FEBS 24251 FEBS Letters 484 (2000) 169–174

An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse

Claire L. Gavaghan[a], Elaine Holmes[a], Eva Lenz[b], Ian D. Wilson[b], Jeremy K. Nicholson[a,*]

[a] *Biological Chemistry, Biomedical Sciences Division, Imperial College of Science, Technology and Medicine, University of London, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ, UK*
[b] *Astra-Zeneca Pharmaceuticals, Drug Metabolism and Pharmacokinetics, Alderley Park, Macclesfield, Cheshire SKN 4TG, UK*

Received 30 August 2000; accepted 28 September 2000

First published online 24 October 2000

Edited by Thomas L. James

Abstract As the human genome sequencing projects near completion, there is an active search for technologies that can provide insights into the genetic basis for physiological variation and interpreting gene expression in terms of phenotype at the whole organism level in order to understand the pathophysiology of disease. We present a novel metabonomic approach to the investigation of genetic influences on metabolic balance and metabolite excretion patterns in two phenotypically normal mouse models (C57BL10J and Alpk:ApfCD). Chemometric techniques were applied to optimise recovery of biochemical information from complex ^{1}H NMR urine spectra and to determine metabolic biomarker differences between the two strains. Differences were observed in tricarboxylic acid cycle intermediates and methylamine pathway activity. We suggest here a new 'metabotype' concept, which will be of value in relating quantitative physiological and biochemical data to both phenotypic and genetic variation in animals and man.

Key words: Metabonomics; ^{1}H NMR spectroscopy; Metabolite variation; Physiological function; Metabotype; Phenotype

1. Introduction

In order to reap the benefits of the human genome project and other major programmes of genetic research, it is necessary to correctly place genetic information in the context of physiological and biochemical mechanisms in the whole organism. It is also imperative to understand the functional consequences of genetic modification in transgenic organisms, both from the point of view of witting and unwitting interventions.

Transcriptomic approaches utilising proprietary gene chips can rapidly generate gene expression data, but genetic up/down regulation needs to be related to physiological processes that occur at various later time points. At one level, proteomics can be used to study protein variation in relation to gene expression [1]. However, proteomic methods are time consuming and expensive, and still do not lead to a direct understanding of physiological consequences of genetic modification because: (a) proteins may have many post translational modifications that complicate understanding the relationships to physiological function and (b) that proteins act 'in concert' to control physiological processes and are spatially and temporally distributed in such a way that, with current scientific knowledge, proteomics is not readily predictive of biological end-points, e.g. drug toxicity.

In consequence, we have developed an efficient nuclear magnetic resonance (NMR)-based metabonomic approach to understanding pathophysiological and toxicological processes [2–9], in which multivariate metabolic events over a period of time can be interrogated using pattern recognition (PR) analysis methods. In principle metabonomics can be extended to the study of genetic variation in experimental animals and man. Thus we have previously defined metabonomics as 'the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification' [2].

High resolution ^{1}H NMR spectroscopy presents an efficient and non-destructive means of simultaneously generating data on hundreds (potentially thousands) of metabolites in biofluids or tissues [10]. The metabolic status of the organism is reflected in the spectral profile of a biofluid and there are species differences observable in the basal NMR spectrum of a biofluid such as blood plasma [11]. The complexity of these information-rich NMR spectral profiles provides a significant analytical challenge, as does the relationship between the profile and biological interpretation. However, information recovery can be maximised by applying multivariate statistical tools to the analysis of these NMR data. We have shown that ^{1}H NMR spectroscopy of biological matrices, coupled with appropriate pattern recognition (PR) and multivariate statistical analysis methods offers a novel in vivo approach to the investigation of drug toxicity. Applications of this metabonomic technology include the identification of biomarkers of toxicity and disease [4–6,12,13], monitoring sequential metabolic perturbations in biofluids and tissues following toxic insult [7,8] and metabolic characterisation of physiological variance in humans under mild physiological stress [14].

Various physiological factors affect the metabolic composition of biological matrices including diet, state of health, diurnal cycles, genetic drift, stress and strain differences and it is necessary to characterise these differences in order to distin-

*Corresponding author. Fax: (44)-207-594 32221.
E-mail: j.nicholson@ic.ac.uk

PII: S0014-5793(00)02147-5

guish between physiological and pathophysiological responses in animal models. Previous work has shown that it is possible to distinguish between two strains of standard laboratory rat (Han Wistar and Sprague Dawley) using NMR-based metabonomic methods [9].

