This is the peer reviewd version of the followng article:

A Novel Gaussian Extrapolation Approach for 2D Gel Electrophoresis Saturated Protein Spots / Natale, M., Caiazzo, A., Bucci, E.M., Ficarra, E.. - In: GENOMICS, PROTEOMICS & BIOINFORMATICS. - ISSN 1672-0229. - 10:6(2012), pp. 336-344. [10.1016/j.gpb.2012.06.005]

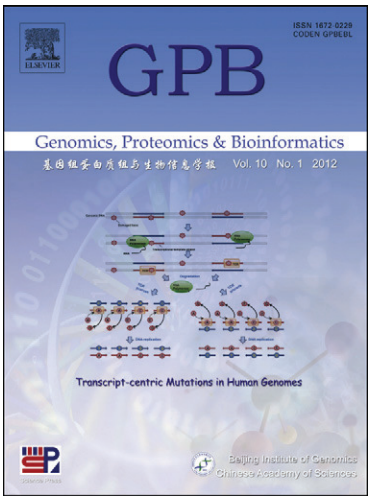
Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

11/10/2026 13:49

(Article begins on next page)

Accepted Manuscript



A Novel Gaussian Extrapolation Approach for 2D Gel Electrophoresis Saturated Protein Spots

Massim Natale, Alfonso Caiazzo, Enrico M. Bucci, Elisa Ficarra

PII: S1672-0229(12)00077-0
DOI: http://dx.doi.org/10.1016/j.gpb.2012.06.005
Reference: GPB 34

To appear in:

Received Date: 23 April 2012
Revised Date: 3 May 2012
Accepted Date: 4 June 2012

Please cite this article as: M. Natale, A. Caiazzo, E.M. Bucci, E. Ficarra, A Novel Gaussian Extrapolation Approach for 2D Gel Electrophoresis Saturated Protein Spots, (2012), doi: http://dx.doi.org/10.1016/j.gpb.2012.06.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

A Novel Gaussian Extrapolation Approach for 2D Gel Electrophoresis Saturated Protein Spots

Massimo Natale[1,2,†,*], Alfonso Caiazzo[3,†], Enrico M. Bucci[2,4], Elisa Ficarra[1]

[1]*Department of Control and Computer Engineering, Politecnico di Torino, Torino 10129, Italy*

[2]*BioDigitalValley S.r.l., Pont Saint Martin 11026, Italy*

[3]*WIAS Berlin,Berlin 10997, Germany*

[4]*Istituto di Biostrutture e Bioimmagini, Naples 80134, Italy*

† Equal contribution.

*Corresponding author.

E-mail: massimo.natale@biodigitalvalley.com

Running title: *Natale M et al / A Novel Gaussian Extrapolation for 2D-GE Saturated Spots*

Abstract

Analysis of images obtained from two-dimensional gel electrophoresis (2D-GE) is a topic of utmost importance in bioinformatics research, since commercial and academic software available currently has proven to be neither completely effective nor fully automatic, often requiring manual revision and refinement of computer generated matches. In this work, we present an effective technique for the detection and the reconstruction of over-saturated protein spots. Firstly, the algorithm reveals overexposed areas, where spots may be truncated, and plateau regions caused by smeared and overlapping spots. Next, it reconstructs the correct distribution of pixel values in these overexposed areas and plateau regions, using a two-dimensional least-squares fitting based on a generalized Gaussian distribution. Pixel correction in saturated and smeared spots allows more accurate quantification, providing more reliable image analysis results. The method is validated for processing highly exposed 2D-GE images, comparing reconstructed spots with the corresponding non-saturated image, demonstrating that the algorithm enables correct spot quantification.

Keywords: Image analysis; Two-dimensional gel electrophoresis; Proteomics; Software tools

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Introduction

Since the pioneer work of O'Farrel [1], two-dimensional gel electrophoresis (2D-GE) has been demonstrated to be the most comprehensive technique for the analysis of proteome, allowing the simultaneous analysis of very large sets of gene products. In the post-genomic era, 2D-GE becomes a powerful tool that is widely used for the analysis of complex protein mixtures extracted from cells, tissues, or other biological samples [2].

In 2D-GE, proteins are visualized on a 2D plane, according to their charge (pI) by isoelectric focusing (on the first coordinate axis) and to their molecular weight (MW) by SDS-PAGE (on the second coordinate axis). As a result, each 2D-GE image contains a few hundred up to several thousand protein spots, whose volume is correlated with the protein expression in the sample. The main goal of comparative proteomics is to match protein spots between gels and define differences in the expression level of proteins in different biological states (*e.g.*, healthy versus diseased, or control versus treated).

An important prerequisite to guarantee optimal performance of automated image analysis packages is good image capture, which should allow for detection of various protein amounts ranging from very low to high abundance. In particular, when a gel image is acquired, it is important to avoid saturation effects, which occur when grey levels exceed the maximum representable level on the used scale (*e.g.*, in an 8-bit image an overexposed or saturated area contains a large number of pixels with gray-level values close to 256).

