research articles Journal of proteome research

Application of Three-Way Principal Component Analysis to the Evaluation of Two-Dimensional Maps in Proteomics

Emilio Marengo,*,† Riccardo Leardi,‡ Elisa Robotti,† Pier Giorgio Righetti,§ Francesca Antonucci,§ and Daniela Cecconi§

Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale, Spalto Marengo 33, 15100 Alessandria, Italy, Dipartimento Chimica e Tecnologie Farmaceutiche e Alimentari, Via Brigata Salerno, 16147 Genova, Italy, and Dipartimento Scientifico e Tecnologico, Università di Verona, Strada Le Grazie 15, 37134 Verona, Italy

Received January 2, 2003

Three-way PCA has been applied to proteomic pattern images to identify the classes of samples present in the dataset. The developed method has been applied to two different datasets: a rat sera dataset, constituted by five samples of healthy Wistar rat sera and five samples of nicotine-treated Wistar rat sera; a human lymph-node dataset constituted by four healthy lymph-nodes and four lymph-nodes affected by a non-Hodgkin's lymphoma. The method proved to be successful in the identification of the classes of samples present in both of the groups of 2D-PAGE images, and it allowed us to identify the regions of the two-dimensional maps responsible for the differences occurring between the classes for both rat sera and human lymph-nodes datasets.

Keywords: three-way principal component analysis • proteomics • multivariate analysis • 2D-maps

Introduction

Two-dimensional electrophoresis is perhaps the most widespread technique for the separation of proteins in the field of *proteomics*.[1,2] Each cell or biological fluid has a rich protein content that can be constituted by thousands of proteins. These proteins can show great differences in structure and size, thus complicating the separation process. Two-dimensional gel electrophoresis is a very powerful tool in this perspective; it consists of two successive electrophoretic runs: the first run (through a pH gradient) separates the proteins with respect to their isoelectric point, whereas the second run (through a porosity gradient) separates them according to their molecular mass. Each two-dimensional map (called "2D-PAGE" from Poly Acrylamide Gel Electrophoresis) thus appears as a "snapshot" of the protein content of the investigated cell, with the proteins represented as spots spread all over the gel matrix.

The study of the protein content of different cell types has become fundamental in the past few years, with the development of *genomics* and *proteomics*.[1,2] The proteins contained in a particular tissue, in fact, are related to its particular physiological state: many differences can occur in the protein content of a cell due to the onset of a given disease.[3−6] These differences can consist of changes in the relative abundance or even in the appearance/disappearance of some proteins. The 2D-PAGE maps can thus be used for both diagnostic and prognostic purposes by comparing maps belonging to healthy subjects with samples belonging to individuals affected by different pathologies.

Unfortunately, the problem concerning the comparison of different groups of 2D-PAGE maps is not a trivial one; maps belonging to replicates of the same sample may, in fact, present numerous differences from one another. The high variability (and the subsequent low reproducibility) of 2D-PAGE maps is to be ascribed to some experimental factors and some problems due to the specimen investigated, which can be listed as follows: (1) a complex sample (generally biological tissues) which can produce maps with thousands of spots; (2) a complex sample pretreatment, with many purification/extraction steps, which greatly contributes to the overall experimental error; and (3) a number of experimental factors that may influence the final result (temperature, polymerization and running conditions) and which have to be closely controlled during the experimental step.

The comparison of 2D-PAGE maps belonging to different classes of subjects (e.g., healthy and pathological) is made even more difficult by the small differences that often occur between healthy and ill subjects and which are much more difficult to recognize in a complex map.

Usually, this comparison is performed with the help of some specific softwares (i.e., Melanie III or PD-Quest);[7−10] these

* To whom correspondence should be addressed. E-mail: marengoe@tin.it.
† Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale.
‡ Dipartimento Chimica e Tecnologie Farmaceutiche e Alimentari.
§ Dipartimento Scientifico e Tecnologico, Università di Verona.

10.1021/pr030002t CCC: $25.00

Published on Web 05/28/2003

methods are all based on a similar approach constituted by three main steps:

(1) The 2D-PAGE images to be compared are aligned so that each image is reduced to the same size as the others. This step requires the choice of at least two spots of sure identification in all of the maps; the maps are then matched one to the other on the basis of the position of these two spots.

(2) The spots present on each map are independently revealed; only the spots for which the optic density distribution fits a Gaussian function are maintained.

(3) The maps are matched one to the another to identify the common information (spots present in all the maps) and the differences (spots detected only on some of the samples). If the comparison is performed on a set of replicate maps, then this step produces a "synthetic" map which summarizes the common information and contains only the spots present in all of the compared maps.

The approach just described, even if very powerful, performs a comparison among the "synthetic" maps produced by the analysis of the different replicates obtained for each class; the comparison is no longer performed on the real samples but on "synthetic" ones, not considering the variability which characterizes each class.

Some studies already present in the literature are concerned with the development of methods for the classification of images based on pixel-data;[11] principal component analysis (PCA)[12−15] has been applied to the study of DNA and RNA fragments of several biological systems[16−19] and to the characterization of proteomic patterns of different classes of tissues.[20−25] PCA has recently been applied also to the characterization of the anticancer activity of bohemine, a new omoleucine-derived synthetic cyclin-dependent kinase inhibitor, by Kovarova et al.[26] These multivariate methods require the previous analysis of the proteomic pattern images by standard softwares (i.e., Melanie III or PD-Quest) to identify the spots, so they present the disadvantage of being submitted to human choice for maps alignment. To avoid this disadvantage, we have considered the problem from another point of view: all of the maps belonging to the same type of sample are maintained during the analysis, and the comparison is performed on each sample. In this way, the information about the variability and the reproducibility of the 2D-PAGE maps is maintained and the comparison is performed on all of the real samples together and not on "synthetic" maps. A similar approach has already been applied by our research group to develop methods for the comparison of different classes of 2D-maps of healthy and pathological individuals.[27,28] In the present paper, a new method based on three-way principal component analysis (three-way PCA)[29−35] is presented. The 2D-PAGE images are first digitalized, transforming each image into a grid containing the value of the average optical density in each cell, revealed in the corresponding area of the map. By doing so, a three-mode dataset is obtained, with the three modes being the isoelectric point, the molecular mass, and the samples, respectively. The digitalized images obtained are then analyzed by three-way PCA.

In this paper, three-way PCA was applied to two different datasets:

(1) a set of 10 2D-PAGE maps: five belonging to control rat serum and five belonging to nicotine treated rat serum;

(2) a set of 8 2D-PAGE maps: four belonging to healthy human lymph-nodes and four belonging to human lymph-nodes affected by a non-Hodgkin's lymphoma.

The approach just described was applied to discriminate the two classes of samples present in both the datasets under investigation and to identify, in both cases, the zones of the maps responsible for the differences occurring between the two classes.

