Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Chromatography B, 833 (2006) 51–62
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Decoding 2D-PAGE complex maps: Relevance to proteomics ☆

Maria Chiara Pietrogrande [a,*], Nicola Marchetti [a], Francesco Dondi [a], Pier Giorgio Righetti [b]

[a] *Department of Chemistry, University of Ferrara, via L. Borsari, 46 I-44100 Ferrara, Italy*
[b] *Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Milano 20131, Italy*

Received 1 September 2005; accepted 26 December 2005
Available online 31 January 2006

Abstract

This review describes two mathematical approaches useful for decoding the complex signal of 2D-PAGE maps of protein mixtures. These methods are helpful for interpreting the large amount of data of each 2D-PAGE map by extracting all the analytical information hidden therein by spot overlapping. Here the basic theory and application to 2D-PAGE maps are reviewed: the means for extracting information from the experimental data and their relevance to proteomics are discussed. One method is based on the quantitative theory of statistical model of peak overlapping (SMO) using the spot experimental data (intensity and spatial coordinates). The second method is based on the study of the 2D-autocovariance function (2D-ACVF) computed on the experimental digitised map. They are two independent methods that are able to extract equal and complementary information from the 2D-PAGE map. Both methods permit to obtain fundamental information on the sample complexity and the separation performance and to single out ordered patterns present in spot positions: the availability of two independent procedures to compute the same separation parameters is a powerful tool to estimate the reliability of the obtained results. The SMO procedure is an unique tool to quantitatively estimate the degree of spot overlapping present in the map, while the 2D-ACVF method is particularly powerful in simply singling out the presence of order in the spot position from the complexity of the whole 2D map, i.e., spot trains. The procedures were validated by extensive numerical computation on computer-generated maps describing experimental 2D-PAGE gels of protein mixtures. Their applicability to real samples was tested on reference maps obtained from literature sources. The review describes the most relevant information for proteomics: sample complexity, separation performance, overlapping extent, identification of spot trains related to post-translational modifications (PTMs).

Keywords: Two-dimensional maps; Spot overlapping; Bioinformatics; Chemometric methods

1. Introduction

The main goal of proteomics is a comprehensive identification and quantification of every protein present in a complex biological sample: 2D-gel electrophoresis (2D-GE) is the classical and principal tool for protein separation prior to mass spectrometry (MS) [1–4]. While MS has developed into a rapid high-sensitivity method for identifying proteins [5,6], a complete separation of proteins prior to MS in a high-throughput manner, that would allow analysis of an entire intact proteome, is still far from being achieved [7,8].

The major advantage of 2D-PAGE is that it enables the simultaneous separation of thousands of unknown proteins, first by charge using isoelectric focusing (IEF) and then by size using SDS–polyacrylamide gel electrophoresis. Each cell or biological fluid has a rich protein content that can be formed by thousands of proteins present in a wide range of relative abundance and displaying great differences in structure and size. As a consequence, a comprehensive separation of all the proteins – each spot is pure, i.e., formed by one pure single component (SC) protein – is not achieved in 2D-PAGE maps. The common condition is a very complex separation pattern in 2D gels, where a single spot can be composed of two or more proteins [2,9]. It must be noted that protein separation is the first stage in protein profiling, prior to mass spectral analysis: the presence of a

Abbreviations: IM, interdistance model; AM, abundance model; SMO, statistical model of peak overlapping; SC, single component; 2D-ACVF, 2D-autocovariance function; 2D-EACVF, experimental 2D-ACVF; 2D-TACVF, theoretical 2D-ACVF; SDO, statistical degree of overlapping

☆ This paper was presented at the 2nd IPSo Congress on Proteomics and Genomics, Viterbo, Italy, 29 May to 1 June 2005.

* Corresponding author. Fax: +39 0532240709.
E-mail address: mpc@unife.it (M.C. Pietrogrande).

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2005.12.051

single or a low number of proteins in a spot to be introduced in the mass spectrometer, increases the reliability and sensitivity of the MS measurement for protein structure identification [4,5].

Moreover, the explosion in proteomics research over recent years has brought with it a very large amount of information generated. An efficient use of the large amount of data produced by each analytical run requires a powerful and user-friendly data analysis by means of computer algorithms [10–20]. An exhaustive interpretation of the plethora of data obtained from each analytical run is still far from being achieved, despite some impressive robotic and chemical technologies introduced by several companies as well as new algorithms for spot detection, comparison, quantification, and statistical analysis of the map data [13–15,19,20]. Therefore, any effort in this direction is very helpful in order to extract information from such complicated separations and solve the complexity of the proteome sample.

Many mathematical–statistical methods have been developed for handling data obtained from 2D-PAGE maps [10–21]. Among them, two chemometric procedures developed by the authors are discussed in the present review [21–26]. Their peculiarity lies in regarding the whole 2D signal as a statistical ensemble whose properties can be estimated by using proper mathematical functions. The following basic parameters, describing the complexity of 2D-PAGE maps, can be extracted:

- the sample complexity expressed by the number of proteins present in the sample (single components, SCs), m;
- the separation performance represented by the average spot dimensions, σ_x and σ_y;
- the separation pattern described by the function (interdistance model, IM) representing the distribution of the spot position (coordinates pI and M_r) in the map. In the 2D-PAGE maps such a distribution can be ordered, disordered or a combination of them.

The first method is the statistical model of peak overlapping (SMO) model for a statistical quantification of the degree of spot overlapping present in a map [21,23,24]. With this model, starting from the experimental map, information on the number of proteins present in the sample and on the separation pattern can be extracted. The second method is based on the study of the 2D-autocovariance function (2D-ACVF) computed on the experimental digitised map [25,26]. From it, the complexity of the mixture (number of components, separation pattern) and the separation performance can be estimated. Moreover, the study of the 2D-EACVF plot allows one to identify ordered separation patterns of SC spots, which can be related to specific protein structures. The methods have been previously developed by some of the authors for mono-dimensional chromatograms [27–42] and subsequently extended to 2D separations [21,23–26].

The basic assumption of both the SMO and 2D-EACVF models is that the SC spots locate randomly in the 2D-PAGE map, i.e., spots position along the two separation axes are independent random variables with Poisson distributions. Attention has to be focused if the experimental 2D-PAGE maps really fulfil this condition: the most common condition in protein maps of living systems, is an uneven spot distribution with a high number of proteins displaying spot overcrowding in the 4–6 pH range of pI coordinates and 20–60 KDa for M_r values. Moreover, specific ordered retention structures (spot trains) can be present, superimposed on, and hidden by, a random spot location [2]. As an example, a map formed by experimental pI and $\log M_r$ coordinates retrieved from the SWISS-2DPAGE database [22] is reported in Fig. 1 . It was generated by using the coordinates of 1956 identified spots in reference maps of human tissues: the pI and $\log M_r$ values were projected into the separation axes and their probability distribution histograms were computed [21]. The obtained distributions follow a Poissonian density function as proved by a chi-square test applied to check the goodness-of-fit (H_0 hypothesis accepted at a significance level of $p = 0.05$). Therefore, the spot positions in the 2D maps result random variables with Poisson distributions, as combination of the two independent random variables with Poisson distributions and the two SMO and 2D-EACVF models can be properly applied [21,25].

Equal and complementary information can be extracted from the 2D-PAGE map using the SMO and 2D-EACVF methods: in particular the most relevant results for proteomics will be discussed in the following.

1.1. Estimation of the map properties

Estimation of the map properties can be estimated by evaluating the sample complexity, m, and the system performance, σ_x and σ_y.

The number of proteins present in the sample is usually higher than the number of detectable spots, as a consequence of the strong spot overlapping [7–9]. A comparison between the values obtained by the two independent SMO and 2D-EACVF procedures is a check of the result reliability. The estimation of σ_x and σ_y values is a powerful tool for selecting proper experimental conditions, i.e., gel structure, immobilized pH gradient, to optimise system performance [2]. Moreover, high σ_x and σ_y values may be diagnostic for overloading effects revealing that a sample excess has been loaded on the gel [39].

