

# Two-dimensional gel electrophoresis (2D-GE) image analysis based on CellProfiler

## *Pseudomonas aeruginosa* AG1 as model

Jose Arturo Molina-Mora, MSc<sup>\*</sup>, Diana Chinchilla-Montero, MSc, Carolina Castro-Peña, MSc, Fernando García, PhD

### Abstract

Two-dimensional gel electrophoresis (2D-GE) is an indispensable technique for the study of proteomes of biological systems, providing an assessment of changes in protein abundance under various experimental conditions. However, due to the complexity of 2D-GE gels, there is no systematic, automatic, and reproducible protocol for image analysis and specific implementations are required for each context. In addition, practically all available solutions are commercial, which implies high cost and little flexibility to modulate the parameters of the algorithms. Using the bacterial strain, *Pseudomonas aeruginosa* AG1 as a model, we obtained images from 2D-GE of periplasmic protein profiles when the strain was exposed to multiple conditions, including antibiotics. Then, we proceeded to implement and evaluate an image analysis protocol with open-source software, CellProfiler. First, a preprocessing step included a bUnwarpJ-Image pipeline for aligning 2D-GE images. Then, using CellProfiler, we standardized two pipelines for spots identification. Total spots recognition was achieved using segmentation by intensity, whose performance was evaluated when compared with a reference protocol. In a second pipeline with the same program, differential identification of spots was addressed when comparing pairs of protein profiles. Due to the characteristics of the programs used, our workflow can automatically analyze a large number of images and it is parallelizable, which is an advantage with respect to other implementations. Finally, we compared six experimental conditions of bacterial strain in the presence or absence of antibiotics, determining protein profiles relationships by applying clustering algorithms PCA (Principal Components Analysis) and HC (Hierarchical Clustering).

**Abbreviations:** 2D-GE = two-dimensional gel electrophoresis, ANOVA = analysis of variance, CIP = Ciprofloxacin, FDR = false discovery rate, HC = hierarchical clustering, IMP = Imipenem, PCA = Principal Component Analysis, pI = isoelectric point, TOB = Tobramycin.

**Keywords:** 2D-GE, bUnwarpJ, CellProfiler, image analysis, proteomics, *Pseudomonas aeruginosa*

## 1. Introduction

Proteomics is a field of study of the omic sciences that focuses on the analysis of the complete set of proteins produced in a cell, tissue, or organism at a given moment, that is, proteomes. The evaluation of protein profiles of biological samples, either by the

presence or absence of proteins, or the measurement of their relative abundance, can help to understand the cellular processes, including associated to pathologies, particular biological conditions or to understand molecular mechanisms of biological relevance.<sup>[1]</sup> However, since cells can produce thousands of proteins, the processing of protein information is complex.

Editor: Duane R. Hospenhal.

Ethics approval and consent to participate: Not Applicable. Since, our study is focused on a bacterial strain, an ethical approval was not necessary, as it was evaluated by the scientific committee of the Research Center in Tropical Diseases (CIET), University of Costa Rica, Costa Rica. Neither humans nor animals were used in this study.

This work was funded by project B7124 Analysis of the extracellular proteome of *Pseudomonas aeruginosa* AG1, registered in the University of Costa Rica (period 2017–2019).

The authors have no conflicts of interest to disclose.

Data availability: Image analysis pipelines for both "Identification of total spots" and "Differential identification of spots" are available at: <https://github.com/josemolina6/2D-GE>. Both need to be opened using CellProfiler software.

The datasets generated during and/or analyzed during the present study are publicly available.

Research Center in Tropical Diseases (CIET), University of Costa Rica, San Pedro, Costa Rica.

\* Correspondence: Jose Arturo Molina-Mora, Research Center in Tropical Diseases (CIET), University of Costa Rica, San Pedro, Costa Rica (e-mail: jose.molinamora@ucr.ac.cr).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Molina-Mora JA, Chinchilla-Montero D, Castro-Peña C, García F. Two-dimensional gel electrophoresis (2D-GE) image analysis based on CellProfiler: *Pseudomonas aeruginosa* AG1 as model. *Medicine* 2020;99:49(e23373).

