Journal of Pharmaceutical & Biomedical Analysis
Vol. 9, No. 8, pp. 663–669, 1991
Printed in Great Britain

0731–7085/91 $3.00 + 0.00
Pergamon Press plc

Interpreting complicated chromatographic patterns*

E. JELLUM,†‡ M. HARBOE,§ G. BJUNE‖ and S. WOLD¶

‡*Institute of Clinical Biochemistry, Rikshospitalet, Oslo, Norway*
§*Institute of General and Rheumatologic Immunology, Oslo, Norway*
‖*Armauer Hansen Research Institute, Addis Ababa, Ethiopia*
¶*Chemometrics, University of Umeå, Umeå, Sweden*

Abstract: GC–MS, HPLC, automatic amino acid analysis, high-resolution two-dimensional electrophoresis and capillary electrophoresis are suitable for the multicomponent analysis of body fluids and tissues. Manual interpretation of the complex metabolite and protein profiles thereby obtained is usually difficult, except in the case of metabolic disorders, where major deviations from the normal profiles often are observed. Implementation of multivariate data analysis makes it possible to retrieve diagnostic information that otherwise may be overlooked, as shown in this report where patients with leprosy have been examined. Urine samples were analysed by ion-exchange chromatography and by GC–MS to obtain profiles of amino acids and organic acids. Qualitative and quantitative information on 68 metabolites were then analysed by principal components analysis (PCA) and by partial least square models (PLS). Three different PLS dimensions were found (cross-validation) corresponding to controls (persons without leprosy), paucibacillary and multibacillary leprosy.

Keywords: *Chromatography–mass spectrometry; amino acids; organic acids; multivariate data analysis; PCA and PLS-models, leprosy classification.*

Introduction

A number of chromatographic and electrophoretic techniques are available for multicomponent analysis of biological samples. The methods include gas chromatography–mass spectrometry (GC–MS), high-performance liquid chromatography (HPLC), high resolution two-dimensional electrophoresis (2D-PAGE) and high-performance capillary electrophoresis (HPCE). These methods are suitable for the study of body fluids and cells in health and disease, and some have proven to be of considerable diagnostic potential [1–6].

Combined use of these techniques may result in the resolution of hundreds, even more than one thousand biological compounds, with a large range of molecular weights. Manual interpretation of metabolic profiles (these comprise amino acids, organic acids, carbohydrates, fatty acids, steroids, bile acids, etc.) is difficult, except perhaps in the case of metabolic disorders. In this type of disease with an inherited enzyme defect, major deviations from the normal profiles are often seen, and may readily be detected by a trained person. Chromatographic profiling techniques have therefore been implemented in all laboratories engaged in the systematic diagnosis of metabolic disorders [1–6].

Most other diseases, however, result in much more subtle alterations of the metabolic profiles, and details that might carry diagnostic information may easily be overlooked. A somewhat similar situation is apparent for high resolution protein-profiling, particularly 2D-PAGE which may separate over 2000 cellular polypeptides [7]. Manual interpretation of complex protein patterns where qualitative changes in proteins occur, i.e. appearance or disappearance of certain spots on the 2D-gels, can after some experience be detected visually (e.g. refs 6–12). Quantitative changes, on the other hand, are difficult to keep track of without a data imaging system.

Common to all multicomponent analytical systems is the large amount of data produced, both qualitative and quantitative. In order to improve the retrieval of information from these data sets, multivariate data analysis may be implemented. Previously the SIMCA programme [13–15] has been used to interpret

*Presented at the Symposium on "Chemometrics in Pharmaceutical and Biomedical Analysis", November 1990, Stockholm, Sweden.

†Author to whom correspondence should be addressed.

metabolic profiles from surgically removed brain tumours, and to differentiate between different types of tumour [16]. Another example of the application of multivariate data analysis to the interpretation of complex chromatographic profiles, namely a study on leprosy patients with different histopathological classification is now reported. Urine samples were analysed by ion-exchange chromatography and by GC–MS to obtain profiles of amino acids and organic acids. Qualitative and quantitative information on a number of metabolites were then analysed by principal components analysis (PCA) and by partial least square models (PLS).