1.2. Estimation of spot overlapping degree

The original statistical degree of peak overlapping (SDO) approach developed by Davis [43–47] allows one to statistically estimate the number of spots formed by one, two, three etc. proteins, i.e., the number of singlets, doublets, triplets etc. present in the map [23]. It requires the knowledge of the m values as estimated from the SMO and 2D-EACVF methods. The common situation – a tissue homogenate under normal sample loading (ca. 1 mg total protein) and standard gel sizes (18 cm × 20 cm, IEF × SDS–PAGE) – is an overcrowded condition, where the singlets would be the least abundant species. For example, it is possible to estimate that in the case of 1500 proteins loaded, the singlets would be only 27%; for a total of 3000 polypeptide chains, the singlets would amount only to 14% [23].

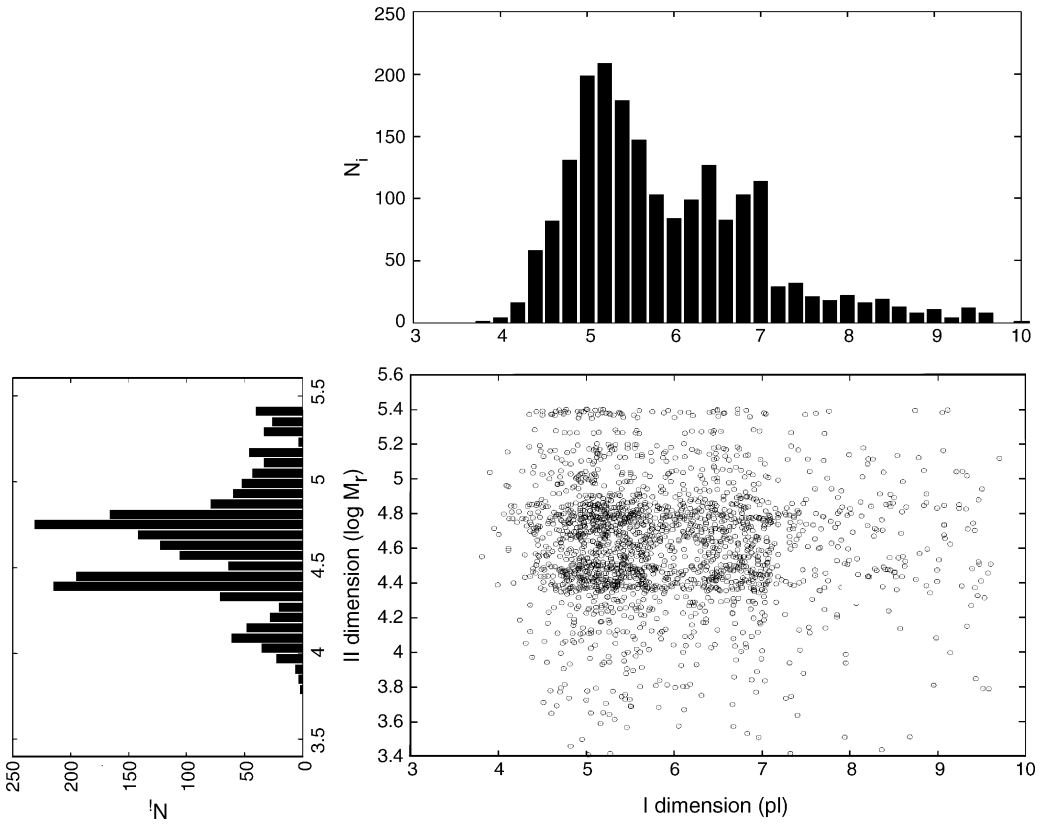

Fig. 1. A computer generated map using 1956 pI and log M_r coordinates retrieved from the SWISS-2DPAGE database [22]. The histograms of the spot locations projected into the pI and log M_r axes are reported: N_j represents the number of protein spots belonging to each class.

1.3. Identification of ordered structures

Ordered sequences of spots can be identified in the complex map by using both the SMO and 2D-EACVF approaches, the last method being particularly helpful in singling them out from the whole complexity of the 2D separation. An example of ordered sequences may be the spot trains consistent with protein isoforms differing in a constant variation of the number of ionogenic groups in the molecule. These protein isoforms suggest the presence of co- and post-translational modifications (PTMs) such as glycosylation, phosphorylation, deamidation [1–4,47–55]. Identification of protein post-translational modifications is quite an important aspect of proteomics, since it has been well established that PTMs occur on almost all proteins and are of extreme biological importance, i.e., they can regulate a variety of protein activities, such as enzymatic activity, ability to interact with other proteins, sub cellular localization, targeted degradation, etc. [1,47–55].

2. Computation

All the programs are written in Fortran [56] and run on a personal computer Pentium III 2Ghz (512 MB RAM) AMD Athlon.

The statistical methods were validated on 2D maps with precisely known properties by comparing the results obtained with the original map properties [21,25]. For this purpose, 2D maps were generated by computer calculations: each spot is considered as a point described by two position coordinates (pI and M_r of the barycentre of the spot) and by a third coordinate describing spot intensity. Different distribution functions were used to generate the position coordinates pI and M_r (IM) and the spot intensity distribution (abundance model, AM). Different combinations of IM and AM yield three different map types.

The *most general maps* were simulated in the simplest model where the pI and M_r coordinates follow a random-Poissonian distribution and the spot intensity distribution is described by uniform (U) or exponential (E) functions (AM = U or E).

Synthetic 2D maps were generated to describe real 2D-PAGE maps [21]. The rejection algorithm [56] was applied for independently generating the pI and M_r coordinates which follow the same position distribution present in experimental maps: such a distribution was computed from the pI and M_r coordinates of 1956 identified spots in reference maps of human tissues (SWISS-2DPAGE database, [22]). Spot intensity values were generated according to an exponential distribution, since it has been demonstrated to be the most probable for a high number of components [7].

Reference 2D maps were build up from the experimental reference maps of human tissues using the pI and M_r values of identified spots retrieved from SWISS-2DPAGE database [22,57–62]. In these maps the spot intensity distribution is described by an exponential function (AM = E).

For each generated map, 50 runs were performed using different random sequences and the reported results (Tables 1 and 2) are the mean values.

Table 1
Results of the SMO method applied to computer generated and experimental maps [21]

m	Strip number	$m_{est} \pm \sqrt{m_{est}}$	ε%	CV%
2000	40	1960 ± 44	2.0	1.7
2000	50	1940 ± 44	3.0	1.5
2000	27 (20)	1972 ± 44	1.4	1.5
2000	43 (10)	1994 ± 45	0.3	1.5
1000	20	1026 ± 32	2.6	2.2
1000	30	1012 ± 32	1.2	1.7
1000	31 (10)	1005 ± 32	0.5	1.7
1000	33 (15)	991 ± 31	0.9	1.5
500	10	509 ± 23	1.8	3.4
500	15	485 ± 22	3.0	2.7
500	9 (10)	511 ± 23	2.2	3.4
500	11 (20)	499 ± 22	0.3	2.3
108[a]	10	101 ± 10	5.6	2.0
108[a]	3 (20)	105 ± 10	2.8	1.9

Maps were divided into different number of strips (2nd column) along the pI axis. m is the number of proteins, m_{est} is the estimated value with the SMO method. Different computation procedures were used: maps were divided into different number of strips (2nd column) with variable or constant (in brackets) number of proteins. Accuracy is expressed as relative error ε%; precision as relative standard deviation CV% (50 runs).

[a] Reference map: colorectal adenocarcinoma cell line [59].

3. Results

3.1. Statistical model of peak overlapping (SMO)

The quantitative theory of SMO has been originally developed to study 1D chromatograms [27–29]. The extension of the procedure to 2D separations implies the division of the 2D map into many strips, considered as 1D separations, on which computations are performed [21]. This approach is based on the assumption that a complex separation can be considered as the superimposition of many simpler separations: the properties of the whole map can be determined by addition of the parameters statistically estimated for each individual separation [21].

The SMO describes the complexity of a multicomponent separation in terms of two probability functions: the first is the interdistance model, which describes the position of SC spots in the separation surface, the second is the abundance model, which represents the distribution of SC abundance (spot intensity) [21,28,29]. The properties of the map are computed by comparing these models to the observable parameters obtained from the software output (i.e., pI and $\log M_r$ coordinates, intensity for each spot).