Theory

Applied Method. The method applied for the comparison of 2D-PAGE maps belonging to different classes consists of four steps:

(1) Digitalization of the Image. Each image is digitalized, producing a grid of (50 × 50) cells: each cell of the grid contains a value ranging from 0 to 1, according to the intensity of the signal in the correspondent position.

(2) Data Transformation. Data transformations have to be applied to the dataset to scale all of the samples and make them comparable. In the present study, maximum scaling was chosen as the most suitable for the dataset under investigation.

(3) Three-Way PCA. This multivariate technique is applied to the digitalized and scaled images to identify the classes of samples present in the dataset and to identify the zones of the maps responsible of the differences occurring between the classes.

(4) Differences Analysis. The maps rebuilt using the relevant factors are compared to identify the differences occurring between the centroids of the two classes of healthy and pathological protein patterns.

Digitalization. A 2D-PAGE appears, as just pointed out in the Introduction, as a transparent polymeric matter with the separated proteins spread all over it as colored spots. The revelation of the spots is performed, in general, with a solution of an organic coloring agent (e.g., Coomassie Blue) or by the deposition of silver onto the protein surface: the use of these staining solutions generates a coloring intensity which is proportional (within given limits) to the protein concentration. Each 2D-PAGE map is previously scanned with a GS-710 densitometer (Bio-Rad Labs, Richmond, CA), which transforms the 2D-PAGE into an image (200 × 200 pixels) in which each pixel corresponds to the value of the average optical density in the correspondent area.

The scanned images are then transformed into a grid of 50 × 50 cells, in which each cell contains a value ranging from 0 to 1. Each numerical value corresponds to the color intensity of the image calculated by averaging the intensities of the pixels which are contained in the correspondent cell. The values smaller than 0.4 were cut off and substituted with null values to eliminate the information about the color intensity of the background.

The actual choice of a 50 × 50 grid is not a constraint but it was suggested by computational and memory requirements.

Data Transformation. A normalization is essential before performing three-way PCA to make all of the samples comparable with each other. The chosen transformation is maximum scaling: the digitalized 2D-PAGE maps are scaled one at a time to the maximum value for each map, according to the following mathematical expression

$$x_k(i, j) = \frac{x_k(i, j)}{\max(x_k)} \quad (1)$$

where $x_k(i,j)$ is the value corresponding to the cell in (i,j) position in the k-th 2D-PAGE map and $\max(x_k)$ is the maximum value in all the cells of the k-th 2D-PAGE map.

By applying such a transformation to each two-dimensional map, the maximum signal intensity value of every 2D-PAGE map becomes a unit value; all of the samples are thus ranged from 0 to 1, and the dataset becomes independent from the intensity differences due to the staining step. This scaling is suggested by the fact that the large variability of the staining procedure causes a "systematic" error (i.e., maps being consistently darker or lighter). If not removed, this error would account for the major amount of the variation.

Three-Mode Principal Component Analysis. Three-way Principal Component Analysis (three-way PCA), based on Tucker3 model,[29−35] has been used for the identification of the classes of samples present in the two datasets. The interest of three-way PCA is that it allows us to take into account the three-way structure of the data set which can be considered as a parallelepiped of size $I \times J \times K$ (conventionally defined as objects, variables and conditions), where, in our case, I is the number of rows of the grid (the x coordinates, i.e., pH), J is the number of columns of the grids (the y coordinates, i.e., molecular mass), and K is the number of samples. The three-way PCA is based on the fact that the observed modes I, J, K can be synthesized in more fundamental modes, each element of a reduced mode expressing a particular structure existing between all or a part of the elements of the associated observation mode. The final result is given by three sets of loadings together with a core array describing the relationship among them. Each of the three sets of loadings can be displayed and interpreted in the same way as a score plot of standard PCA. Mathematically, this is expressed as follows

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \qquad (2)$$

where x_{ijk} denotes the elements of the initial matrix $\boldsymbol{X}$, a_{ip}, b_{jq}, and c_{kr} denote reduced elements of the component matrixes **A**, **B** and **C** of order $I \times P$, $J \times Q$ and $K \times R$ respectively, g_{pqr} denotes the elements (p, q, r) of the $P \times Q \times R$ core array $\boldsymbol{G}$, and e_{ijk} denotes the error term for element x_{ijk} and is an element of the $I \times J \times K$ array $\boldsymbol{E}$.[29−35]

In the case of a cubic core array (i.e., if $P = Q = R$), a series of orthogonal rotations can be performed on the three spaces of the three modes, looking for the common orientation for which the core array is as much as possible body-diagonal. If this condition is sufficiently achieved, i.e., if the elements g_{111}, g_{222} ... are the only elements of the core matrix being significantly different from 0, then the rotated sets of loadings can also be interpreted jointly by overlapping them.

The datasets were analyzed with a program developed by the authors in the MATLAB 6.1 (The mathworks, Natick, MA) environment.

Experimental Section

Chemicals and Materials. Urea, thiourea, CHAPS, iodoacetamide (IAA), tributyl-phosphine (TBP), and sodium dodecyl sulfate (SDS) were obtained from Fluka Chemie (Buchs, Switzerland). Glutaraldehyde, sodium acetate trihydrate and formaldehyde were from Sigma (St. Louis, MO). Ampholines, bromophenol blue and agarose were from Pharmacia-LKB (Uppsala, Sweden). Ethanol, methanol, acetone, acetic acid, silver nitrate, and citric acid monohydrate were from Merck (Darmstadt, Germany). Acrylamide, *N*,*N*′-methylenebisacrylamide, ammonium persulfate, TEMED, the Protean IEF Cell, Protean II xi cell, GS-710 Densitometer, Mini Trans Blot Electrophoretic Transfer Cell, linear Immobiline dry strips pH gradient 3−10 (17 cm) were from Bio-Rad Labs (Richmond, CA).

Rat Serum and Human Lymph-Nodes. Rat Serum. The first investigated dataset consists of 10 samples belonging to two different groups: 5 samples of Wistar rat serum belonging to healthy individuals and 5 samples of Wistar rat serum belonging to nicotine treated individuals

The 10 2D-PAGE maps obtained are represented in Figure 1(a). Looking at the 2D-PAGE images, it is not very easy to distinguish the healthy individual from the nicotine treated one by a visual inspection of the image. Moreover, the 2D-maps belonging to the same group show a large variability of the spots number, position, shape, and size.

Sample Treatment. Five Wistar rats were treated for 14 days with a saline solution (control samples), and the other five were treated for the same 14 days with nicotine. The nicotine was administered subcutaneously by injecting 1 mL/Kg of a 0.4 mg/mL nicotine solution.