Received: 25 October 2019 / Received in final form: 20 October 2020 / Accepted: 27 October 2020

<http://dx.doi.org/10.1097/MD.00000000000023373>

In this sense, two-dimensional gel electrophoresis (2D-GE) has become a method of choice for proteomic studies since its introduction more than 40 years ago.<sup>[2]</sup> Its current use in part is explained due to its high performance in terms of the separation of complex protein mixtures.<sup>[3]</sup> The use of 2D-GE gels allows the comparison of complex protein profiles, first separating them by isoelectric point (pI) and then by molecular weight.<sup>[4]</sup> With this, the proteins are separated as spots, which are revealed with stains such as Coomassie blue or silver stain, to then capture images of the gel. These images are then analyzed to identify the points and study the protein content, as well as continue with subsequent proteomic studies by other strategies.<sup>[1]</sup>

However, due to the anomalies present in the images of 2D-GE gels, there is still no reliable, automatic and highly reproducible pipeline for 2D-GE image analysis.<sup>[4]</sup> At a strictly experimental level, the challenges of this type of technique include experimental variation (reagents, running conditions, etc), particular mobility of the proteins, deformation of the gel and the high probability of finding several proteins in the same space of the plane of the gel.<sup>[5]</sup> At the level of image analysis, the difficulties are greater, including anomalies such as the presence of vertical and horizontal stripes, noise around protein spots, diffuse spots and background noise, fusions of spots, artifacts due to the presence of dust or bubbles, saturation of certain spots and lack of linear intensity of protein spots.<sup>[1,3]</sup>

At the preprocessing level, one of the basic tasks is the alignment of images, in which one of the images is intentionally deformed to match the spots with the other image. This is done with a transformation that optimizes the measure of similarity and in turn quantifies the quality of alignment.<sup>[6]</sup> Then, algorithms are implemented to detect protein spots, that is, the recognition of objects by segmentation to define the limits of each spot, many of them with methods based on intensity, form, or hybrid strategies.<sup>[3]</sup> In a subsequent step, the quantification of the level of protein expression is performed according to the intensity and the number of pixels.<sup>[1]</sup> If required, a differential expression analysis can be performed by comparing conditions, in which multivariate statistical criteria are used, including analysis of variance (ANOVA) according to the size and intensity of the spot, strategies of correction of *P* values such as FDR (false discovery rate) or machine learning algorithms for clustering or classifying protein profiles.<sup>[5]</sup>

For the implementation of these analysis modules, there are software packages, practically all commercially available. This has the disadvantage that many are for a particular proteomics market, subject to purchase of equipment and that makes it even more expensive. Within these commercial solutions are PDQuest, ImageMaster2D, ProteomeWeaver, ProteinMine, Delta2D and Melanie, among others,<sup>[1]</sup> which generally contain modules that include the alignment of images to be compared, automatic identification and edition of spots, counting, quantification of intensity, and area calculation by spot. Within the options of free software, ImageJ<sup>[7]</sup> has been widely used for analysis of images of biological origin, but automation is limited, given that its approach is of individual analysis, as has been described.<sup>[8,9]</sup> In the approach of Natej and collaborators, a protocol was implemented with ImageJ for the study of spots in 2D-GE gels, applicable to pairs of images but with a strategy of limited scalability to large sets of images.<sup>[4]</sup>

Thus, due to all above, the aim of this work was to implement and evaluate an image analysis protocol with open-source software for identifying spots in 2D-GE gels images, which also

includes the potential application to automatically analyze a large number of images and, due to the computational requirements, that is potentially parallelizable. For this, we standardized experimental protocols for the study of the periplasmic proteins of *Pseudomonas aeruginosa* AG1 under various conditions of exposure to antibiotics. This bacterium is an opportunistic pathogen that survives diversity of environments, including hospital environments.<sup>[10]</sup> Specifically, our study model is the strain *P. aeruginosa* AG1, a Costa Rican isolate<sup>[11]</sup> with a multiresistance profile to antibiotics and with clonal MLST (<https://pubmlst.org/>) categorized as ST111, which implies a high risk for public health because of its resistance to therapy and association with nosocomial infections.