In practice, to avoid saturation during the scanning procedure, the operators should assure that the more abundant protein spots are represented by pixels slightly below the maximum intensity available, while keeping the dynamic range as wide as possible and acquiring images at 16-bit rather than 8-bit, because a 16-bit image has 65,536 gray levels while a 8-bit has only 256 gray levels.

The use of fluorescent protein stains can extend the linear dynamic range by 3-5 orders of magnitude [3]. However, the amplitude of the dynamic range is still a relevant issue in biological applications, since protein concentrations in biological systems may vary by seven or more orders of magnitude [3]. Currently, there is no acquisition system that can cover this wide range. In practice, the differences in protein concentration in biological systems compromise the weak and low abundant protein spots, which are often more biologically relevant, but generally too faint to be accurately quantified [4].

The intensity of a pixel is expressed within a given range, whose depth indicates how accurately the grey value can be represented. 2D-GE images are usually acquired at 16 bit resolution in the grey scale, resulting in maximum grey values of 2^{16}. The bit depth together with the type of staining influences the linear dynamic range. Some of the most widely-used staining methods are summarised in **Table 1**. The linear dynamic range refers to the number of orders of magnitude where the grey scale values increase proportionally with protein amount. In other words, the orders of magnitude define the volume difference between smallest and largest spots. For example, 1 order of magnitude

ACCEPTED MANUSCRIPT

means a difference from 1 to 10, while 5 orders from 1 to 100,000. The linear dynamic range is necessary to examine these 2D-GE spots containing varied concentrations of the target proteins, so that the correlation between protein concentration and spot volume can be generated and evaluated.

Moreover, due to the highly complex nature of biological samples, in most cases several proteins have similar pI and MW values, and the corresponding spots migrate to the same portion of the gel image, resulting in complex regions with multiple overlapping spots. In this context, commercial software available currently often requires manual revision of spot detection and refinement of computer generated matches [5-7].

The aim of the spot detection step is to define the location, the true boundary and the intensity (usually the total pixel volume within the boundary) of each spot. In particular, in order to achieve fast and reliable image analysis, the algorithm used for image segmentation must provide well-defined and reproducible spot detection, quantification and normalization [8]. However, despite the availability of several spot detection algorithms, the accurate definition of protein spot might be critical in certain regions, where defining the boundaries is challenging due to variable background, spots overlapping and spot saturation [9-11].

In a saturated spot, intensity values are truncated, preventing the resolution of high intensity pixel, since the missing area of the spot will not be measured. In other words, a saturated spot does not provide any reliable quantitative data, and it might also compromise the overall normalization. In particular, when comparing different experimental states for expression changes representative of a particular condition, the inclusion of highly to medium abundant and saturated spots might bias normalization, especially if their variance is a significant portion of the total spot volume [12]. For these reasons, several authors recommend manually deleting the saturated spots, before analysing the gels [13]. In fact, currently available commercial software (as Delta2D, ImageMaster, Melanie, PDQuest, Progenesis and REDFIN) or academic packages (as Pinnacle and Flicker) are not able to deal with specific protein spot distortions found in the gel images [14].

In a recently-proposed approach to improve spot detection, Kim and Yoon [15] noticed that the gradients of the valley under a plateau spot orient the peak. They therefore suggested a spot separation method that computes the accumulated gradient of each point in the potential complex regions. The accumulated gradient image is then segmented with a watershed segmentation algorithm. Using this method, they can detect 75%-96% over-saturated protein spots. However, this method is not able to provide quantitative information.

Since automatic detection of oversaturated protein spots is, to date, infeasible, software developers suggest recovering missing data by rescanning the gel at a lower exposure (http://www.nonlinear.com/support/progenesis/samespots/faq/saturation.aspx). Although theoretically possible, it must be noted that this is often prohibitive, since gel staining is photosensitive and colors fade away in a few minutes upon light exposure, or because image analysis is performed a long time after the acquisition of the gels so that the expensive 2DE procedure has to be repeated for another scanning [15]. Moreover, acquiring images with long exposure might provide

ACCEPTED MANUSCRIPT

more information, since the acquired images contain a larger number of spots and thus also the lower-abundance protein spots.

In this work, we present a novel two-step algorithm for detection and reconstruction of over-saturated protein spots. Firstly, it reveals overexposed areas, where spots may be truncated, and plateau regions caused by smeared and overlapping spots. Secondly, it processes the greyscale distribution of saturated spots, reconstructing the saturated region by a generalized Gaussian approximation. The method yields a highly-refined and fully-automatic spot detection that does not need further manual corrections. Furthermore, the pixel correction in saturated and smeared spots according to the generalized Gaussian curve allows more accurate quantification, providing more reliable results from 2D-GE analysis. To validate the proposed method, we processed a highly-exposed 2D-GE image containing saturated spots and compared the reconstructed spots to the corresponding spots of the unsaturated image. Results indicate that the method can effectively reconstruct the saturated spots and improve spot quantification.