Blood samples were collected on the 14th day (when it is known that nicotine administration begins to induce dependence on treated rats) on rats which were fasted for 12 h prior to collection to avoid interferences due to high concentrations of lipids in the blood. All samples were centrifuged at 4°C to separate from each clot the serum samples (about 200 μL for each blood sample) and they were preserved at −20 °C until the analysis was performed. One hundred μL of serum was added with 0.4 mL of a denaturing solution containing 7 M urea, 2 M thiourea, 5 mM TBP (tributyl-phosphine), and 40 mM Tris. A 20-mM IAA (iodoacetamide) portion was then added, and alkylation was continued for an hour. The samples were then submitted to dialysis to eliminate the salts present in sera, and then the reagents eliminated by the dialysis process were added (7 M urea, 2 M thiourea and 20 mM Tris); 2% CHAPS (3-[(cholamidopropyl)-dimethylammonium]-1-propane-sulfonate) was added as a surfactant.

Human Lymph-Nodes. The proposed method was applied to a set of eight real samples, divided in two classes: four samples belonging to healthy human lymph-nodes, and four samples belonging to human lymph-nodes affected by Mantle Cell Lymphoma. Figure 1(b) represents the 8 experimental 2D-PAGE maps obtained.

Sample Treatment. Sample preparation and solubilization for biopsies was performed essentially as described by Sanchez et al.[36] for human lymphoma biopsies; the same approach was used also for human healty lymph-node biopsies (control biopsies). Briefly, 10 frozen slices (about 20 μm 5 mm × 10 mm) of a human lymphoma biopsy were mixed with 100 μL of 2D- solubilizing solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, 1% Ampholines, protease inhibitor, pH ca. 9. After centrifugation, for removal of particulate material[37], 20 mM IAA was added to perform complete alkylation of proteins.[38,39] Salts, which can interfere with the 2D separation process and visualization of 2-D result, were removed by dialysis. Protein estimation for each sample was carried out with the Bio-Rad DC Protein Assay in order to load

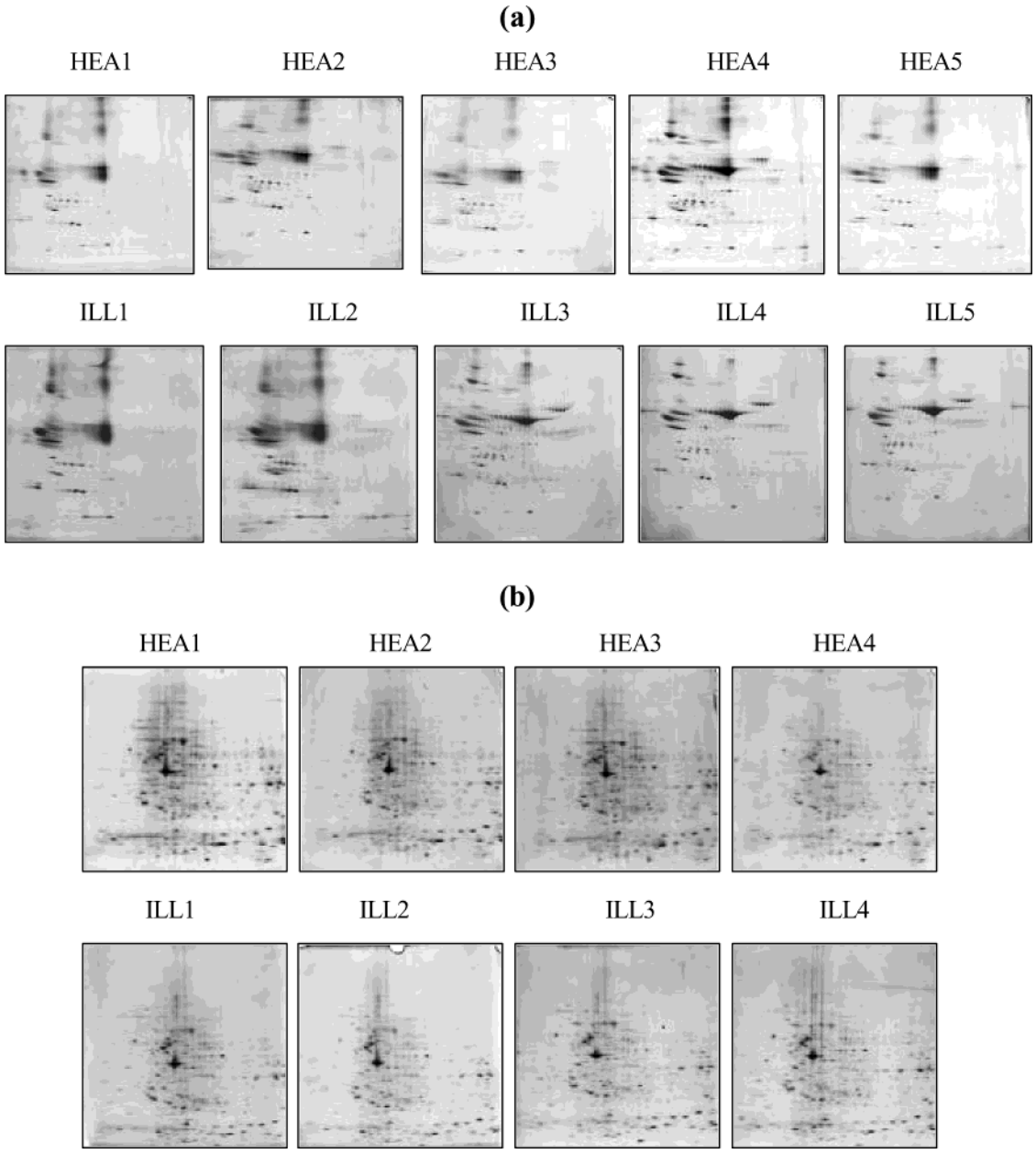

Figure 1. 2D-PAGE maps of the two investigated datasets: 10 rat serum samples (a) and 8 human lymph-node samples (b).

on the IPG strips always the same amount. Samples were stored at −20 °C until used. Samples from three control subjects (1 mg/ml of protein from each sample) were mixed to obtain a representative sample (pool), used to generate four control maps.

2D-PAGE Analysis. The first dimensional run was performed for rat serum samples on nonlinear IPG strips pH 3−10 (18 cm, Amersham Pharmacia Biotech, Uppsala, Sweden), and for human lymph-nodes samples on linear IPG strips pH 3−10 (17 cm, Biorad, Hercules, CA).

The IPG strips were re-hydrated with 450 μL of the sample solution. The passive gel re-hydration was allowed to continue for 8 h before the focusing step. Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Biorad, Hercules, CA). The total product time × voltage applied was 75.000 Vh for each strip.