With this bacterial model, from the experimental assays, separation of the proteins was achieved using 2D-GE gels and it was revealed with silver staining. After capturing the respective images, we implemented a pre-processing step that included an initial phase of image alignment using the script bUnwarpl<sup>[12]</sup> in the program ImageJ<sup>[7,13]</sup>; this package has the ability to align hundreds of images to the same reference in one step. Subsequently, we made the spots identification with two protocols using the program CellProfiler.<sup>[8,14]</sup> A first protocol was established to identify total spots in the images of the gels, and that was contrasted with a reference analysis with the commercial program Melanie (<https://2d-gel-analysis.com/>). In a second implementation, spots differential identification in experimental conditions was made, separating the common spots from the exclusive ones. Finally, a comparison of several experimental conditions was carried out with two clustering algorithms, showing the similarity of protein profiles of *P. aeruginosa* AG1 exposed to antibiotics. To the best of our knowledge, CellProfiler program has not been used for the identification of spots on 2D-GE gels, although it has been implemented to for recognizing biological objects (cells, complete organisms, tumors, colonies of microorganisms and others) in hundreds of images, making this implementation as promising for the analysis of hundreds of gels in proteomics studies.

## 2. Methods

### 2.1. Experimental assays for 2D-GE gels

For the extraction and analysis of periplasmic proteins of *P. aeruginosa* AG1, cultures were used at exponential phase in LB medium (Luria Bertani, 2 clones) and LB medium added with subinhibitory concentrations of antibiotics ciprofloxacin (CIP, 12.5 µg/mL), tobramycin (TOB, 62.5 and 125 µg/mL), and imipenem (IMP, 25 and 50 µg/mL). The marker “IEF 3–10 SERVA liquid mix” (with proteins of size and known isoelectric point) was used as migration control. After pre-cultivation for 16h under the corresponding conditions, the bacteria were cultured for 6h at 37°C under agitation. After verifying their exponential growth by optical density, the samples were centrifuged at 10,000 rpm for 30 min and the supernatant was discarded.

For the extraction of periplasmic proteins with chloroform, pellets were washed with sterile PBS 1× and then 0.01 M Tris-hydrochloride pH 8.0 filtered and chloroform were added. After an incubation, the sample was centrifuged and the supernatant stored at –80°C. For protein precipitation, the supernatants were treated with methanol and chloroform. After vigorous stirring and a strong centrifugation, the separation was achieved in

2 phases, an upper one of methanol/water and a lower one of chloroform. The periplasmic protein fraction was found in the middle of both phases, which was finally precipitated with more alcohol and centrifugation. After the supernatant was removed, the protein pellet was dried and resuspended in 0.05% SDS lysis buffer, obtaining the protein extract of interest. Modified protocol of Ames et al.<sup>[15]</sup>

Finally, the protein separation in two-dimensional gel was performed by adding the proteins to Isoelectric Focusing (IEF) strips and hydrated for 24h at room temperature. Then, the proteins were separated using a non-linear 3 to 11 pH gradient, following the manufacturer's instructions (GE HealthCare Immobiline Dry Strip Gels™). For the second dimension (molecular weight), the IEF strips were incubated in equilibrium buffer (50 mM Tris-HCl, 6 M Urea, 30% glycerol and 2% SDS) with 4-dithiothreitol (DTT), for 10 min, before separation into a SDS-GE gradient of 4% to 20% for 90 min at 150 V. PageRuler Protein Ladder (Fermentas) was used as a molecular weight marker. All gels were visualized with silver stain. The bands were observed in the ChemiDoc photo viewer (BioRad).

## 2.2. Preprocessing of 2D-GE images by alignment

Due to the conditions inherent in the assembly of 2D-GE gels, the images require preprocessing alignment (Fig. 1). Thus, the detailed protocol was implemented by Natale and collaborators<sup>[4]</sup> using the bUnwarpJ package in the ImageJ program.<sup>[12]</sup> Using 5 reference points, with spots known as common between the images, we proceeded to the deformation of the larger images to align with the spots of the smaller image, using the parameter of "degree of deformation" as fine. After the deformation, the aligned images were saved for the following analysis steps.

## 2.3. Identification of total spots

In order to identify the totality of visible protein spots in the gels, an image analysis protocol was implemented using the CellProfiler program (<https://CellProfiler.org/>). As detailed in Figure 1 (middle-left) the protocol consisted of 5 steps:

1. the inversion of the images to enable recognition,
2. the implementation of an object recognition, evaluating different parameters and recognition algorithms and segmentation,
3. improving the identification by manual editing,
4. calculating different metrics by object and, finally,
5. visualizing the recognition in the images.

Similarly, the automatic protocol of a specialized program for 2D gels, Melanie (<https://2d-gel-analysis.com/>), was used to compare the performance of our protocol, contrasting the number of recognized elements and the intensity measured with a linear regression.