Mouse models are becoming increasingly important in risk assessment and as models of human disease, particularly with respect to the development of transgenic technology [15,16]. In this study we have applied ^{1}H NMR-based metabonomic techniques to characterise the urinary profiles of two experimental mouse models viz. C57BL10J and Alpk:ApfCD, which are both phenotypically normal, and to explore the extent of physiological variation and metabolic differences between the two strains. This work shows the future potential for the use of metabonomic studies in the investigation of transgenic animals and other animal models of human disease and the effects of genetic manipulation on whole organ physiology.

2. Materials and methods

2.1. Animal handling and collection of urine for ^{1}H NMR spectroscopy

Urine samples were obtained from 6 week old C57BL10J mice ($n = 24$) and Alpk:ApfCD mice ($n = 28$) by manipulation of the lower abdomen. Samples were collected into glass vials and were stored at −20°C pending NMR spectroscopic analysis. Prior to the study, animals were housed collectively in polycarbonate solid bottom cages, according to strain, and were allowed free access to water and food. Mice were fed pelleted irradiated rat and mouse diet 1 (Special Diet Services, Witham, Essex, UK), from weaning until the end of the experiment. Animals were maintained at room temperature with artificial 12 h dark/12 h light cycles

2.2. Sample preparation and ^{1}H NMR spectroscopy of whole mouse urine

In order to control the sample pH to a narrow range, 300 mM sodium phosphate buffer (pH 7.4) was added to the urine samples. Due to the small volume of mouse urine available (typically 100 μl), sodium phosphate buffer was added to make the overall sample up to 200 μl volume. This differential dilution of the samples does not affect the PR analysis because of the normalisation processes used (see below). ^{1}H NMR spectra were acquired on a Bruker DRX-500 spectrometer at ambient temperature using the following standard pulse sequence:

RD-90°-t_1-90°-t_m-90°-acquire FID

Secondary irradiation of the water signal was achieved during the mixing time t_m (100 ms). Typically 256 free induction decays (FIDs) were collected into 32k data points using a spectral width of 5995.204 Hz, utilising a 8.7 ms pulse width (90° pulse angle), an acquisition time of 2.733 s, a relaxation delay of 2.3 s and a total pulse recycle time of 5.033 s. The FIDs were multiplied by an exponential weighting function corresponding to a line broadening of 0.3 Hz prior to FT. The acquired NMR spectra were phased, baseline-corrected and referenced to hippurate (*o*-CH; δ7.84).

2.3. Data reduction of the NMR spectra

The ^{1}H NMR spectra (δ0.2–9.7) were automatically data-reduced to 193 integral segments using the AMIX (Analysis of MIXtures software package, version 2.5, Bruker Analytische Messtechnik, Karlsruhe, Germany) with the regions δ6.0–4.52 set to zero integral. The region δ6.0–4.52 was excluded from the analysis in order to remove the effects of variation in the suppression of the water resonance and the effects of variation in the urea signal caused by partial cross solvent saturation via solvent exchanging protons. Each segment consisted of the integral of the NMR region to which it was associated. The data were normalised in Excel (Microsoft®, Excel 97, SR-2) and then exported into the SIMCA-P (version 8.0, Umetrics, Umea, Sweden) software package. All subsequent analysis used centred scaling.

2.4. Principal components analysis (PCA) of the ^{1}H NMR spectral data

PCA is a data visualisation method that is useful for overviewing relationships or 'groupings' within multivariate data [17]. The data are represented in K-dimensional space (where K is equal to the number of variables) and are subsequently reduced into a few principal components which are descriptive dimensions that describe the maximum variation within the data. Any prior knowledge relating to class membership is not used in the PCA analysis. The PCs can be displayed graphically to yield a 'scores plot', which displays any patterns or groupings within the data and can also be useful in highlighting outliers. A PCA model was constructed using all samples. The scores plot of PC1 versus PC2 was examined for separation between the two strains of mice: C57BL10J and Alpk:ApfCD (referred to as classes).