Each focused strip was equilibrated, for 25 min, with an SDS denaturing solution containing: 0. 375 M Tris HCl (pH 8.8), 6 M urea, 20% glycerol, 2% SDS. In the second dimension, homemade vertical acrylamide-bisacrylamide gradient gels (8−18%; dimensions: 182 × 190 × 1.5 mm) were used. Electrophoresis was performed by using a PROTEAN II xl Multi-Cell (Biorad). After separation in SDS-PAGE gels, the proteins were visualized by Colloidal Coomassie stain. The images of stained gels were scanned with GS-710 Calibrated Imaging Densitometer (Biorad, Hercules, CA).

Software. The software "MATLAB" (The Mathworks Inc, ver. 6.1) was applied for performing all of the data treatments: digitalization of the images and data scaling. Three-way PCA was performed by a self-made software developed with MATLAB, using the Tucker 3 algorithm. Most of the graphic representations were built with "UNSCRAMBLER" (Camo Inc.,

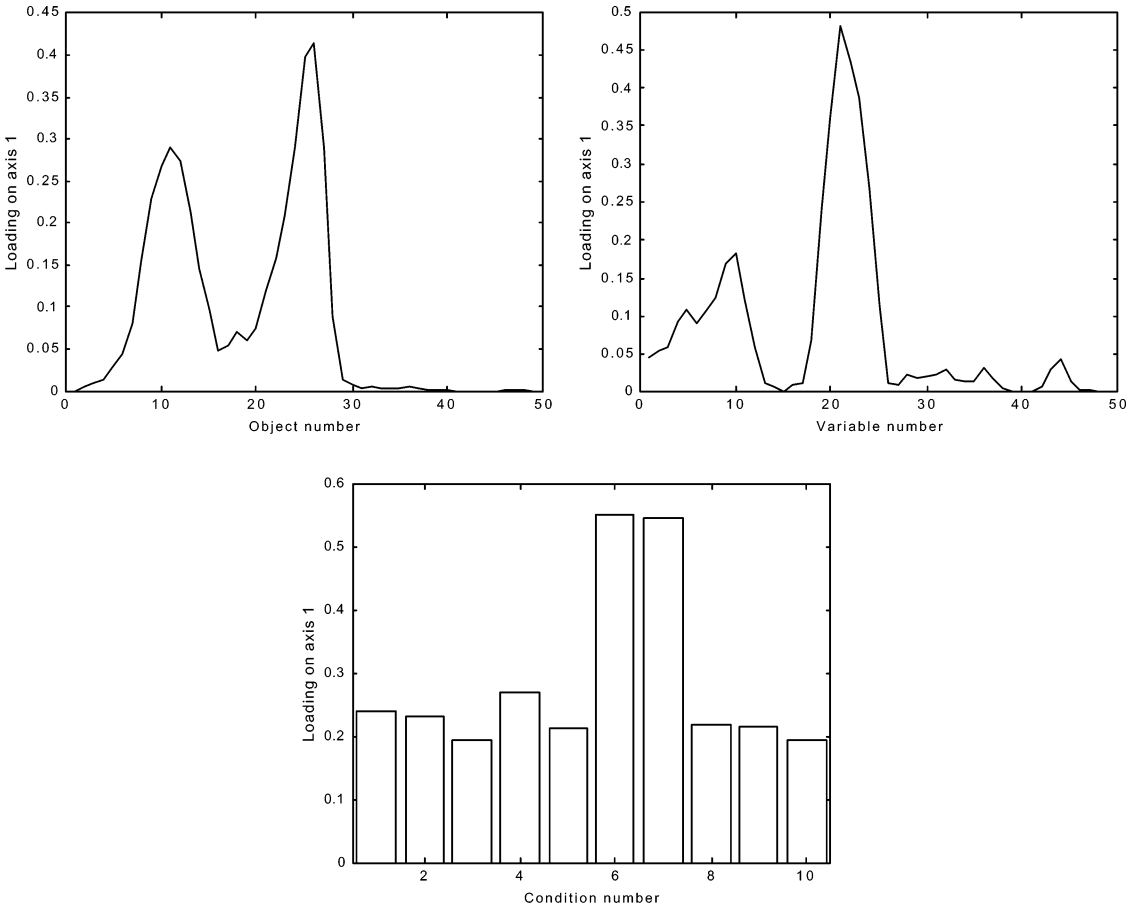

Figure 2. Plots of the three modes for the first entry of the core matrix g_{111}.

ver. 7.6), "MATLAB", "STATISTICA" (Statsoft Inc., ver. 5.1) and "EXCEL 2000" (Microsoft Corporation).

Results and Discussion

Rat Serum Dataset. When looking at the original 2D-PAGE images in Figure 1(a), sample HEA2 appears shifted with respect to the other samples of the same class. This peculiarity of sample HEA2 would lead to an incorrect analysis with three-way PCA because the major amount of explained variance would be probably led by this characteristic. Sample HEA2 was then shifted and matched to the other samples belonging to the same class.

Three-Way PCA. Before performing 3-way PCA, a maximum scaling procedure was applied to the dataset; this scaling technique produces some changes in the dataset: in the maximum-scaled patterns, some minor spots can be detected with respect to the original ones; in addition, maximum scaling is fundamental for obtaining a complete independence of the analysis from the staining step, often scarcely reproducible.

Three-way PCA performed on the normalized dataset gave the results represented in Figures 2 and 3 (two factors were retained for each mode); samples from 1 to 5 correspond to the control rat serum samples, while samples from 6 to 10 correspond to the nicotine-treated rat sera. The first two factors explain 59.8% of the total variance.

After body diagonalization[32] the following core matrix is obtained; the cubic core matrix is reported according to the following unfolding

$$\begin{bmatrix} g_{111} & g_{121} & g_{112} & g_{122} \\ g_{211} & g_{221} & g_{212} & g_{222} \end{bmatrix}$$

$$\begin{bmatrix} 12.76 & -0.22 & 0.13 & 5.10 \\ 0.61 & 0.99 & -2.44 & 1.51 \end{bmatrix}$$

Since it is not super-diagonal, the plots of objects, variables and conditions can be interpreted jointly only for the first factor (the term g_{111} being by far the largest one).

The analysis is performed on one series of combinations at a time: this corresponds to examine one entry of the core matrix at a time. The first one, g_{111}, explains 51.6% of the total variance. The three plots, one for each mode, are represented in Figure 2. The first axis of the mode of conditions clearly discriminates samples 6 and 7 (ILL1 and ILL2) from the others; looking at the plots of the other two modes, it is possible to state that these two samples are characterized by larger values of the objects around 11 and 26 and of the variables around 21 and, at a lesser extent, around 10. By looking at the maps, it is easy to see that these samples are characterized by very intense spots in the rows around 11 and 26 (corresponding to pH 4.5−5 and 6−7) and in the columns around 21 (molecular mass 100−50 KDa).