## 2.4. Differential identification of spots

To compare the differential expression of proteins between experimental conditions, we proceeded to implement an analysis of pairs of images (Fig. 1 middle-right, also see Figure 4A for case of two clones of control condition). The steps for this process included:

1. the inversion of aligned images,

2. creating a new image of spots commonly shared by the images, preserving the minimum value of pixels in the same location,
3. automatic identification and manual edition of primary objects (same as protocol of total spots), and
4. the elimination of common spots of each image.

With this, we obtained images of gels with common spots eliminated, so in a next step we performed

5. the identification and edition of primary objects of the exclusive spots of each gel,
6. calculation of metrics for each spot, and finally,
7. the representation of common and exclusive spots for each image.

With this, each image of each condition identified spots present in both conditions (configured to be marked in red), or, exclusive of each gel (blue or green colors in each image).

## 2.5. Comparison of gels from multiple experimental conditions

In order to compare different profiles of periplasmic proteins in various conditions of antibiotic exposure in *P. aeruginosa* AG1, we proceeded to run two machine learning algorithms for clustering: a Principal Component Analysis (PCA) and a Hierarchical Clustering (HC) analysis (Fig. 1 down). To address this, the images were first aligned (as previously described) and then the images were divided into 121 sectors (11 × 11 quadrants) and, given that the location was in coordinates, the counting of spots was made for each of the zones. This information was used to implement the clustering algorithms, which used Euclidean distance for the dissimilarity and default parameters of the Caret package (<http://caret.r-forge.r-project.org/>) in the R program (<https://www.r-project.org/>).