Results

In order to test the algorithm for localizing plateau areas, we examined 12 2D-GE gels with images acquired at 3 different exposures for each gel (36 images in total). An example of the plateau areas detected using our algorithm is shown in **Figure 1**. In this case, only the image in Figure 1C (acquired at the lowest exposure) could be properly analysed by commercial software available, while the other two images would be discarded because of the large saturated areas. However, in several cases researchers would be interested in processing also the image acquired with the highest exposure (Figure 1A), since it contains the highest number of spots (and thus also the lower-abundance protein spots). It must be said that, in practice, operators only acquire a single image per gel, choosing then the image that can give the largest amount of useful information. In the example considered here, image shown in Figure 1A provide more information due to more spots revealed. However, without an extrapolation approach, this image would be discarded due to the higher number of saturated spots. We therefore used Melanie (GeneBio - Geneva) to perform the spot detection on these three gel images and detected 254, 213 and 172 spots in images shown in Figure 1A, B and C, respectively. Hence, being able to analyse the higher exposure images would yield 30% more spots, which represent low abundant proteins and relatively faint spots.

As mentioned earlier, the commercial software is not able to detect or report saturated areas and, furthermore, does not warn the operator of the presence of saturated regions. In fact, saturated spots can only be viewed using the 3D layout, while it would be desirable that the 2D-GE could be detected automatically by the software and the saturated spot were reported before performing an analysis. The proposed method is valid both for over-stained spots and for highly-abundant protein spots. An example of a highly-abundant protein spot is provided in Figure 1, where albumin is the largest spot in the middle of the images. The albumin spots are saturated in all the three images (A, B and C), and our algorithm is able to detect these areas correctly. In Figure 1, we showed that the same gel can be

ACCEPTED MANUSCRIPT

analyzed with or without saturated spots, by acquiring the images at different exposures. These images provide us with the opportunity to see how the grey values are distributed in the same spot.

The effect of overexposure on a single spot is shown in detail in **Figure 2**, which refers to the saturated spot contained in the blue box in Figure 1. In particular, Figure 2A-C show the spot acquired at three different exposure conditions from a greater to lower opening. From the 3D plot of the spot (Figure 2D-F), one can identify a plateau zone for the higher exposures (Figure 2D and 2E), where the red color denotes the maximum possible value of the grey scale.

Finally, Figure 2G-I show the result of the plateau-area finder (red contour) within the spot detection result using a watershed algorithm (blue contour). In particular, we observe that the spot in Figure 2G has a larger area, but also a wide plateau area. The effect of overexposure on spot volume quantification is shown in **Figure 3**, where the spot intensity distributions are truncated (Figure 3A and 3B).

Before adopting the Gaussian extrapolation to reconstruct the saturated spots, we validated the fitting model on unsaturated spots. As described by Matuzevičius et al.[17] , we found a significant correlation between our mathematical model and value distribution for unsaturated spots (Figure 4A and 4C), with an error in volume estimation below 5%. Two reconstructed spots were shown to validate the reconstruction properties of the Gaussian extrapolation method (Figure 4B and 4D).

Next, to evaluate the effect of saturation on a quantitative proteomic analysis, we perform 2D-GE analysis on 12 gels with images acquired at three different exposures (36 images in total). The image analysis was performed using the method describe by Natale et al. [18]. In most cases saturated spots are found with the maximum or medium exposure. The goal of this analysis was to quantify how the spot volume increases with the three different exposures (low, medium and high). We hypothesized that the volume of the same spot grows linearly with increasing exposure, but with a different proportion depending on whether the spot was saturated or not. We performed spot detection on the 12 images at higher exposure and then exported all the spots acquired from the images at high exposure to those from the medium and low exposures. We chose the high exposure because of the highest number of spots detected. From all images, once matched, we extracted the raw volume of the four spots shown in Figure 3 (two small, one medium and one big spot). In order to compare the spots among the different gels, we normalized raw volumes by dividing each spot volume in the medium and higher exposure by the spot volume detected in the gel at the lower exposure.

In **Table 2** we reported the variation in the volume increase of the saturated and unsaturated spots at different exposures. At the low exposure, all spots have been normalized with respect to themselves (100%). At the maximum exposure, volume of unsaturated spots was increased by more than two times (239%), whereas the volume of saturated spot was increased only 1.5 times (163%), thus losing a considerable amount of information. This problem, if not reported, could lead to errors in statistical analysis, producing inconsistent results.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

The method developed in this work allows us to automatically identify the saturated pixels, and recalculate the correct distribution of grey values. In Figure 4B we showed a saturated spot, while in Figure 4D the same spot was shown after applying the algorithm of Gaussian extrapolation.

Conclusion

The method introduced in this paper allows us to extend the linear dynamic range acting on the bit depth. When the largest spots are saturated because they reach the maximum grey value permitted by the bit resolution, the algorithm is able to reconstruct the saturated spots, recalculating a grey value of individual pixels well above 65,536. In particular, we demonstrated that our algorithm can successfully reconstruct the spot volume up to 2-3 times the maximum grey value. Therefore the values of the pixels of the reconstructed images should be stored as a values array. Of course, the linear dynamic range is strongly influenced by the type of staining used and by the acquisition tool.

Saturation of abundant spots is an important issue in 2-DE evaluation, in particular when dealing with complex samples like serum or plasma, where protein distribution might be highly heterogeneous. Since commercial software currently available is not yet able to perform 2D-GE analysis in the presence of saturated spots, the only alternatives are based on gel image acquisition or gel rescanning with different parameters.

