

# An Integrated Strategy in Two-Dimensional Electrophoresis Analysis Able to Identify Discriminants Between Different Clinical Conditions

LINDA PATTINI,\*<sup>1</sup> SAVERIA MAZZARA,\* ANTONIO CONTI,† SERGIO CERUTTI,\* AND MASSIMO ALESSIO†<sup>1</sup>

*\*Department of Bioengineering, IIT Unit, Politecnico di Milano, Milan, Italy; and †Department of Proteome Biochemistry and ‡Department of Neurology, San Raffaele Scientific Institute, Milan, Italy*

Two-dimensional gel electrophoresis (2DE) is an indispensable tool in proteomics for the analysis of protein expression in complex biological systems such as cells and tissues. However, the automatic extraction of information from gel images is still a challenging task. In this paper we propose a strategy that represents a computational procedure of support to the discrimination of different clinical conditions associated with the samples. The analyzed gel images were acquired within the framework of a study of peripheral neuropathies: twenty-four 2DE maps generated from cerebrospinal fluid (16 pathologic and 8 control subjects) were processed. Quantitative features were defined to describe each image and treated with a method of dimensionality reduction. The informativeness of the descriptors allowed us to see the gel of the data set as items in a three-dimensional space, segregating according to the clinical conditions. Moreover, information with prognostic value was obtained for a single outsider gel of a patient who was included in a clinical subgroup at the first diagnosis but whose disease progressed with clinical features belonging to a different clinical subgroup. The method developed may represent an effective tool of classification that can be used repeatedly to capture the essential impression from separation images. *Exp Biol Med* 233:483–491, 2008

**Key words:** 2DE; bioinformatics; image analysis; proteomics; neuropathies

---

This work was supported by Fondazione CARIPLO (Nobel GUARD Project and 2006.0537/105411 Project) and Lega Italiana per la lotta contro i tumori.

---

<sup>1</sup> To whom correspondence should be addressed at Department of Bioengineering, Politecnico di Milano, Piazza L. da Vinci, 32, 20133, Milan Italy; E-mail: pattini@biomed.polimi.it (L.P.) and Proteome Biochemistry, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy; E-mail: alessio.massimo@hsr.it (M.A.)

---

Received July 11, 2007.  
Accepted December 11, 2007.

---

DOI: 10.3181/0707-RM-187  
1535-3702/08/2334-0483\$15.00  
Copyright © 2008 by the Society for Experimental Biology and Medicine

## Introduction

Two-dimensional electrophoresis (2DE) is the technique of choice that provides the separation of the proteins contained in a biological sample (1). The maps obtained from protein migration are acquired as gray level images that can be processed to allow the analysis and the comparison of the experimental outcome for different samples. However, the complexity of the maps makes this task difficult and time-consuming. The result of the separation, indeed, still suffers from a low degree of reproducibility as a result of both local and global alterations of the migration in the gel. In the case of clinical studies, reproducibility is also affected by the fact that samples are collected sequentially along 1 or 2 years, and consequently electrophoresis is performed at different times, introducing heterogeneity into the experimental outcomes. Furthermore, there is often a limited amount of specimen that can be obtained from patients, and this limited quantity does not allow one to generate replicates of the gels. Thus, it is mandatory to include in the analysis suboptimal 2DE gels; otherwise, subjects would be excluded from the study. Consequently, the comparison and the classification of the gel images, which are usually based on the matching of the maps, may be a difficult task, especially if an automatic and objective procedure is needed.

The computational aspects of image processing play a central role in the analysis of 2DE gels (2). This step is very labor-intensive and involves considerable expertise to properly extract information. The process, in general, takes advantage of dedicated software packages for quantitative analysis and comparison of the gels considered as collections of identified spots (3). An automatic strategy in which the classification is done in a repeatable fashion but is not dependent on the choices of the operator can be a useful complement to the routine differential proteomic analysis.

In the present study, gel images obtained from

cerebrospinal fluid samples acquired in a study of peripheral neuropathy were considered. Results of proteomic differential analysis have been published (4); here we followed a complementary approach aimed at discriminating the categories involved and avoiding the step of matching through the application of a strategy that considers the overall impression of the gel.

In recent years a similar classification approach that considers the overall pattern that emerges from gel images without spot detection has been explored for many different data sets (5–8). This method tries to capture the essential information of a gel image through the use of a set of descriptors obtained from the assessment of the pixel values in different quadrants of the image obtained by a regular grid. In this way the gel is described by a vector of values that can be treated with the approaches of multivariate statistics with the aim of discriminating the samples corresponding to different biological conditions. The dimensionality of this kind of representation is effectively reduced in most of the cases through the use of principal component analysis (PCA) that allows one to see the gels as positions in a useful space where clustering of biologically homogeneous items can be appreciated. This approach allows a person to discriminate between the classes of the data set, and this discrimination leads to the correct classification of each sample.

Starting with this kind of approach, we developed an integrated strategy with the aim of implementing a classification tool that is able to show discriminating patterns within a pool of gels from a clinical study in which quality and reproducibility are challenged by the presence of samples not obtained *ad hoc*. We acted on the definition of the descriptors of the gel, and then the features extracted for the experimental subregions of the separation image were analyzed through the use of PCA (9) as the dimensionality reduction technique.

The adopted computational strategy is a useful complement to the experimental routine and is capable of automatically capturing the essential information from the images, often saving low-quality samples.

## Materials and Methods

**Patients and Two-Dimensional Electrophoresis.** We analyzed twenty-four 2DE maps generated from cerebrospinal fluid (CSF) obtained from three groups: patients with neuropathy and pain (PN,  $n = 8$ ), patients with neuropathy but without pain (NPN,  $n = 8$ ), and healthy controls (CN,  $n = 8$ ). Patient features and 2DE gel generation were described previously (4). Eligibility criteria included the classic diagnostic methods used for disease identification. All patients were initially hospitalized and drug-free. Quantitative evaluation for disability was calculated according to the method of Appel *et al.* (10). All patients underwent the normal procedure of differential diagnosis including hematologic examination (serum im-

muno-electrophoresis, vitamin B12, folic acid, rheumatoid factor, immunocomplexes, cryoglobulins, extractable nuclear antigen ENA, and anti-mitochondria, anti-smooth muscle and anti-DNA antibodies), electroneurographic (ENG) and electromyographic (EMG) evaluations based on standard criteria (11), and lumbar subarachnoid CSF collection and examination. At neurologic examination all patients showed neurogenic changes and no motor or sensory conduction abnormalities (as revealed by EMG and ENG studies, respectively). The disease course was chronically progressive in all cases.