Experimental

Patient specimens

Urine samples were obtained from 10 Ethiopian persons referred to the Armauer Hansen Research Institute, Addis Ababa. Four were classified histopathologically as paucibacillary leprosy, two as multibacillary leprosy, whereas the remaining four samples were from persons without leprosy, serving as controls. The samples were kept frozen and hand-carried under identical conditions by airplane from Addis Ababa to Oslo, Norway, for analysis.

Organic acid analysis by GC–MS

This was carried out as described previously [6]. In brief, the urine samples (5 ml) were acidified and the organic acids were extracted with diethyl ether and methylated with diazomethane before analysis. A Finnigan (Sunnyvale, CA, USA) GC–MS, model 4021C with an Incos/Nova 4 data system was used. The GC-column was a fused silica capillary, 30 m, 0.25 mm i.d. coated with SP-1000. The major GC-peaks were identified by mass spectral library search [17] and peak heights were measured as arbitrary units for the quantity of each identified compound. *n*-eicosane was used as internal standard.

Amino acid analysis

The urinary amino acids were quantitatively determined using a Kontron Liquimat III automatic ion exchange analyser with post-column ninhydrin detection [6]. Values were expressed as micromoles of amino acid per mmole of creatinine.

Data analysis

The chromatographic data comprised qualitative and quantitative information on 74 metabolites, of which six were constant in all 10 specimens ($n = 10$), leaving 68 compounds ($p = 68$) with variable concentration. Principal components analysis (PCA) and partial least square models (PLS) were therefore preferred, as p exceeded n [18].

PCA-analysis. This extracts the dominating pattern from the data table X in terms of the first few eigenvectors of the association matrix [19]. This analysis can be seen as a projection of the n object points in a p-dimensional plane or hyper-plane. With $n = 10$ two dimensions in the projection can be used to obtain an ordinary plane. The resulting projection should give the most reliable overall picture of the object configuration in the $p = 68$ dimensional space. If the dominant patterns in the data are related to the histopathological classification of leprosy, this will be seen as a grouping of similar samples in the plane. In the analysis, the data are first standardized by dividing each variable (table column) by its standard deviation, thereby giving each variable the same influence in the projection. Thereafter the data are centred by subtracting from each column its average. Then the table X is subjected to PCA, where the parameters, vectors t and b, are calculated to make the elements in the residual matrix E minimal. In matrix notation:

$$X = 1\bar{x} + t_1 b_1 + t_2 b_2 + E.$$

PLS-analysis. This method of analysis [14, 15, 19] is similar to PCA in that it projects the X-matrix by a bilinear equation like that shown above. It differs, however, in that auxilliary information can be included in the projection, i.e. in this case the classification of leprosy. This is expressed as an auxilliary vector, y, with one element per sample. This y-variable is here chosen as a 'diagnostic index' with the values 0 for controls (no leprosy), 1 for paucibacillary, and 2 for multibacillary leprosy. With this, two criteria are simultaneously optimized in the projection, namely the residual sum and squares *and* the correlation between the vectors t and y. In this way the degree of approximation of X is somewhat relaxed to improve the expression of the desired information. Expressed in another way, PLS may retrieve information from X that is less obvious than the dominant patterns.

The significance of the projection is checked by cross-validation. Data from a small number of samples is kept out of the calculations, the PLS projection is computed from the remaining data, and the y-values of the deleted are thereafter predicted from the projection. The differences in square between predicted and actual y-values for the deleted samples are summed to form PRESS (predictive sum of squares). Subsequently data from other samples are kept out, a new projection is calculated from the remaining set of data, etc. This sequence is repeated until each sample has been deleted once and PRESS has one term from each.

