Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis 36 (2004) 823–833
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Evaluation of metabolic variation in normal rat strains from a statistical analysis of ^{1}H NMR spectra of urine

Timothy M.D. Ebbels*, Elaine Holmes, John C. Lindon, Jeremy K. Nicholson

Biological Chemistry, Biomedical Sciences Division, Faculty of Medicine, Sir Alexander Fleming Building, Imperial College London, South Kensington, London SW72AZ, UK

Received 30 June 2004; accepted 3 August 2004
Available online 3 October 2004

Abstract

NMR spectroscopic and statistical methods have been applied to investigate the biochemical variations within and between two phenotypically normal rat strains. The 600 MHz ^{1}H NMR spectra of urine were acquired as part of a series of drug toxicity studies from 450 control rat urine samples from each of two strains of laboratory rat (Han Wistar and Sprague Dawley). The spectra were data-reduced to 256 intensity descriptors over a range of δ 0.2–10.0. The spectral variation was analysed both within and between strains in terms of the mean, standard deviation, skewness and kurtosis of each descriptor. It is demonstrated that spectral intensities corresponding to a number of endogenous metabolites do not show Gaussian distributions and there is evidence for bimodality for some metabolites. Additionally, despite the visual similarity of the NMR spectra from the two strains of rat, the descriptor distributions and the statistics derived from them revealed differences in the metabolite profiles, which clearly distinguished the two populations. This work is of value in the determination of biochemical normality and variability, and thus can be used to investigate, and place confidence limits on the biochemical deviations, which arise as a consequence of genetic modification or pathophysiological events.

Keywords: Sprague Dawley; Han Wistar; ^{1}H NMR spectroscopy; Gene function; Metabolite profile; Physiological variance

1. Introduction

The increased interest in the use of multivariate metabolic profiling methods for characterising disease states, toxicity screening and understanding the biochemical consequences of genetic modification necessitates a greater understanding of the statistical features of the measurement variables that are used to classify samples. ^{1}H NMR spectroscopic analysis of biofluids generates complex metabolic profiles, which can be related to the physiological states or pathological condition of an organism [1]. In combination with multivariate statistical analysis, this technique has been applied to the characterisation of endogenous metabolites in biofluids, proving effective for monitoring disease states and for studying the toxicity of pharmacologically active agents [2,3]. Furthermore, pattern recognition analysis of NMR spectroscopic data (NMR-PR) provides a means for classifying and predicting the physiological and pathophysiological status of complex organisms from time-related metabolic changes, a topic now designated as metabonomics [4–6]. High-resolution ^{1}H NMR spectra of biofluids may contain resonances from hundreds or thousands of low molecular weight metabolites, but pattern recognition (PR) can be performed on the NMR spectral signals without the need to assign all of the spectral peaks to specific metabolites before analysis. In order to reduce the complexity of such data, the spectra have often been segmented into discrete regions prior to PR analysis [7,8]. Typically, the level of reduction has been from 64k data points to ∼250 integrated spectral regions for a standard one-dimensional spectrum [9]. Standard chemometric techniques such as principal components analysis (PCA) or cluster analyses can then be used to

* Corresponding author. Present address: Bioinformatics Unit, Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK. Tel.: +44 20 7679 7981.
E-mail address: t.ebbels@cs.ucl.ac.uk (T.M.D. Ebbels).

0731-7085/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2004.08.016

map these spectra in multivariate space where each sample occupies a position in space based on its metabolite composition, which in turn is related to its physiological status. Thus, samples which derive from inherently similar states, e.g. two control samples or two samples obtained from rats treated with a renal cortical toxin at a particular time, would be expected to occupy a similar position in this multivariate space, based on the similarity of their biochemical composition. The multivariate area occupied by each sample class of toxicity or disease can be defined by mathematical models. In addition, chemometric techniques such as PCA or cluster analysis [10–12] allow the compression of multivariate data into 2 or 3 dimensions, thus allowing the human eye to distinguish clustering within the data.

Previous studies have shown that a number of different target organ toxicities can be classified using these NMR-PR based methods using urine samples and, moreover, that novel combinations of urinary biomarkers can be derived for each toxicity or disease state. For example, perturbation of trimethylamine-*N*-oxide, dimethylamine, dimethylglycine, citrate, 2-oxoglutarate, *N*-acetyl glycoproteins and succinate is a biomarker pattern indicative of renal medullary toxicity [13]. Although biomarkers of region-specific toxicity can be clearly defined in many instances, the importance of evaluating the scope of normal physiological variation within control populations of experimental animals and humans has been shown [8,14,15]. For example, spectral profiles can be influenced by physiological factors such as diet, time of sampling, hormonal status, strain of animal model and level of physical activity. NMR-based metabonomic approaches have been shown to be capable of defining a metabolic signature of body fluids such as urine and plasma characteristic of strain or species in laboratory models. This metabolic signature or metabotype can be related to differences in the genetic composition of organisms and can be used to interpret the functional consequences of genetic modification [16]. Previous chemometrics studies have demonstrated clear metabolic differences in urine samples obtained from Sprague Dawley (SD) and Han Wistar (HW) rats and also from two genetically distinct strains of mice [16,17]. Additionally this metabolic profiling approach has been successfully applied to phenotypically differentiating species of *Eisenia* (oligochaetes) based on coelomic fluid samples [18].

In order to make inferences regarding variations in spectral profile that relate to toxic or disease episodes, it is necessary to understand and account for this natural physiological and genetic variation in endogenous metabolite profiles. In this study, a range of statistics (namely the mean, standard deviation, skewness and kurtosis) has been employed to analyse the distribution of the urinary spectral descriptors for two different but phenotypically normal rat strains. This extends an earlier study, which investigated the statistical variability within HW rats only [15]. This study also used the principal components of the NMR data (up to 20 were necessary) to carry out a linear discriminant analysis to attempt to search for small differences in the spectra which were related to the fact that they arose from different experimental studies over a period of time and also included the effects of the time of collection of the urine samples. Small but distinct differences between studies could be resolved. The study has now been extended to investigate variability in the two most commonly used strain of laboratory rat. Having, a priori, established the nature and extent of natural variation for each spectral region, it is then possible to determine the degree of confidence with which toxin- or disease-induced metabolic variations in urine profiles can be used for classification and biomarker identification. In addition, with the completion of the genetic sequences of a number of species, attention is turning to the consequences of alterations in gene expression (functional genomics). Since biochemical metabolic endpoints are the ultimate consequences of such changes, it is suggested that metabonomic approaches will prove useful in the interpretation and understanding of genetic modification.

2. Materials and methods

2.1. Animal handling and sample collection

As part of ongoing toxicology studies, control male Sprague Dawley (SD) and Han Wistar (HW) rats were housed in metabolism cages and urine samples collected daily from each animal, totalling 450 samples from each strain of rat. All studies were conducted using the same basic protocol in that a standard diet, Rat and Mouse Diet No. 1 (Special Dietary Services) was given to all animals and free access to food and water was permitted throughout the study. A temperature of $21 \pm 2\,^{\circ}C$ and a relative humidity of $50 \pm 10\%$ was maintained with a 12 h light/12 h dark cycle. On collection, urine samples were centrifuged at 3000 rpm for 10 min, in order to remove any solid debris, and were subsequently stored at $-70\,^{\circ}C$ pending ^{1}H NMR spectroscopic analysis.