We presented a new approach for the treatment of over-saturated protein spots in 2D-GE, which combines a spot detection algorithm to find the oversaturated spots and a Gaussian extrapolation approach to approximate the unsaturated spot volume (and thus the correct protein expression). The proposed method has been validated by comparing the reconstructed spots in the same gel acquired at different exposures, showing that the method allows us to automatically extract the pixel data within the spot boundaries, and in the presence of a plateau area, recalculate the correct distribution of grey values.

Methods

In order to collect a set of images useful to test the algorithm, we run 10 plasma samples using 2D-GE. 2D-GE was performed according to Maresca et al. [19], with each sample run in duplicate. For the first-dimension of electrophoresis, 200µg of plasma protein samples (approximately 3µl) were applied to 18-cm linear IPG strips 4-7 (GE Healthcare, Uppsala, Sweden) and focused until 72000 V/hr were reached. Prior to SDS-PAGE, the IPG strips were equilibrated twice for 15min in equilibration buffer (50 mM Tris-HCl pH 8.8, 6 M urea, 30% glycerol(v/v), 2% SDS(w/v) and traces of bromophenol blue) containing 1% DTT (w/v)for the first equilibration step and 2.5%iodoacetamide(w/v) for the second step, respectively. SDS-PAGE was performed on 12.5% polyacrylamide gels at 60 mA/gel at 16°C and terminated when the dye front reached the lower end of the gel. After SDS-PAGE, the gels were stained overnight with RuBPs in fixing solution (30% ethanol and 10% acetic acid) according to the manufacturer's instructions. The gel images were acquired by

ACCEPTED MANUSCRIPT

Gel-Doc-it 310 Imaging system (UVP, Upland, CA), with the following setting— exposure 7 s, gain 3 and three different apertures: 6, 7 and 8. Aperture is expressed as F-stop (*e.g.*, F2.8 or f/2.8): the smaller is the F-stop number (or f/value), the larger is the aperture.

Using a larger exposure yields a larger number of spots and, at the same time, more saturated areas. In our case, only the image acquired at the lower exposure (Figure 1C) could be properly analysed by commercial and academic software currently available, while the saturation in some areas in the other two images (Figure 1A and B) prevented accurate evaluation of protein expression. Using the algorithm described in this work, we have been able to approximate the intensity values of the non-saturated spots within the saturated area.

Despeckling

After image scanning, the 2D-GE images are noisy and usually affected by several artefacts such as scratches, air bubbles and spikes. Scratches and air bubbles are unintentionally introduced by the operator, while spikes are caused by the presence of precipitated staining particles or imperfect polyacrylamide gel matrix. In particular, spikes are a sort of impulsive noise because of the small area of high intensity.

The aim of reducing speckles in a 2D-GE image (despeckling) is to remove the noise without introducing any significant distortion in quantitative spot volume data. Noise suppression methods employed in commercial software packages for 2D-GE image analysis are based on spatial filtering, particularly the most widely used techniques are based on local median filter [20].

Although several other methods have been developed [21], the advantage of using the median filter is the easy application, since it normally requires the setting of only one main parameter, the filter kernel, which is generally related to the noise size [21]. At the same time, the size of the filter matrix also accounts for the major impact exerted on the resulting image. We evaluated the performance of filter size on a set of test images from 3×3 to 7×7 and observed remarkable deterioration for larger filter size, due to the decreased the contrast of the image. Therefore, a 3x3filter was chosen, which was capable of removing most of the spikes, while preserving the shape of the spots. The result of a 3×3 median filter on a sample image is shown in **Figure 5**. A portion of gel with spikes in 2D and 3D view was shown in Figure 5A and 5C, respectively. The results of a 3x3 median filter are shown in Figures 5B and 5D accordingly.

Identification of plateau regions

Saturation effects in 2D gel images could occur for several reasons: i) intense staining, causing complete absorption of light passing through the spot, ii) long exposure time during the acquisition step, iii) too high protein concentration, or iv) accumulation of pigment on top of high intensity spots, leading to the loss of dynamic linear range. Saturation effects might also result from manipulation of image data with image processing software that is not designed for quantitative image analysis (*e.g.*,

ACCEPTED MANUSCRIPT

Photoshop). For instance, visually-enhanced images often show plateau areas in the high and low intensity regions [13].

The commercial software is able to detect a saturated area when the intensity reaches the maximum value of grey-scale. However, in most cases saturated areas are very similar to plateau zones, *i.e.*, where the pixels have similar intensity without reaching the maximum value of grey-scale. In these cases, the software might also fail to detect the plateau regions, causing the operator to underestimate the problem.

In order to identify the plateau regions, we implemented a morphological filter, inspired by the rolling-ball algorithm described by Sternberg [22], which allows segmentation of the plateau zones. This method is based on a structural element (SE) defined by a circle of given radius (RD) and a grey-scale value tolerance (GVT). In particular, for each pixel (x,y) of the image I, the SE is defined as a circular neighbourhood of RD:

$$SE = \{(s,t) \text{ in } I \mid \sqrt{(s-x)^2 + (t-y)^2} \leq RD \} \quad (1a)$$

where

$$I = [0, image\ width] \times [0, image\ height] \quad (1b)$$

denotes the spatial domain of the image.