Where clusters are not distinctly separated in the scores plots and groups overlap, it is desirable to fine-tune the models to allow complete distinction between groups. Furthermore, it can often arise that the maximum variation, as represented by the PCs in PCA, may not always produce the maximum separation between the classes. In these cases it is possible to use methods which use other directions in multivariate space for establishing separations between classes. One such method is partial least squares-discriminant analysis (PLS-DA).

2.5. Prediction of classes by PLS-DA

Having established the presence of strain-related clusters, data were analysed by PLS-DA with a view to optimising the separation between the C57BL10J and Alpk:ApfCD strains.

PLS-DA is a linear regression method whereby the multivariate variables corresponding to the observations (spectral descriptors) are related to the class membership for each sample [18]. The method is supervised, i.e. the class membership of the samples is included in the calculation. PLS-DA provides an easily understood graphical means to identifying the spectral regions of difference between the classes and also allows statistical evaluation as to whether the differences between classes are significant.

A PLS-DA model was calculated including all samples (C57BL10J = class 1 and Alpk:ApfCD = class 2). For the purpose of cross-validation, the samples were divided into a training set and a validation set, each of equal size and comprising of ~50% C57BL10J and 50% Alpk:ApfCD, and a PLS-DA model was calculated for both sets. The PLS-DA coefficients were examined and compared for all three PLS-DA models (whole set, training set and validation set) to establish the spectral regions of variance between the C57BL10J and Alpk:ApfCD strains.

To establish whether the differences were significant between strains, the PLS-DA model for training set was used to predict the strain of samples in the validation set and vice versa. Class (strain) membership was predicted using a value of 0.5 dividing line between the two classes including values predicted using a class membership probability value > 0.01.

Table 1
Summary of prediction results investigating two strains of mice by PLS-DA

PLS-DA model	Percentage of correct predicted classifications for		Percentage of predictions with significance < 0.01 for	
	one component model	two component model	one component model	two component model
Alpk:ApfCD vs. C57BL10J (training set)	87	100	81	94
Alpk:ApfCD vs. C57BL10J (validation set)	87	98	85	94

Calculated using class membership probability > 0.01.