Patients submitted to neurologic examinations in which strength evaluation (MRC grading system) and Verbal - Numerical Pain Rating Scales were performed at the time of the first hospitalization and 6 and 12 months later by the same neurologist. CSF was collected by lumbar puncture by using a procedure typically done for diagnostic purposes. Exclusion criteria included the following: human immunodeficiency virus- or hepatitis C virus-positive, neurodegenerative disease or previous cerebral ischemic event, and deep metabolic deficit. Control CSF was obtained from age-matched and gender-matched patients for whom neurologic disease was suspected but who were later shown by testing to be free of pathologic alterations. For each sample a 2DE map was generated, and corresponding images were analyzed.

**Image Analysis and Descriptors.** Gel images were acquired at a resolution of 100  $\mu\text{m}$  by using a Molecular Dynamics Personal SI Laser Densitometer (Molecular Dynamics, Sunnyvale, CA) and saved as 12-bit gray-level images in “.tif” format (12). The migration maps covered the ranges from 3.2 to 10.4 in pI and from 10 to 150 kDa in molecular weight (MW).

Spot detection was performed by using Progenesis Workstation v2004 software (Nonlinear Dynamics, Newcastle, UK; Ref. 13). The procedure was standard and left at default settings to limit user intervention. This software implemented the most recent strategies of image processing exploited in the analysis of gels and gave back for each analyzed image the collection of the identified protein spots provided with a set of quantitative parameters such as volume, area, maximal intensity, and position. The option of “total spot volume normalization” was adopted to make quantitation comparable despite possible fluctuations in signal intensities between gels, and the automatic “background subtraction” was included in the process of quantitation. The step of spot detection was included in the adopted strategy to improve the signal-to-noise ratio and to allow the emergence of only the useful signal; the idea is that the descriptors that are then derived refer only to areas of the image that are segmented as spots but lack quantitative description artifacts and background signal. Conversely, the extraction of features directly from the images, meant as matrices of pixel intensity without the step of spot detection, could include information from areas that do not correspond to areas that contain real spots. The

quantitative features obtained on the basis of this standard protocol were used in the successive exploratory data analysis. Gels were calibrated, in the Progenesis environment, to obtain the position of each identified spot in terms of the experimental coordinates: apparent relative molecular mass ( $M_r$ ) was estimated by comparison with MW reference markers (Precision, Bio-Rad, Hercules, CA), and pI values were assigned to detected spots by calibration as described in the GE-Healthcare guidelines. This crucial step was obtained separately for each gel; the position of the detected spots was expressed in terms of pI and MW by using some reference markers and interpolating (by means of a cubic spline) these values to obtain the calibration curve that empirically defines the relation between geometrical and physicochemical coordinates for every gel image. We used 8 markers of MW (10, 15, 25, 37, 50, 75, 100, and 150 kDa) and 11 markers of pI. The pI markers were positioned according to the percentage of IPG-strip gel length, as indicated by GE-Healthcare table of reference for nonlinear gradients, and then fine-tuned to the experimental electrophoretic migration of protein spots clearly visible in all gels, which were previously identified (4) by mass spectrometry analysis and matching with CSF reference maps. The pI values used were 3.2 (beginning of the IPG-strip gel), 4.0 (A1AG1 protein, accession number P02763), 4.8 (A1AT protein, accession number P01009), 5.3 (RETB protein, accession number P02753), 5.6 (ALBU protein fragment, accession number P02768), 6.2 (TRFE protein, accession number P02787), 6.8 (A1AT protein fragment, P01009), 8.0 (IGLC protein, accession number P99007), 9.2 (CYTC protein, accession number P01034), and 10.4 (end of the IPG-strip gel). This calibration was accomplished without spot matching through the gel collection, which is a critical and time-consuming step and one in which a considerable fraction of detected spots are lost and thus excluded from calibration. Of course, this introduced a certain degree of approximation, but the coarse resolution at which the proposed strategy works compensated for it, as the descriptors referred to macroregions of the gel, not to the single protein spot. In this way, each gel was treated as a collection of identified spots whose positions were expressed in terms of the experimental coordinates of pI and MW rather than in terms of pixels. Only at this point was the migration space of 3.2–10.4 in pI and 20–100 kDa ideally partitioned into subquadrants at a resolution of 0.3 unit in pI and 3 kDa in MW.

The subdivision was linear in the physicochemical coordinates but did not correspond to a regular grid on the gel image. The subquadrants were individuated consistently and tracked the ideal separation area in the pI and MW space in each gel image, despite contingent alterations of the single gel. This step was crucial to make the samples comparable in the absence of canonical image matching by means of registration techniques.