The SE is depicted in **Figure 6A**. The *RD* and the *GVT* are defined by a single parameter. The *GVT* represents the rate of grey values of the pixel in the centre of the SEs (Figure 6B). For instance, setting the parameter to 10, the *RD* is 10 pixels and the *GVT* in the 10% of grey values of the pixel at position (x,y).

The centre of SE is moved along each pixel of the image and the maximum and minimum grey values of the pixels for each point (x,y) within the given *RD* are calculated. When the difference between maximum and minimum is less than *GVT*, the area defined by the local operator is considered as a plateau area.

It is worth noting that, when the value of the parameter is low, the RD of the ball is small and only few pixels are included in SE. We examined the performance with RD values from 10 to 25 and for this analysis we selected a value of 15.

Detection of saturated spots

After localizing the plateau regions, we segmented the image to identify the isolated protein spots containing plateau areas on each gel image.The segmentation procedure yields a set of image segments, consisting of connected neighbouring pixels enclosed by a spot boundary. Each image segment represents the spot of one single isolated protein.

A watershed algorithm was used in this segmentation step. In this approach, the image is considered as a *landscape*, and the segmentation is to identify all the local minima in the landscape and then find the *catchment basins* associated with each local minima. The boundary between several

ACCEPTED MANUSCRIPT

catchment basins is called a watershed [22]. For 2D-GE each separated catchment basin is considered as an isolated protein spot. Unlike the usual watershed segmentation, only the plateau areas in our method have been assigned to a catchment basin.

Finally, the grey values of the pixels inside the spot, excluding the region identified as a plateau area, were used in the Gaussian extrapolation step to recover the distribution of the unsaturated spot.

Gaussian extrapolation

The final step consists of reconstructing the saturated spots resulting from high exposure images and approximating the unknown grey values in the plateau region. This has been done considering the unsaturated spot to be described by an analytical function, depending on a restricted set of parameters. In particular, we assumed each cross section of the spot intensity along both vertical and horizontal axes to be approximated by a generalized Gaussian distribution. Namely, for each value of the Y-coordinates we considered a function of the form:

$$f(x, M(y), \sigma(y), x_0, b) = \frac{M(y)}{\sigma(y)} \exp\left(-\frac{|x - x_0(y)|^b}{b\sigma(y)^b}\right) \quad (2)$$

For b = 2, equation (2) defines the kernel of a standard Gaussian distribution centred in $x_0(y)$, where $\sigma(y)$ and $M(y)$ is the standard deviation and the maximum of intensity values, respectively. Note that, unlike the other parameters, *b* does not depend on y, assuming that the approximating Gaussian can have different maximum, center and variance in different sections.

The reconstruction problem can be formulated as follows. Given

$$\{(x_i, y_j), i = 1, \ldots, N_x, j = 1, \ldots, N_y\} \quad (3)$$

and the corresponding intensities $\{I_{ij}\}$, we determine the set of parameters (M, σ, x_0, b) for which the function defined in (2) fits at best the values of the intensity in the unsaturated region. In practice, we have to minimize an error function that defines how good a particular parameter set is. For example, a standard least-squared criterion can be used

$$E(M(y_j), \sigma(y_j), x_0(y_j), b) = \sum_{i=1}^{N_x} \| f(x_i, M(y_j), \sigma(y_j), x_0(y_j), b) - I_{ij} \|^2 \quad (4)$$

However, the error measure can be defined in different ways, *e.g.*, according to different norms or including different weighs for the different parameters and/or the different pixels. In particular, we modified (4) in order to control the variation of the parameters for different sections. In our case the error function (5) can be formulated as follows

$$E(M(y_j), \sigma(y_j), x_0(y_j), b) = E(M(y_j), \sigma(y_j), x_0(y_j), b) + \left(M(y_j) - M(y_{j-1})\right)^2 + \ldots (\sigma(\quad (5)$$

ACCEPTED MANUSCRIPT

For positive values of the parameters $(\theta_M, \theta_{cv}, \theta_{N_0})$, the problem is then reduced to finding the parameters yielding the minimum of the selected error function. One possibility is to perform an exhaustive search on all the values of a pre-defined parameters space. However, if the size of the parameter space is large, a more effective Newton-Raphson algorithm to find the zero of the gradient could be employed.

Authors' contributions

MN conceived the study, developed the algorithm for the detection of overexposed areas and plateau regions. AC designed the mathematical model for the reconstruction of over-saturated protein spots. EMB participated in study design and performed the 2D-GE analysis. EF participated in study design and coordination. MN and AC drafted the manuscript with the help of EF. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

Acknowledgements

We extend thanks to Prof. Mauro Fasano (University of Insubria) for providing the 2D-GE images. This work was partially funded by the Valle d'Aosta Regional Government (http://www.regione.vda.it/) in the frame of the regional law n.84-07/12/1993 (project ParIS - Parkinson Informative System).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

[1] O'Farrel PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem 1975;250:4007-21.

