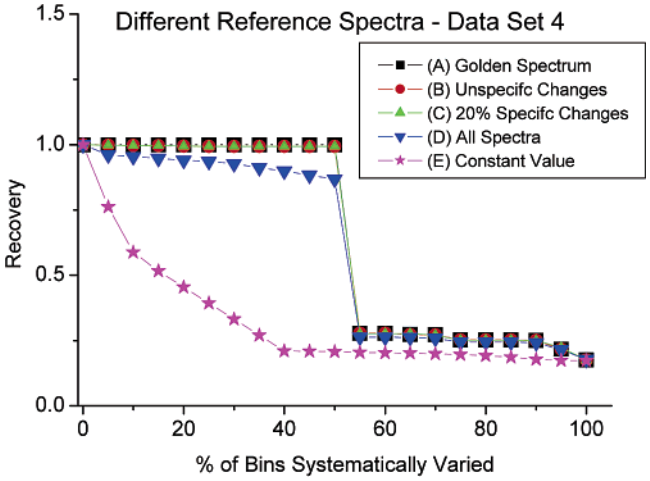


**Figure 4.** Performance of quotient normalization of simulated data set 4 for 5 different methods of calculating a reference spectrum.

(E) The spectrum with a constant value of 1 in each bin has nothing to do with the spectra of the simulations and shows a totally different shape.

For the 5 different reference spectra, quotient normalizations of the four noise-free data sets described before were performed. For the first three data sets, no significant differences between the quotient normalizations with different reference spectra could be observed (data not shown). For the fourth data set, differences were visible, which are shown in Figure 4. A constant value reference spectrum results in a very bad performance. This is self-evident as the distribution of the quotients of the spectrum and reference spectrum corresponds to the distribution of the spectrum itself and consequently is flat and broad. Therefore, an increase of several peaks will significantly shift the median. As is visible, the most stable normalization is obtained if the reference spectrum is as close as possible to a representative spectrum without specific changes (unspecific changes are allowed). In real-world studies, this corresponds to a median spectrum of control animals. It is also apparent that specific changes of 20% of the total intensity do not influence the normalization significantly. If the reference spectrum is calculated as the median spectrum of all simulated spectra, the performance of the quotient normalization drops for the fourth data set. Yet it must be kept in mind that this reference spectrum is calculated from spectra with highly unrealistic changes.

As a conclusion, it can be said that the choice of reference spectrum is rather uncritical, as long as the shape of the reference spectrum roughly represents the shape of the spectra to be normalized. From the theoretical point of view, it is clear that the optimal reference spectrum should come as close as possible to spectra without specific changes. Thus, from the practical point of view, the optimal spectrum is a mean or median spectrum of the spectra of control animals, of which the median spectrum has the advantage of being more robust.

**Data from Metabonomic Study.** In this section, the performance of the three normalization procedures is compared using data of a complete metabonomic study with cyclosporin-A as active compound. In this study, single animals showed extreme excretions of glucose because of glucosuria. The strong increase of the glucose signals are a typical challenge for normalization procedures in real-world metabonomic studies.

The performance of the normalization methods is measured by two methods. The first measure is based on the integral regions between 1.54 and 1.84 ppm. By extensive manual inspection of all spectra, it was determined that the relative intensity of all signals in this region is highly constant for all animals at all time points. This means that the metabolites behind these signals are highly constant and are not influenced by metabolic changes. Therefore, the spectral integral of this region should be very similar between different samples after an optimal normalization. Thus, the relative standard deviations of the integrals of this part of the spectra between the different samples are used as the first quality criterion for the normalization procedure.

The second criterion of the measurement of the performance is the concentration of creatinine, which has already been discussed for normalization before. The analysis of creatinine by clinical chemistry and by multivariate analysis of the spectra revealed that in this study the relative creatinine levels are not changed by dosing cyclosporin A. Thus, the concentration of creatinine represents the dilution of the corresponding sample. Therefore, concentrations of creatinine (determined by the creatinine peak at 4.05 ppm before normalization) and the normalization factors (reciprocal of scaling factors) should be highly correlated for optimal normalizations.