For each subquadrant the relative collection of spots was determined; to this aim the parameters of pI and MW

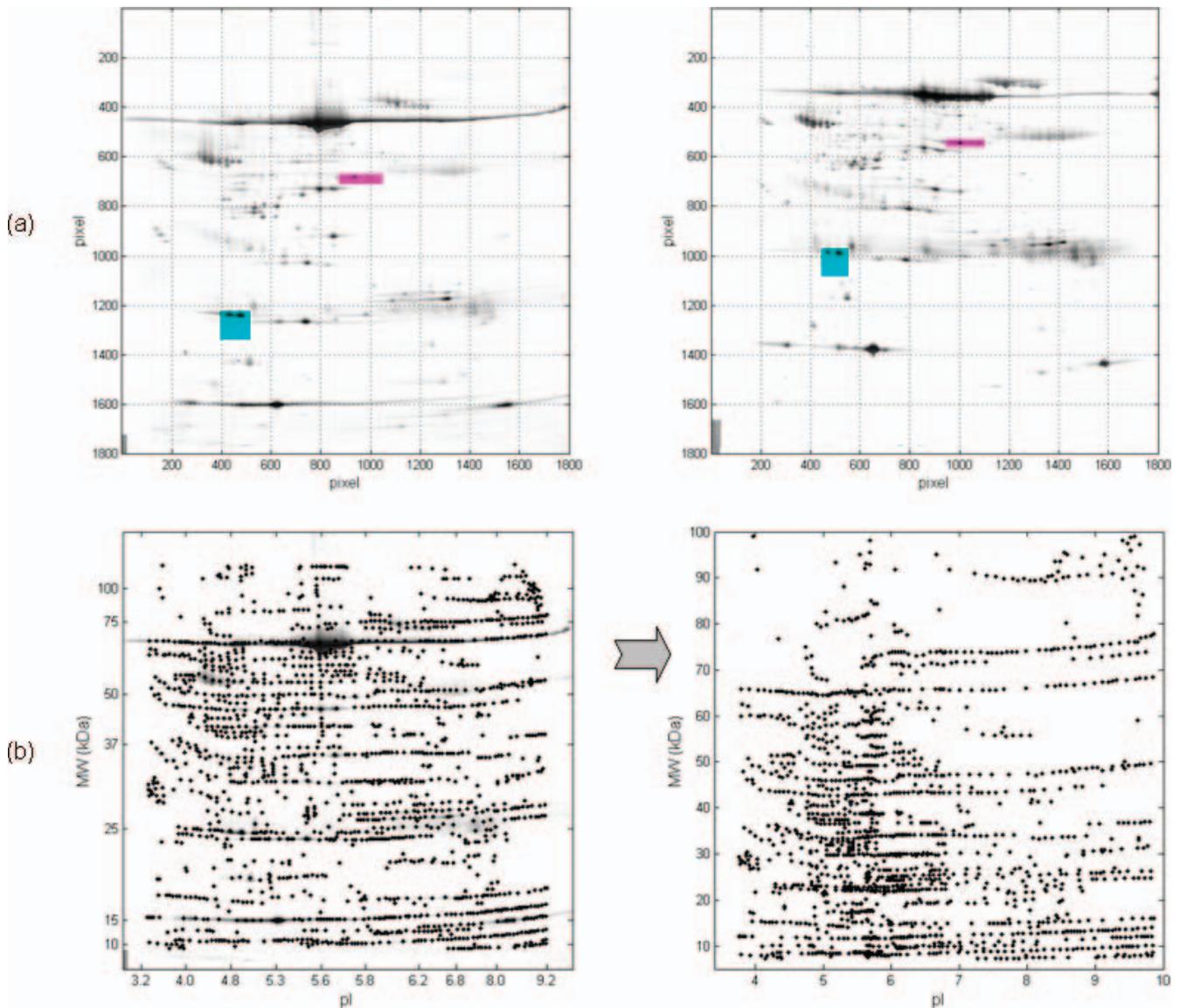
associated with each spot were approximated at the second decimal digit, because a greater number of significant digits, although provided by the commercial tool, were redundant with respect to the intrinsic accuracy of the technology. Given the collection of all the spots of a subquadrant, we were able to determine a cumulative descriptor. By summation of the spot volumes, the integral of the intensities of the pixel segmented as useful signal was obtained and considered to be a quantitative feature of the single subquadrant. At the established resolution mentioned above, 648 descriptors, which were equal to the number of the subquadrants, were extracted for each gel image.

The samples, at that point, were described as vectors of sorted features, depending on the cumulative spot volumes in the different gel areas. Thus, they might be explored through an approach of dimensionality reduction such as PCA with the aim of visualizing them and verifying the informativeness of their description with respect to the corresponding clinical conditions.

**Application of PCA.** The main purpose of PCA is dimensionality reduction while accounting for as much of the variation in the data as possible. To the original data a transformation is applied that leads to a new space whose variables (principal components) are not correlated and are obtained as a linear combination of the original ones. The weights of the original variables in the principal components are referred to as loadings. By selecting the first  $f$  components ( $f \ll$  number of descriptors), it is possible to represent a significant quote of the total amount of information contained in the data.

To assess the separability of the groups considered in the pairwise comparisons, standard discriminant analyses (linear [LDA] or quadratic [QDA]) were applied and obtained the surface representing the decision region boundaries. The adequacy of the descriptors was objectively assessed by resubstitution (i.e., counting the number of items misclassified with respect to the decision region retrieved [error rate]). Also, a leave-one-out cross-validation was performed in which the surface was determined by excluding one item from the analysis and then verifying the correctness of its classification; iterating the process for all the samples of the data set, a measure of the predictive accuracy was obtained as the percentage of successes.

By looking at the gels in the new three-dimensional space of the first three principal components, it is possible to appreciate sample positions that segregate consistently with the corresponding biological conditions; therefore, we re-examined the gel images to identify the regions (the subquadrants that were quantified) that most contributed to the representation where samples of different clinical conditions appeared to be separable. Thus, the obtained loadings may be used to evaluate which original variables are important (large loadings) and most concur with the new representation. The loadings of the principal components, meant as the weight of the original coordinates projected onto the new dimensions, were measured: the subquadrants



**Figure 1.** Comparison of images of multiple samples. (a) The general pattern of the spot separation is recognizable in both gels, but the two maps are not superimposable (equal distances in terms of pixels do not correspond to the same range in pI and MW). Through the calibration of each gel image in terms of biochemical coordinates (pI and MW), it is possible to recover the correspondence between equivalent areas (cyan: 5–5.3 units [pI], and 22–25 kDa; magenta: 5.8–6.1 units [pI] and 47–50 kDa). (b) After the calibration step, each identified spot was provided with coordinates in the pI–MW space (panel on the left); this step allowed the compilation of a virtual linear map of the separation outcome (panel on the right).