[2] Görg A, Weiss W, Dunn MJ.Current two-dimensional electrophoresis technology for proteomics. Proteomics 2004;4:3665-85.

[3] Miller I, Crawford J, Gianazza E. Protein stains for proteomic applications: which, when, why? Proteomics 2006;6: 5385-408.

[4] Hortin GL, Sviridov D.The dynamic range problem in the analysis of the plasma proteome. J Proteomics 2010;73:629-36.

[5] Millioni R, Puricelli L, Sbrignadello S, Iori E, Murphy E, Tessari P. Operator- and software-related post-experimental variability and source of error in 2-DE analysis. Amino Acids 2012;42:1583-90.

[6] Clark BN, Gutstein HB. The myth of automated, high-throughput two-dimensional gel analysis. Proteomics 2008;8:1197–1203.

[7] Wheelock ÅM, Buckpitt AR,.Software-induced variance in two-dimensional gel electrophoresis image analysis.Electrophoresis2005;26:4508–20.

[8] dos Anjos A, Møller AL, Ersbøll BK, Finnie C, Shahbazkia HR. New approach for segmentation and quantification of two-dimensional gel electrophoresis images. Bioinformatics 2011;27:368-75.

[9] Srinark T, Kambhamettu C. An image analysis suite for spot detection and spot matching in two-dimensional electrophoresis gels. Electrophoresis 2008;29:706–15.

[10] Rashwan S, Faheem T, Sarhan A, Youssef BAB. A fuzzy-watershed based algorithm for protein spot detection in 2DGE images. Proceedings of the 9th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision (ISCGAV'09) 2009:35-40. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA

[11] Daszykowski M, Bierczynska-Krzysik A, Silberring J, Walczak B. Avoiding spots detection in analysis of electrophoretic gel images. Chemometr Intell Lab Syst 2010;104:2–7.

[12] Miller I, Crawford J, Gianazza E. Protein stains for proteomic applications: Which, when, why? Proteomics 2006;6: 5385–408.

[13] Berth M, Moser FM, Kolbe M, Bernhardt J. The state of the art in the analysis of two-dimensional gel electrophoresis images. Appl Microbiol Biotechnol 2007;76:1223–43.

[14] Maurer MH. Software analysis of two-dimensional electrophoretic gels in proteomic experiments. Curr Bioinform 2006;1:255-62.

[15] Kim MA, Yoon YW. Spot detection of complex regions by accumulated gradient in two dimensional electrophoresis images. Proceedings of the 11th International Conference on Advanced Communication Technology (ICACT'09) 2009;3:1982-5. IEEE Press, Piscataway, NJ, USA.

[16] Serackis A, Navakauskas D. Treatment of over-saturated protein spots in two-dimensional electrophoresis gel images. Informatica 2010; 21:409–24.

[17] Matuzevičius D, Serackis A, Navakauskas D. Mathematical models of oversaturated protein spots. Electron Electr Eng 2007;1:63-8

[18] Natale M, Maresca B, Abrescia P, Bucci EM. Image analysis workflow for 2-D electrophoresis gels based on ImageJ. Proteomics Insights 2011;4:37-49.

[19] Maresca B, Cigliano L, Corsaro MM, Pieretti G, Natale M, Bucci EM, et al. Quantitative determination of haptoglobin glycoform variants in psoriasis. Biol Chem 2010;391:1429-39.

[20] Kaczmarek K, Walczak B, de Jong S, Vandeginste BGM. Preprocessing of two-dimensional gel electrophoresis images. Proteomics 2004;4: 2377–89.

[21] Cannistraci CV, Montevecchi FM, Alessio M. Median-modified Wiener filter provides efficient denoising, preserving spot edge and morphology in 2-DE image processing. Proteomics 2009;9: 4908–19.

[22] Sternberg S. Biomedical image processing. IEEE Computer 1983;16:22-34.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure legends

Figure 1 Saturation zones in 2D-GE images acquired at different exposures

The figure shows images of the same gel acquired at three different exposures as high (**A**), medium (**B**) and low (**C**). The saturation zones detected from our algorithm are indicated in red. Spots boxed in blue were further analyzed in following figures.

Figure 2 Visualization of plateau areas in 3D view and by our algorithm

The spot shown in panel A-C was the 3rd spot in the blue box in Figure 1 with high, medium and low exposure, respectively. **D-F**. The 3D visualization of the spot in panel A-C was shown in D-F, respectively, according to the color scale on the left. Red areas in D and E indicated saturated areas containing pixels with similar intensity (the saturated areas contains pixels with gray values close to 65,536 in 16-bit gray scale). **G-I.** The spots were detected using watershed segmentation, with the spot boundary in blue and the plateau boundary in red as detected by our algorithm. The red areas in D and E coincide with those in panels G and H, respectively.

Figure 3 Visualization of unsaturated and saturated spots in 2D view

Upper panel indicated the regions in the blue box as shown in Figure 1 A-C with high, medium and low exposure. These regions were visualized in 2D view in the lower panels accordingly. Arrows indicated the spots analyzed in Table 2. The profiles in this figure clearly show the Gaussian distribution of unsaturated spots.