The 2D map is divided into many strips by choosing 1 dimension (pI or $\log M_r$). On each strip a critical interdistance value x_0 (the smallest distance by which the centers of two adjacent non-overlapping spots can be separated) is selected and used for counting the spots: if two or more spots fall inside the same x_0 interdistance, they will be counted as one spot formed by two or more proteins. The average value computed from the area of the counted spots, is the average observed peak area, y_{obs}. Choosing increasing x_0 values (related to the required separation resolution), a decreasing number of spots is counted, yielding increasing y_{obs} values. Many (y_{obs},x_0) couples can be obtained from the 2D map signal, with an experimental limit due to the instrumental minimum distance between two resolved spots [28]. A theoretical expression has been derived to relate the observed y_{obs} and x_0 values experimentally computed from the map to the theoretical values m and the real mean intensity,

Table 2
Results of the 2D-ACVF method applied to computer generated and experimental maps [25,26]

AM	m	σ_x	σ_y	σ_h^2/a_h^2	m_{est}	$\sigma_{x,est}$	$\sigma_{y,est}$	σ_m^2/a_m^2
U	250	0.75	0.75	0.33	243 ± 7	0.76 ± 0.01	0.75 ± 0.01	0.33 ± 0.03
E	250	0.75	0.75	1	241 ± 6	0.75 ± 0.01	0.75 ± 0.01	0.99 ± 0.11
U	750	0.75	0.75	0.33	723 ± 23	0.75 ± 0.01	0.75 ± 0.01	0.31 ± 0.02
E	750	0.75	0.75	1	700 ± 20	0.75 ± 0.01	0.75 ± 0.01	0.90 ± 0.08
U	750	1.00	0.75	0.33	716 ± 24	1.00 ± 0.02	0.76 ± 0.02	0.30 ± 0.02
E	750	1.00	0.75	1	684 ± 22	1.01 ± 0.02	0.75 ± 0.01	0.87 ± 0.08
U	1500	0.75	0.75	0.33	1404 ± 40	0.76 ± 0.01	0.76 ± 0.01	0.28 ± 0.01
E	200	0.025	0.0006	1	183 ± 13	0.025	0.0006	0.92
E	200	0.025	0.0006	1	203 ± 14	0.026	0.0006	0.92
E	200	0.016	0.0004	1	193 ± 14	0.016	0.0004	0.96
E	200	0.009	0.0002	1	196 ± 14	0.009	0.0002	0.97
E	200	0.008	0.0002	1	197 ± 14	0.008	0.0002	0.98
E	500	0.025	0.0006	1	459 ± 21	0.025	0.005	0.92
E	500	0.016	0.0004	1	479 ± 22	0.017	0.005	0.93
E	500	0.009	0.0002	1	485 ± 22	0.009	0.002	0.97
E	750	0.009	0.0002	1	710 ± 26	0.009	0.003	0.95
E	750	0.008	0.0002	1	736 ± 27	0.008	0.002	0.96
E	1000	0.009	0.0002	1	919 ± 30	0.009	0.003	0.93
E	1000	0.008	0.0002	1	764 ± 31	0.008	0.002	0.96
E	HEPG2 99	0.009	0.0002	1	100 ± 10	0.009	0.0002	0.99
E	DL-1 108	0.009	0.0002	1	104 ± 10	0.009	0.0002	0.99
E	PLASMA 626	0.009	0.0002	1	601 ± 24	0.009	0.0002	0.97

AM: abundance distribution model [25,26]. Number of proteins (m) and spot shape (σ_x,σ_y) are the values used for map computer simulation; m_{est}, $\sigma_{x,est}$, $\sigma_{y,est}$ are the separation parameters estimated by the 2D-ACVF method.

$\bar{y}$, which is the true mean peak intensity value corresponding to well separated SC peaks. A linear relationship was obtained:

$$\ln y_{obs} = \ln \bar{y} + m\frac{x_0}{X} \quad (1)$$

where X is the total separation length computed from the gel dimension. The experimental values of the (y_{obs}, x_0) couples can be fitted by a straight line, whose slope represents a statistical estimation of m, the number of single components.

Eq. (1) is strictly true in the simplest and most general case, where separation interdistances exhibit an Exponential distribution. Any deviation from linearity in Eq. (1), may be related to the specific feature of the experimental separation pattern. In fact, a general relationship has been derived, showing that the 1st derivative of the relationship between the quantity $1/y_{obs}$ and the critical interdistance x_0 gives the function describing the interdistance distribution between subsequent spots, the IM [28]:

$$IM = -\bar{y}\frac{d(1/y_{obs})}{dx_0} \quad (2)$$

The SMO method was applied to numerically simulated and experimental maps: the most relevant results are reviewed in the following.

3.1.1. Estimation of the separation parameters and spot overlapping degree

The SMO procedure was validated on numerically simulated maps with known number of proteins describing real experimental 2D-PAGE maps (results in Table 1) [21]. The number of components is estimated for each 1D strip by using Eq. (1); the values for each strip were added together to compute the total number of proteins (m in 3rd column). The maps were divided along the pI axis using two computation procedures: a fixed number of strips was selected or a variable number was computed by fixing the minimum number of proteins present in each strip (values in brackets in 2nd column of Table 1). A simple interpretation on the results can be visualized in a 3D plot displaying the accuracy of the estimated results, expressed as relative error ε%, as a function of the number of components and the number of strips (Fig. 2a). All the reported results show a correct estimation of m, within the accuracy of m estimation, i.e., $\pm\sqrt{m}$ (compare 1st and 3rd columns in Table 1). The accuracy of the m estimation is particularly good (relative error ε% lower than 3%) if the division procedure into variable number of strips is used (larger red points in Fig. 2a), even if the number of proteins present in the map is low.

The SMO procedure was applied for studying reference 2D maps retrieved from the SWISS-2DPAGE database [22]. As an example, the map of colorectal adenocarcinoma cell line (DLD1_HUMAN) [59] is reported in Fig. 3a . The SMO method allows to estimate m with good precision and accuracy (last two rows in Table 1). Looking into the details of the colorectal adenocarcinoma map, we can see that some regions exhibit a random pattern, described by an exponential IM (disordered region, interdistance histogram in inset in Fig. 3a), while others show an order in spot position (ordered region, interdistance histogram in inset in Fig. 3a). However, the whole map, i.e., superimposition of many ordered and disordered sequences, exhibits a completely disordered separation pattern, which can be properly described by an exponential IM [21].

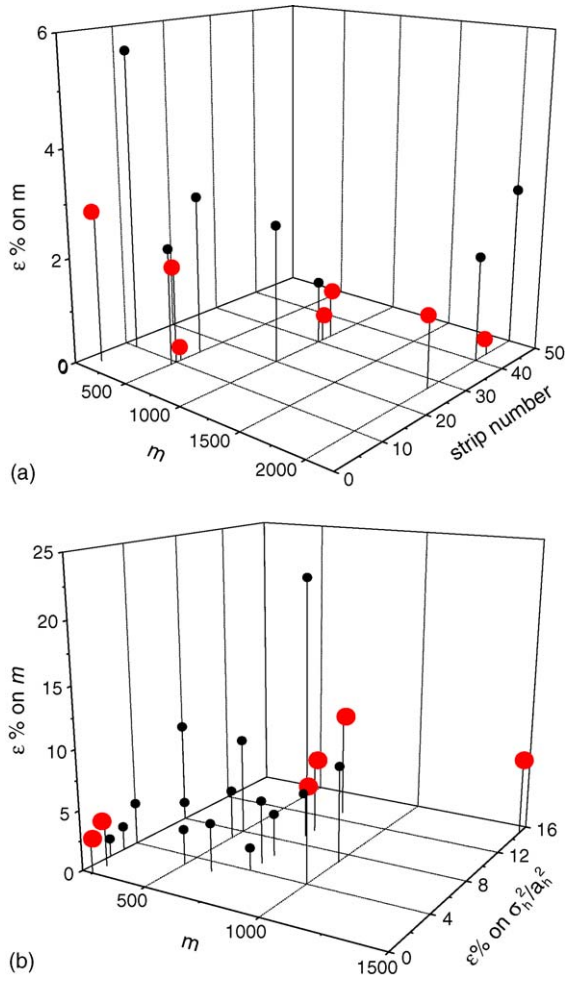

Fig. 2. Accuracy of the m estimated values, expressed as relative error, ε%: (a) SMO method: ε% as a function of the number of components, m, and the number of strips. Black points: the map is divided into a fixed number of strips; larger red points: the map is divided into a variable number of strips by fixing the minimum number of spots present in each strip; (b) 2D-ACVF method: ε% as a function of the number of components, m, and the relative error affecting the σ_h^2/a_h^2 estimation. Black points: the simplified 2D-ACVF method; larger red points: the original 2D-ACVF method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

By using the estimated m value, the SDO procedure [27,43–46] can be applied to statistically estimate the degree of spot overlapping present in a 2D-PAGE map, i.e., the purity extent of each spot can be evaluated as the percentage of spots formed by one, two, three or more proteins [23]. This information is useful for estimating the influence of different experimental conditions (strip dimension, detection system performance, pI range) on spot overlapping. Computations were performed on synthetic 2D maps describing experimental 2D-PAGE gels [23]: the degree of error associated with identification and quantification of each protein can be predicted and the best experimental strategies can be set-up to reduce spot overlapping and achieve the highest resolution in protein separation.