Results

Metabolite profiling

The urinary organic acid profiles (GC–MS trace) from two of the urine samples (multibacillary leprosy and control) are shown in Fig. 1. Differences in the patterns may be seen, as was the case with the other eight samples analysed (GC-profiles not shown). Manual inspection of these chromatograms, however, did not reveal any correlation with the clinical condition.

Table 1 shows qualitative and quantitative information on 68 organic acids and amino acids. The data in this table was subjected to PCA and PLS analysis as described above.

Data analysis

Principal components analysis. PCA of the scaled and centred metabolite matrix X (10 × 68) was first carried out as a blind test, where the only information given to the data system was that 'one half of the samples was from controls, the remaining half from leprosy patients'. If indeed the dominant pattern of the metabolite profile had relation to leprosy, this should be apparent in a principal component plot of the 10 × 68 table. This was not the case, as shown in Fig. 2. The sample codes were then uncovered and included in the figure. This now shows that the main result appears to be that sample 9 is different from the others. No relation is apparent between controls, multi- and pauci-bacillary leprosy and position of the sample points. It may therefore be assumed that other factors such as age, sex, weight, eating habits, drug intake, etc. play more important roles for the metabolite profiles.

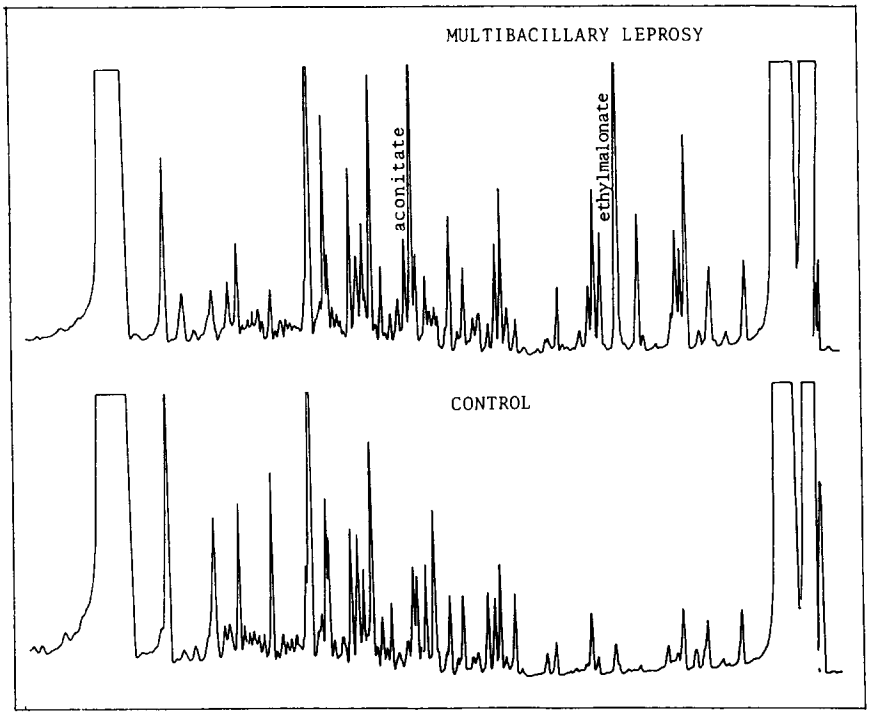

Figure 1
Organic acid profile of urine from a patient with multibacillary leprosy (top) and a control (bottom). The urinary organic acid extracts were methylated before separation by capillary GC–MS as described in the text.