For the integral normalization, the linear regression of the creatinine concentrations versus the normalization factors has a correlation coefficient of 0.87, which looks quite decent on the first sight. Yet, the normalization of samples containing strong specific changes of single metabolites was suboptimal. An example of a suboptimal normalization is shown in the first row of Figure 5. Two spectra of samples from the high-dosed animal 28 are shown. The gray spectrum represents a predose sample collected at 0 h, and the black spectrum represents a postdose sample collected at 48 h. This postdose sample contains massive amounts of glucose resulting from a drug-induced glucosuria of the animal. Therefore, the black spectrum contains very intensive glucose signals, which account for approximately 50% of the total integral of the spectrum. Consequently, the integral normalization incorrectly downscales the complete spectrum by a factor of 2. As an erroneous result, it can be seen that the black spectrum seems to differ from the gray spectrum for the complete spectral range. Also, in the focused view of several spectra (−16 to 72 h) of animal 28 shown in Figure 6, it is visible that the spectra of the two samples containing massive amounts of glucose (48 and 72 h) are incorrectly downscaled. In addition, the spectral range between 1.54 and 1.84 ppm looks highly scattered in the plot. As the spectral intensity within this range is mainly a function of dilutions of the corresponding samples, the scattering shows that the integral normalization does not compensate for differences of the overall concentrations of these samples correctly. This holds true not only for the samples of animal 28 but also for the complete study, noticeable as a rather high relative standard deviation of 10.3% for the spectral integral between 1.54 and 1.84 ppm.

The vector length normalization even performs worse than the integral normalization. The correlation between the normalization factor and the creatinine concentration is poor ($r = 0.62$). The integral between 1.54 and 1.84 ppm shows a very high relative standard deviation of 15.0%, which is also visible as high scattering of the spectra shown in Figure 6. Similar to the normalization of