2.2. One-dimensional ^{1}H NMR spectroscopy of urine samples and data reduction

An aliquot of urine (400 μL) from each sample was placed in a 5 mm outer diameter NMR tube together with 200 mM sodium phosphate buffer (200 μL) in order to minimize variance in metabolite NMR chemical shifts arising from differences in urinary pH. An internal reference standard, 3-trimethylsilyl-[2,2,3,3-$^{2}H_4$]-propionate (TSP) made up to a final concentration of 1 mM in D_2O solution (50 μL), was added to each sample. All samples were measured on a Bruker DRX-600 NMR spectrometer (Bruker Biospin, Coventry, UK) at 300 K operating at 600.13 MHz for ^{1}H observation. For each sample, a one-dimensional NMR spectrum was acquired with water peak suppression using a standard pulse sequence (noesypresat, Bruker), using 64 free induction decays (FIDs), 64k data points, a spectral width of 12 376 Hz, an acquisition time of 2.65 s and a total pulse recycle delay of 7.68 s. The FIDs were multiplied by an exponential

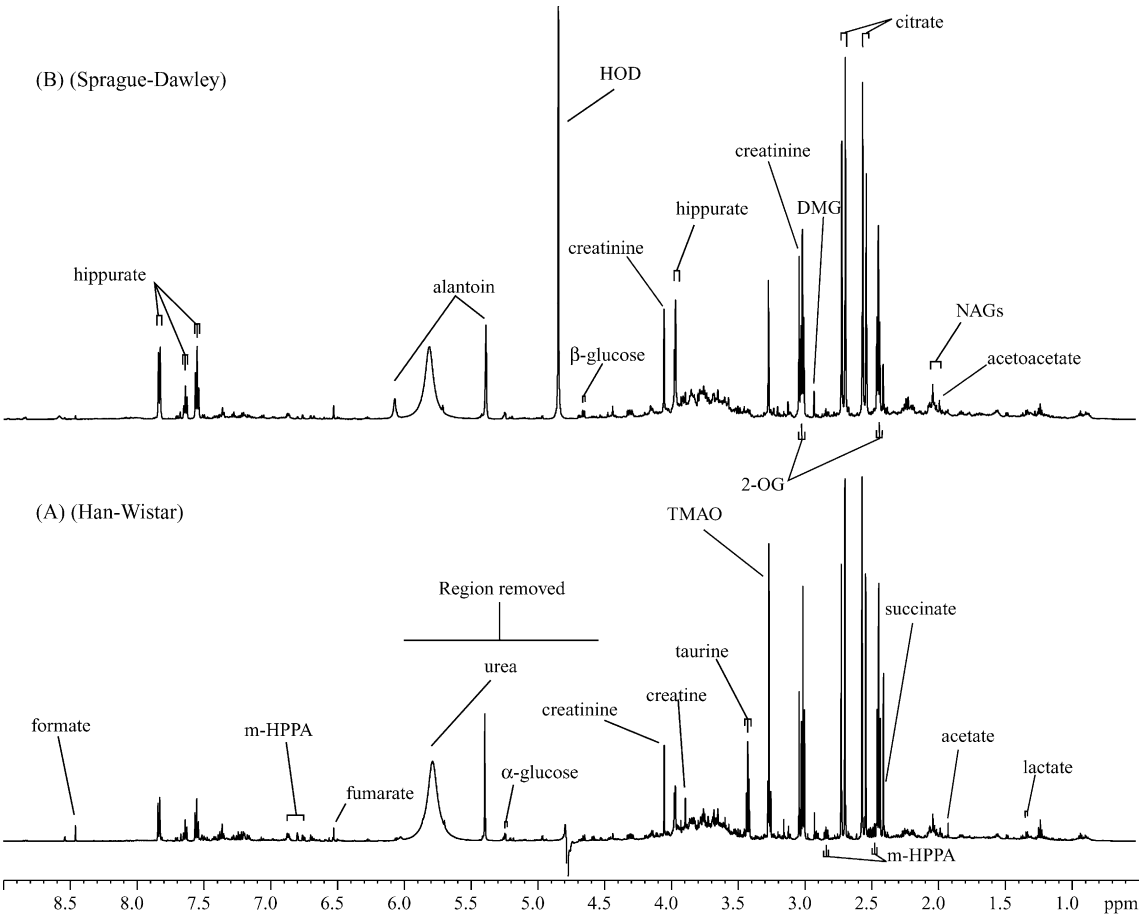

Fig. 1. Representative 600 MHz ^{1}H NMR spectra of urine from Sprague Dawley (upper trace) and Han Wistar (lower trace) rats. DMG: dimethylglycine; TMAO: trimethylamine-*N*-oxide; 2-OG: 2-oxoglutarate; HOD: residual water; NAGs: *N*-acetyl glycoprotein fragments; *m*-HPPA: *meta*-hydroxyphenylpropionic acid.

weighting function corresponding to a line broadening of 0.3 Hz prior to Fourier transformation, phasing and base line correction.

The ^{1}H NMR spectra were reduced into consecutive integrated spectral regions of width 0.04 ppm using AMIX software (Bruker). The region δ 4.52–6.00 was excluded from the analysis in order to remove the effects of variations in the suppression of the water resonance and variations in the urea signal caused by cross saturation from exchanging protons. In addition the region δ −0.2 to 0.2 containing the internal reference (TSP) was excluded from statistical analyses. The data were then imported into Excel (Microsoft®, Excel 97, SR-2) and the integrated spectral regions normalised to the total sum of the spectral regions. Since the integrals are thus expressed in terms of relative intensities, it is difficult to distinguish between an absolute increase in one metabolite as opposed to an absolute decrease in another. However, since in this study all urine samples were taken from control animals, it was expected that relative differences between samples would be small compared to the overall spectral intensities and any such effects would be minor. Nevertheless, this procedure of normalisation to unit area partially removes any concentration differences, and hence is particularly useful in the case of urine samples, as excretion volumes and hence metabolite concentrations are highly variable.

2.3. Statistical analysis of data-reduced ^{1}H NMR spectra

Standard statistics were computed for each spectral region using the software package MATLAB (The MathWorks Inc., Natick, MA, Version 5.3.1 R11.1) The mean, standard deviation, skewness and kurtosis for each spectral region were examined for each strain of rat and comparisons between SD and HW rats were also made. The actual distributions of values for each of the descriptors were examined as histograms over 10 intervals for all 900 samples. Distributions showing evidence for bimodality were selected using three criteria: (a) that the two peaks in the distribution were present at sufficient signal to noise ratio (SNR) assuming Poisson counting statistics (SNR_{crit} = 0.5), (b) that each peak contained at least 25% of the samples and (c) that this evidence was present when the distributions were computed using both 10 and 20 intervals. A further selection of multi-modal distributions was made by relaxing the last criterion to use only 20 intervals (giving an expected Poisson counting error in each interval of ∼15%).