Table 1
Qualitative and quantitive data on organic acids and amino acids in urine specimens from 10 Ethiopian persons: four controls, four with paucibacillary leprosy and two with multibacillary leprosy. The organic acids (from 2-OH-isobutyrate to cinnamoylglycine) are expressed in arbitrary units (peak heights relative to citrate) and the amino acids (taurine to arginine) are expressed as micromoles amino acid per mmole creatinine

Compound name	Sample number 1	2	3	4	5	6	7	8	9	10
2-OH-isobutyrate	0.19	0.24	0.19	0.12	0.23	0.30	0.11	0.43	0.64	0.23
Lactate	1.5	0.40	0.19	0.10	0.10	0.40	1.08	0.14	1.37	2.9
3-OH-isovalerate	0.42	0.96	0.50	0.13	0.84	1.18	0.36	0.75	0.98	1.2
Glykolate	0.29	0.51	0.24	0.04	0.21	0.54	0.15	0.14	0.75	0.50
2-OH-valerate	0	0.25	0.28	0.02	0.06	0	0	0.06	0.21	0.11
Oxalate	0.22	0	0.08	0.05	0.04	0.12	0.10	0.05	0	0.25
Metylmalonate	0.13	0.05	0.03	0.01	0.13	0.12	0.08	0.17	0.18	0.08
3-OH-butyrate	0	0.11	0.3	0.01	0.02	0	0	0	0.04	0.03
2-Methyl-3-OH-butyrate	0.15	0	0.04	0.02	0.06	0.05	0.06	0.05	0.06	0.06
Ethylmalonate	0.20	0.47	0.94	0.06	0.29	0.37	0.18	0.42	0.64	0.23
Metylsuccinate	0	0.16	0.27	0.03	0.05	0.12	0.07	0.09	0.25	0.10
3-OH-isobutyrate	0.73	0.33	0.37	0.13	0.13	0.25	0.28	0.26	0.83	0.39
Succinate	0.17	0.18	0.15	0.02	0.07	0.25	0.14	0.20	1.93	0.19
Benzoate	0.64	0.34	0.04	0.01	0.02	0	0.15	0.06	0.30	0.22
Unknown compound A	0.27	0.22	0.15	0.07	0.04	0.05	0.14	0.05	0.19	0.09
Methylglutarate	0.14	0.05	0.03	0.05	0.01	0.58	0.04	0.05	0.14	0.05
Unknown compound B	0.44	0.09	0.08	0.17	0.04	1.54	0.08	0.14	0.25	0.07
Adipate	0.08	0.16	0.38	0.24	0.03	0.09	0.21	0.06	0.64	0.12
Methyladipate	0.28	0.25	0.25	0.16	0.16	0.23	0.22	0.15	0.79	0.12
Unknown compound C	0.38	0.11	0.07	0.17	0.03	1.39	0.08	0.09	0.29	0.05
Unknown compound D	0	0.25	0	0	0.16	0	0	0	2.1	0
Pimelate + 3-OH-3-Me-glutarate	0.31	0.44	0.20	0.17	0.15	0.21	0.25	0.29	1.21	0.17
Cyclopropane dicarboxylate	0.36	0.09	0.31	0.17	0.06	0.25	0.10	0.14	0.10	0.10
Suberate	0.13	0.18	0.17	0.23	0.06	0.11	0.14	0.11	0.33	0.13
p-Cresol	1.66	1.25	1.12	0.22	1.56	0.77	1.54	0.11	3.81	0.36