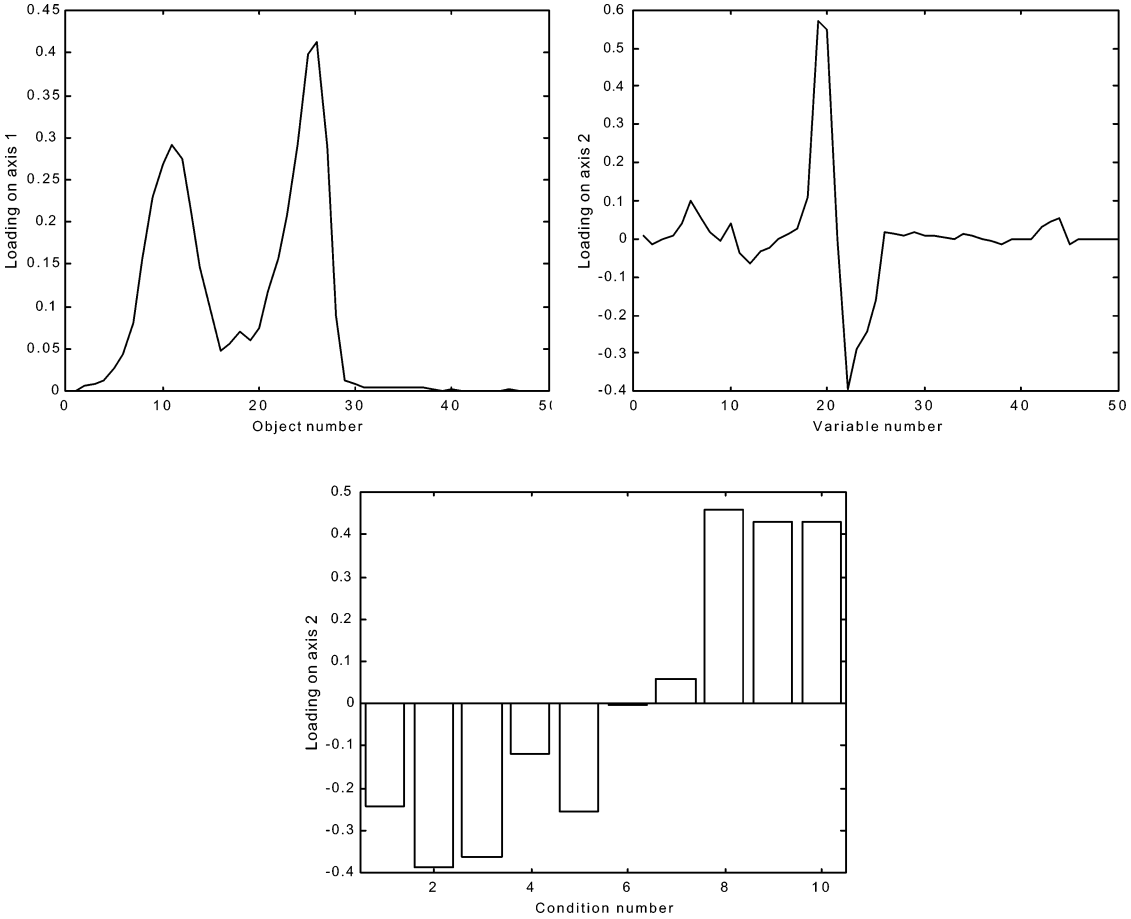

Figure 3. Plots of the three modes for the second entry of the core matrix g_{222}.

The second entry of the core matrix, g_{122}, explains 8.2% of the total variance. The three plots of this entry are represented in Figure 3. The second axis of the mode of conditions clearly separates the two classes of samples: the control ones at negative loadings and the treated ones at positive loadings. Looking at the plots of the objects and variables modes, the control samples appear characterized by small values of the objects around 11 and 26 and by large values of the variables around 22; the treated samples appear instead characterized by large values of the objects around 11 and 26 and of variables around 19.

It is therefore possible to notice that the information is mainly concentrated in objects around 11 and 26 (corresponding to pH 4.5−5 and 6−7)) and in variables around 10 and 21 (molecular mass 200−150 KDa and 90−47 KDa).

Differences Analysis. The fundamental aim of the present study is to identify the regions of the maps responsible for the discrimination of the two classes of samples: the control and the treated ones. The identification of the differences occurring between the two classes of samples is performed by using the centroids (the averaged samples) of each class, in the space defined by variables and objects obtained from three-way principal component analysis. These centroids represent the average information concerning control and treated individuals contained in the first two factors which permit the class discrimination. The centroids can be re-projected in the original space, thus obtaining the corresponding 2D-map images containing only the information accounted for by the first two 3-way factors. The images rebuilt in this way can be compared to identify the discriminant regions of the 2D-maps. This procedure allows a sort of filtering of the useful (discriminant) information contained in the 2D-maps.

The centroid of the control class is obtained on the basis of all of the five original maps belonging to this class, whereas the centroid of the treated class is calculated on the basis of only the three samples (ILL3, ILL4, ILL5) which appear the most different from the control class, as pointed out by three-way PCA. The two re-projected maps and their difference are represented in Figure 4(a). In Figure 4(b), the positive values (toward red) refer to regions which characterize the control sample, while the negative ones (toward blue) refer to regions which characterize the treated specimens. The control samples appear thus richer in spots than the diseased ones and the two classes show differences due both to the presence/absence of spots and to different relative intensities of the spots.

Human Lymph-Nodes Dataset. Three-way PCA was performed on the dataset of human healthy lymph-nodes (samples 1−4) and human lymph-nodes affected by a non-Hodgkin lymphoma (samples 5−8) after the application of a maximum scaling procedure.

Three-Way PCA. As in the previous dataset, a maximum scaling technique was applied: this allows for the appearance of minor spots in the maximum-scaled patterns, and it is