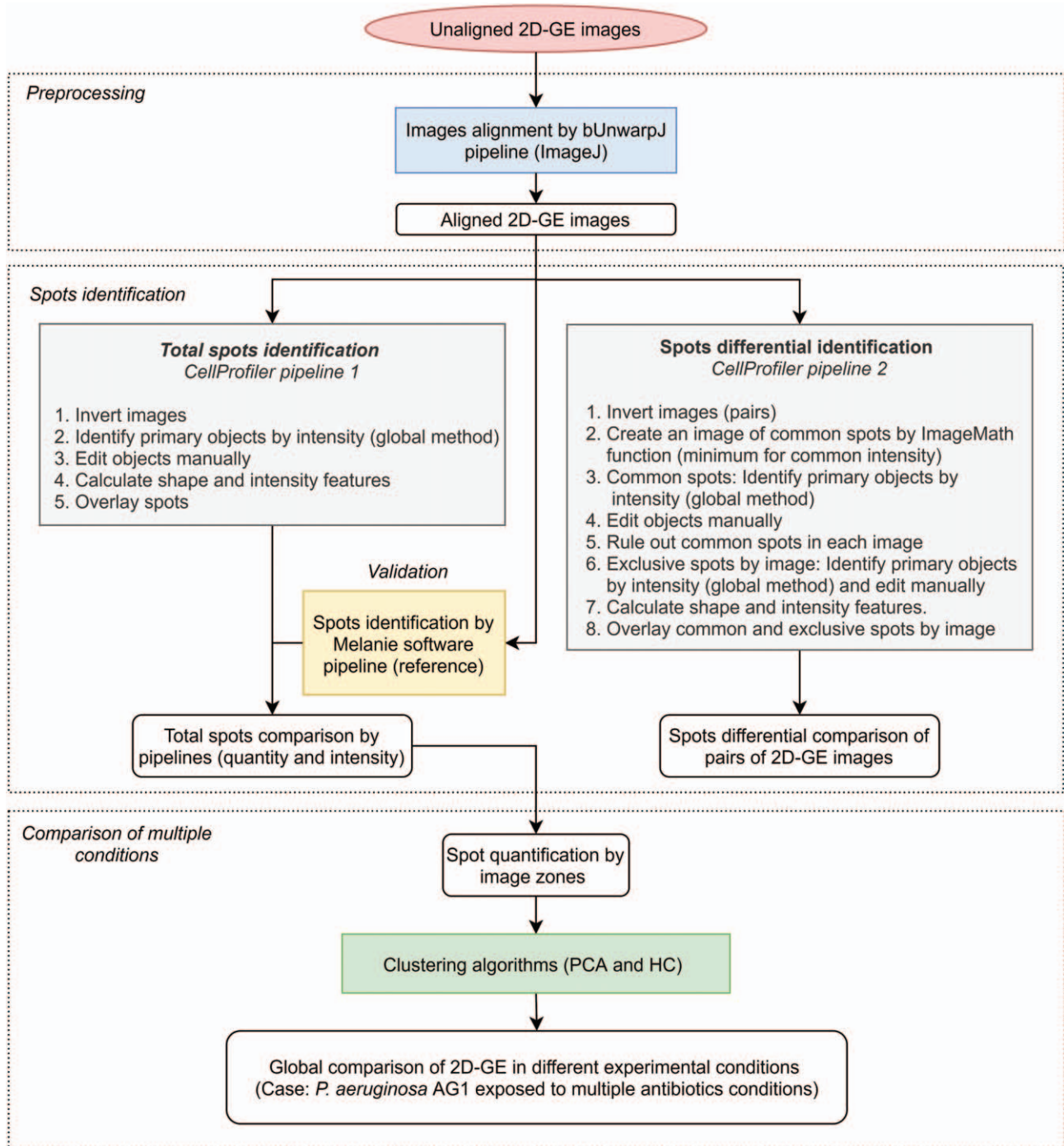
## 3. Results

In order to establish an automatic procedure for the identification of spots of proteins in 2D-GE gels, we first proceeded with the generation of images from experimental assays with the periplasmic proteins of *P. aeruginosa* AG1, in conditions with or without antibiotics. Then, we proceeded with the analysis of images, including alignment, identification of total spots and validation, differential identification when comparing pairs of conditions and finally analysis by clustering, as summarized in Figure 1.

To align and compare the protein migration profile in 2D-GE gels, the bUnwarpJ package was used to deform the larger images and align them to a reference. In the case presented in Figure 2A, which starts with two images of different sizes (two clones of the strain in control condition), five points of reference or common denominator are established between the images, which are used by the algorithm to optimize the alignment by calculating a field and network of deformation (Fig. 2B). With this, the larger image is reduced to align and make the spots comparable between conditions (the image was cropped to visualize the distribution, Fig. 2C).

Using the CellProfiler software, two spots recognition protocols were implemented. In the first one, with the identification of total spots, it was established that the optimal conditions were the use of a global algorithm (assuming relatively homogeneous background pixels and other parameters with default values),