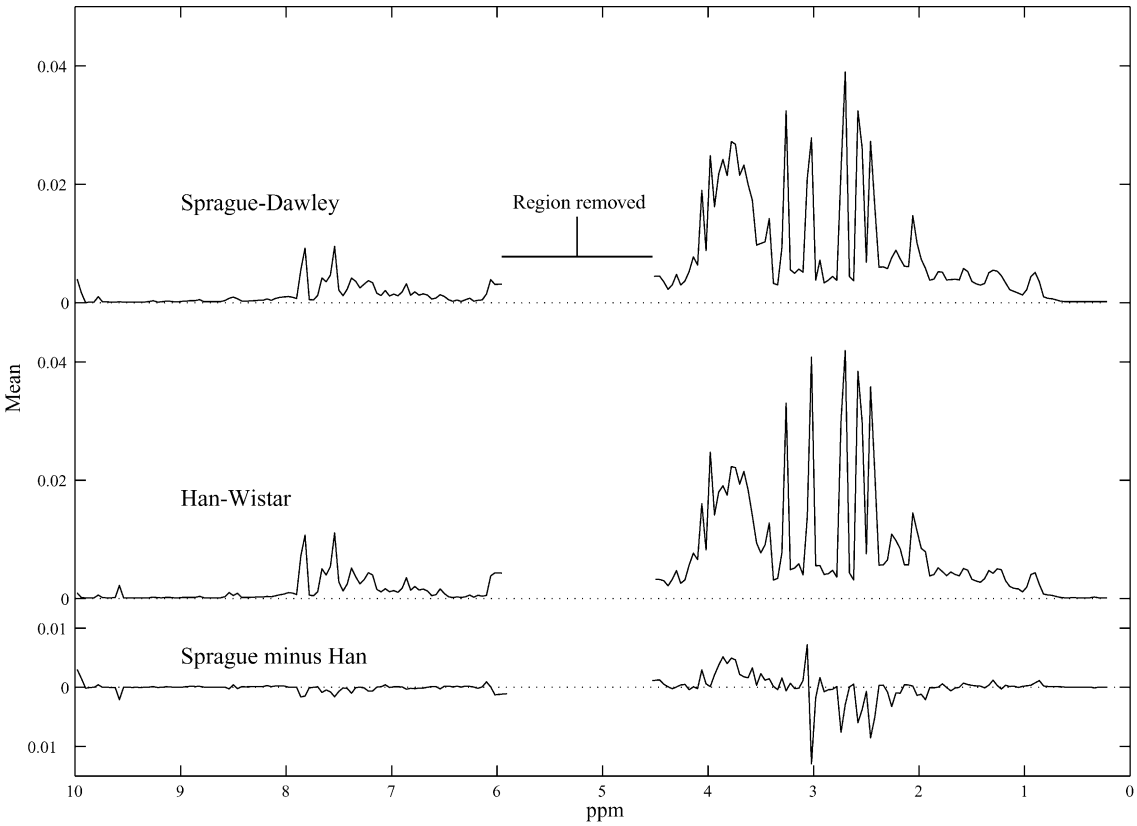

Fig. 2. Mean data-reduced ^{1}H NMR spectra of urine for the two rat strains. The lower trace shows the difference between the two mean spectra. The region δ 4.5–6.0 has been excluded.

3. Results and discussion

3.1. Visual comparison of the ^{1}H NMR spectra

Representative ^{1}H NMR urine spectra for HW and SD rats were are shown in Fig. 1(A) and (B), respectively and found to be visually similar. Typically, control urine spectra were dominated by resonances from glucose, taurine, creatinine, hippurate, citrate, 2-oxoglutarate, succinate and trimethylamine-*N*-oxide. Resonances deriving from amino acids, organic acids, creatine, acetate, acetoacetate, dimethylglycine, *N*-acetyl glycoprotein fragments and *m*-hydroxyphenylpropionate (*m*-HPPA) were also reasonably prominent in the urine spectra. Although consistent strain-related differences in the spectra are difficult to detect visually, previous work has shown that it is possible to partially discriminate between these two strains using soft independent modelling of class analogy (SIMCA), a PCA-based technique [17], where specific strain-related differences in the urinary profile were characterised. It was shown that HW rats excreted higher concentrations, than SD

Table 1
The NMR spectral regions with the 10 highest mean values

Rank	Han Wistar			Sprague Dawley			Difference	
	Spectral region[a]	Mean ($\times 10^{-2}$)	Metabolite(s)	Spectral region[a]	Mean ($\times 10^{-2}$)	Metabolite(s)	Spectral region[a]	Metabolite(s)
1	2.70	4.19	Citrate	2.70	3.90	Citrate	3.02	2-OG
2	3.02	4.08	2-OG	2.58	3.24	Citrate	2.46	2-OG
3	2.58	3.84	Citrate	3.26	3.24	TMAO/taurine	2.74	Citrate
4	2.46	3.58	2-OG	3.02	2.79	2-OG	3.06	Creatinine
5	3.26	3.30	TMAO	2.46	2.72	2-OG	2.58	Citrate
6	2.74	3.00	Citrate	3.78	2.72	Complex[b]	3.86	Creatine
7	2.54	2.99	Citrate	3.74	2.67	Complex[b]	2.42	2-OG/succinate
8	3.98	2.47	Hippurate	2.54	2.62	Citrate	3.78	Complex[b]
9	3.78	2.23	Complex[b]	3.98	2.48	Hippurate	3.74	Complex[b]
10	3.74	2.21	Complex[b]	3.86	2.42	Complex[b]	3.82	Complex[b]

TMAO: trimethylamine-*N*-oxide, 2-OG: 2-oxoglutarate.

[a] Each spectral region comprises an integrated bin of width 0.04 ppm.

[b] A complex region with many overlapping resonances from glucose, other sugars and α-protons of amino acids, especially Glu, Gln and Ala.