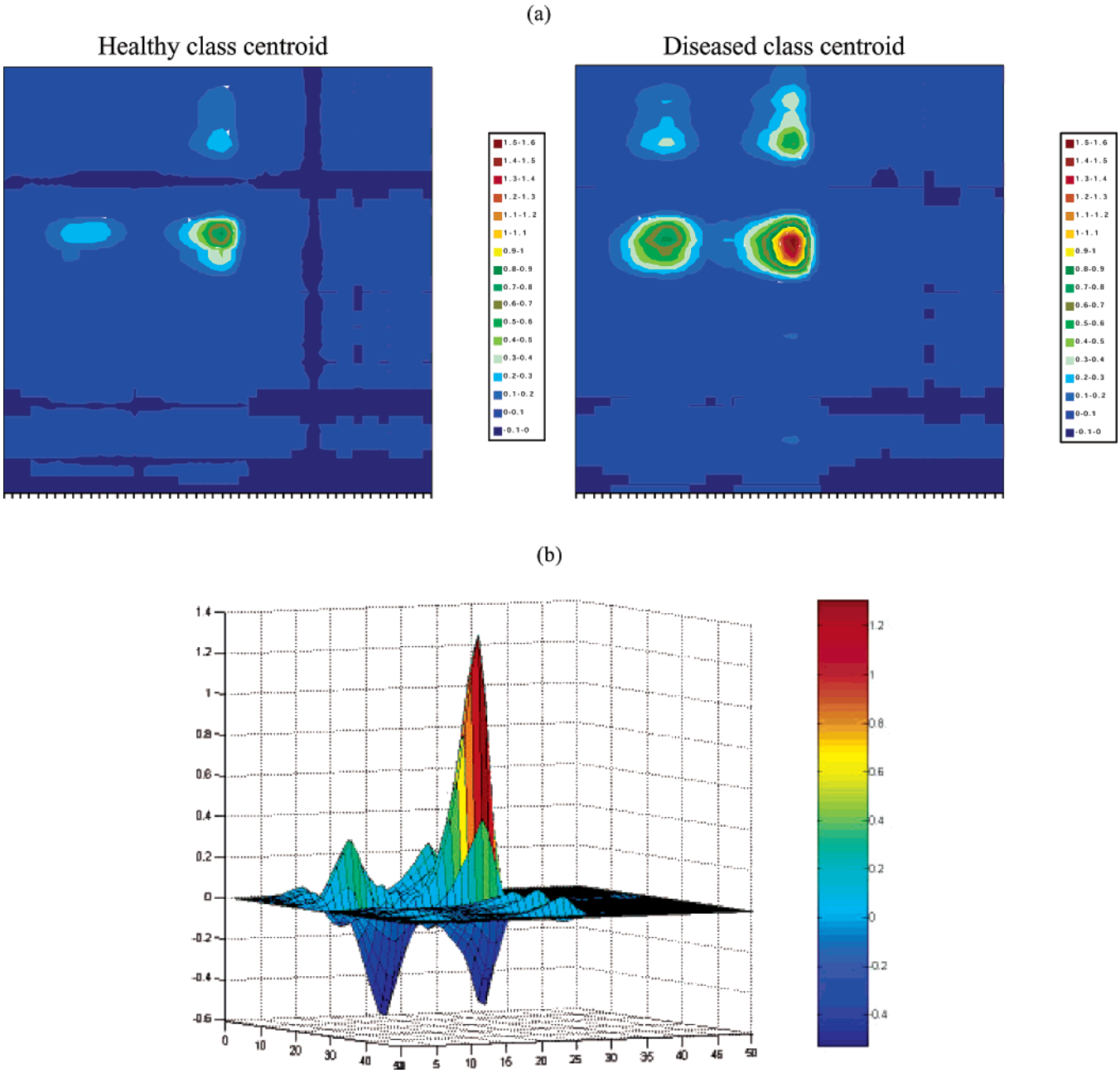

Figure 4. Contour plots of the centroids for the control and treated classes (a) and the correspondent map of the differences (b).

necessary because it makes the statistical analysis independent from the staining step.

The results of Three-way PCA performed on the normalized dataset are represented in Figure 5. Also, in this case, two components for each mode have been retained, and the model explains 50.8% of the total variance. After body-diagonalization, the following core matrix is obtained

$$\begin{bmatrix} 10.06 & 2.29 & -0.14 & 1.86 \\ 0.38 & -0.37 & 1.82 & 1.66 \end{bmatrix}$$

As for the previous dataset, it does not show a super-diagonal structure. However, the element $g_{1,1,1}$ is by far the largest one, and the element $g_{2,2,2}$ is quite close to the second largest. This behavior permits the joint interpretation of the three plots for the first component and, at least to a certain degree, also for the second component.

From the plot of conditions (Figure 5), it can be seen that the data are grouped into three clusters, the control samples (samples 1−3, HEA1−3), the pathological ones (samples 6−8, ILL2−4), and two samples that lay at intermediate values along the second axis (samples 4 and 5, HEA4 and ILL1). Both classes show a sample (samples 4 and 5) which lays far from the others belonging to the same class.

The analysis of the plots of variables and of objects allows the identification of the variables and of the objects discriminating the three clusters and giving the greatest contribution to the model.

Samples 1−3 are characterized by large values of objects 22−27 (pH 5−6) and of variables 18−23 (mol.m. 62−50), whereas