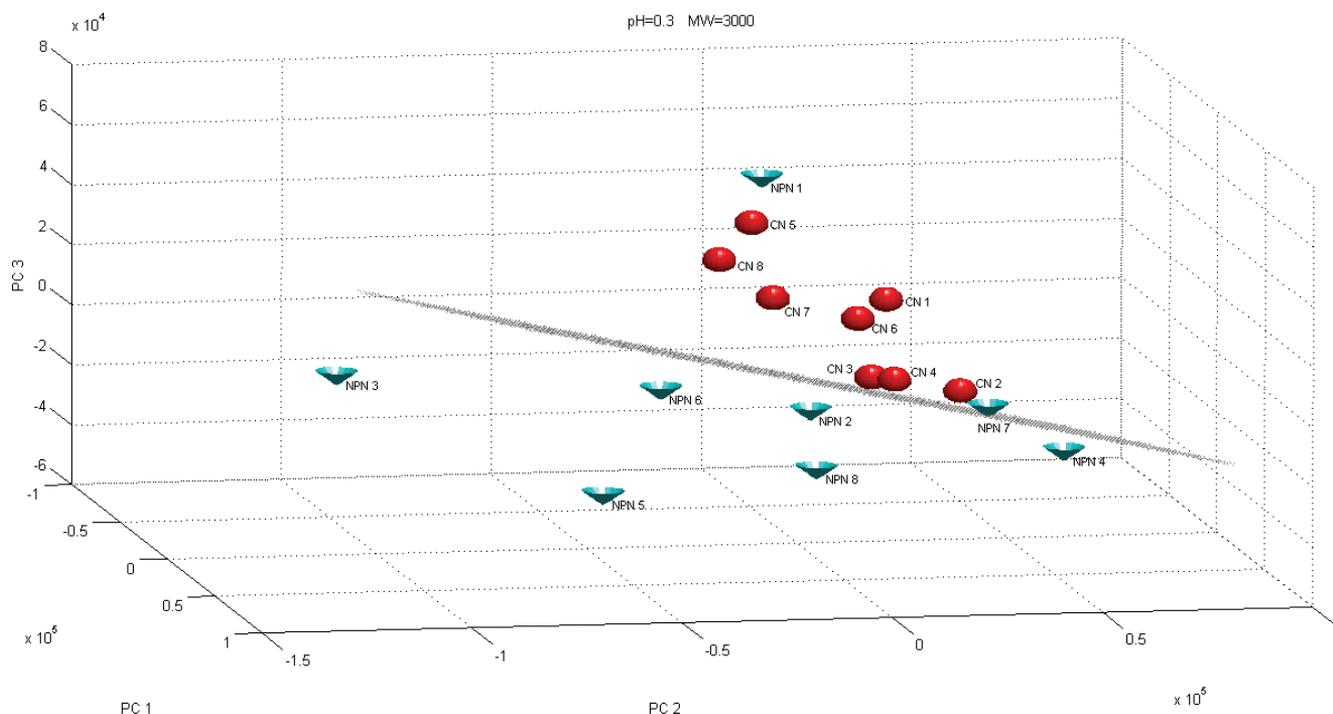
that had the highest weights were considered. Such an indication can be a useful clue in the identification of regions where spots most differ through the clinical categories.

The method proposed was implemented in Matlab v. 7.0 (The Mathworks, Inc., Natick, MA).

## Results and Discussion

In the previously proposed method (5), the image was divided with a regular grid in many different quadrants, and for each of them a descriptor was extracted on the basis of the pixel intensities; eventually a fuzzyfication step was applied to smooth the signal and consider the imprecision of

the electrophoretic run. In general, this strategy works well in an ideal case in which images of protein migration in the different samples cover equivalent pixel areas; otherwise, a selection of homogeneous samples is necessary. Considering the entire ensemble of 2D gels of a general experiment related to real clinical samples may become impractical. In the data set of our study, even at a visual inspection, macroscopic differences in the exit of the migration due to the experimental variability were evident; for example, the areas covered by protein migration (in terms of pixels) were not constant. In general the positions in pixels between gels are not equivalent with respect to the separation, and proportions are not conserved through the collection of



**Figure 2.** Samples from patients with neuropathy but without pain (NPN, cones) and from healthy controls (CN, spheres) visualized in the three-dimensional space of the first three principal components; their positions allow one to identify a linear decision region to discriminate between the two classes. A color figure is available in the online version of the journal.

samples. In Figure 1a, two gel images are reported as an example: the general pattern of the samples is recognizable in both images, but it appears evident that the protein separation covers areas whose proportions are quite different (see colored areas in Fig. 1). These same issues make the comparison step and the implementation of the canonical differential analysis critical. In general, the application of registration techniques and warping strategies are needed to map an image onto another (13) and make comparison possible. Moreover, some gel images can be definitively excluded if they lack the necessary homogeneity. This can be a problem in the case of clinical samples because many times the amount of biological material available for laboratory investigations is very limited because of technical or ethical reasons and it is not possible to perform further sample collection. From most patients in our study, we obtained about 0.5 ml of CSF with a protein concentration of 0.2 mg/ml that was used for a single 2D gel (100  $\mu$ g). Thus, it was mandatory to be able to include suboptimal gels in our analysis.