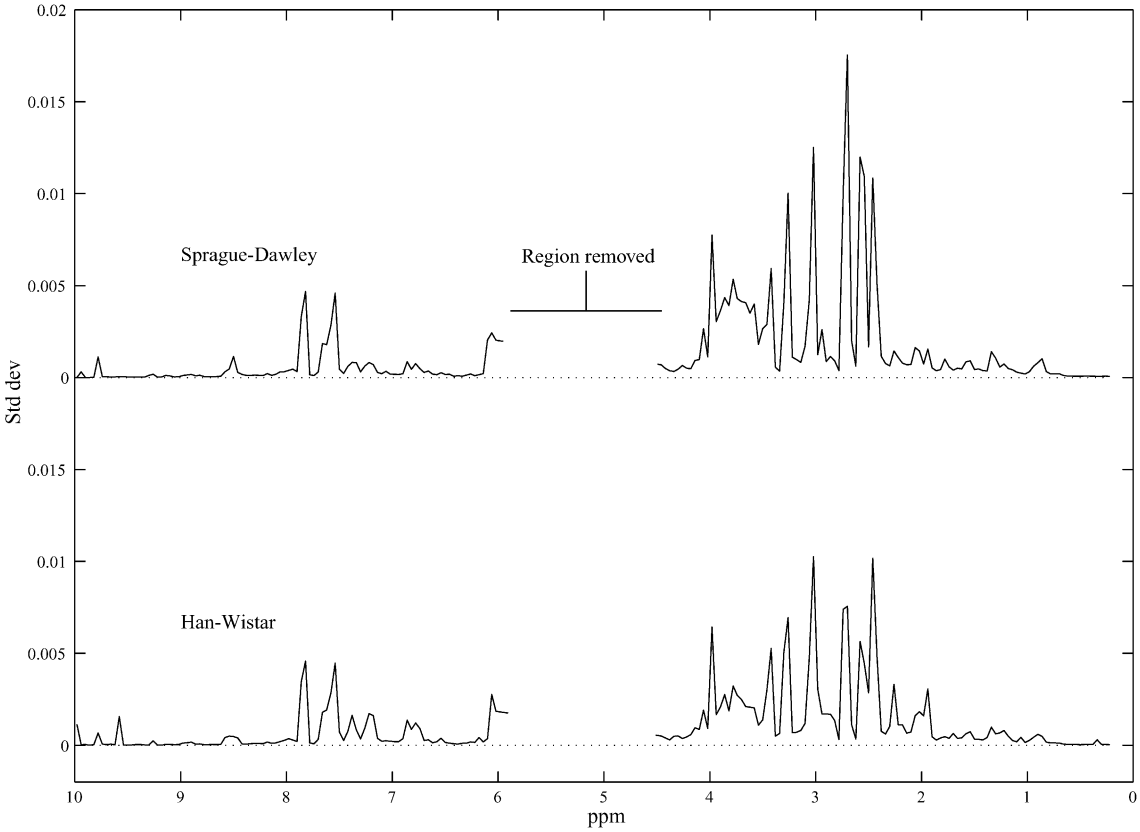

Fig. 3. Standard deviation data-reduced spectra of urine for the two strains of rat.

rats, of lactate, acetate and taurine together with lower concentrations of hippurate.

3.2. Statistical analysis of the spectra

In order to better define the distribution of values for each of the spectral descriptors for these two strains of rat, a series of univariate statistics were calculated. To define a population of control urine NMR spectra in multivariate space, it is possible to compute statistics based on individual spectral descriptors. The mean of each spectral region gives the position occupied by the average spectrum in the multivariate space, while the standard deviation of each region gives the extent of variation in each dimension. However, some statistics (such as relative standard deviation, skewness and kurtosis) are severely affected by a low mean value for a given descriptor. The effect of this is to emphasise certain regions, particularly the extreme edges of the spectrum where there are

Table 2
The spectral regions with the 10 highest standard deviations

Rank	Han Wistar			Sprague Dawley		
	Spectral region	Standard deviation ($\times 10^{-2}$)	Metabolite(s)	Spectral region	Standard deviation ($\times 10^{-2}$)	Metabolite(s)
1	3.02	1.03	2-OG	2.70	1.75	Citrate
2	2.46	1.02	2-OG	3.02	1.25	2-OG
3	2.70	0.76	Citrate	2.58	1.20	Citrate
4	2.74	0.74	Citrate	2.54	1.09	Citrate
5	3.26	0.69	TMAO/taurine	2.46	1.08	2-OG
6	3.98	0.64	Hippurate	3.26	1.00	TMAO/taurine
7	2.58	0.56	Citrate	2.74	1.00	Citrate
8	3.42	0.53	Taurine	3.98	0.77	Hippurate
9	3.30	0.51	Taurine	3.42	0.59	Taurine
10	2.42	0.48	2-OG	3.78	0.53	Complex[a]

Abbreviations as in Table 1.

[a] A complex region with many overlapping resonances from glucose, other sugars and α-protons of amino acids, especially Glu, Gln and Ala.

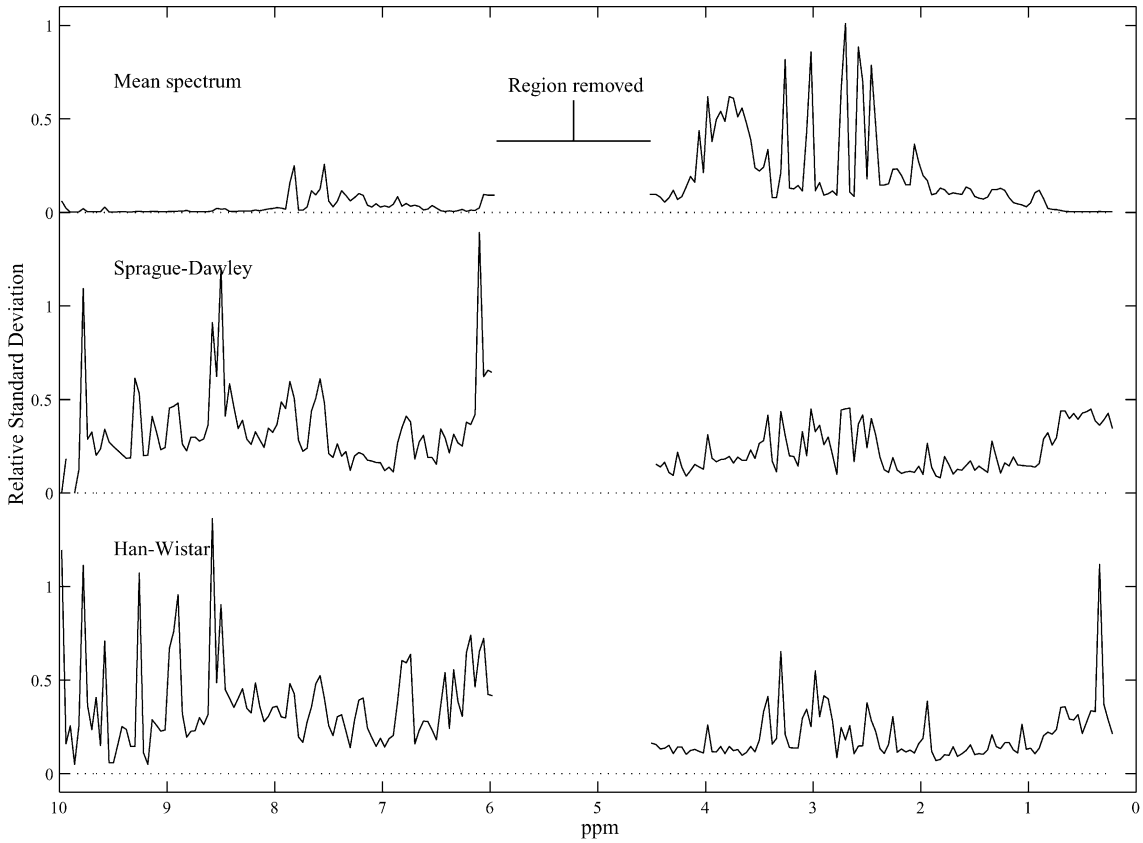

Fig. 4. Relative standard deviation data-reduced spectra of urine for the two strains of rat (lower two panels). For reference the overall mean spectrum is also shown (top panel).

no peaks and where the variation is caused entirely by noise or residual baseline distortions. To circumvent this problem the statistics for each descriptor were computed using only those spectra which had values above a given threshold for that descriptor. The threshold used (0.01% of the total spectral integral) corresponded to approximately 0.2% the value of the highest peak in the mean spectrum.

3.3. Mean NMR spectra

The mean segmented spectra for the two control rat populations are shown in Fig. 2. They are clearly alike, although not identical, as can be seen from the difference spectrum plotted beneath the two means. The metabolites represented by the 10 largest mean values, have been assigned for both the HW and SD groups and a list of the main differences between the mean spectra for the two strains is given in Table 1. As expected from simple visual analysis of the two sets of spectra, citrate, glucose, 2-oxoglutarate, and hippurate all have high mean values. Although the spectral regions demonstrating the highest mean values were similar for both strains of rat, slight differences between the two strains were observed. For example, SD rats excreted relatively greater concentrations of trimethylamine-*N*-oxide and glucose than did HW rats, and this is consistent with earlier studies [17].