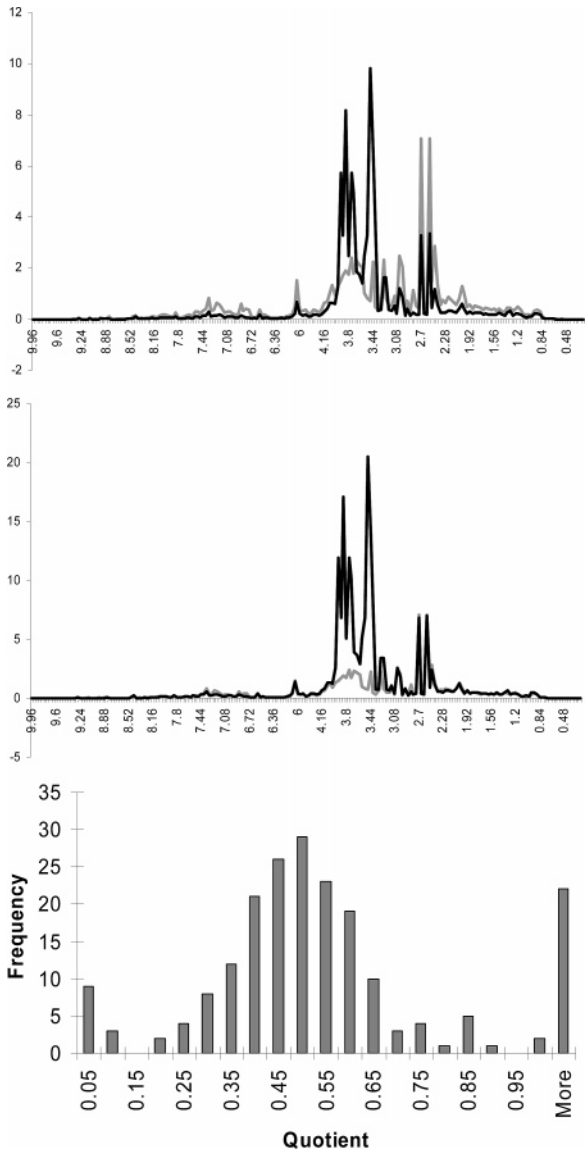


**Figure 5.** Two spectra of samples of animal 28 from the metabonomic study and the histogram representing the quotients of the two spectra. The gray spectrum represents the sample collected at 0 h, and the black spectrum represents the sample collected at 48 h. The sample collected at 48 h contains massive amounts of glucose. The glucose signals account for approximately 50% of the total integral of the spectrum. The integral normalization, which is shown in the first row, incorrectly down-scales this spectrum by a factor of 2 resulting in two spectra, which look different for the complete spectral range. In contrast, the quotient normalization, which is shown in the second row, reveals that only very few differences, in addition to the glucose peak, exist in these two samples. The histogram of the quotients of the integral normalized spectra shows that the most probable quotient is around 0.5. This corresponds to the incorrect down-scaling of the integral normalization by a factor of 2. The dominance of the glucose peaks is only visible as a bar for quotients higher than 1 (last bar on the right side); therefore it does not influence the most probable quotient.

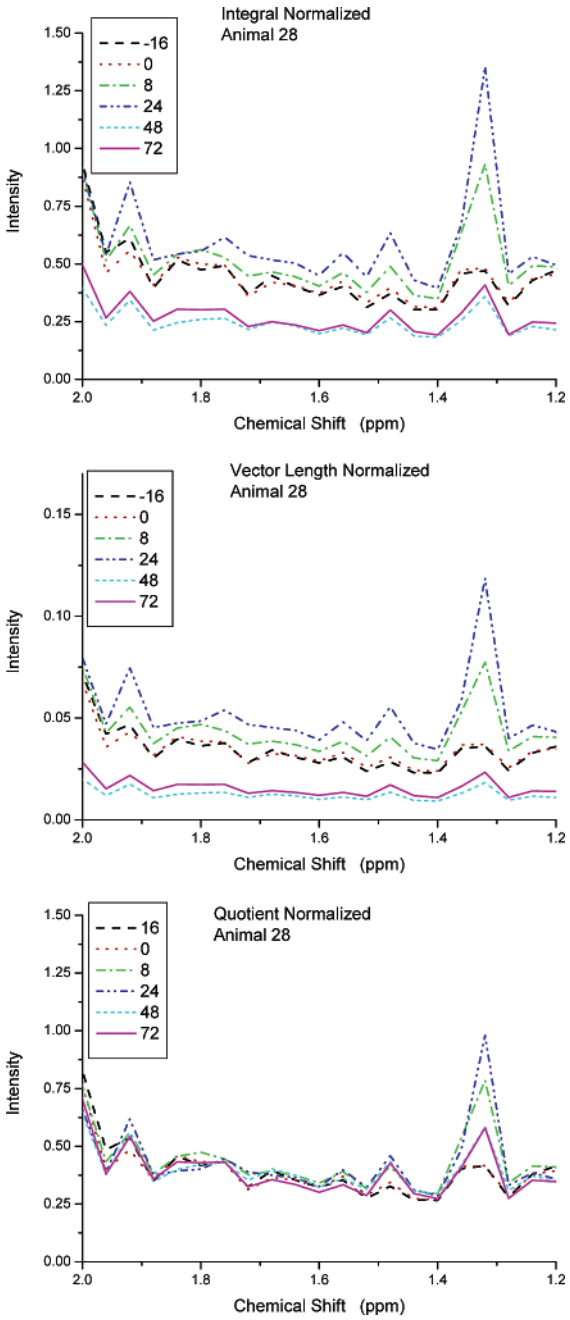


**Figure 6.** Clipping of the normalized spectra of animal 28 at time points −16, 0, 8, 24, 48, and 72 h for the integral normalization (top row), the vector length normalization (middle row), and quotient normalization (bottom row). The samples at 48 and 72 h for this animal contain extreme amounts of glucose and challenge the normalization procedures. The spectral range between 1.54 and 1.84 ppm does not contain peaks which are significantly influenced by metabolic changes in this study. Thus, the signal intensities of the various spectra within this spectral range should be comparable in the case of an optimal normalization.

the third simulated data set, the vector length normalization is hampered by strong changes influencing the normalization with a quadratic term.