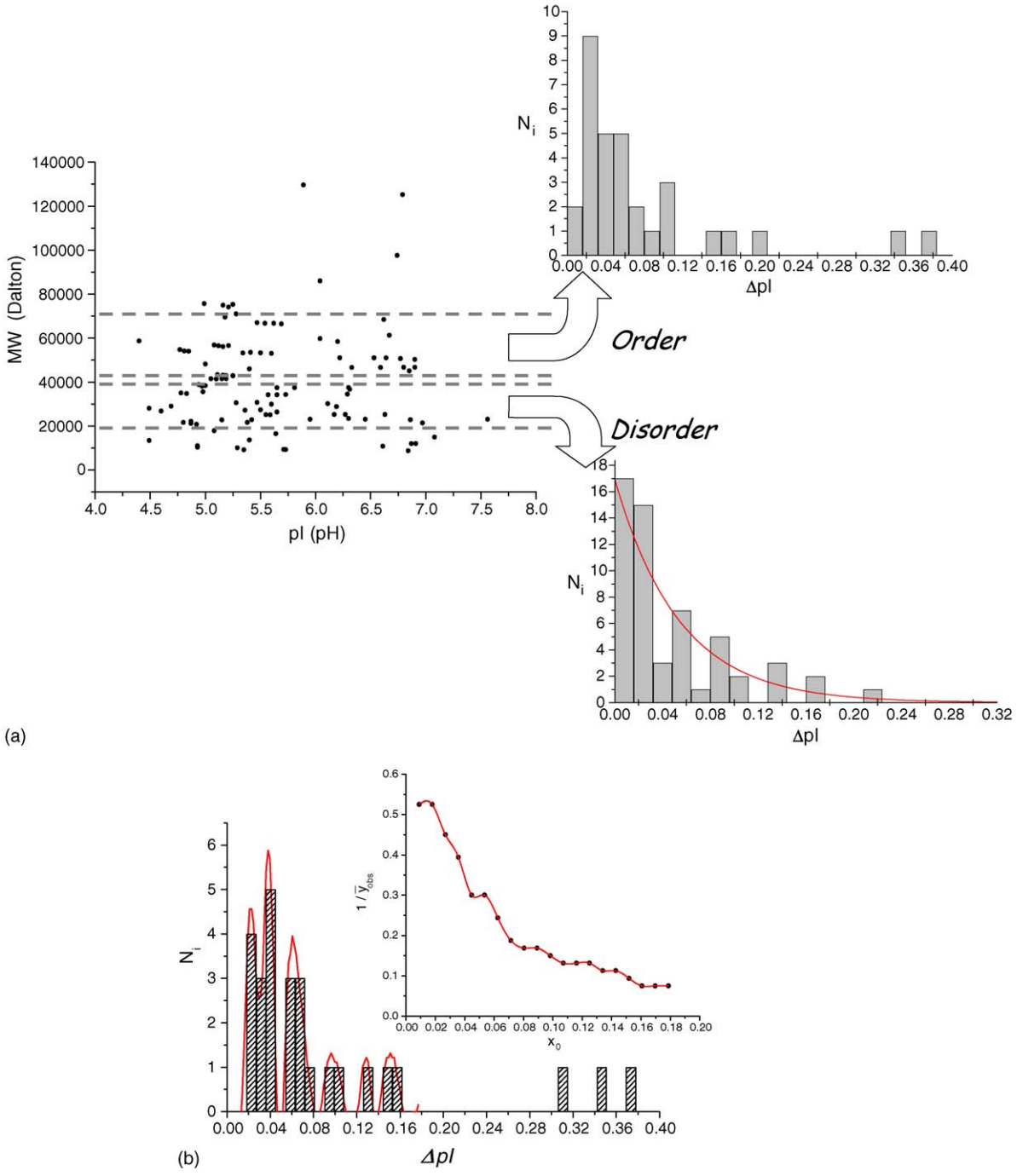

Fig. 3. Study of colorectal adenocarcinoma cell line map using the SMO method: (a) experimental 2D-PAGE map in a digitised form, showing ordered (top) and disordered (bottom) regions. (insets) ΔpI interdistance distribution in the ordered and disordered regions and (b) histogram of ΔpI interdistance distribution in an ordered region: real values (histogram) and the distribution (solid red line) estimated by using the plot reported in the inset (reprinted from [21], by permission). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

Recently, the SMO and SDO models were, for the first time, applied to experimental 2D-PAGE maps: the reliability of the obtained results was checked by comparing the estimated overlapping degree with the experimental MS data [24]. The studied sample was a neuroblastoma xenograft implanted in mice, submitted to 2D-PAGE separation. By the MS analysis using a quadrupole-TOF mass spectrometer 74 proteins were identified: 52 (71%) of them were found to be singlets, 14 (19%) were doublets, 6 (8%) were triplets and 1 each were quadruplet and quintuplet. The computation of the SMO and SDO procedures on the selected spots yielded the following results: 74 proteins formed 54 singlets, 14 doublets, 6 triplets, 1 quadruplet and 1 quintuplet. The excellent agreement found between theoretical and experimental data proves that the SDO model is a powerful and robust method to accurately predict the overlapping degree present in a map. In the case studied the number of

singlets would be in majority over the other spots (71%): this is an exceptional situation in 2D separations, since only 264 total spots were detectable in the map.

3.1.2. Identification of ordered structures

The SMO method is a powerful tool for identifying the presence of ordered structures in the spot positions, even if it might escape detection since it is hidden in the spot overcrowding [21]. In the example of the colorectal adenocarcinoma cell line map (Fig. 3a), the distribution of position interdistance along the p*I* separation axis was computed in ordered and disordered regions of the map (histograms in enlarged details in Fig. 3a). The presence of an ordered separation pattern can be visually detected by a significant deviation from linearity of Eq. (1) characterized by an higher slope of the y_{obs} versus x_0 or $1/y_{obs}$ versus x_0 plots (insert in Fig. 3b). The same ordered structures can also be identified and quantified by the presence of some maxima in the distribution function estimated by using Eq. (2) (red line in Fig. 3b): they correspond to the interdistance values repeated in the map, e.g. 0.04 and 0.06 ΔpI values (Fig. 3b). The reliability of the obtained results is shown by the good agreement between the distribution histogram computed from the spot position coordinates (histogram in Fig. 3b) or estimated by using Eq. (2) (red line in Fig. 3b). It must be underlined that the Eq. (2) makes it possible to detect the presence of ordered patterns in a map, but not to give any information on the map region where such patterns are present. Only a detailed analysis of the position coordinates of each spot can give this information.