3.4. Standard deviation and relative standard deviation of the NMR spectra

The standard deviations of each chemical shift region, plotted in the form of spectra are shown in Fig. 3. The first impression from this plot is that the standard deviation spectra clearly resemble closely the mean spectra. This is largely because the natural variation within each population causes the larger spectral peaks to vary by larger amounts. However, some interesting regions of higher variation can be seen. The standard deviation for spectral regions containing taurine resonances (mid-region corresponding to δ 3.26 and 3.42) was high for both HW and SD rats. Taurine excretion has been shown to vary according to diurnal and hormonal cycles [19], which can be problematical since increased excretion of taurine in episodes of hepatotoxicity has been well documented [20]. The regions that contain NMR peaks, with the 10 highest values of standard deviation are listed in Table 2.

The fact that higher concentration metabolites are varying to a larger extent implies that the important variation is actually how the amount of the metabolite varies with respect to its normal (mean) value. Therefore, the relative standard deviation spectra are plotted in Fig. 4, defined as the standard deviation divided by the mean for each spectral region. This plot is distinct in form to that shown in Figs. 2 and 3

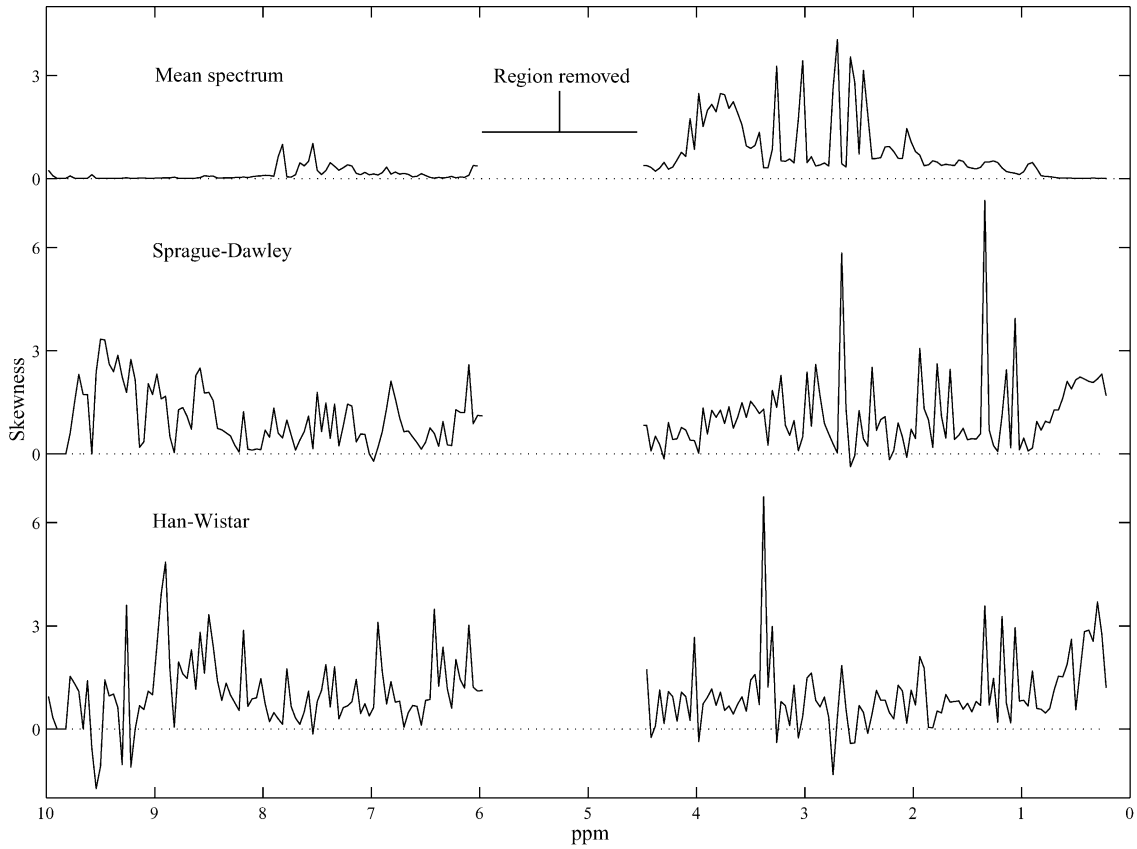

Fig. 5. Skewness data-reduced spectra of urine for the two populations (lower two panels). For reference the overall mean spectrum is also shown (top panel).

and hence the mean spectrum has been included in Fig. 4 for reference.

Many of the regions with the highest relative variability, as shown in the plots of the relative standard deviation, are not associated with large peaks in the mean spectrum. Here, and in the following identifications, the chemical shift of the centre of each integrated spectral region is quoted and the regions in decreasing order of variability as selected by each statistic are discussed. Of the 10 regions having the highest relative standard deviations, the most changeable are those at δ 9.78, 9.26, 8.90 and 8.94 in SD and δ 6.10, 9.78, 8.58 and 9.30 in HW, identified with aromatic protons from nucleoside bases. Following these, the δ 8.50 region containing formate shows high variability, although more so for SD rats than for HW rats. In both strains an unassigned NH resonance at δ 9.78 is highly variable, as are the NH resonances around δ 6.02 and 6.06 from allantoin; this is not surprising since variable cross saturation is possible in different samples arising from the suppressed water resonance. In HW rats, further variability arises from an unidentified singlet resonance at δ 6.18. Finally, the regions at δ 7.54, 7.58 and 7.86 assigned to hippurate are shown to be more variable in SD rats than in HW rats. Hippurate excretion has been proved to be particularly susceptible to changes in populations of gut microflora [21]. In particular alteration of diet has been found to result in the development of a bimodal distribution of hippurate and chlorogenic acid metabolites [22].

3.5. Skewness and kurtosis of the NMR generated metabolic profiles

To examine the departure of the populations from statistical normality, the skewness and kurtosis spectra have also been computed as shown in Figs. 5 and 6, respectively. The skewness statistic is a measure of the asymmetry of the distribution; a positive skewness indicates a tail extending toward positive values while a negative skewness indicates a negative going tail. The kurtosis, on the other hand, measures the importance of the wings of a distribution relative to that of a normal distribution. Large wings relative to a normal distribution cause a positive kurtosis, while distributions with relatively light tails result in negative kurtosis. For a normal distribution, both the skewness and kurtosis are by definition zero.