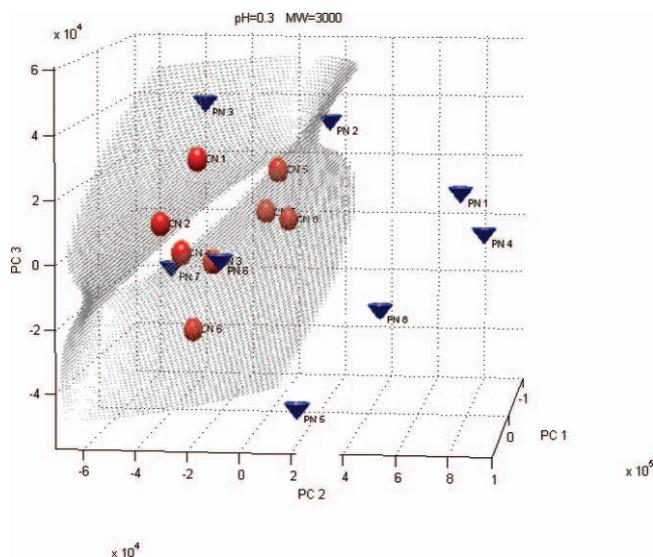
To bypass these problems, the positions of the detected spots were considered in terms of pI and MW, after the *ad hoc* calibration. Through this simple step, an acceptable correspondence between different gels was restored (see Supplemental Material for a collection of representative gel images provided with the markers positioned to obtain continuous gradation in pI and MW). For example, in Figure 1a, the subquadrant of 5–5.3 units (pI) and 22–25 kDa (cyan) and the subquadrant of 5.8–6.1 units (pI) and

47–50 kDa (magenta) are highlighted in both gels. The positions in pixels are quite different, but the migration areas are consistent as a result of the calibration. The outcome of the spot detection can be then visualized in a linear virtual map of the separation (Fig. 1b), whose partition yields the subquadrants that contain the spot ensembles whose integral volumes were adopted as descriptors.

The procedure was applied to the considered data set. To allow specific discriminant patterns, we considered categories that were the most homogenous; subjects with neuropathy were further classified as those with algic symptoms and those without pain; with the addition of control subjects, three categories were considered and the three possible pairwise comparisons were accomplished.

Figure 2 reports the result of the first comparison, which was between NPN and CN samples. The visualization of the gels in the space of the first three principal components showed the samples grouped consistently in separable regions with their corresponding clinical condition (specific neuropathy diagnosis). The cumulative percentage of variance of the original data considered in this representation was 74%. The decision region obtained by LDA provided an error rate of 6.25%; the position of the sample 1 of the NPN group was clearly contained in the controls' "cloud," but supplemental information about this item of the data set will be provided below. Leave-one-out cross-validation provided an accuracy of 81.25%.

In the second analysis, the PN and CN groups were



**Figure 3.** The result of the analysis of samples from patients with neuropathy and pain (PN, cones) and from healthy controls (CN, spheres) is reported. There was separation of the two types of subjects, except PN 3 and PN 7; however, the latter belongs to the groups of samples that had a greater number of detected spots with respect to the other gels in the data set. A color figure is available in the online version of the journal.

considered (Fig. 3). The three-dimensional representation accounted for 70% of the data variance. The positions that corresponded to the two kinds of subjects were separable in this case by a quadratic surface. The error rate was 12.5%: the two samples PN 3 and PN 7 were not consistently positioned as indicated by the results of the DA analysis; however, the first one was on the frontier of the decision region and the second one was the sample that in the spot detection provided a number of spots sensibly greater than the rest of the collection (Table 1). Because the descriptors used as inputs of this multivariate statistics approach were based on cumulative spot volumes, this problem may alter the classification. Leave-one-out cross-validation provided accuracy of 68.75%.

Finally we explored the data in relation to subjects with neuropathy to assess our approach's ability to discriminate

between subjects with pain and those without pain. Effectively the items that corresponded to the two categories segregated consistently (Fig. 4; percentage of explained variance, 78%) and were separable by means of a quadratic surface. The only exception was represented by the sample NPN 1, which was clearly positioned in the cluster of the opposite category. What was surprising was that although the initial information about the condition of this patient corresponded to an asymptomatic neuropathy (stage 1 according to criteria [14]) and an absence of pain (neuropathic pain scale [NPS] score of 0 [15]), the disease course led to the development of algic symptoms (NPS score of 7) and worsening of neuropathy (stage 3 [disabling neuropathy]). What was considered initially as a misclassification in fact became proof of the robustness of the discrimination power of the proposed method with prognostic potential. The information about this particular sample, initially annotated as NPN but effectively belonging to the PN group, may explain its mispositioning also in the graph concerning the first comparison (NPN vs. CN), leading eventually to a null error rate. For this comparison we obtained a leave-one-out cross-validation accuracy of 62.5%. The poor results of cross-validation may be due to the small number of the samples in the data set.

To see whether our approach can properly classify the samples from multiple conditions, we analyzed all samples present in the data set (except those that lacked homogeneity with respect to spot number) in an unbiased fashion. As shown in Figure 5, the three groups can be correctly classified. Moreover, the sample NPN1 (indicated by the arrow), which was clustered with both PN and CN in the previous comparisons (Figs. 2 and 4), was more closely related to the PN category, a result that confirmed the clinical reclassification.

It is worth noting that the samples of our data set are not technical replicates (i.e., gels obtained from fractions of the same biological sample, unique or even pooled). In this work, "biological" replicates were processed (i.e., each gel image was representative of a different human subject); this approach increased the level of variability and complexity.