3.2. Autocovariance function method

The method is based on the study of the 2D experimental autocovariance function (2D-EACVF) computed on the experimental map acquired in digitised form [25]. The digitised map consists of a gridded surface $N_x \times N_y$, where all the nodes are equally spaced; on it the 2D-EACVF is computed as:

$$\text{2D-EACVF}_{k,l} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x - kN_y - l} \sum_{j=1} (f_{i,j} - \bar{f})(f_{i+k,j+l} - \bar{f}) \tag{3a}$$

$$k = -L_{\max,x}, \ldots, -1, 0, 1, \ldots, L_{\max,x} \tag{3b}$$

$$l = -L_{\max,y}, \ldots, -1, 0, 1, \ldots, L_{\max,y} \tag{3c}$$

where $f_{i,j}$ represents the map intensity at the point (i,j), $\bar{f}$ is the average intensity calculated over all the sampled points, $L_{\max,x}$ and $L_{\max,y}$ are the maximum spans of the p*I* and $\log M_r$ values over which 2D-EACVF is calculated. A cyclic calculation procedure was used (the beginning and the end of the separation axes are merged by using negative k or l indices): in this way, each point of the 2D-EACVF is computed using the same number of points and thus it is estimated with the same degree of precision (see ref 36 for the details). Each point used for computation can be converted into $\Delta x = \Delta pI$ and $\Delta y = \Delta \log M_r$ on the basis of the sampling inderdistances between subsequent points on the X- and Y-axes. The 2D-EACVF can be plotted versus the k and l interdistance points along the two separation axes to obtain the 2D-EACVF plot. An example of a map containing 50 SCs is reported in Fig. 4a ; 25 of them are randomly distributed in the separation space, 25 SCs form five ordered sequences of five spots. A 3D plot of the 2D-EACVF computed on the map (Fig. 4a) is reported in Fig. 4b (the inset shows an enlarged view of the region close to the origin 2D-EACVF(0,0)). The 2D-EACVF represents intercorrelation between positions of subsequent spots: if some constant interdistances are repeated in different regions of the map, the 2D-EACVF corresponding to the repeated interdistances assumes a value significantly higher than 1 and the 2D-EACVF plot shows well defined deterministic cones (Fig. 4b). The study of the 2D-EACVF is based on theoretical expressions of 2D-ACVF (2D-TACVF) as a function of the separation parameters (quantities m, σ_x, σ_y and retention pattern structure). Two limit examples of retention patterns are discussed, since they are the basis for studying any experimental 2D separation [25]: a disordered separation (Poissonian retention pattern) and an ordered map containing ordered sequences of spots. For both the cases, the SC spot shape is represented by a bivariate Gaussian distribution described by the standard deviation along the two separation axes, σ_x and σ_y: both circular ($\sigma_x = \sigma_y$) and elliptical ($\sigma_x \neq \sigma_y$) spots are assumed.

The Poissonian retention pattern describes a completely disordered separation where SC positions are uniform randomly distributed over the X,Y area. The 2D-TACVF is given by [25]:

$$\text{2D-TACVF}(\Delta x, \Delta y) = \frac{V_T^2(\sigma_h^2/a_h^2 + 1)}{4\pi m \sigma_x \sigma_y XY} e^{-\left[\frac{(\Delta x)^2}{4\sigma_x^2} - \frac{(\Delta y)^2}{4\sigma_y^2}\right]} \tag{4}$$

where

$$V_T = 2\pi m a_h \sigma_x \sigma_y \tag{4a}$$

is the total volume of the signal computed on the three coordinates (x, y, f); a_h and σ_h^2 are, respectively, the mean and the variance of SC spot abundance.

An ordered pattern in a 2D-PAGE map is formed by SCs displaying ordered sequences of spots in the separation space: the position of the nth term of the series is described by:

$$x(n) = a_x + b_x n \tag{5a}$$

$$y(n) = a_y + b_y n \tag{5b}$$

where a_x, a_y, b_x and b_y are constants. In this case, the 2D-TACVF is expressed by:

$$\text{2D-TACVF}(\Delta x, \Delta y) = \sum_{k=0}^{k=n_{\max}} \frac{V_T^2}{4\sigma_x \sigma_y \pi XY (n_{\max} - k + 1)} \left(\frac{\sigma_h^2}{a_h^2} + 1\right) \times e^{-[(\Delta x - b_X k)^2/4\sigma_x^2] - [(\Delta y - b_Y k)^2/4\sigma_y^2]} \tag{6}$$

where $n_{\max}$ is the highest value of n, i.e., the SC number of the series. Note that the a_x, a_y parameters are missing in this expression, since the ACVF expresses only the recursivity of an ordered structure [25]. According to Eq. (6), the 2D-TACVF plot, and therefore the 2D-EACVF plot, shows well defined peaks located

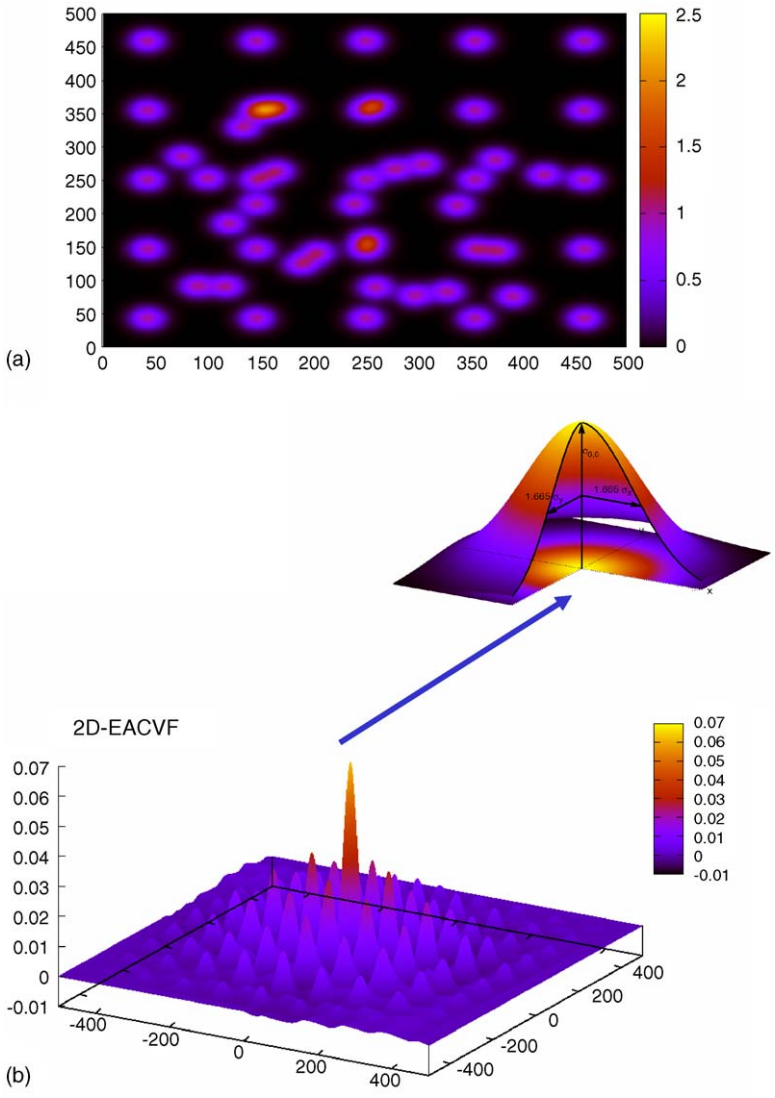

Fig. 4. Computation of the Experimental 2D-ACVF (2D-EACVF) on the experimental map and the resulting 2D-EACVF plot: (a) experimental map in the digitised form and (b) 2D-EACVF plot vs. interdistance along the two separation axes; enlarged detail: 2D-EACVFplot for interdistances lower than $4\sigma_x$ and $4\sigma_y$.

at interdistances kb_x and kb_y, corresponding to repeated interdistances among the terms of the SC series. These peaks are called deterministic since they reflect the order of the sequence.

In addition to the original method based on non-linear fitting of 2D-EACVF to 2D-TACVF [25], a simplified version of the 2D-ACVF method has been recently developed: it is based on a graphic measurement of the two main regions of the 2D-EACVF [26].

Short-term correlation (interdistance lower than $4\sigma_x$ and $4\sigma_y$, enlarged detail in inset in Fig. 4b). The first part of 2D-EACVF – interdistance lower than $4\sigma_x$ and $4\sigma_y$, short-term correlations – resembles the mean spot size averaged on all the spots present in the map: from its shape it is possible to estimate the mean standard deviation, σ_x and σ_y, related to the separation performance.