Azelate	0.55	0.60	0.06	0.14	0.17	0.44	0.49	0.15	0.50	0.33
Aconitate peak 1	2.08	1.64	0.62	0.48	1.00	1.42	0.89	1.94	2.71	0.57
Aconitate peak 2	0.91	0.42	0.27	0.20	0.40	0.58	0.40	0.82	0.95	0.23
Aconitate peak 3 + FDA	0.48	0.55	0.19	0.28	0.20	0.46	0.25	0.39	1.0	1.2
Unknown compound E	0.78	0.13	0.39	0.29	0.12	0.40	0.31	0.12	0.35	0.13
Aconitate peak 4	1.34	1.24	0.51	0.34	0.80	1.12	0.69	1.63	1.73	0.50
Citrate	1	1	1	1	1	1	1	1	1	1
Isomer of isocitrate	0.84	0.33	0.12	0.40	0.17	0.26	0.35	0.22	1.92	0.20
Homovanillate + furoylglycine	0.92	0.87	0.22	0.34	0.32	0.79	0.56	0.68	1.01	0.52
Isobutyrophenon	0.29	0.31	0	0	0.08	0	0.14	0	0.56	0
p-OH-phenylacetate	>0.50	5.38	0.43	0.66	0.84	2.0	>6.9	5.08	5.71	>3.33
Hippurate	>3.5	>9	>1.5	<1.3	>4.6	>8.8	>6.9	>7.7	>6	>3.3
Indolacetate	0.15	0.24	0.04	0.09	0.13	0.68	3.63	0	0.60	0.04
p-OH-phenyllactate	0.03	0.02	0.05	0.01	0	0	0.08	0	0.05	0.01
Cinamoylglycine	0.06	0.24	0.04	0.08	0.06	0	0.11	0	0.77	0.10
Taurine	97	27	119	44	73	97	10	98	5.0	88
Phosphoethanolamine	?(0)	2.5	?(0)	?(0)	trace	?(0)	trace	?(0)	0.8	15
Aspartate	6.5	3.5	3.0	2.0	3.0	3.0	4.5	3.0	2.0	30
Hydroxyproline	0	0	0	0	trace	0	trace		trace	trace
Threonine	20	11	19	14	10	24	7.5	12	9.5	34
Serine	65	26	68	48	31	66	37	31	31	76
Asparagine	7.5	trace	7.5	7.5	2.0	3.0	0	3.0	trace	11
Glutamate	2.5	trace	0.7	trace	0.8	1.0	2.5	0.3	trace	2.0
Glutamine	64	38	102	84	35	78	55	40	28	94
Proline	2.5	0	trace	trace	trace	0	0	0	trace	trace
Glycine	261	68	214	225	144	520	156	183	76	548
Alanine	85	35	45	62	17	69	56	16	56	146
Citrulline										
α-Aminobutyrate	8.5	3.0	8.0	2.5	3.5	5.5	4.0	2.5	3.5	3.5
Valine	4.5	2.5	6.5	3.5	2.5	0	3.5	4.0	2.0	7.0
Cystine	16	5.5	7.0	9.5	5.0	4.5	7.5	5.0	4.0	8.5
Methionine	<8	trace	2.09	<2.5	<1.5	<4	−2	<2.0	<2	<5
Cystathionine	8.0	3.5	3.5	4.0	4.0	4.5	3.5	4.0	4.0	5.0
Isoleucine	2.0	0.9	2.0	1.5	1.0	2.0	2.0	1.5	1.0	2.5
Leucine	8.0	5.5	7.5	5.5	4.5	7.5	6.5	5.0	8.5	9.5
Tyrosine	18	5.5	18	9.0	4.0	18	13	12	6.5	32
Phenylalanine	9.5	<8	9.0	10	<10	<15	5.0	<11	2.5	<20
β-Aminoisobutyrate	19	180	95	28	75	30	8.5	19	8.5	66