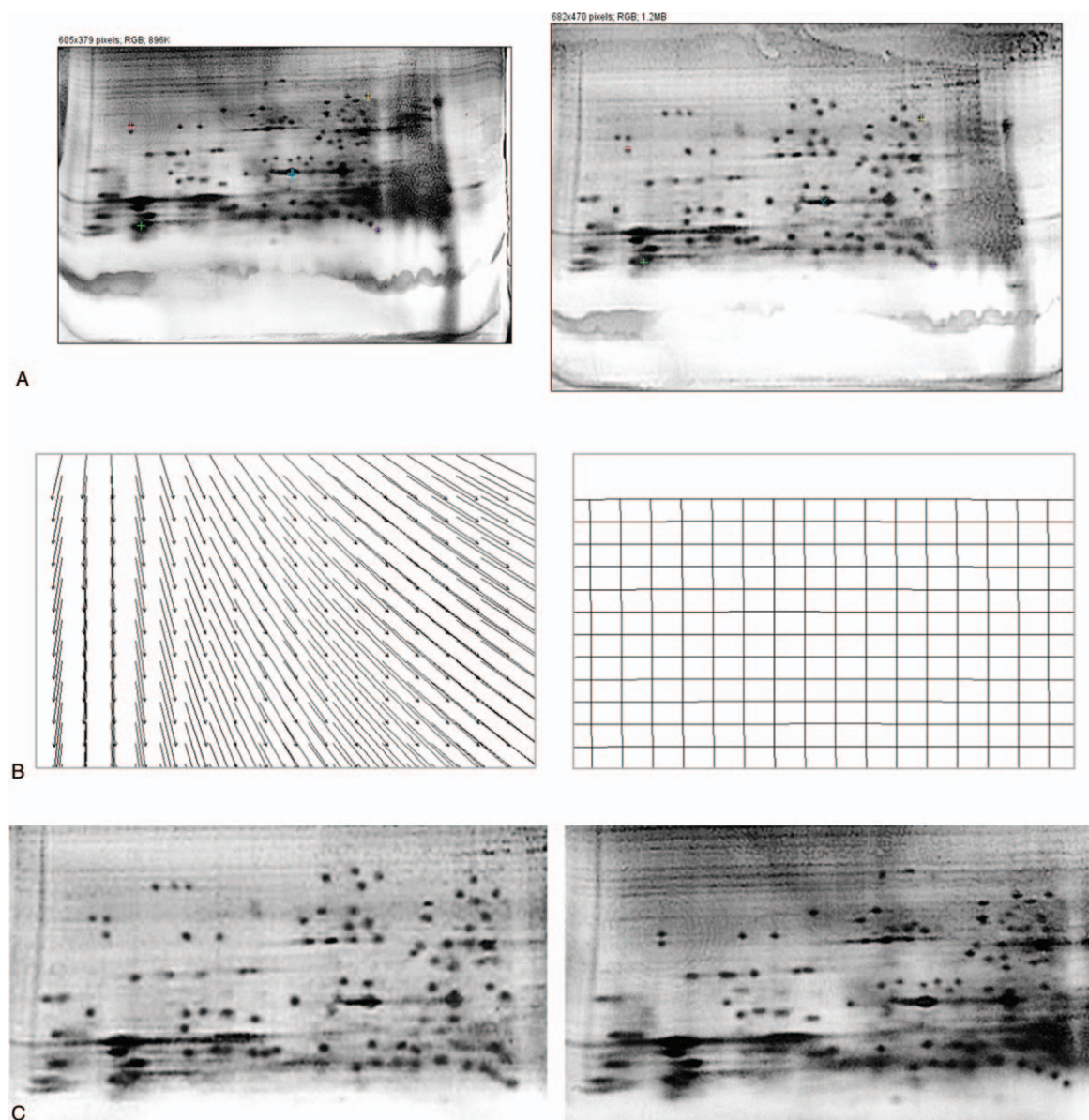


**Figure 1.** General workflow by image analysis for identifying and comparing spots in 2D-GE gel images.

sizes of 40 to 100 pixels for the objects and the use of intensity to recognize and segment objects. Thus, after the inversion of the image and the recognition of objects, the recognized objects were presented on the original image (Fig. 3A left). When performing the comparison with an automatic protocol with the Melanie program (used as a reference for validation), it was verified that the resolution capacity of the protocol we implemented had the same ability to identify spots (Fig. 3A, right). The number of spots was counted in 124 for both protocols (this value was controlled with the manual edition available both in our protocol and in

Melanie pipeline and that includes cases of proteins grouped as a single spots in cases of large spots). Given that the boundaries or edges of recognition of an object varied between protocols, we proceeded to perform a linear regression between the intensity values, determining that the intensity behavior between the algorithms is linear (Fig. 3B).

In a second protocol with the same program, we proceeded with the differential identification of spots when comparing pairs of gels, obtained from two clones of the same strain *P. aeruginosa* AG1 in LB medium condition. The identification of objects was