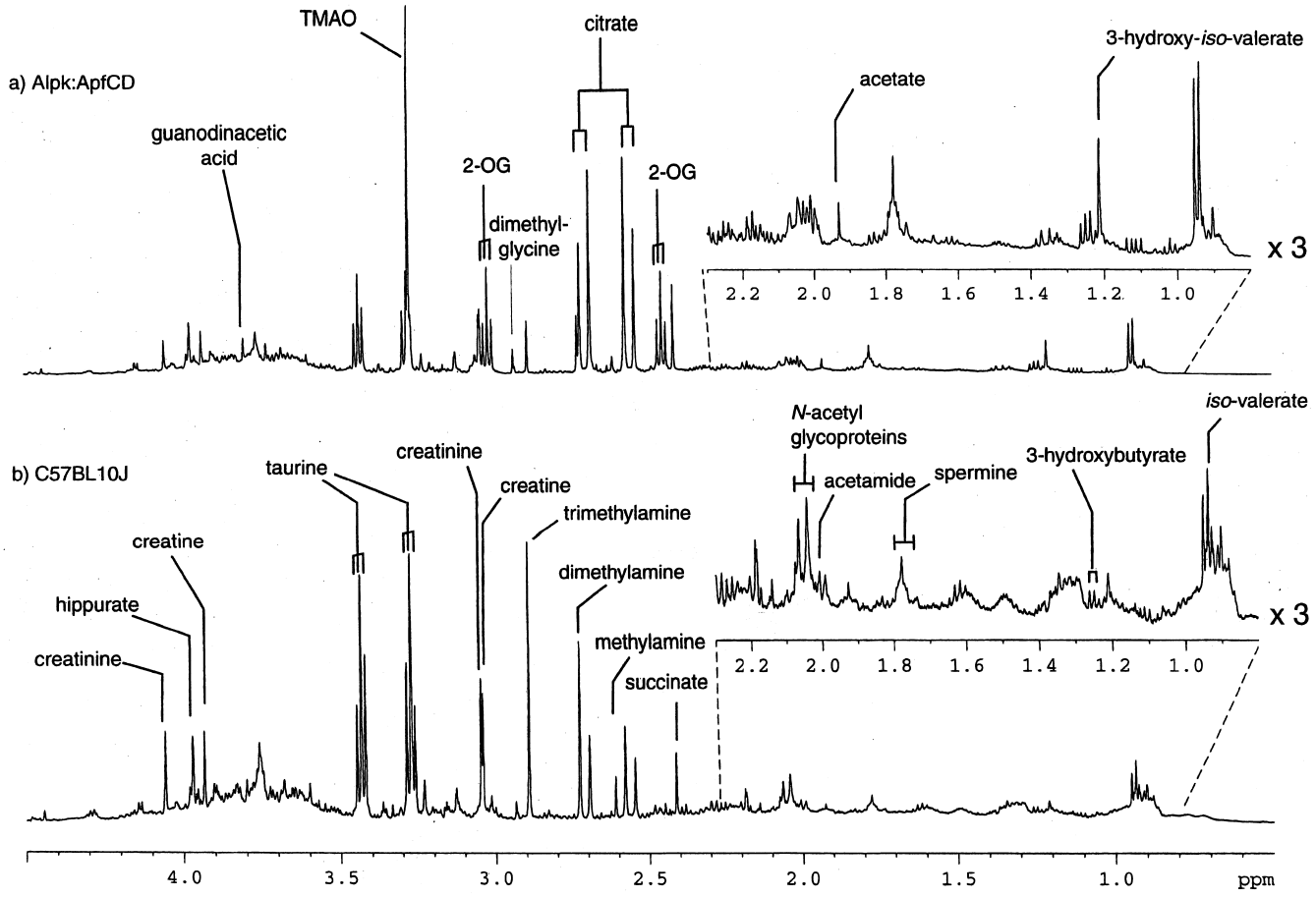

Fig. 1. 500 MHz ^{1}H NMR spectra of typical urine samples obtained from (a) Alpk:ApfCD mouse and (b) C57BL10J mouse. The first two principal components of all the data described 65.5 and 52.9%, respectively, of the overall variance.

3. Results

3.1. ^{1}H NMR spectroscopy of urine samples from C57BL10J and Alpk:ApfCD mice

Visual inspection of the ^{1}H NMR urine spectra revealed changes in the patterns associated with a variety of differences in the two strains (Fig. 1). Inter-animal variation in urine composition and the complexity of the spectra, coupled with differences in the volume of urine excreted, makes visual comparison of spectra difficult, hence chemometric methods of spectral analysis were employed.

3.2. PCA of ^{1}H NMR urinary data

The scores plot of PC1 versus PC2 showed that, with the exception of one sample obtained from a Alpk:ApfCD mouse, the samples were separated into two distinct clusters relating to the strain in the first PC (Fig. 2). Samples obtained from C57BL10J mice formed a tighter cluster than did those from Alpk:ApfCD mice.

3.3. Investigation of the differences between strains of Alpk:ApfCD mice and C57BL10 by PLS-DA

PLS-DA was applied to these data to investigate whether Alpk:ApfCD and C57BL10 mice were significantly different in terms of their urinary metabolite profile, and to determine which metabolites showed the greatest variation between the two strains. The prediction results are summarised in Table 1 and show that for a two component model, the strain of mouse can be predicted accurately in 98% of cases. Furthermore, strain can be predicted with a significance level of $\geq 81\%$ for a one component model and with a significance level of $\geq 94\%$ for a two component model using a 99% confidence limit.

The regression coefficients were plotted for the PLS-DA models to investigate and indicate the major spectral variations between two strains (Fig. 3). The metabolites contained within the selected regions (as determined from regression coefficient plots) were identified and have been summarised in Table 2. The most significant metabolites distinguishing between strains included 2-oxoglutarate, citrate, trimethylamine-*N*-oxide, and guanidinoacetic acid (relatively higher concentration in Alpk:ApfCD mice) together with taurine and creatinine (relatively higher concentration in C57BL10 mice). Minor strain differences in the urinary excretion of hippurate and creatine were observed. In addition, the urinary levels of both dimethylamine and trimethylamine were found to be greater in C57BL10J mice, in contrast to trimethylamine-*N*-oxide which was present in relatively higher concentrations in Alpk:ApfCD mice. Resonances from the methyl groups of unassigned organic acids made a minor contribution to the discrimination between the two strains (Fig. 1, Table 2).