Figure 4 Visualization of unsaturated and saturated spots in 3D view before and after reconstruction

The 3D visualization of the unsaturated spot and saturated spot was shown in panel **A** and **B**, respectively. Pixels are shown on X and Y axes and the gray values are shown on Z axis. Shape reconstruction of the unsaturated spot shown in panel A and saturated spot shown in B was shown in panel **C** and **D**, respectively.

Figure 5 Despeckling using median filter

A. Portion of 2D-GE gel where the spikes are marked with black arrows. **B.** Portion of 2D-GE gel after despeckling using 3×3 median filter. 3D view of the portion of gels in A and B was indicated in panel **C** and **D**, respectively. The red arrow in panel C indicates a spike,

ACCEPTED MANUSCRIPT

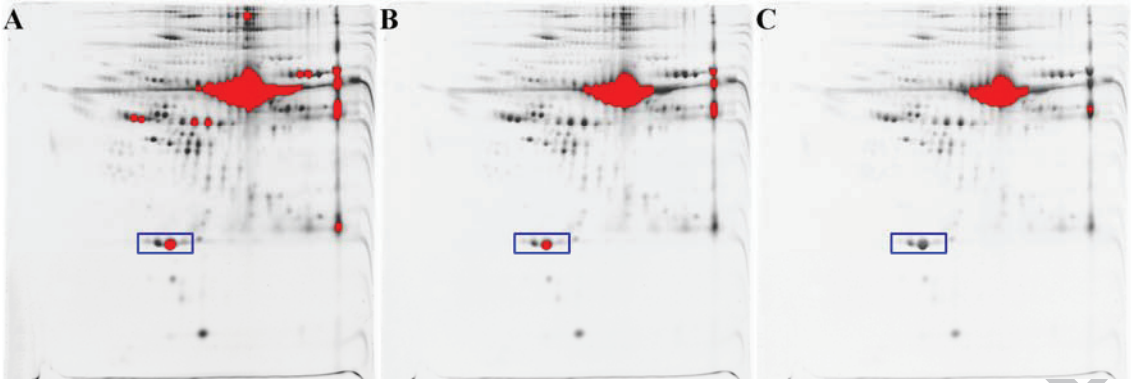
which completely disappeared in panel D after despeckling. The increase in gray values from low to high is shown in color gradient (blue to red).

Figure 6 Structural element used to identify the plateau regions

A. Structural element on a portion of image. Each square represents a pixel and the radius (RD) defines the pixels that are included in the structural element. **B.** The greyscale value tolerance (GVT) represents the difference between the minimum and maximum grey value of the pixels within the structural element. The red spot indicates the projection of the center of the structural element above the portion of image in 3D view.

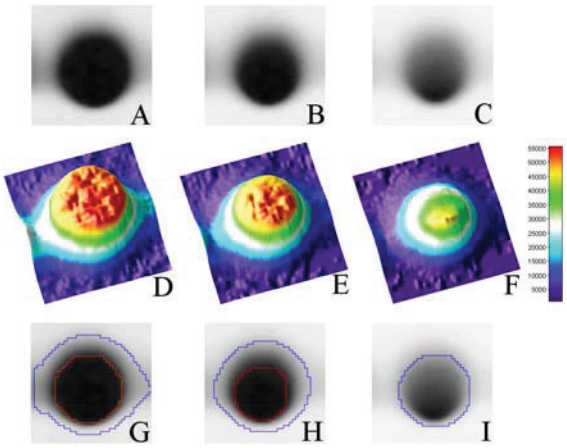
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



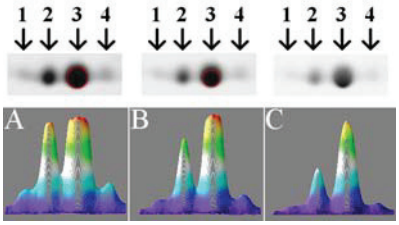
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



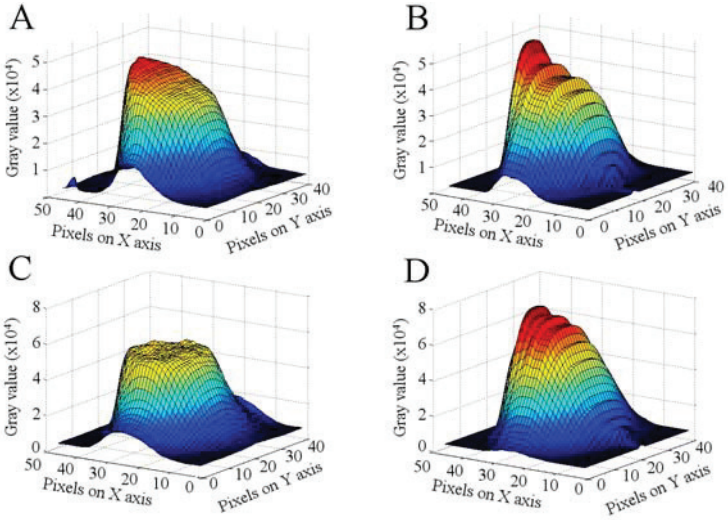
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