The quotient normalization was performed using the median spectrum of the control samples of this study as the reference spectrum. The quotient normalization shows by far the best performance in terms of correlation of the normalization factors with the NMR-determined creatinine concentration ($r = 0.99$). Also the relative standard deviation of the integral between 1.54 and 1.84 ppm is very low (4.5%). This excellent performance is especially visible for the clipping of the spectra of animal 28 shown in Figure 6, which hardly shows any scattering. When looking at the two complete spectra of animal 28 shown in the second row of Figure 5, it is apparent that there are only a few differences

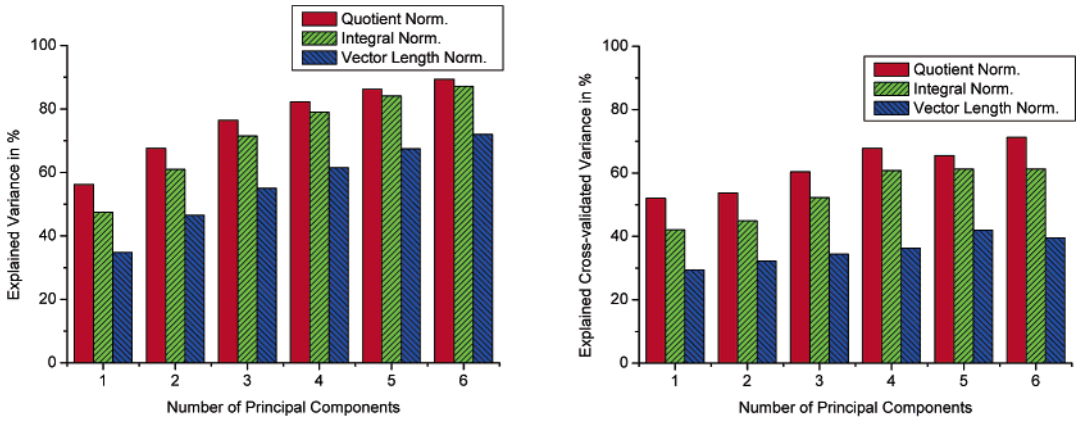


**Figure 7.** Explained variance (left side) and explained variance estimated by cross-validation (right side) for the three PCA models of 4023 control urine samples normalized by quotient normalization, integral normalization, and vector length normalization.

between the samples in addition to the extreme glucose peak. The corresponding distribution of the quotients of signals of the black spectrum (after integral normalization and before quotient normalization shown in the first row of Figure 5) and of the reference spectrum is shown on the histogram of Figure 5. It is obvious that the intensive glucose signals have little influence on the distribution of quotients. These signals are visible as quotients higher than 1 (bar on the right side), but they hardly influence the most probable quotient estimated as median. This median of quotients, which is the normalization factor of the quotient normalization, is about 0.5. This corresponds to the incorrect downscaling of the spectrum by the integral normalization by a factor of 2 because of the 50% contribution of the glucose peaks to the total integral. When the results of the integral normalization (first row of Figure 5) and the quotient normalization (second row of Figure 5) are compared, it is obvious that, for the quantification of metabolites, an optimal normalization, such as the quotient normalization, is crucial. When the black spectrum and gray spectrum normalized by the integral normalization are compared, the differences look like an up-regulation of glucose and a down-regulation of all other metabolites. Compared with the quotient normalization, it is evident that the constraint of a total integral of the integral normalization results in an underestimation of the concentration changes of glucose and in false positive changes of all other metabolites. For subsequent multiviarate data analyses such as PLS or MCR, it is impossible to differentiate between these true positive and false positive changes of integral normalized data. In addition to the negative impact of signals of significantly increased endogenous metabolites on the integral normalization, signals from drug metabolites, which might not have been identified and excluded, can even worsen the results of the integral normalization.

**Data from Control Samples.** Integral and vector length normalization were shown to be hampered by strong metabonomic responses, using the metabonomic study as an example. An interesting point is looking at the other extreme of small normal metabonomic changes of control animals only. As the animals, which were not dosed or which were dosed with a vehicle only, should have constant relative creatinine levels (see discussion concerning the creatinine normalization), the performance of the three normalization methods was assessed by the relative standard deviation of the creatinine peaks. As the selected 4023 nondosed samples do not contain strong metabonomic responses, it was expected that all three normalization procedures should show a similar performance.