The first point to note from Fig. 5 is that the skewness values are almost all positive. This indicates that the majority of descriptor distributions have positive going tails, which reflects the occasional increase of some endogenous metabolites above their normal low level. This phenomenon could arise where an alteration in baseline physiological status occurred during the collection period. For example, if a urine

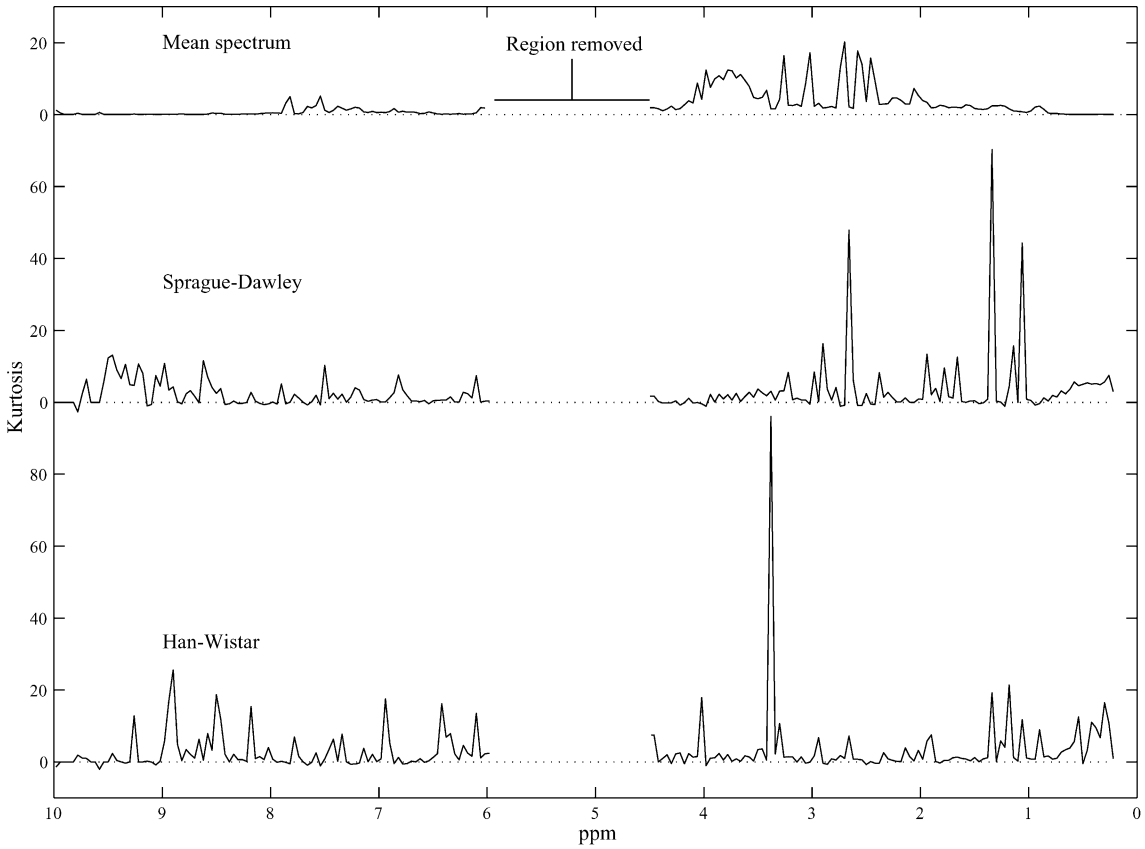

Fig. 6. Kurtosis data-reduced spectra of urine for the two populations (lower two panels). For reference the overall mean spectrum is also shown (upper panel).

sample was obtained from an animal that had not eaten for a significant time, then a slight increase in lactate and ketone bodies would be expected due to a physiological shift towards triglyceride metabolism.

A number of significant peaks were identified in the skewness spectrum, the highest of which in the SD strain arise from lactate (δ 1.34), citrate (δ 2.66) and valine (δ 1.06). Also positively skewed in SD samples, but to a lesser extent are acetate (δ 1.94) and lysine (δ 1.78). In HW rats, lactate is also skewed (less strongly than in SD animals) along with fumarate (δ 6.42) and formate (δ 8.50). Additionally for HW urine samples, spectral regions relating to nucleoside and *N*-methylnicotinamide resonances at δ 8.90, 8.94, 9.26 have high positive skew whereas similar molecular species are highly skewed at δ 9.50, 9.34 and 9.22 in SD urine spectra. Other highly skewed regions in HW samples included those at δ 3.38 (partially assignable at least, to *scyllo*-inositol), δ 1.18 (β-hydroxybutyrate) and δ 6.94 (probably attributable to a nucleoside resonance). However, the spectral integrals in these regions arise from complex overlap of several signals and therefore cannot be unequivocally assigned to a specific metabolite.

Turning to the kurtosis spectrum as shown in Fig. 6, one again observes that most of the values are positive. This indicates distributions differing significantly from Gaussian, in agreement with the skewness results. A number of very strong peaks are seen in the kurtosis spectrum. In SD urine spectra, the three highest peaks are seen to arise from lactate (δ 1.34), citrate (δ 2.66) and valine (δ 1.06) as in the skewness spectra. In contrast, by far the strongest peak in HW spectra was at δ 3.38 which has been assigned to *scyllo*-inositol. In general, similar regions to those identified by the skewness statistic are highlighted. However, in addition for HW spectra, creatinine (δ 4.02) also has a large kurtosis, while dimethylglycine (δ 2.90), and lysine/arginine at δ 1.66 are also seen to have high kurtosis in spectra generated from SD urine samples.

Citrate resonances contributed highly to the kurtosis spectrum for SD rats. The chemical shifts of citrate resonances are particularly susceptible to changes in the normal physiological pH range in urine. Although all the samples were buffered to compensate for pH variation, calcium phosphate precipitation in urine is not uncommon and this can affect the pH after the buffer adjustment has been made. Studies of the effect of buffering urine samples on citrate resonance positions showed that it was possible for these to span more than one integral region [23], and this could therefore lead to highly kurtosed distributions.

3.6. Descriptor distributions

The statistical description of the control populations is further enhanced by comparing the actual distributions (rather