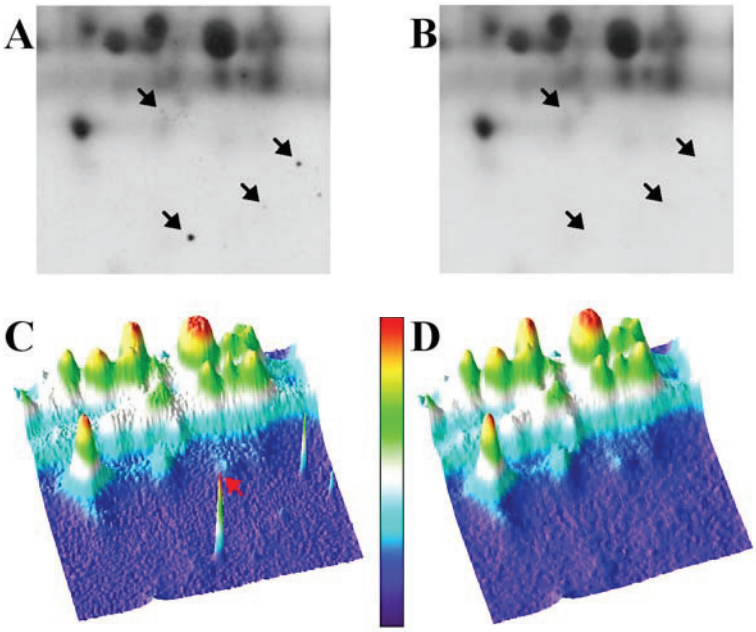
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



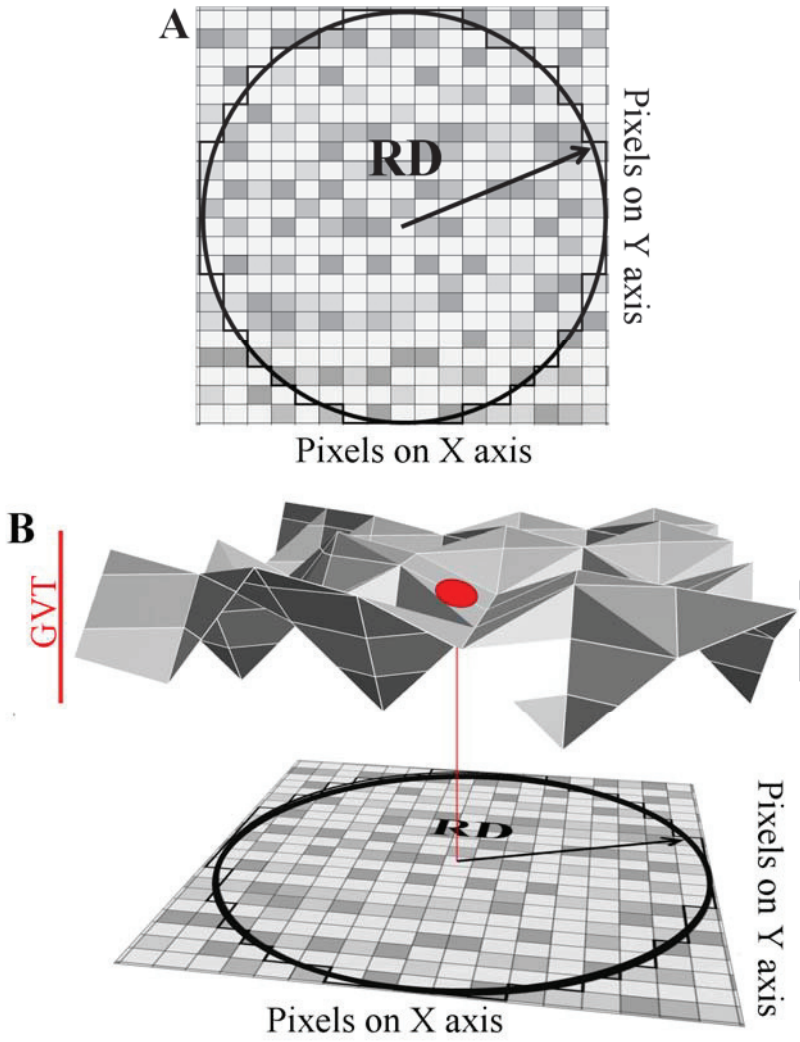
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1 Some of the most widely used staining methods

Stain/Dye	Residues associated	Sensitivity (ng)	Linear dynamic range (order)	MS compatibility
Coomassie Brilliant Blue	Arginine, lysine	8–10	1	Good
SYPRO Ruby	Primary amines	1	3-4	Good
Deep Purple	Primary amines	>1	4	Good
Minimal Cy Dyes	Lysine	0.1–0.2	4-5	Challenging

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2 Variation in the volume increase of the saturated and unsaturated spots at different exposures

	Low exposure	Medium exposure	High exposure
Unsaturated Spots	100%	150%±21%	239%±40%
Saturated Spots	100%	129%±12%	163%±14%

Note: Data was shown as average ± standard deviation.

ACCEPTED MANUSCRIPT