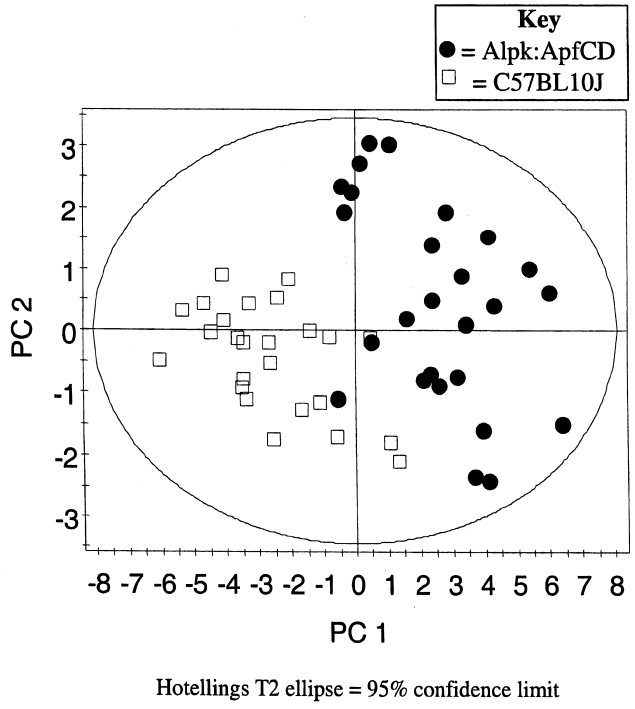

Fig. 2. PC scores plot derived from the ^{1}H NMR spectra of urine samples obtained from Alpk:ApfCD and C57BL10J mice.

4. Discussion

4.1. Strain variation in the biochemical composition of urine

In this study C57BL10J mice were found to excrete proportionally higher concentrations of dimethylamine and trimethylamine, together with lower concentrations of trimethylamine-*N*-oxide in their urine (Figs. 1 and 3). The methylamines are present in the urine, plasma and gastric fluid of rats. One of the principle roles of the methylamines is to act as non-perturbing renal osmolytes. They are present in high concentrations in the kidney and serve to protect against high concentrations and balance osmotic pressure between extra- and intra-cellular fluid [19]. The major route of biosynthesis of the methylamines derives from the degradation of dietary choline which is metabolised to trimethylamine by the enzymes of the gut microflora (Fig. 4) [20,21]. Betaine is also a degradation product of dietary choline, and in cases where the intestinal microflora has been reduced, the urinary excretion of betaine is elevated at the expense of dimethylamine and trimethylamine-*N*-oxide [22]. However, choline is not the sole precursor of dimethylamine and synthesis can also occur via endogenous pathways [23] and direct conversion to trimethylamine-*N*-oxide can also occur [24]. Since both the C57BL10J and Alpk:ApfCD mice used in this study were obtained from colonies which have been maintained in the same laboratory since 1967, it is likely that the difference in dimethylamine: trimethylamine-*N*-oxide and trimethylamine:trimethylamine-*N*-oxide ratio in these murine models is due to strain differences in enzyme activity rather than the presence of different gut microflora.

Urine samples from Alpk:ApfCD mice contained relatively higher urinary levels of creatine and lower urinary concentrations of guanidinoacetic acid in comparison with urine samples from C57BL10J mice. One route of biosynthesis for creatine involves the transfer of a methyl group from *S*-adenosylmethionine to guanidinoacetic acid, catalysed by the

Table 2
Summary of differences between Alpk:ApfCD and C57BL10 strains analysed by PLS-DA

Integrated spectral region	Ratio of metabolite concentration Alpk:ApfCD to C57BL10	δ^1H	Metabolite(s) present in selected region
7.84	+	7.84, d	hippurate
7.64	+	7.65, t	
7.56	+	7.56, t	
4.04	− −	4.06, s	creatinine
4.00–3.96	++	3.97, d	hippurate
		3.94, s	creatine
3.92–3.8	++	3.80, s	guanidinoacetic acid
3.44	−−	3.43, t	taurine
3.32	−	3.28, t	taurine
3.28	++	3.28, t and 3.27, s	taurine* and TMAO
3.04	+	3.04, s and 3.05, s	creatine and creatinine**
3.00	++	3.02	2-oxoglutarate
2.92	+		dimethylglycine
2.88	− −	2.88	trimethylamine
2.48–2.72	+++	2.80, d and 2.67, d	citrate
***	(decrease)	2.73, s	dimethylamine
2.44	++	2.45, t	2-oxoglutarate
2.36–1.88	−		overlapping resonances including amino acids and *N*-acetyl glycoproteins
1.78	−	1.76, m	spermine
1.92	−	1.93	acetate