The simplified procedure, based on the first part of the 2D-EACVF (interval $\pm 4\sigma$), is a general method, independent of the specific features of the retention pattern [26]. When the 2D-TACVF is computed at the origin, i.e., $\Delta pI = 0$, $\Delta \log M_r = 0$, the same expression – pre-exponential terms in Eqs. (4) and (6) – is obtained for the Poisson case (Eq. (4)) as well as for the ordered pattern (Eq. (6)) for $k = 0$ (m is substituted by $(n_{max} + 1)$). From 2D-EACVF(0,0) the number of proteins displaying a disordered, m, or an ordered pattern, n_{max}, can be estimated. Some separation parameters in Eq. (4) and (6) are calculated by the developed algorithm: the total volume of the separation, V_T, was computed by numerical integration. The true average SC abundance, a_h^2, its standard deviation σ_h^2 and the relative dispersion of the SC abundance σ_h^2/a_h^2 are not experimentally accessible parameters because of SC spot overlapping and must be approximated by their estimate, i.e., by the experimental values computed from the observed spot maximums. An algorithm based on the comparison of seven successive points for each dimension was used to detect the spot maxima. The area of each detected spot was computed and from these values the average spot maximum abundance, a_m^2, its standard deviation, σ_m^2, and the relative disper-

sion ratio of the maxima, σ_m^2/a_m^2, were obtained and substituted in Eqs. (4) and (6) in the place of the true theoretical value σ_h^2/a_h^2 [25].

The quantities σ_x, σ_y are obtained by a simple graphical inspection of 2D-EACVF plot as:

$$\sigma_x = \frac{d_x}{1.665} \tag{7a}$$

$$\sigma_y = \frac{d_y}{1.665} \tag{7b}$$

where d_x and d_x are the half widths at half height of 2D-EACVF plot along the separation directions x and y (see inset in Fig. 4b) [26].

Long-term correlation (interdistance higher than $4\sigma_x$ and $4\sigma_y$, Fig. 4b). This part contains information for detection and characterization of ordered retention patterns present in the complex map: the presence of deterministic peaks is diagnostic for the existence of the sequences and their position is related to its b_x and b_y parameters (Eqs. (5a) and (5b)). This property is due to two concomitant abilities of the 2D-ACVF: it cancels the effect of the randomness of SC spot positions while it amplifies the recursivity of the repeated interdistances, since it assumes values significantly different from zero only for repeated spot positions. Therefore, the visual inspection of the 2D-EACVF plots or the analysis of the 2D-EACVF values make it possible to identify the presence of ordered retention patterns singling them out from the random pattern of a complex 2D-PAGE map. This behavior is more simply and clearly shown by the intersection of 2D-EACVF with the separation axes. Since the order (constant interdistance repetitiveness) may be related to constant changes in the molecular structure of the proteins, the 2D-EACVF plot can be regarded as a simplified comprehensive picture of the map, still retaining information on the chemical structure of compounds present in the mixture. If more precise and detailed information on specific proteins is required, the 2D-EACVF can be computed and analyzed for selected regions of the 2D map, i.e., those containing the proteins of interest.

The 2D-EACVF method was applied to numerically simulated and experimental maps: the most relevant results are reviewed in the following.

3.2.1. Estimation of the separation parameters

The 2D-EACVF method was validated on computer-generated maps describing experimental 2D-PAGE maps (data in Table 2) [25,26]. The 2D-bed dimensions (X and Y) refer to standard 180 mm × 200 mm gel sizes; the spot shapes are assumed both circular ($\sigma_x = \sigma_y = 0.75$) and elliptical ($\sigma_x \neq \sigma_y$) (3rd and 4th columns in Table 2); the number of proteins varies from $m = 200$ to 1500 (2nd column); spot abundance is described by uniform (U) and exponential (E) distributions (1st column) yielding $\sigma_h^2/a_h^2 = 0.\bar{3}$, and 1.0, respectively (5th column). The results presented in Table 2 were obtained from 50 repeated simulations. The most general 2D separations, assuming Poissonian distribution of spot position – exponential IM – were studied by using the original EACVF method [25]: the obtained results are reported in the 1st–7th rows in Table 2. Some synthetic 2D maps describing experimental 2D-PAGE gels were studied (8th–19th rows in Table 2): the generated pI and M_r coordinates follow the same position distribution present in experimental maps, the elliptical spot shape represents experimental conditions, i.e., $\sigma_x = 0.009$ pH and $\sigma_y = 0.0002 \log M_r$ correspond to the standard case of a 18-cm strip of broad pH range (pH 3–7) with standard (1 mm) scanner resolution [2,23]. The simplified 2D-EACVF method [26], was used for these maps (results reported in Table 2, 8th–19th rows).

With both the methods, the most critical parameter in Eqs. (4) and (6) is the SC abundance dispersion ratio (σ_h^2/a_h^2, 5th column in Table 2): in fact as a consequence of the SC spot overlapping, it is not experimentally accessible and it must be approximated by the maximum spot dispersion ratio computed on the map (σ_m^2/a_m^2, 9th column). The relative error ε% computed in estimating σ_h^2/a_h^2 can be computed by comparing σ_m^2/a_m^2 to σ_h^2/a_h^2 values. A simple interpretation of the results can be visualized in a 3D plot displaying the accuracy of the estimated m values, expressed as relative error ε%, as a function of the number of components and the ε% in estimating σ_h^2/a_h^2 (Fig. 2b). The obtained data show that the number of proteins, m, present in the sample can be correctly estimated within a bias of 10%, using both the original (larger red points in Fig. 2b) and the simplified method (black points in Fig. 2b), also in the case of the most overcrowded maps (compare 2nd and 6th columns in Table 2). The results in the Table also show that the 2D-EACVF plot gives a correct estimation of the mean spot shape for all the simulated maps (σ_x and σ_y in 3rd–4th and 7th–8th columns).

Moreover, some reference 2D maps retrieved from the SWISS-2DPAGE database were studied with both the original and simplified methods [22]: a hepatoblastoma-derived cell line (HEPG2_HUMAN) [57,58], a colorectal adenocarcinoma cell line (DL-1) (DLD1_HUMAN) [59] and a human plasma (PLASMA_HUMAN) [60–62]. The values $\sigma_x = 0.009$ pH and $\sigma_y = 0.0002 \log M_r$ were assumed for spot dimension since they represent the standard case for experimental 2D-PAGE maps [2,23]. The results obtained (last rows in Table 2) show a correct estimation of the separation parameters, m, σ_x and σ_y (compare columns 3-1 and columns 7–8 with columns 3–4), even if the number of proteins considered is low and the theoretical model describing the distribution of SC positions is far from being thoroughly known [26].

A comparison between the SMO and 2D-EACVF methods in estimating the number of proteins, m, was performed on the 2D-PAGE map of colorectal adenocarcinoma cell line (DL-1) [59]. The SMO method estimates the values $m = 101 \pm 10$ and $m = 105 \pm 10$ (13th, 14th rows in Table 1), and the 2D-EACVF procedure yields $m = 104 \pm 10$ (21st row in Table 2). The excellent agreement between the values obtained by the two independent procedures makes it possible to verify the reliability of the results obtained.

3.2.2. Identification of ordered structures

The great strength of the 2D-EACVF in identifying ordered sequences was tested in the case of spot trains. As an example, spot trains showing a mono-dimensional shift parallel to the pI axis are studied, since they represent a common feature in 2D