**Table 1.** Data Set: Number of Detected Spots in Each 2DE

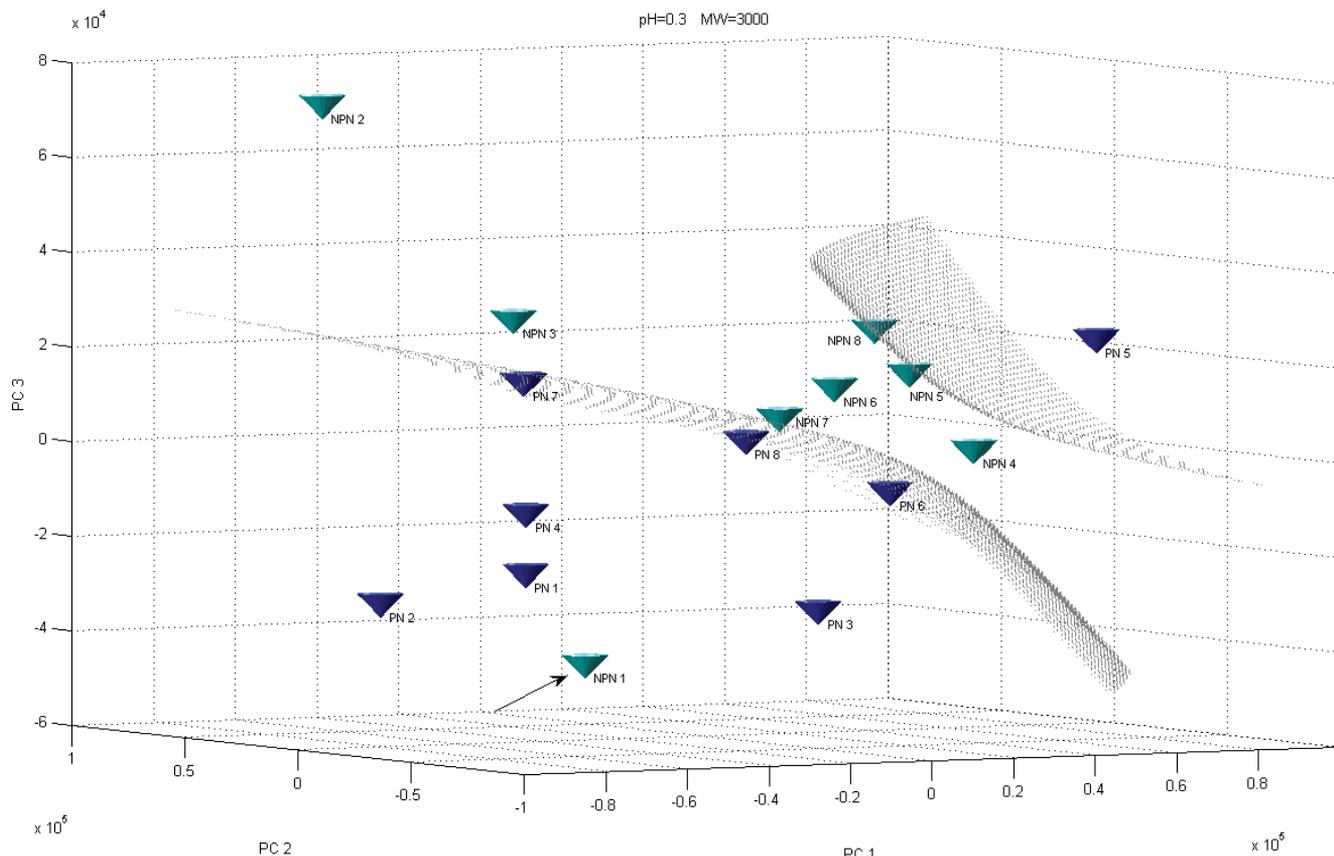
CN <sup>a</sup>	No. of Spots	PN <sup>b</sup>	No. of Spots	NPN <sup>c</sup>	No. of Spots
1	2018	1	1939	1	1517
2	1494	2	1739	2	1443
3	1812	3	1675	3	1447
4	1737	4	1844	4	1547
5	<i>1297<sup>d</sup></i>	5	<i>1069<sup>d</sup></i>	5	1660
6	2436	6	1945	6	1323
7	2367	7	<i>3158</i>	7	1348
8	2422	8	2018	8	1314

<sup>a</sup> CN, healthy controls.

<sup>b</sup> PN, patients with neuropathy and pain.

<sup>c</sup> NPN, patients with neuropathy but without pain.

<sup>d</sup> Italic numbers refer to map with a nonhomogeneous number of spots.



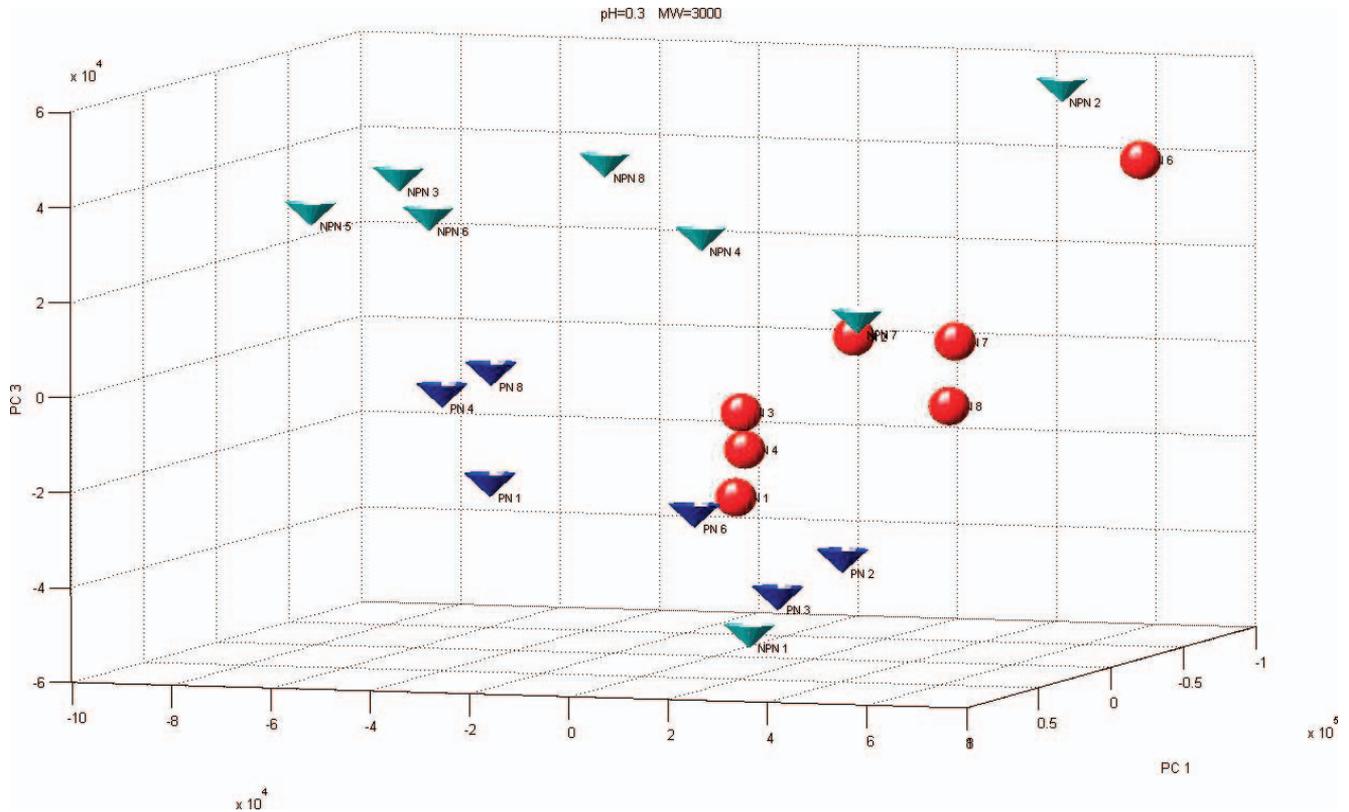
**Figure 4.** Subjects with disease and pain (PN, dark cones) and those with disease but without pain (NPN, light cones) were compared in terms of their 3D positioning. Two groups of items were well-separated by a quadratic surface; the exception was the sample from NPN 1 (indicated by the arrow) that, although initially annotated as NPN, has developed algic symptoms; this results revealed a prognostic potential of the approach. A color figure is available in the online version of the journal.