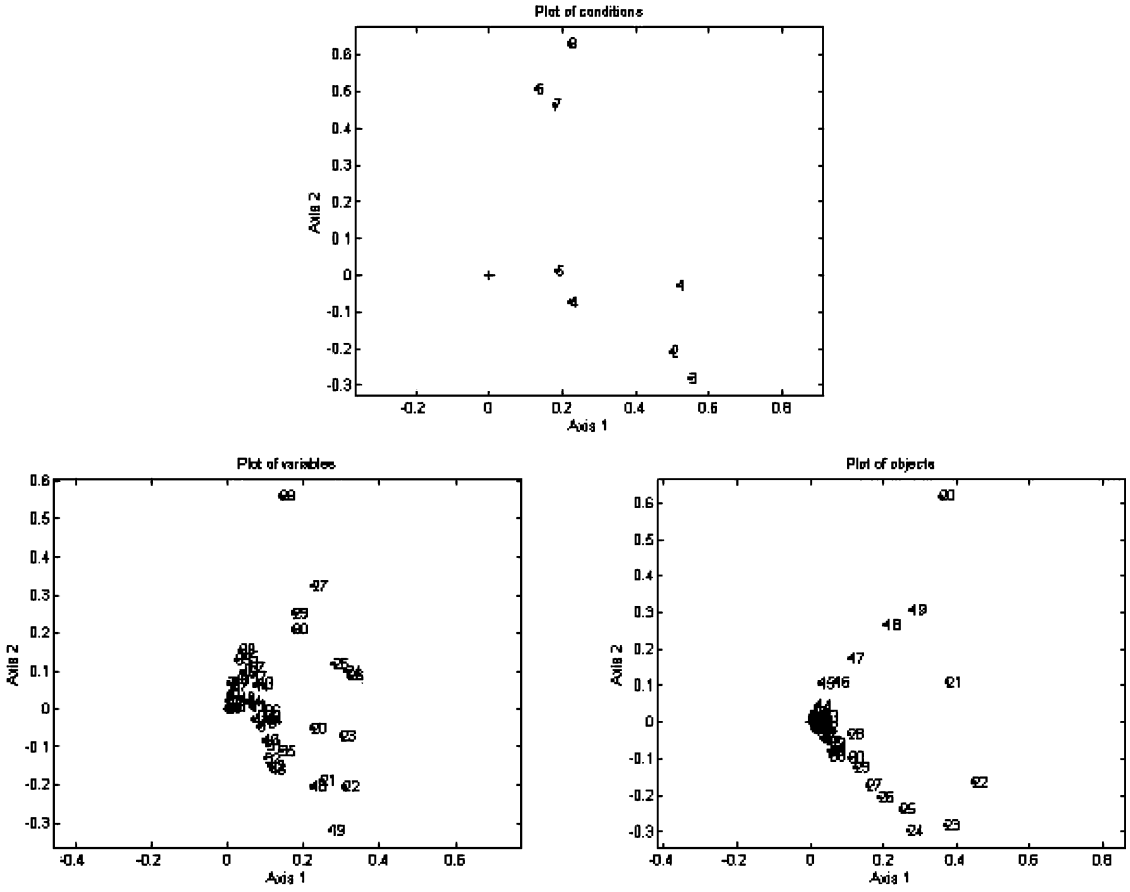

Figure 5. Results of three-way PCA applied to the human lymph-node dataset.

samples 6−8 are characterized mainly by large values of objects 18−20 and by large values of variables 27−30.

As in the previous case study, these findings are confirmed by the visual examination of the maps.

The plots of the objects and of the variables also suggest that the information is mainly resident in objects 15−30 (pH 4.5−7), whereas all the variables give some contribution to the model.

Differences Analysis. As for the previous dataset, the identification of the differences occurring between the two classes is performed on the basis of the centroids calculated for each class in the factors space. Samples HEA4 and ILL1 are excluded from the calculation of the centroids because these samples appeared intermediate between the groups of control and pathological samples, as pointed out by three-way PCA. So, the centroid of the control class was calculated on the basis of samples HEA1−3, whereas the centroid of the pathological one was calculated on the basis of samples ILL2−4. The two reprojected maps are reported in Figure 6(a). Figure 6(b) reports instead the difference between the two averaged samples: the regions toward red contain a signal intensity which is higher in the healthy samples, whereas colors toward blue refer to regions which contain a larger signal for the pathological samples. From this analysis, it appears clear that the differences between healthy and pathological samples are due entirely to the presence of larger signal intensities and a greater number of spots in the healthy samples.

Conclusions

In this paper, three-way PCA was performed on two different datasets: the first constituted by 10 samples of rat sera and the second constituted by 8 samples of human lymphonodes. Maximum scaling was applied to each dataset before performing three-way PCA to eliminate the influence of the staining procedure on the statistical analysis.

The applied method resulted a successful tool for the discrimination of the classes of samples present and for the identification of the zones responsible for the differences occurring between the samples belonging to the different classes. This last goal was realized by "differences analysis", which allowed the identification of the regions which characterize each class of samples for both the considered dataset. It is necessary to stress that the method just presented represents a preliminary approach to the problem of comparing 2D-PAGE maps belonging to different classes: further studies are necessary before being able to use this method for diagnostic/prognostic purposes. Work is in progress to compare the present results with those obtained from the traditional approach.

Acknowledgment. The authors gratefully acknowledge financial support from MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca, Rome, Italy; COFIN 2000). Professor P.G.R. is also supported by Grant No. I/R/294/02 from ASI (Agenzia Spaziale Italiana). The authors also thank N. Busti (GSK) for technical assistance in animal treatment and serum sample collection.

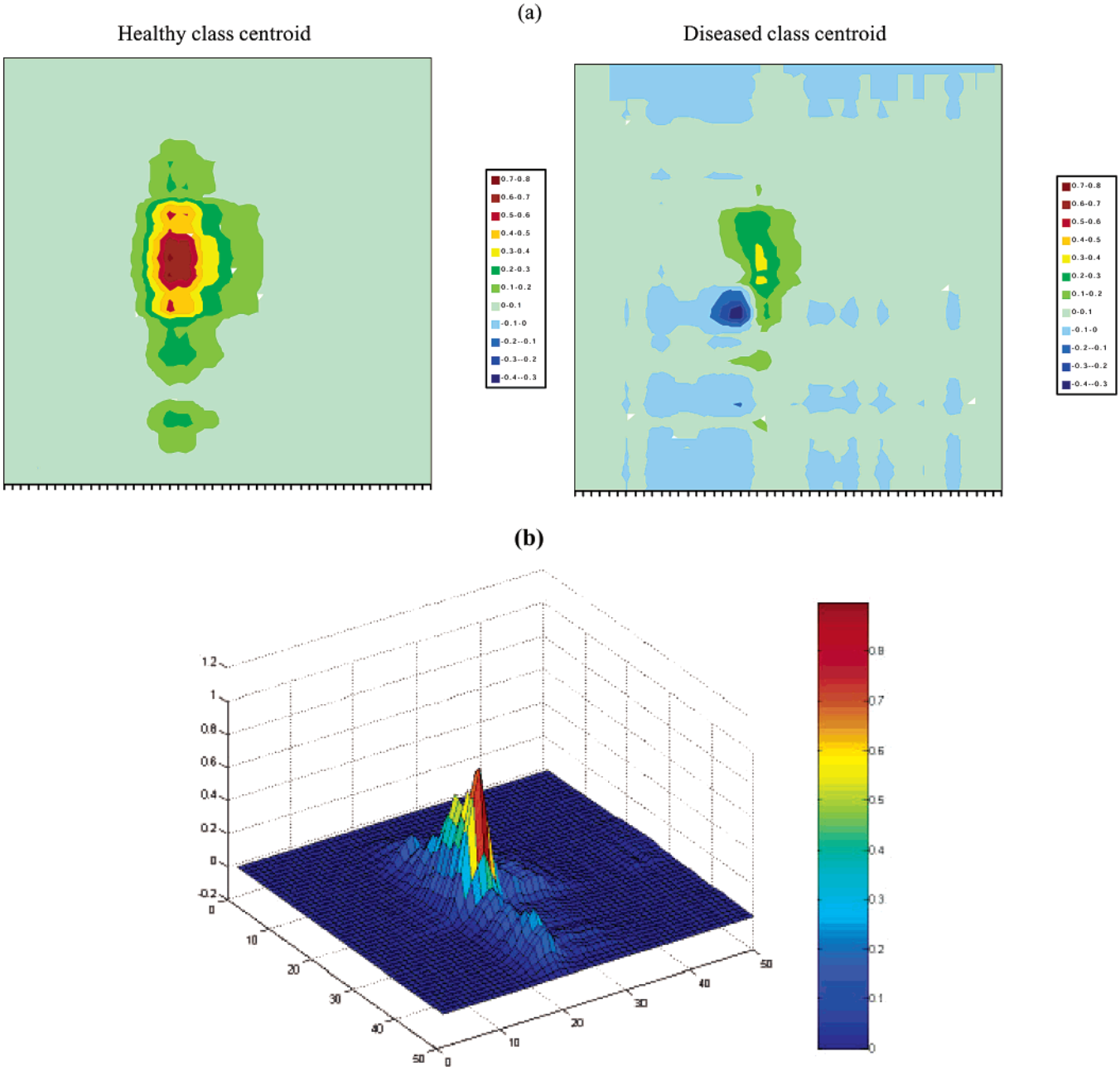

Figure 6. Contour plots of the centroids for the healthy and pathological classes (a) and the correspondent map of the differences (b).