Table 1
Continued

Compound name	Sample number 1	2	3	4	5	6	7	8	9	10
Gamma aminobutyrate	trace	trace	trace	trace	trace	trace	trace	trace	trace	trace
Ornithine	2.0	2.5	2.0	3.5	2.0	2.0	2.5	1.5	1.5	2.0
Lysine	37	10	17	22	12	19	23	18	8.5	33
1-Methylhistidine	trace	trace	trace	trace	trace	22	trace	trace	trace	16
Histidine	100	63	120	131	62	183	97	120	53	219
3-Methylhistidine	31	20	20	20	21	34	19	23	14	22
Arginine	5.5	trace	2.5	2.0	1.0	1.5	0.8	1.0	trace	3.0

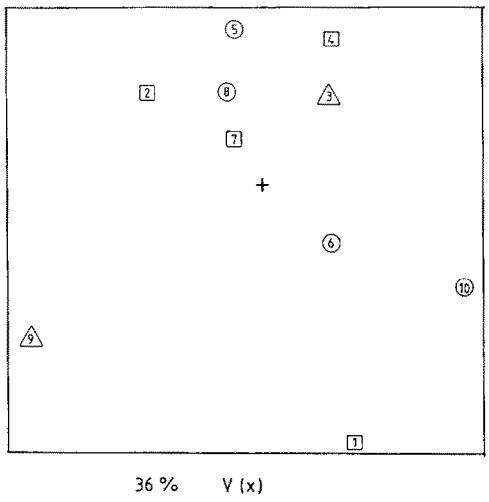

Figure 2
The two first principal components scores of chromatographic data from leprosy samples (Table 1). Open circles, controls; squares, paucibacillary leprosy; triangles, multibacillary leprosy.

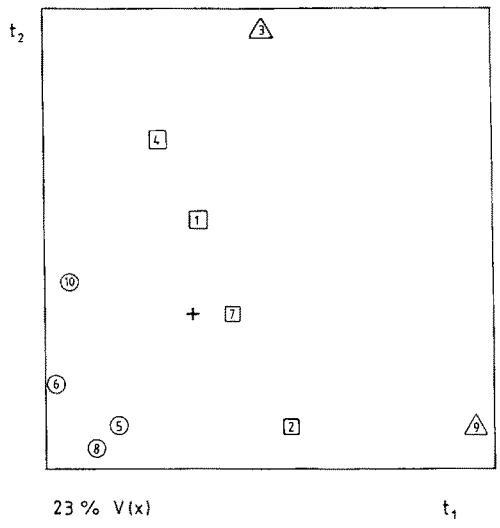

Figure 3
The two first PLS scores of the same chromatographic data as in Fig. 2 and Table 1. The PLS analysis was made with a y-vector with values 0 for controls, 1 for paucibacillary leprosy and 2 for multibacillary leprosy.

PLS-analysis. With the additional information that four of the samples were from controls, four from paucibacillary leprosy and two from multibacillary leprosy, the more exhaustive PLS-data analysis was carried out. A y-vector (diagnostic index) with the values 0 for controls, 1 for paucibacillary and 2 for multibacillary leprosy was set up, and the resulting $X + y$ data set was analysed by two-block PLS. Three significant PLS dimensions were found (cross-validation), describing 74, 24 and 1% of the variance in y. The X-score plot of the dominating two first dimensions (t_1 vs t_2) is shown in Fig. 3, with the corresponding loading plot in Fig. 4.

Figure 3 demonstrates that the PLS analysis of the chromatographic data distinguishes between controls, paucibacillary and multibacillary leprosy. The projection also indicates a systematic second factor or combination of factors (unknown) varying within both groups

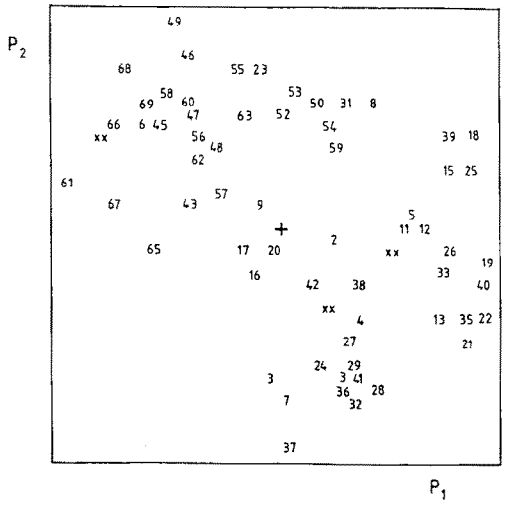

Figure 4
Loading map corresponding to Fig. 3. Note, for example, that variables 15, 18, 25 and 39 are high in leprosy samples and low in the controls.

of leprosy. From the resulting PLS model one can compute coefficients similar to regression coefficients. However, it must be remembered that these coefficients are *not* independent, but should be seen as weights describing the joint contribution of the variables to the modeling of y (here the diagnostic index).