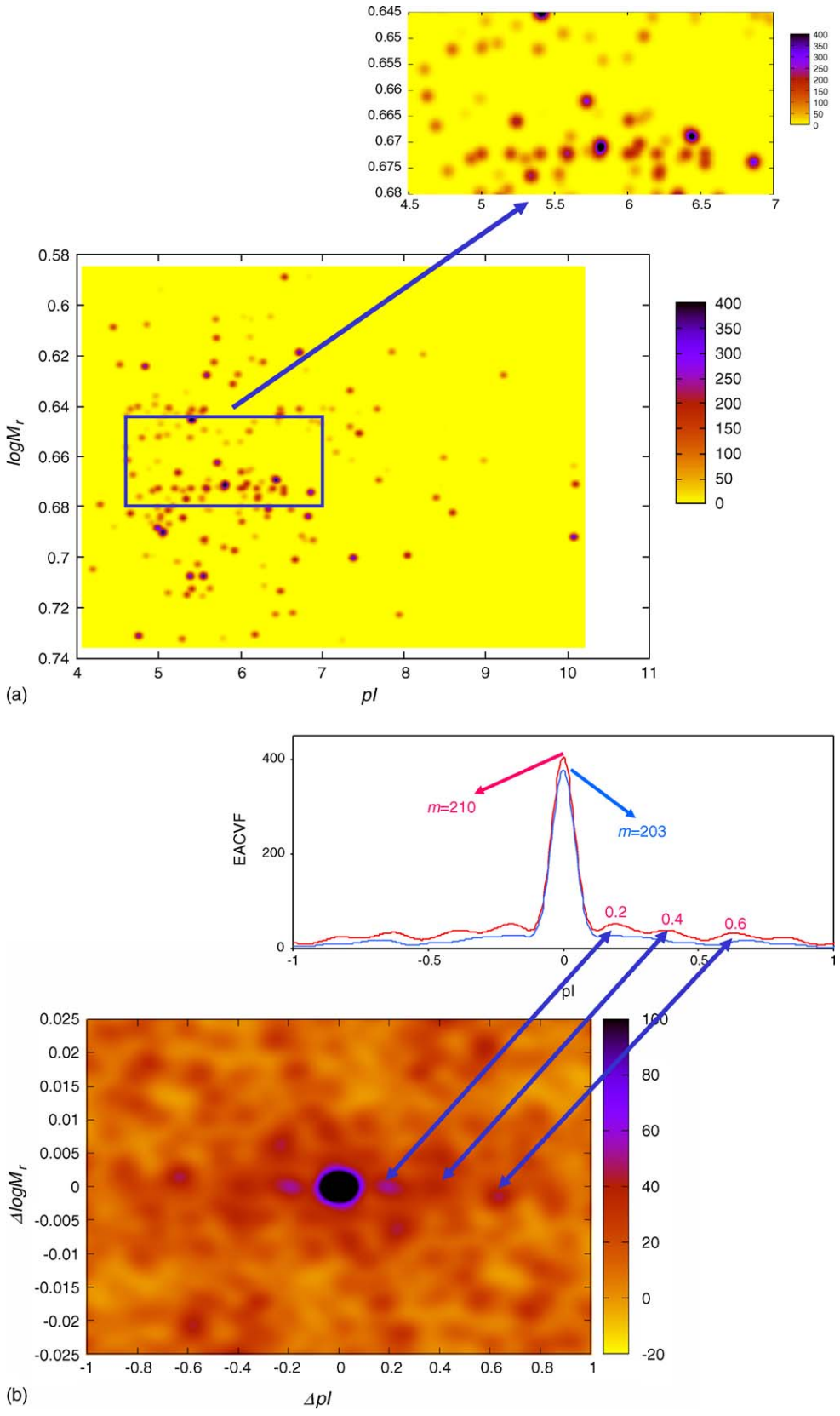

Fig. 5. Identification of a train of spots by the 2D-EACVF method: (a) simulated map where a train of 8 spots in the p*I* range 5–6.4, with a constant $\Delta pI = 0.2$ pH and at a constant $\log M_r$ value of 0.67, was superimposed on a map containing 200 SCs (Table 2, row 9). Enlarged detail: selected 0.6–0.74 $\log M_r$ region of the map containing the train of 8 spots and (b) plot of 2D-EACVF computed on the 2D map. (Inset) 2D-EACVF values over the p*I* separation axis: comparison between the 2D-EACVF plots computed on the map with (upper red line) and without the train of spots (lower blue line) (reprinted from [26], by permission). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

gels. They may be the consequence of proteins post-translational modifications yielding a change in amino acid charges with a consequent alteration in p*I*, while not necessarily in M_r [1–3]. For studying this effect, a computer-simulated map was generated (Fig. 5a), where a train of eight spots in the p*I* range 5–6.4 pH, with a constant ΔpI of 0.2 pH at a constant $\log M_r$ value of 0.67 was superimposed to the original map containing 200 SCs (enlarged inset in Fig. 5a). The 2D-EACVF was computed on the whole map (line 9 in Table 2) and the 2D-EACVF plot reported in Fig. 5b. We observe that the 2D-EACVF exhibits a C_2 symmetry (Eqs. (4) and (6)): correlations in positions $(\Delta x, \Delta y)$ are equal to those in $(-\Delta x, -\Delta y)$, that means that both positive and negative ΔpH shifts give the same 2D-EACVF values. In Fig. 5b well-defined deterministic cones are evident along the p*I* axis at values ΔpH 0.2, 0.4, 0.6 pH: they are related to the constant interdistances repeated in the spot trains. This behaviour is more clearly shown by the intersection of 2D-EACVF with the p*I* separation axis: the inset in Fig. 5b reports the 2D-EACVF plots computed on the same map with (upper red line) and without (lower blue line) the spot train. A comparison between the two lines shows that the 2D-EACVF peaks at 0.2, 0.4, 0.6 ΔpH (upper red line) clearly identify the presence of the spot train singling out this ordered pattern from the random complexity of the map (lower blue line, from map without the spot train). The difference between the two lines identifies the contribution of the two components to the complex separation: the blue line corresponds to the random separation pattern present in the map, the red line describes the order in the 2D map due to the superimposed spot train. It must be noted the high sensitivity of the 2D-EACVF method in detecting order: in fact it is able to detect the presence of only seven-fold repetitiveness hidden in a random pattern of 200 proteins.

4. Concluding remarks

At present, increasing accuracy and precision are achieved in complex 2D gel image acquisition and spot detection due to the development of computer assisted analysis as well as in protein spot identification and quantification due to the advancement of specific software programs. However, the complexity of the plethora of data obtained requires proper signal processing procedures for a complete extraction of the whole analytical information.

The mathematical–statistical methods here reviewed are proved to be powerful tools for proteomics to study 2D-PAGE maps of complex mixtures of proteins. Both methods allow one to extract similar information on the sample complexity and the separation performance and to single out ordered patterns present in spot positions. The availability of two independent procedures to compute the same separation parameters is a powerful tool to check the reliability of the obtained results.

Moreover, the two procedures display different and complementary properties so that each of them is the method of choice to obtain specific information on the 2D-PAGE maps. The SMO procedure is an unique tool to quantitatively estimate the degree of spot overlapping present in a map as well as to predict the influence of different experimental conditions (strip dimension, detector system performance, p*I* range) on spot overlapping. These informations are useful to quantitatively estimate the degree of error associated with identification and quantitation of each protein and to set-up experimental conditions, which will increase resolution and separation performance. For example, an extraordinary improvement in spot purity is obtained if narrow range strips are used, or if the sample is pre-fractionated, using chromatography or electrophoresis, to reduce the initial sample polydispersity.

The strength of the 2D-ACVF method lies in its ability to simply display a comprehensive description of the whole map since the 2D-EACVF plot retains, in a simplified form, information on the quali/quantitative composition of the complex mixture. In fact, the deterministic peaks in the 2DEACVF plot are diagnostic for the presence of ordered spot sequences, which can be related to specific chemical composition of the sample, the height of the 2D-EACVF peaks provides information on the repetitivity abundance, i.e., the number of repeated interdistances and/or the intensity of the repeated spots. In particular, the 2D-EACVF plot, as a fingerprint of the whole map, can be helpful for protein expression profiling and for comparing maps in order to determine qualitative (appearance or disappearance of spots) and quantitative (intensity of those spots) differences in protein expression.

These tools are general and can be also applied for understanding the system molecular complexity in the growing area of multidimensional identification technologies, such as in the multidimensional separation science and technology.

Acknowledgements

This work was supported by the Italian University and Scientific Research Ministry (grants #2001033797_001, #2003039537_005; FIRB 2001, Cod. RBNE01KJHT; PRIN 40%, 2003), Rome; by the University of Ferrara and by CariVerona-2004.

Azzurra Tosi is gratefully acknowledged for map simulation and processing.

References

[1] M.R. Wilkins, K.L. Williams, D.F. Hochstrasser (Eds.), Proteome Research: New Frontiers in Functional Genomics, Springer, New York, 1997.
[2] P.G. Righetti, A. Stoyanov, M.Y. Zhukov, The Proteome revisited: Theory and Practice of All Relevant Electrophoretic Steps, Elsevier, Amsterdam, 2001, p. 275.
[3] H. Hamdan, P.G. Righetti, Proteomics Today, Wiley, Hoboken, 2005.
[4] A. Vlahou, M. Fountoulakis, J. Chromatogr. B 814 (2005) 11.
[5] R.S. Morrison, Y. Kinoshita, M.D. Johnson, T. Uo, J.Y. HoMcBee, T.P. Conrads, T.D. Veenstra, Mol. Cell. Proteomics 25 (2002) 553.
[6] D. Chelius, T. Zhang, G. Wang, R.-F. Shen, Anal. Chem. 75 (2003) 6658.
[7] J.C. Giddings, in: H.J. Cortes (Ed.), Multidimensional Chromatography, Marcel Dekker, New York, 1990.
[8] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 9390.
[9] P. Lescuyer, D.F. Hochstrasser, J.-C. Sanchez, Electrophoresis 25 (2004) 1125.