The results are rather astonishing. The vector length normalization has an unacceptably high relative standard deviation of the creatinine peak of 12.2%, whereas the integral normalization has a low relative standard deviation of 7.6%, and the quotient normalization has the best performance with 6.7%. Although the difference between the integral normalization and the quotient normalization does not appear to be tremendous at first glance, it must be noted that the samples are from control animals. Therefore, the variation and relative population of bins, which change significantly, is rather small resulting in a small expressed difference. But this means that even when looking at control animals, small specific variations due to normal metabonomic fluxes have a significant impact on the quality of different normalization procedures; the quotient normalization again shows the best performance.

The consistency of a set of samples can also be judged when looking at the convergence of a PCA model.[30] Ideally, a PCA model should describe the variation in the set of samples in as few PCs as possible. Therefore, a PCA of the 4023 spectra of nondosed samples was performed after application of the three normalization methods. The first three components of a PCA for the vector length normalized data explain a rather low cumulative variance of the spectra of 55.0% ($R^2$) and 34.4% variance estimated by cross-validation ($Q^2$). A PCA of the integral normalized spectra can explain a variance of 71.5% ($R^2$) and a cross-validated variance of 52.2% ($Q^2$) with three principal components. The PCA of the quotient-normalized spectra can explain 76.4% of the cumulative variance ($R^2$) and 60.4% of the cross-validated variance ($Q^2$) with only three components. The plots of the explained variance and of the explained variance estimated by cross-validation for the first 6 principal components of these 3 PCA models, which are shown in Figure 7, also demonstrate how the quotient normalization scales the spectra in a way that the data set is by far most consistent.

**Further Aspects.** Normalization procedures generally depend on the baseline corrections of spectra and on the phasing of NMR

(30) Lefebre, B.; Ryan, S.; Golotvin, S.; Nicholls, A.; Price, P.; Megyesi, J.; Safirstein, R. Intelligent Bucketing for Metabonomics. Part 2; Poster; http://www.acdlabs.com/download/publ/2004/intelbucket2.pdf.

spectra. In the case of NMR, high variations of the concentration of samples and large signals can cause baseline fluctuations and phases, which are not fully corrected by NMR spectra processing software. In the case of incorrect phasing of spectra, the probabilistic quotient normalization is typically not influenced. Incorrect phasing of spectra results in wider distributions of the quotients but does not influence the most probable quotient in the absence of baseline distortion. The influence on the quotient normalization, as well as on the integral normalization, depends on the type of baseline fluctuation. For example, for systematic offsets of the baseline, all normalization methods incorrectly try to compensate the offset by scaling the spectra. Further ongoing investigations in our laboratory aim to combine normalization, baseline-correction, and phasing of the NMR spectra into one preprocessing step. In our experience, the routinely applied quotient normalization has not been influenced by suboptimal phasing and baseline corrections and has outperformed the integral normalization in every case. Nevertheless, we recommend visual inspection of spectra for artifacts.

## CONCLUSIONS

In this work, the probabilistic quotient normalization for metabonomic studies was proposed. In contrast to the integral normalization, which is the de facto standard in the field of metabolic profiling of biofluids, and in contrast to the vector length normalization, the quotient normalization has no constraints such as a total integral or a total vector length. Therefore, the quotient normalization does not fail if these constraints are not valid as demonstrated by several simulations. A comparison of the three normalization methods for a real-world data set from a metabonomic study showed that the quotient normalization performs by far the best in compensating for different dilutions of urine. Both the integral normalization and the vector length normalization, in particular, are hampered by extreme amounts of metabolites, such as glucose, in the samples. It was also demonstrated that the quotient normalization is not only more robust but that the normalization of samples from control animals with only low metabolic variations is more exact when using quotient normalization instead of integral normalization. Thus, for all possible scenarios of NMR spectra from metabonomic studies, the probabilistic quotient normalization is the superior preprocessing method, and its application is advantageous for subsequent multivariate data analyses and quantifications of metabolites.

## ACKNOWLEDGMENT

The authors want to acknowledge the other partners of the COMET project for providing the samples and the group of Prof. J. Nicholson for measuring the samples and providing the program NMRPROC.

Received for review September 13, 2005. Accepted April 18, 2006.

AC051632C