Abbreviations: trimethylamine-*N*-oxide (TMAO). Scale: a net increase/decrease in the integrated spectral regions for Alpk:ApfCD mice is indicated by +/− = > 0.005 and < 0.01, ++/− − = > 0.05 and < 0.1, +++/− − − = > 0.1 as determined by the regression coefficients. *TMAO decreased in white overlapped with taurine therefore only 3.32 buckets indicates that taurine is decreased, **resonances overlapped. Effects are counter-acted, ***dimethylamine is present in decreased concentrations in Alpk:ApfCD mice. However, because the resonance for dimethylamine is overlapped with citrate, the net change from the PLS statistics indicates a net increase for this spectral region in Alpk:ApfCD mice.

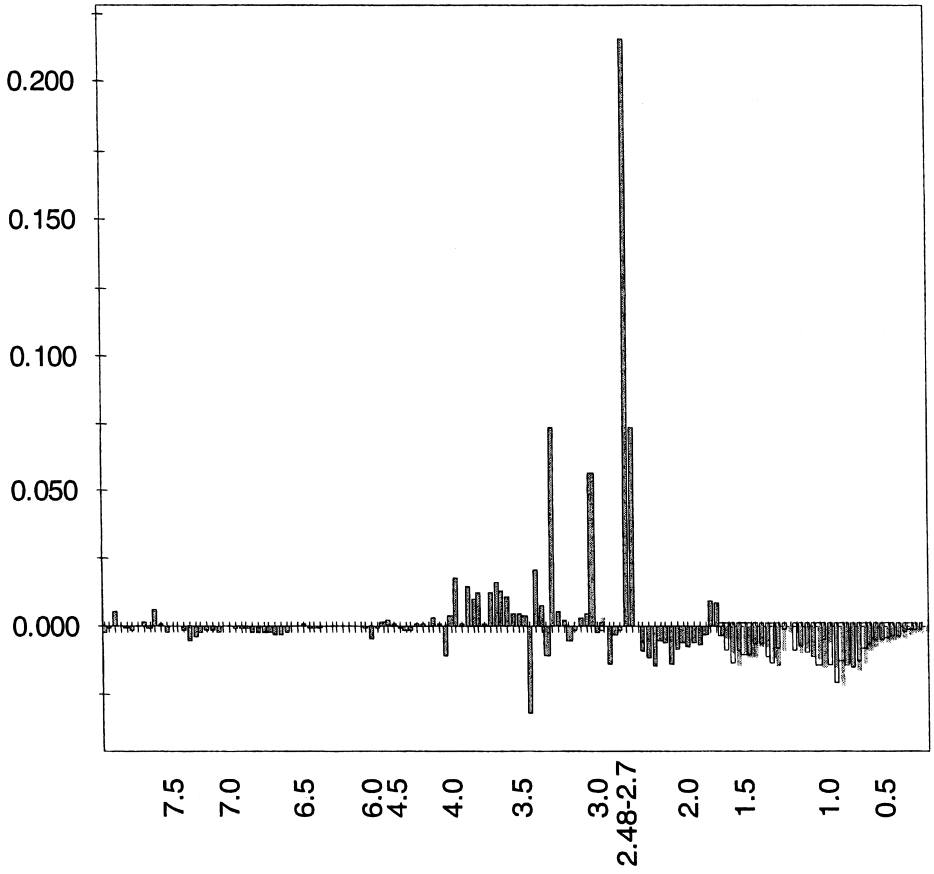

Fig. 3. Regression coefficients of the ^{1}H NMR variables for Alpk:ApfCD mouse strain derived from the PLS-DA model of Alpk:ApfCD and C57BL10J mouse strains. Each bar represents a spectral region covering 0.04 ppm. This plot shows how the ^{1}H NMR metabolite profile of the Alpk:ApfCD strain differs to that for the C57BL10J strain where a positive value indicates that relatively greater urinary concentrations of a metabolite are present in the Alpk:ApfCD mouse.

enzyme guanidinoacetate methyltransferase (GAMT) [25]. Therefore it is likely that the enzyme GAMT is slightly up-regulated in the Alpk:ApfCD mouse.