For this reason the results that we obtained are very significant and robust. Also the numerosity of the data set supports these considerations. The previously applied fuzzy-logic approach to compensate for migration variation between gels works well with largely superimposable images, as it happens for technical replicates or gels obtained *ad hoc* in a unique trial. Our proposed strategy is different because we used, from the original gel images, quadrants delimited in the ideal space of pI and MW, not in terms of pixels as done in previous studies (5–8). This kind of calibration allowed us to recover correspondence between gels where the outcome of the migration was very different (Fig. 1). In such a situation the original method with fuzzy logic would not be sufficient. Moreover, to improve the signal-to-noise ratio, we used the optical density volumes of the spots identified by Progenesis software. In this way we avoided the need to consider pixel intensities that do not refer to spot areas but may refer to background.

If the representation in the space of the principal components shows that the gels belonging to different categories as having different privileged regions of the space, it may be interesting to return to the gel images and identify the different weights of the gel regions, (in terms of pI and MW) that allowed us to visualize the groups of gels

as “separable.” By associating each of the three PCs to one of the three channels of the RGB code for color images (red for PC1, green for PC2, blue for PC3), it was possible to visualize the linearized maps of the corresponding loadings for the considered comparisons (CN vs. NPN, CN vs. PN, PN vs. NPN [Fig. 6]). This kind of information may support the identification of proteins or groups of proteins differentially expressed in different conditions. At the moment the chosen resolution (pH = 0.3 and MW = 3000 Daltons) seems to provide the best results. It represents a good compromise between the detail at the spot level that may be subject to fluctuations between gels, also in the calibrated data, and a more coarse but even more robust information about the overall pattern.

The discriminant features seen in our previous work (4) do not belong to the largest loading descriptors, but this result is not surprising. Indeed, the quadrants of the maps in Figure 6 correspond to relative large regions of the physical gel that correspond to areas that generally include many spots, and the proteins identified in previously published work are not necessarily in the largest loading quadrants. The approaches are quite different: the traditional one is based on the values of a single spot matched through the different images, whereas the proposed strategy considers



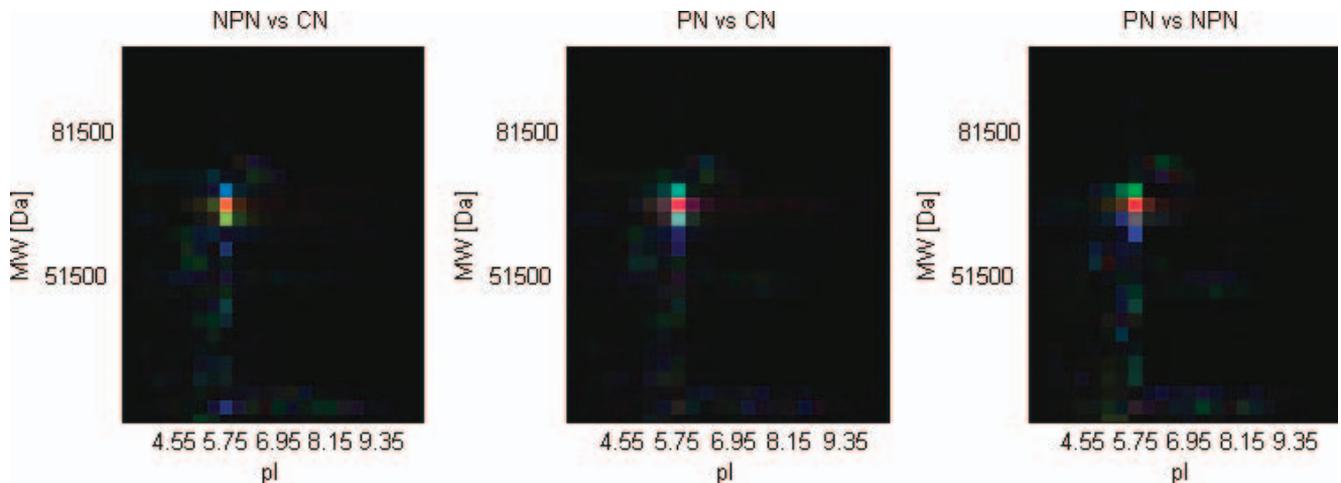
**Figure 5.** Results of the analysis performed on the three groups of subjects (PN, NP, and CN) except those whose gel images did not have a homogeneous spot number. The data for the three subject groups appeared separable, and the sample NPN was positioned in the “PN region.” A color figure is available in the online version of the journal.

the overall impression of the sample on the gel and quantified as integral volumes in macroareas; therefore, there may be no correspondence.