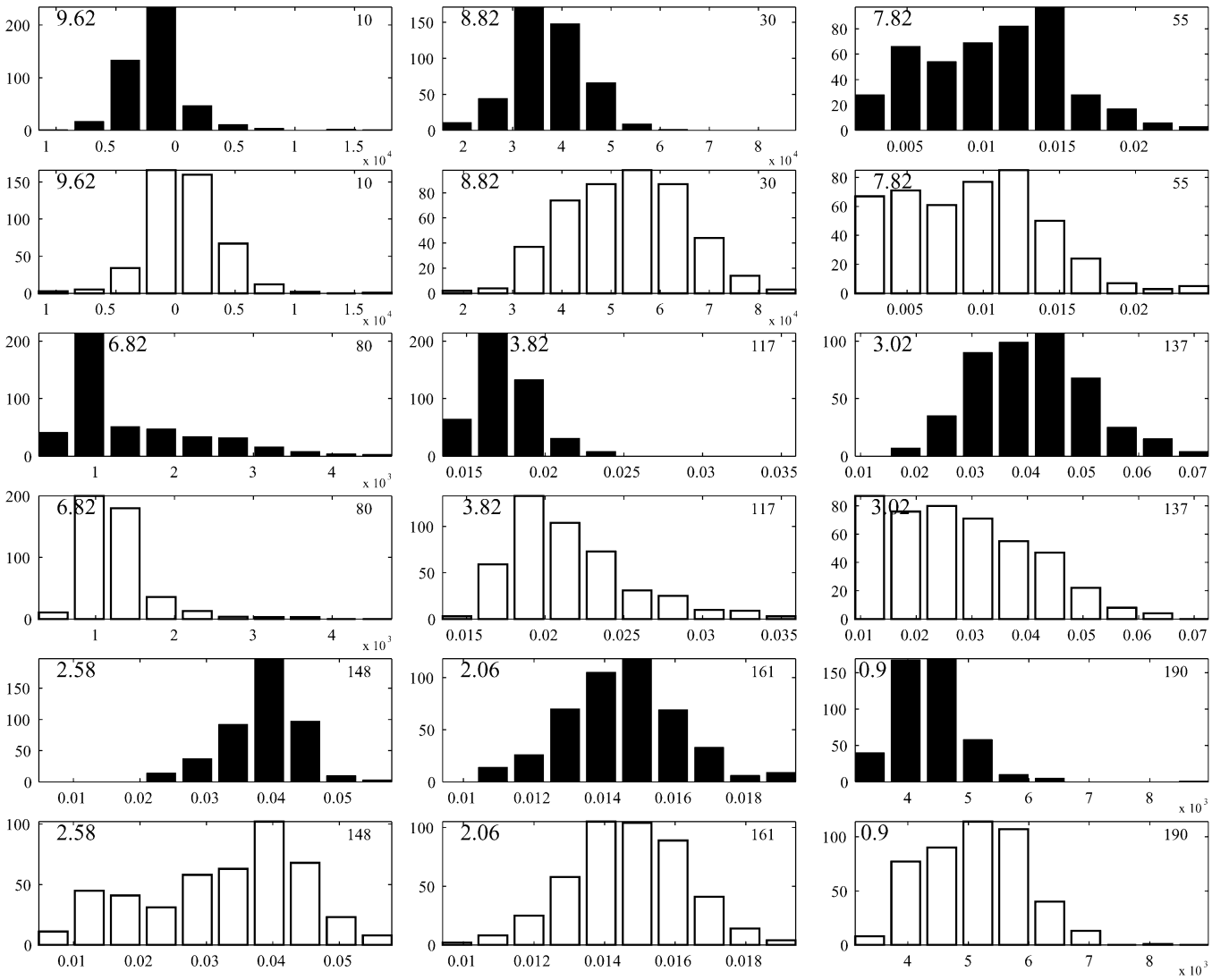

Fig. 7. Examples of the descriptor distributions for the two populations (filled: Han Wistar; open: Sprague Dawley). The number and chemical shift (in ppm) of the spectral region for each histogram are shown in the top right and top left corners, respectively.

than statistics based on those distributions). The distributions were formed *without* thresholding, i.e. all regions from all spectra were included. Some representative regions taken from across the spectra are compared in Fig. 7. The clearest message from this figure is that the distributions are not always normal, confirming the statistical implications from the preceding sections. The distributions often have tails extending to positive values, some are almost uniform across a wide range of values and others hint at bimodality.

Comparing the pairs of distributions for both strains of rat, one can see that they are often quite dissimilar, with the differences not just in the mean or spread of the distribution, but in the overall shape. For example, this is seen in the region at δ 2.58 (containing citrate), where the SD distribution appears to be similar in shape to the HW one, but with an additional tail extending to low values. At δ 3.02, corresponding to the region containing the tail of the creatinine resonance, the HW distribution appears to be almost normal with perhaps a slight positive skew. However, the SD distribution is radically different with a relatively smooth descent from high values around zero toward positive values.

3.7. Bimodal distributions

Distributions for the six regions showing the strongest evidence for bimodality are shown in Fig. 8 (selected with histograms using both 10 and 20 intervals). There is clear objective evidence that many of the distributions are not unimodal. In particular, the regions at δ 7.82, 7.54 and also 3.98 (not shown) corresponding to hippurate show very similar distributions with at least two different populations contributing. As discussed previously, hippurate excretion is highly variable and is partially dependant upon the gut microflora in the host organism. Dimethylamine, with a peak at δ 2.74 (slightly overlapped with one of the citrate resonances), shows further clear evidence for bimodality. Although the exact amount

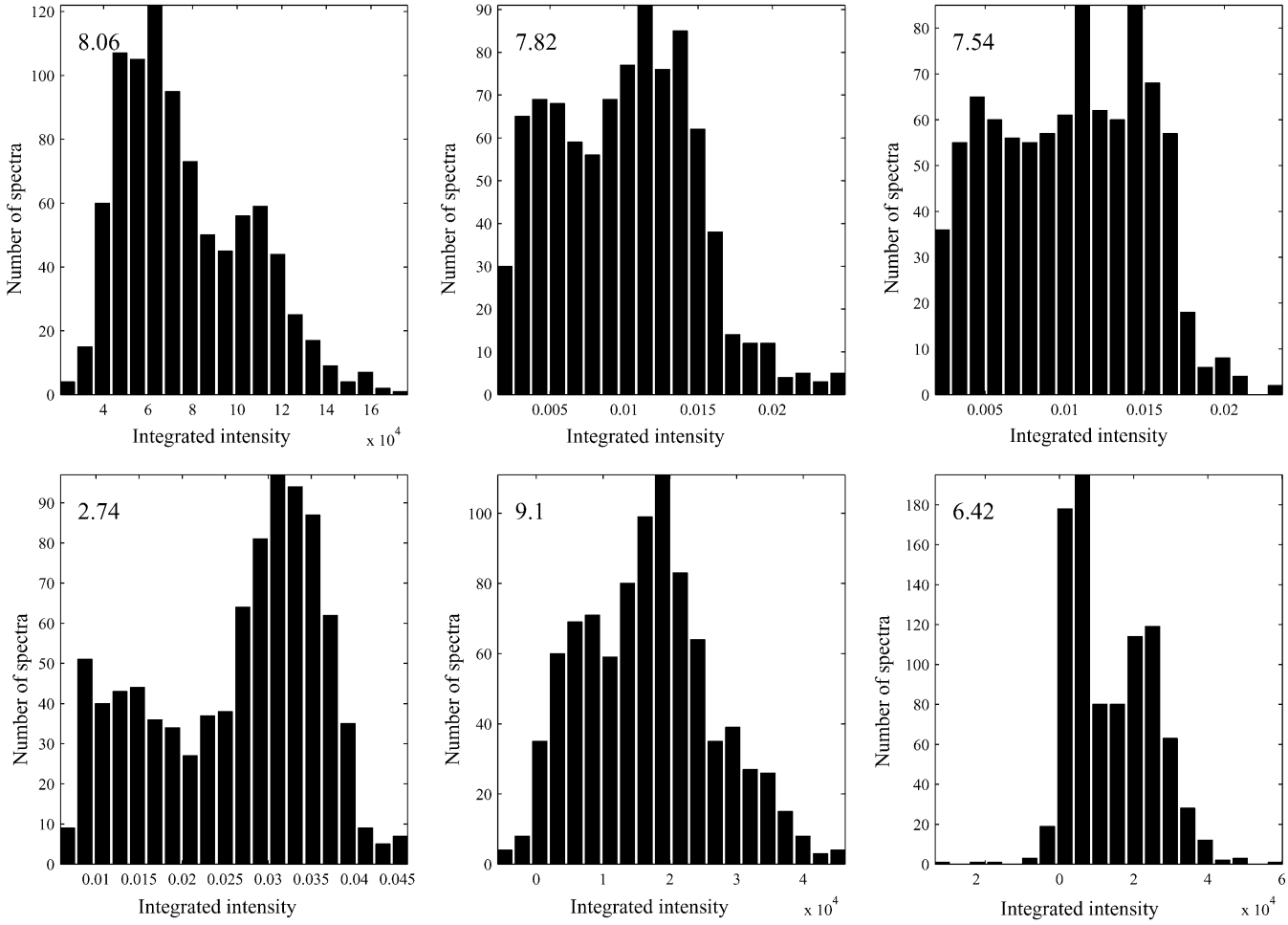

Fig. 8. Distributions of the six descriptors showing the best evidence for bimodality. The chemical shift of each spectral region is shown in the top left of each plot.