References

(1) Righetti, P. G.; Stoyanov, A.; Zhukov, M. *The Proteome revisited: Theory and Practice of the Relevant Electrophoretic Steps*; Elsevier: Amsterdam, 2001.

(2) Wilkins, M. R.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F. *Proteome Research: New Frontiers in Functional Genomics*; Springer-Verlag: Berlin, Germany, 1997.

(3) Crenshaw T. R.; Cory J. G. *Adv. Enzyme Regul.* **2002**, *42*, 143−157.

(4) Aksenov M. Y.; Aksenova M. V.; Butterfield D. A.; Geddes J. W.; Markesbery W. R. *Neuroscience* **2001**, *103*(*2*), 373−383.

(5) Levine R. L.; Stadtman E. R. *Exp. Gerontol.* **2001**, *36*(*9*), 1495−1502.

(6) Toda T. *Exp. Gerontol.* **2000**, *35*(*6*−*7*), 803−810.

(7) Westergren-Thorsson G.; Malmstrom J.; Marko-Varga G. *J. Pharmaceut. Biomed. Anal.* **2001**, *24*(*5*−*6*), 815−824.

(8) Bhatia K.; Lord R.; Stanton P. *Eur. J. Cancer* **2002**, *38*(*S3*), S156.

(9) Hoffmann, F.; Kriegel, K.; Wenk, C. *Discrete Appl. Math.* **1999**, *93*, 75−88.

(10) Lee, K. H. *Trends Biotechnol.* **2001**, *19*(*6*), 217−222.

(11) Swets D. L.; Weng J. J. *IEEE T. Pattern Anal.* **1996**, *18*(*8*), 831−836.

(12) Davis, J. C. *Statistics and Data Analysis in Geology*; John Wiley & Sons: New York, 1986.

(13) Brereton, R. G. *Chemometrics: Application of Mathemathics and Statistics to Laboratory Systems*; Ellis Norwood: New York, 1990.

(14) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; Wiley: New York, 1978.

(15) Massart, D. L.; Vanderginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: a Textbook*; Elsevier: Amsterdam, 1988.

(16) Leonard, R. B.; Mayer, J.; Sasser, M.; Woods, M. L.; Mooney, B. R.; Brinton, B. G.; Newcombgayman, P. L.; Carroll, K. C. *J. Clin. Microbiol.* **1995**, *33*(*10*), 2723−2727.

(17) Johansson, M. L.; Quednau, M.; Ahrne, S.; Molin, G. *Int. J. Syst. Bacteriol.* **1995**, *45*, 670−675.

(18) Couto, M. M. B.; Vogels, J. T. W. E.; Hofstra, H.; Husiintveld, J. H. J.; Vandervossen, J. M. B. M. *J. Appl. Bacteriol.* **1995**, *79*, 525−535.

(19) Boon, N.; De Windt, W.; Verstraete, W.; Top, E. M. *FEMS Microbiol. Ecol.* **2002**, *39*, 101−112.
(20) Alika, J. E.; AkenOva, M. E.; Fatokun, C. A. *Genet. Res. Crop Evol.* **1995**, *42*, 393−399.
(21) Anderson, N. L.; EsquerBlasco, R.; Richardson, F.; Foxworthy, P.; Eacho, P. *Toxicol. Appl. Pharm.* **1996**, *137*, 75−89.
(22) Vohradsky, J. *Electrophoresis* **1997**, *18*, 2749−2754.
(23) Kovarova, H.; Radzioch, D.; Hajduch, M.; Sirova, M.; Blaha, V.; Macela, A.; Stulik, J.; Hernychova, L. *Electrophoresis* **1998**, *19*, 1325−1331.
(24) Lacroix-Desmazes, S.; Bayry, J.; Misra, N.; Horn, M. P.; Villard, S.; Pashov, A.; Stieltjes, N.; d'Oiron, R.; Saint-Remy, J.; Hoebeke, J.; Kazatchkine, M. D.; Kaveri, S. V. *New Engl. J. Med.* **2002**, *34*, 662−667.
(25) Dewettinck, K.; Dierckx, S.; Eichwalder, P.; Huyghebaert, A. *Lait* **1997**, *77*, 77−89.
(26) Kovarova, H.; Hajduch, M.; Korinkova, G.; Halada, P.; Krupickova, S.; Gouldsworthy, A.; Zhelev, N.; Strnad, M. *Electrophoresis* **2000**, *21*, 3757−3764.
(27) Marengo, E.; Robotti, E.; Gianotti, V.; Righetti, P. G. *Ann. Chim. (Rome)* **2003**, *93*, in print.
(28) Marengo, E.; Robotti, E.; Gianotti, V.; Righetti, P. G.; Cecconi, D.; Domenici, E. *Electrophoresis* **2003**, in print.
(29) Tucker L. R. *Psychometrika* **1966**, *31*, 279−311.
(30) Kroonenberg P. M. *Three-Mode Principal Component Analysis*; DSWO Press: Leiden, 1983.
(31) Geladi P. *Chemometr. Intell. Lab.* **1989**, *7*, 11−30.
(32) Kroonenberg P. M. *Chemometr. Intell. Lab.* **1989**, *15*, 143−157.
(33) Henrion R. *Chemometr. Intell. Lab.* **1993**, *6*, 477−494.
(34) Henrion R. *Chemometr. Intell. Lab.* **1994**, *25*, 1−23.
(35) Henrion R.; Andersson C. A. *Chemometr. Intell. Lab.* **1999**, *47*, 189−204.
(36) Sanchez, J. C.; Appel, R. D.; Golaz, O.; Pasquali, C.; Ravier, F.; Bairoch, A.; Hochstrasser, D. F. *Electrophoresis* **1995**, *16*, 1131−1151.
(37) Rabilloud, T. *Electrophoresis* **1996**, *17*, 813−829.
(38) Herbert, B.; Galvani, M.; Hamdam, M.; Olivieri, E.; McCarthy, J.; Pedersen, S.; Righetti, P. G. *Electrophoresis* **2001**, *22*, 2046−2057.
(39) Galvani, M.; Hamdan, M.; Herbert, B.; Righetti, P. G. *Electrophoresis* **2001**, *22*, 2058−2065.

PR030002T