Our method can be very useful and close to the real proteomics process because it is very reproducible and does not need any *a priori* information; thus, it may represent an effective and rapid classification approach, providing complementary information and supporting differential

analysis. The information visualized through the application of this strategy may also highlight the presence of outsider(s) in the data set or group, as demonstrated at least with the set of data used in this study.

The results that we obtained are very encouraging, especially for such a complex data set of gels that was collected over a long period of time and showed a high degree of heterogeneity. In particular the separability



**Figure 6.** Pseudocolor maps of the loadings that corresponded to the three PCs are shown for the considered comparisons; the evident regions are the most informative in regards to the discriminant potential between the clinical groups.

between subjects with or without algic symptoms underlines the effectiveness of the approach and the informativeness of the adopted descriptors. If the strategy is applied correctly, it may lead to important indications for the identification of specific biomarkers, retrieving the constraints that allow one to see the groups of samples as “separable.”

From the clinical point of view, peripheral neuropathies are characterized by asymmetric, slowly progressive weakness and can occur either with or without pain. In the absence of a complete understanding of molecular disease development, the therapeutic approach to pain neuropathy is limited to the traditional tricyclic antidepressants and anticonvulsants (16, 17). In fact, peripheral neuropathies are not impairing diseases, but the quality of life is seriously affected in patients with pain. No single pain measure has sufficient reliability and validity because pain is a multi-dimensional experience (18); therefore, finding putative pain biomarkers useful for early diagnosis might be of great interest for therapeutic strategies.

With the computational strategy proposed, the chance to distill synthetic information about the protein content of biological fluids may be a valid support, in particular if the procedure is rapid and repeatable. It may lead to the definition of a protocol of automatic classification that represents a useful complementary analysis in the proteomics laboratory to perform tests capturing the essential impression of the gel image. Moreover, it may lead to the identification of samples that present an anomalous pattern and that can be reasonably excluded from statistical descriptions.

Accurate information extraction in the processing of gel images is an important topic in computational biology, and the treatment of patterns emerging from separation images intended as fingerprints of the corresponding clinical conditions may provide an interesting and fruitful point of view.

1. Rabillaud T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* 2:3–10, 2002.
2. Dowsey AW, Dunn MJ, Yang GZ. The role of bioinformatics in two-dimensional gel electrophoresis. *Proteomics* 3:1567–1596, 2003.
3. Marengo E, Robotti E, Antonucci F, Cecconi D, Camprostrini N, Righetti PG. Numerical approaches for quantitative analysis of two-dimensional maps: a review of commercial software and home-made systems. *Proteomics* 5:654–666, 2005.
4. Conti A, Ricchiuto P, Iannaccone S, Sferrazza B, Cattaneo A, Bachi A, Reggiani A, Beltramo M, Alessio M. Pigment epithelium-derived factor is differentially expressed in peripheral neuropathies. *Proteomics* 5:4558–4567, 2005.
5. Marengo E, Robotti E, Righetti PG, Antonucci F. New approach based on fuzzy logic and principal component analysis for the classification of two-dimensional maps in health and disease. Application to lymphomas. *J Chromatogr A* 1004:13–28, 2003.
6. Marengo E, Robotti E, Gianotti V, Righetti PG, Cecconi D, Domenica E. A new integrated statistical approach to the diagnostic use of two-dimensional maps. *Electrophoresis* 24:225–236, 2003.
7. Marengo E, Leardi R, Robotti E, Righetti PG, Antonucci F, Cecconi D. Application of three-way principal component analysis to the evaluation of two-dimensional maps in proteomics. *J Proteome Res* 2:351–360, 2003.
8. Marengo E, Robotti E, Cecconi D, Scarpa A, Righetti PG. Application of fuzzy logic principles to the classification of 2D-PAGE maps belonging to human pancreatic cancers treated with trichostatin-A. FUZZ-IEEE 2004, IEEE International Conference on fuzzy systems, July 25–29, 2004. Budapest, Hungary.
9. Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. *IEEE Trans Pattern Analysis Machine Intelligence* 22:4–37, 2000.
10. Appel V, Stewart SS, Smith G, Appel SH. A rating scale for amyotrophic lateral sclerosis: description and preliminary experience. *Ann Neurol* 22:328–333, 1987.
11. Kimura J. Anatomy and physiology of the neuromuscular junction. In: Kimura J, Ed. *Electrodiagnosis in Diseases of Nerve and Muscles*. (3rd ed.) New York: Oxford University Press, pp227–249, 2001.
12. Scielzo C, Ghia P, Conti A, Bachi A, Guida G, Geuna M, Alessio M, Caligaris-Cappio F. HS1 protein is differentially expressed in chronic lymphocytic leukemia patient subsets with good or poor prognoses. *J Clin Invest* 115:1644–1650, 2005.
13. Rosengren AT, Salmi JM, Attokallio T, Westerholm J, Lahesmaa R, Nyman TA, Nevalainen OS. Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels. *Proteomics* 3:1936–1946, 2003.
14. Dick PJ. Detection characterization and staging of polyneuropathy: assessed in diabetics. *Muscle Nerve* 11:21–32, 1988.
15. Galer BS, Jensen MP. Development and preliminary validation of a pain measure specific to neuropathic pain. *Neurology* 48:332–338, 1997.
16. Dworkin RH, Backonja M, Rowbotham MC, Allen RR, Argoff CR, Bennett GJ, Bushnell MC, Farrar JT, Galer BS, Haythornthwaite JA, Hewitt DJ, Loeser JD, Max MB, Saltarelli M, Schumacher KE, Stein C, Thompson D, Turk DC, Wallace MS, Watkins LR, Weinstein SM. Advances in neuropathic pain: diagnosis, mechanisms, and treatment recommendations. *Arch Neurol* 60:1524–1534, 2003.
17. Sindrup SH, Jensen TS. Efficacy of pharmacological treatments of neuropathic pain: an update and effect related to mechanism of drug action. *Pain* 83:389–400 1999.
18. Wall PD, Melzack R, Eds. *Text Book of Pain*. Edinburgh: Churchill-Livingstone, 1999.