[10] C.A. Orengo, D.T. Jones, J.M. Thorton, Bioinformatics. Genes, Proteins and Computers, BIOS, Oxford, UK, 2003, p. 245.
[11] M. Randic, J. Chem. Inferface Comput. Sci. 41 (2001) 1339.
[12] N. Farriol-Mathis, J.S. Garavelli, B. Boeckmann, S. Duvaud, E. Gasteiger, E. Gateau, A.-L. Veuthey, A. Bairoch, Proteomics 4 (2004) 1537.
[13] E. Marengo, E. Robotti, P.G. Righetti, F. Antonucci, J. Chromatogr. A 1004 (2003) 13.
[14] E. Marengo, R. Leardi, E. Robotti, P.G. Righetti, F. Antonucci, D. Cecconi, J. Proteome Res. 2 (2003) 351.
[15] E. Marengo, E. Robotti, F. Antonucci, D. Cecconi, N. Campostrini, P.G. Righetti, Proteomics 5 (2005) 654.
[16] R. Stanislaus, L.H. Jiang, M. Swartz, J. Arthur, J.S. Almeida, Bioinformatics 5 (2004) 9.
[17] R. De Knikker, Y. Guo, J.-I. Li, A.K.H. Kwan, K.Y. Yip, D.W. Cheung, K.-H. Cheung, Bioinformatics 5 (2004) 25.
[18] M. Vracko, S.C. Basak, Chem. Int. Lab. Syst. 70 (2004) 33.
[19] Melanie II, Geneva BioInformatics, GeneBio S.A., http://www.genebio.com.
[20] PD Quest, Bio-Rad Laboratories Inc., http://www.biorad.com/.
[21] M.C. Pietrogrande, N. Marchetti, F. Dondi, P.G. Righetti, Electrophoresis 23 (2002) 283.
[22] http://www.expasy.ch/.
[23] M.C. Pietrogrande, N. Marchetti, F. Dondi, P.G. Righetti, Electrophoresis 24 (2003) 217.
[24] N. Campostrini, L.B. Areces, J. Rappsilbe, M.C. Pietrogrande, F. Dondi, F. Pastorino, M. Ponzoni, P.G. Righetti, Proteomics 5 (2005) 2385.
[25] N. Marchetti, A. Felinger, L. Pasti, M.C. Pietrogrande, F. Dondi, Anal. Chem. 76 (2004) 3055.
[26] M.C. Pietrogrande, N. Marchetti, F. Dondi, P.G. Righetti, Electrophoresis 26 (2005) 2739.
[27] M.C. Pietrogrande, F. Dondi, A. Felinger, J.M. Davis, Chem. Int. Lab. Syst. 28 (1995) 239.
[28] F. Dondi, A. Bassi, A. Cavazzini, M.C. Pietrogrande, Anal. Chem. 70 (1998) 766.
[29] M.C. Pietrogrande, A. Cavazzini, F. Dondi, Rev. Anal. Chem. 19 (2000) 123.
[30] A. Felinger, M.C. Pietrogrande, Anal. Chem. 73 (2001) 619A.
[31] A. Felinger, L. Pasti, F. Dondi, Anal. Chem. 62 (1990) 1846.
[32] F. Dondi, A. Betti, L. Pasti, M.C. Pietrogrande, A. Felinger, Anal. Chem. 65 (1993) 2209.
[33] M.C. Pietrogrande, L. Pasti, F. Dondi, M.H. Bollain Rodriguez, M.A. Carro Diaz, J. High Resol. Chromatogr. 17 (1994) 839.
[34] M.C. Pietrogrande, F. Dondi, A. Felinger, J. High Resol. Chromatogr. 19 (1996) 328.
[35] F. Dondi, M.C. Pietrogrande, A. Felinger, Chromatographia 45 (1997) 435.
[36] A. Felinger, Data Analysis and Signal Processing in Chromatography, Elsevier, Amsterdam, 1998.
[37] M.C. Pietrogrande, P. Coll, R. Sternberg, C. Szopa, R. Navarro-Gonzalez, C. Vidal-Majar, F. Dondi, J. Chromatogr. A 939 (2001) 69.
[38] M.C. Pietrogrande, P. Coll, R. Sternberg, C. Szopa, R. Navarro-Gonzalez, C. Vidal-Majar, F. Dondi, J. Sep. Sci. 26 (2003) 569.
[39] M.C. Pietrogrande, I. Tellini, L. Pasti, F. Dondi, C. Szopa, R. Sternberg, C. Vidal-Madjar, J. Chromatogr. A 1002 (2003) 179.
[40] C. Szopa, M. de Pra, I. Tellini, R. Sternberg, M.C. Pietrogrande, C. Vidal-Madjar, F. Raulin, J. Sep. Sci. 27 (2004) 495.
[41] M.C. Pietrogrande, M.G. Zampolli, F. Dondi, Ann. di Chim. 94 (2004) 721.
[42] M.C. Pietrogrande, M.G. Zampolli, F. Dondi, C. Szopa, R. Sternberg, A. Buch, F. Raulin, J. Chromatogr. A 1071 (2005) 255.
[43] J.M. Davis, J.C. Giddings, J. Chromatogr. 289 (1984) 277.
[44] J.M. Davis, Anal. Chem. 63 (1991) 2141.
[45] J.M. Davis, Anal. Chem. 65 (1993) 2014.
[46] J.M. Davis, Chromatographia 42 (1996) 367.
[47] H. Sariouglu, F. Lottspeich, T. Walk, G. Jung, C. Eckerskorn, Electrophoresis 21 (2000) 2209.
[48] A.A. Alaiya, U.J. Roblick, B. Franzen, H.-P. Bruch, G. Auer, J. Chromatogr. B 787 (2003) 207.
[49] M. Mann, O.L. Jensen, Nat. Biotechnol. 21 (2003) 255.
[50] N. Farriol-Mathis, J.S. Garavelli, B. Boeckmann, S. Duvaud, E. Gasteiger, E. Gateau, A.-L. Veuthey, A. Bairoch, Proteomics 4 (2004) 1537.
[51] P. Lescuyer, D.F. Hochstrasser, J.-C. Sanchez, Electrophoresis 25 (2004) 1125.
[52] N. Blom, T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft, S. Brunak, Proteomics 4 (2004) 1633.
[53] A.J. Forbes, S.M. Patrie, G.K. Taylor, Y.-B. Kim, L. Jiang, N.L. Kelleher, Proc. Natl. Acad. Sci. U.S.A. 9 (2004) 2678.
[54] G.T. Cantin, J.R. Yates III, J. Chromatogr. A 1053 (2004) 7.
[55] J. Zhang, H. Hu, M. Gao, P. Yang, X. Zhang, Electrophoresis 25 (2004) 2374.
[56] W.H. Press, S.A. Teukosky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in Fortran, Cambridge, UK, 1986.
[57] J.-C. Sanchez, R.D. Appel, O. Golaz, C. Pasquali, F. Ravier, A. Bairoch, D.F. Hochstrasser, Electrophoresis 16 (1995) 1131.
[58] J.-C. Sanchez, P. Wirth, S. Jaccoud, R.D. Appel, C. Sarto, M.R. Wilkins, D.F. Hochstrasser, Electrophoresis 18 (1997) 638.
[59] I. Demalte-Annessi, J.-C. Sanchez, C. Hoogland, V. Rouge, P.-A. Binz, R.D. Appel, D.F. Hochstrasser, SWISS-2DPAGE database (1999).
[60] N.G. Anderson, N.L. Anderson, Clin. Chem. 28 (1982) 739.
[61] M.-C. Blatter, R.W. James, S. Messner, F. Barja, D. Pometta, Eur. J. Biochem. 211 (1993) 871.
[62] C. Eckerskorn, K. Strupat, D. Schleuder, D.F. Hochstrasser, J.-C. Sanchez, F. Lottspeich, F. Hillenkamp, Anal. Chem. 69 (1997) 2888.