The largest coefficient (for variable 18) is 0.104, and an additional 17 variables have values algebraically larger than 50% of this largest value, i.e. algebraically larger than 0.05. The variables with positive weights, i.e. that tend to be large for leprosy patients, are in decreasing order: 18 (adipate), 39 (*p*-OH-phenyllactate), 25 (*p*-cresol), 8 (3-OH-butyrate), 31, 15, 19, 23, 53, 11, 5 and 12. Their coefficient values multiplied by the corresponding scaling weights are: 0.451, 1.021, 0.729, 0.431, 0.389, 0.497, 0.414, 0.431, 0.327, 0.306, 0.219 and 0.250.

The variables with negative weights tending to be low for leprosy patients are, in order of decreasing importance: 61, 37, 3, 65, 7 and 43. Their coefficient values multiplied by the corresponding scaling weights are: −0.278, −0.223, −0.257, −0.116, −0.412 and −0.161. If we multiply these values with the corresponding data of Table 1 and adding 4.16, we get the following values of the 'diagnostic index' for the 10 samples: 1.03, 0.83, 1.83, 0.87, −0.15, 0.15, 0.86, −0.02, 1.70 and 0.02. The cross-validation indicates that 79% of the variation of y is predicted by the model. This gives the conclusion that samples will have values of the 'diagnostic index' based on these dominating 18 variables as follows: controls, $-0.35 < y < 0.35$; paucibacillary, $0.55 < y < 1.25$; and multibacillary leprosy, $1.4 < y < 2.1$. The reason for the expected y-values less than 1.0 (0.9) and 2.0 (1.75), respectively, is that only 18 of the 68 variables are included in this calculation.

Figure 4 shows the loadings of the variables. This plot indicates which variables contribute to the separation of the groups in Fig. 3. The variables that tend to be high for leprosy patients are seen up to the right (39, 18, 15, 25, 8, etc.), and those that are low down to the left (61, 67, 65, 3, 7 and 37). Variables far out in the upper left and lower right corners (49, 68, 21, 28, 32, etc.) do not contribute to the separation of controls from the leprosy samples, but may be relevant for other properties of the patients.

Discussion

The study illustrates that chromatographic profiles may contain information which is easily overlooked when manual interpretation is employed. This was previously exemplified by the multicomponent chromatographic analysis of brain-tumour biopsies [16], and now of urine samples from leprosy patients with different histopathological classification. Visual inspection of the organic acid and amino acid profiles did not reveal information that could be used to differentiate between controls, paucibacillary and multibacillary leprosy. Multivariate data analysis, on the other hand, using the PLS-models distinguished between these three groups. This was in contrast to data analysis using principal components projection, indicating that the dominating patterns in the chromatographic profiles are related to factors (e.g. diet, age, sex, drug intake, etc.) other than leprosy.

In this type of data a majority of variables are correlated with each other. Therefore it is less meaningful to construct diagnostic intervals for individual variables (compound concentrations). Rather, it is more practical to construct weighted combinations of the most informative variables (here 18) providing estimators of 'diagnostic indices'. Since the indices are based on multitudes of variables, they are more precise and informative than any individual variable in itself.

In this study only amino acids and organic acids were determined. Modern separation techniques which include in addition to GC–MS, also HPLC, capillary electrophoresis [20] and two-dimensional protein electrophoresis, opens up the possibility of determining many other compounds, e.g. a large number (2000) of peptides and proteins. One is thereby approaching a situation where a considerable fraction of the biochemical content of urine, blood and tissue may be separated and analysed. In order to retrieve as much information from all these quantitative and qualitative biochemical data, implementation of multivariate data analysis seems to be a logical approach.