Other significant strain differences included the presence of relatively lower urinary concentrations of citrate and succinate in the C57BL10J mouse. These metabolites are intermediates in the tricarboxylic acid cycle and may therefore indicate a strain difference in mitochondrial function or activity.

4.2. Validity of the NMR-based metabonomic approach

We have shown that ^{1}H NMR-based metabonomic approaches are an efficient means of generating metabolic biofluid profiles for animal models, and indeed are sensitive enough to distinguish between two closely related strains, the Alpk:ApfCD mouse and the C57BL10 mouse. In this study we were able to predict the correct strain of mouse for 87% of the samples analysed with ≥81% of these samples being significantly different from the other strain at the 99% confidence limit (Table 1). Previous work using this technology to differentiate between the urinary profiles of two commonly used experimental rat strains (Han Wistar and Sprague Dawley) also supports the contention that genetic strain variation between animals markedly affects their biochemical profiles in physiologically normal animals [9]. This leads us to suggest the novel concept of the 'metabotype' which we define as 'a probabalistic multiparametric description of an organism in a given physiological state based on analysis of its cell types, biofluids or tissues'. The metabotype concept could be widely applicable to define the metabolic consequences of a genetic modification or disease process that may not

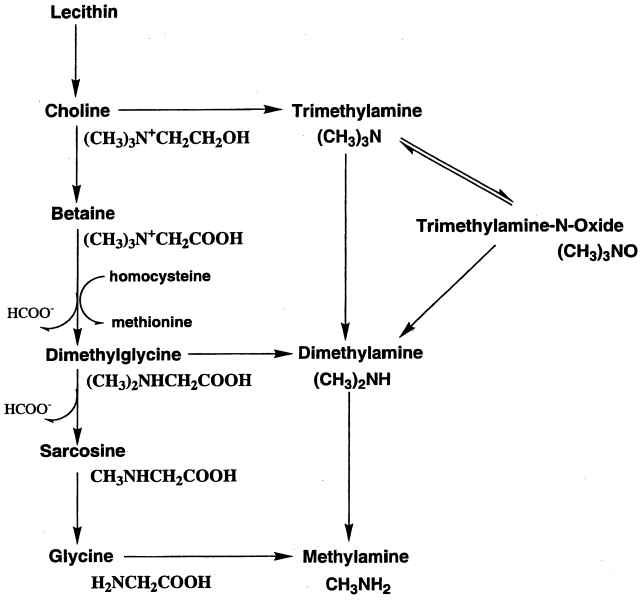

Fig. 4. Schematic representation of the biosynthesis of the methylamines.

be directly observable from gross examination of the phenotype and has the advantage of being quantitative in *n* dimensions, thus allowing statistical comparisons of genetic modification on physiological status.

The metabolic profiling of murine biofluids also offers particular potential in the field of evaluating transgenic models of disease and drug efficacy. The correlation between genomic and proteomic data has not, as yet, been defined and little is known regarding the effect of gene expression on the holistic function of the organism at a metabolic level. Metabonomics thus offer the possibility of establishing a link between gene expression and altered physiological function.

A further benefit of NMR-based metabonomics of biofluids is the capability of obtaining a sequential uninterrupted set of biofluid samples collected non-invasively over time [8], thereby allowing continuous analysis of the metabolic profile, unlike genomic and proteomic analysis which can only provide a 'snapshot' picture of the status of a particular cell type. Since the development of disease or toxicity is generally a dynamic process, often with several phases each represented by a suite of unique biochemical markers, the ability to monitor dynamic metabolic profiles is useful.

We have shown in this study that the metabolite excretion profiles in the Alpk:ApfCD mouse and the C57BL10 mouse, which demonstrate no phenotypic abnormalities, are significantly different. Many of the metabolites that exhibit differences between strains are constituents of the same metabolic pathways. Thus, we have shown that in principle NMR-based metabonomics can be an effective means of relating genetic profile of an organism to physiological function.

Acknowledgements: We are grateful to Erik Johansson for discussions. This work was supported by Astra-Zeneca.