of protein in rat chow is carefully regulated, the protein source can vary from batch to batch. Variation in methylamines is likely to result from dietary differences in the protein source as fish contain a high concentration of these compounds [24]. As these studies were conducted over a period of several weeks, it is possible that dietary variation in fish protein is responsible for the bimodal distribution in dimethylamine. Other regions showing less significant evidence for multi-modality are *N*-methylnicotinic acid (δ 9.10), aromatic protons from nucleoside bases at δ 8.98 and 8.90, fumarate (δ 6.42), 2-oxoglutarate (δ 3.02), methyl groups from isoleucine/leucine together with unidentified resonances at δ 8.06 and 6.08.

3.8. Implications of strain-related differences in metabolite distributions to toxicological and functional genomic studies

This work shows the value of the combination of NMR spectroscopy and statistical evaluation as a tool in the generation and analysis of multivariate biochemical information that can be used to investigate strain and individual variation in experimental mammals. Specifically, the statistical variations in specific spectral regions of the ^{1}H NMR urine spectra obtained from control SD and HW rats have been defined. Although visually the ^{1}H NMR urine spectra for the two strains showed a high degree of similarity in the excretion patterns of major metabolites, the strains were clearly distinguishable. This concurs with previous studies showing significant differences in pathway flux and excretion between closely related strains or species of animal model [16,17]. Such metabolic pattern differences in animals kept in rigorously controlled environments are likely to be strongly reflective of differences in gene expression or protein activity in selected tissues due to polymorphisms of phenotypically silent genes in each strain. Diet, hormonal and diurnal variations can also contribute to variability, but although the samples come from a variety of studies, these are all control samples and such differences are small compared to changes caused by toxins and diseases. In addition, the gut microflora might well be different across the samples. These data also confirm that certain spectral regions corresponding to particular metabolites, and hence pathways, are hyper-variable even within strain. However, even for regions showing relatively large variations in

profile, such statistical analysis enable placement of confidence limits on the significance of toxin or disease-induced variations in metabolic profile.

We have show that, in addition to the variation in the quantitative metabolite pattern, the statistical distribution of the pattern differences can convey information regarding polymorphic variability in gene functions relative to metabolic fluxes. Moreover, this variation due to polymorphism in animal populations may be important in understanding the variability in response of a given population to a stressor such as toxic insult or disease processes. Differences in the distributions of metabolite levels demonstrate that NMR-based metabonomics can be used to characterise the metabotype, which relates to the genotypic differences. Hence it can be envisaged that there will be an increase in the use of metabonomic technology in functional genomics, with applications including metabotyping of transgenic and knockout organisms, characterising both witting and unwitting effects of genetic modification [6], understanding the biochemical consequences of genetic modification and evaluating the response of genetically modified organisms to disease or therapy.

References

[1] J.K. Nicholson, I.D. Wilson, Prog. NMR Spectrosc. 21 (1989) 444–501.
[2] K.P.R. Gartland, C.R. Beddell, J.C. Lindon, J.K. Nicholson, Mol. Pharmacol. 39 (1991) 629–642.
[3] M.L. Anthony, B.C. Sweatman, C.R. Beddell, J.C. Lindon, J.K. Nicholson, Mol. Pharmacol. 46 (1994) 199–211.
[4] J.K. Nicholson, J.C. Lindon, E. Holmes, Xenobiotica 11 (1999) 1181–1189.
[5] J.C. Lindon, J.K. Nicholson, E. Holmes, J.R. Everett, Concepts Magn. Reson. 12 (2000) 289–320.
[6] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Nat. Drug Discuss. 1 (2002) 153–161.
[7] R.D. Farrant, J.C. Lindon, E. Rahr, B.C. Sweatman, J. Pharmaceut. Biomed. Anal. 10 (1992) 141–144.
[8] E. Holmes, P.J.D. Foxall, J.K. Nicholson, G.H. Neild, S.M. Brown, C.R. Beddell, B.C. Sweatman, E. Rahr, J.C. Lindon, M. Spraul, P. Neidig, Anal. Biochem. 220 (1994) 284–296.
[9] E. Holmes, S.C. Connor, A.W. Nicholls, S. Polley, J.K. Nicholson, J.C. Lindon, J. Connelly, Chemom. Intell. Lab. Syst. 44 (1998) 251–261.
[10] B. Kowalski, D. Sharaf, D. Illman, Chemometrics, Wiley, New York, USA, 1986.
[11] W. El-Deredy, NMR Biomed. 10 (1997) 99–124.
[12] J.C. Lindon, E. Holmes, J.K. Nicholson, Prog. NMR Spectrosc. 39 (2001) 1–40.
[13] E. Holmes, S. Caddick, J.C. Lindon, I.D. Wilson, S. Kryvawych, J.K. Nicholson, Biochem. Pharmacol. 49 (1995) 1349–1359.
[14] E. Holmes, A.W. Nicholls, J.C. Lindon, S.C. Connor, J.C. Connelly, J.N. Haselden, S.J.P. Damment, M. Spraul, P. Neidig, J.K. Nicholson, Chem. Res. Toxicol. 13 (2000) 771–778.
[15] A.R. Tate, S.J.P. Damment, J.C. Lindon, Anal. Biochem. 291 (2001) 17–26.
[16] C.L. Gavaghan, E. Holmes, E. Lenz, I.D. Wilson, J.K. Nicholson, FEBS Lett. 484 (2000) 169–174.
[17] E. Holmes, J.K. Nicholson, G. Tranter, Chem. Res. Toxicol. 48 (2001) 182–191.
[18] J.G. Bundy, D.J. Spurgeon, C. Svendsen, P.K. Hankard, D. Osborn, J.C. Lindon, J.K. Nicholson, FEBS Lett. 521 (2002) 115–120.
[19] M.E. Bollard, E. Holmes, J.C. Lindon, S.C. Mitchell, D. Branstetter, W. Zhang, J.K. Nicholson, Anal. Biochem. 295 (2001) 194–202.
[20] C.J. Waterfield, J.A. Turton, M.D.C. Scales, J.A. Timbrell, Arch. Toxicol. 67 (1993) 244–254.
[21] A.N. Phipps, J. Stewart, B. Wright, I.D. Wilson, Xenobiotica 28 (1998) 527–537.
[22] J. Azmi, J. Griffin, H. Antti, R.F. Shore, E. Johansson, J.K. Nicholson, E. Holmes, Analyst 127 (2002) 271–276.
[23] C.L. Gavaghan, J.K. Nicholson, S.C. Connor, I.D. Wilson, B. Wright, E. Holmes, Anal. Biochem. 291 (2001) 245–252.
[24] S. Lowis, M.A. Eastwood, W.G. Brydon, Br. J. Nutr. 54 (1985) 43–51.