Almost all measurements made on clinical samples are correlated. Chemometric methods like PLS extract information also from correlated data, even when there are many more variables than samples. This should make these

methods useful in several clinical problems similar to the present one [21, 22].

References

[1] E. Jellum, *J. Chromatogr.* **143**, 427–462 (1977).

[2] S.I. Goodman and S.P. Markey, *Laboratory and Research Methods in Biology and Medicine*, Vol. 6. Alan R. Liss, New York (1981).

[3] R.A. Chalmers and A.M. Lawson, *Organic Acids in Man*. Chapman and Hall, New York (1982).

[4] Z. Deyl and C.C. Sweeley (Eds), Profiling of Body Fluids and Tissues, *J. Chromatogr.* **379** (1986).

[5] Z. Deyl and P.M. Kabra (Eds), Chromatography and Electrophoresis in Routine Clinical Analysis, *J. Chromatogr.* **429** (1988).

[6] E. Jellum and A.K. Thorsrud, *J. Chromatogr.* **488**, 105–127 (1989).

[7] J.E. Celis and R. Bravo (Eds), *Two-Dimensional Gel Electrophoresis of Proteins*. Academic Press, Orlando, Florida (1984).

[8] N.G. Anderson and N.L. Anderson, *Anal. Biochem.* **85**, 331–338 (1978).

[9] N.L. Anderson and N.G. Anderson, *Proc. Natl. Acad. Sci. USA* **74**, 5421–5425 (1977).

[10] A.K. Thorsrud, M.H. Vatn and E. Jellum, *Clin. Chem.* **28**, 884–889 (1982).

[11] K.E. Willard, A.K. Thorsrud, E. Munthe and E. Jellum, *Clin. Chem.* **28**, 1067–1073 (1982).

[12] R.P. Tracy and D.S. Young, in *Two-Dimensional Gel Electrophoresis of Proteins*, (J.E. Celis and R. Bravo, Eds), pp. 193–240. Academic Press, Orlando, Florida (1984).

[13] C. Albano, W.J. Dunn, U. Edlund, E. Johansson, B. Norden, M. Sjöström and S. Wold, *Anal. Chim. Acta Comput. Techn. Optim.* **103**, 429–443 (1978).

[14] S. Wold, H. Martens and H. Wold, in *Lecture Notes in Mathematics* (A. Ruhe and B. Kågström, Eds), pp. 286–293. Springer, Heidelberg (1983).

[15] W. Lindberg, J.-Å. Persson and S. Wold, *Anal. Chem.* **55**, 643–648 (1983).

[16] E. Jellum, I. Bjørnson, R. Nesbakken, E. Johansson and S. Wold, *J. Chromatogr.* **217**, 231–236 (1981).

[17] E. Jellum, E.A. Kvittingen, O. Thoresen, G. Guldal, L. Horn, R. Seip and O. Stokke, *Scand. J. Clin. Invest.* **46** (Suppl. 184), 11–20 (1986).

[18] M.G. Kendall, *A Course in Multivariate Analysis*. Griffin, London (1957).

[19] B.R. Kowalski and S. Wold, in *Classification, Pattern Recognition and Reduction of Dimensionality* (P.R. Krishnaiah and L.N. Kanal, Eds), pp. 673–697. North-Holland, Amsterdam (1982).

[20] J. Jorgenson and K.D. Lukacs, *Anal. Chem.* **53**, 1298–1304 (1981).

[21] D. Coomans, I. Broeckaert, M.P. Dearde, A. Tassin, D.L. Massart and S. Wold, *Comput. and Biomed. Res.* **17**, 1–14 (1984).

[22] D. Coomans and S. Wold, *Scand. J. Clin. Lab. Invest.* **44** (Suppl. 171), 247–258 (1984).

[Received for review 26 November 1990; revised manuscript received 15 January 1991]