References

[1] Anderson, N.L., Taylor, J., Hofmann, J.P., Esquer-Blasco, R., Swift, S. and Anderson, N.G. (1996) Toxicol. Pathol. 24, 72–76.
[2] Nicholson, J.K., Lindon, J.C. and Holmes, E. (1999) Xenobiotica 11, 1181–1189.
[3] Gartland, K.P.R., Sanins, S.M., Nicholson, J.K., Sweatman, B.C., Beddelll, C.R. and Lindon, J.C. (1990) NMR Biomed. 3, 166–172.
[4] Anthony, M.L., Rose, V.S., Nicholson, J.K. and Lindon, J.C. (1995) J. Pharm. Biomed. Anal. 13, 205–211.
[5] Holmes, E., Nicholls, A.W., Lindon, J.C., Ramos, S., Spraul, M., Neidig, P., Connor, S.C., Connelly, J. and Nicholson, J.K. (1998) NMR Biomed. 11, 235–244.
[6] Howells, S.L., Maxwell, R.J., Peet, A.C. and Griffiths, J.R. (1992) Magn. Res. Med. 28, 214–236.
[7] Holmes, E., Bonner, F.W., Sweatman, B.C., Lindon, J.C., Beddell, C.R., Rahr, E. and Nicholson, J.K. (1992) Mol. Pharmacol. 42, 922–930.
[8] Beckwith-Hall, B.M., Nicholson, J.K., Nicholls, A.W., Foxall, P.J.D., Lindon, J.C., Connor, S.C., Abdi, M., Connelly, J. and Holmes, E. (1998) Chem. Res. Toxicol. 11, 260–272.
[9] Holmes, E., Nicholls, A.W., Lindon, J.C., Connor, S.C., Connelly, J.C., Haselden, J.N., Damment, S.J.P., Spraul, M., Neidig, P. and Nicholson, J.K. (2000) Chem. Res. Toxicol. 13, 771–778.
[10] Nicholson, J.K. and Wilson, I.D. (1989) Prog. NMR Spectrosc. 21, 444–501.
[11] Nicholson, J.K., Buckingham, M.J. and Sadler, P.J. (1983) J. Biochem. 211, 606–615.
[12] Tate, A.R., Foxall, P.J.D., Holmes, E., Moka, D., Spraul, M., Nicholson, J.K. and Lindon, J.C. (2000) NMR Biomed. 13, 1–8.
[13] Holmes, E., Foxall, P.J.D., Spraul, M., Farrant, R.D., Nicholson, J.K. and Lindon, J.C. (1997) J. Pharm. Biomed. Anal. 15, 1647–1659.
[14] 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. and Neidig, P. (1994) Anal. Biochem. 220, 284–296.
[15] Williams, R.S. and Wagner, P.D. (2000) J. Appl. Physiol. 88, 1119–1126.
[16] Rao, S. and Verkman, A.S. (2000) Am. J. Physiol. Cell Physiol. 279, C1–C18.
[17] Beebe, K.R., Pell, R.J. and Seahsholt, M.B. (1998) in: Chemometrics: A Practical Guide, John Wiley and Sons, New York.
[18] Wold, S., Sjostrom, M. and Eriksson, L. (1998) Partial least squares projections to latent structures (PLS), in: Chemistry, Encyclopedia of Computational Chemistry, Elsevier, Amsterdam.
[19] Gullans, S.R., Heilig, C.W., Stromski, M.E. and Blumenfeld, J.D. (1989) Ren. Physiol. Biochem. 12, 191–201.
[20] Asatoor, A.M. and Simenhoff, M.I. (1965) Biochim. Biophys. Acta 111, 384–392.
[21] Baker, J.R. and Chaykin, S. (1962) J. Biol. Chem. 237, 1309–1313.
[22] Zeisel, S.H., Wishnok, J.S. and Blusztajn, J.K. (1983) J. Pharm. Exp. Ther. 225, 320–324.
[23] Zeisel, S.H., DaCosta, K. and Fox, J.G. (1985) Biochem. J. 232, 403–408.
[24] Zhang, A.Q., Mitchell, S.C. and Smith, R.L. (1998) Food Chem. Toxicol. 36, 923–927.
[25] Takeda, M. (1990) Nippon Jinzo Gakkai Shi 32, 117–126.