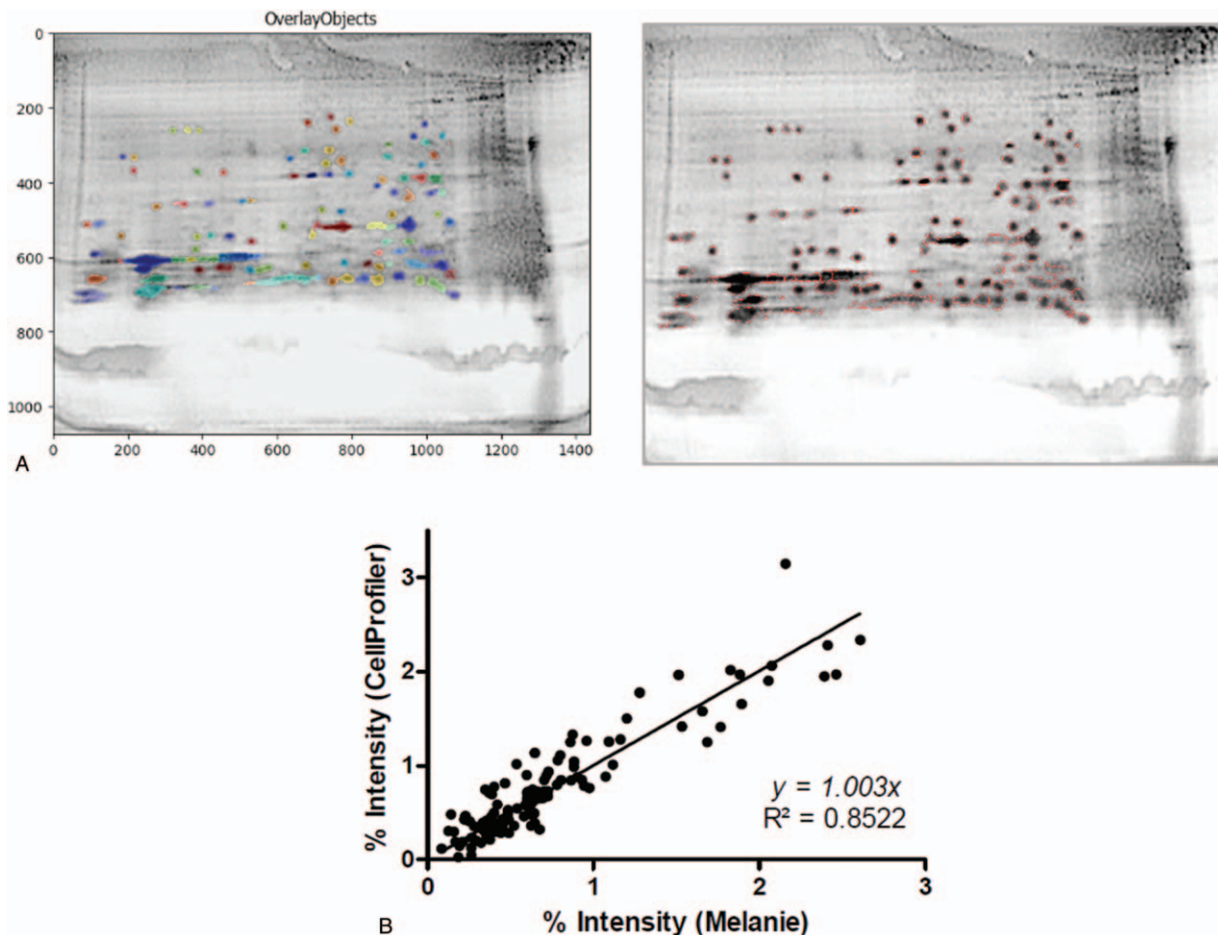


**Figure 2.** Alignment of 2D-GE gel images by warping method with bUnwarpJ pipeline (gels of proteic profiles from two clones from same strain *Pseudomonas aeruginosa* AG1). (A) Raw images showing differences by size and scale. Color marks define the reference points for warping. (B) Deformation field (left) and deformation grid (right) of larger image to align to the small one. (C) Aligned 2D-GE gel images after warping and cropping (two clones).

done with the algorithm and intensity conditions described for the previous case, both for common spots and exclusive spots. Obtaining common spots was achieved by creating a new image, preserving the lower pixel value for the two images (so if a dot was present in both conditions, the image created would have a high value). Then, the spots were identified and they were labeled as proteins common to both conditions. Using the MaskImage function, the elimination of these common objects was achieved and, in a new recognition for each image, it was possible to identify the exclusive elements of each gel. Using colors, each type of object, common spots (red) or exclusive spots (green or blue) were marked on the images, showing that for this case the majority of proteins were shared by the two clones of the bacteria (Fig. 4B and C).

Finally, with the identification of spots made for each gel in different conditions including antibiotics, we proceeded to the comparison of the protein profiles. First, a division of the images into zones was carried out, and the number of spots was counted. Then, the PCA and HC clustering algorithms were evaluated, obtaining that the profiles given by different antibiotics generate more differences than the concentration of the antibiotic. In the case of the PCA (Fig. 5A), using first two components (with a cumulative variation between both  $>60\%$ ), they show a similar relationship between the control with LB medium and the case of ciprofloxacin. This relationship is maintained when evaluating HC (Fig. 5B), but the relationship between imipenem and its two concentrations shows minor differences. In addition, for this same case, the division by zones shows the sectors of gels with





**Figure 3.** Total spots identification by a CellProfiler pipeline and comparison with Melanie pipeline. (A) CellProfiler pipeline (left) vs Melanie software (right) for segmentation of objects and final identification after manual edition. (B) Comparison of spots intensity using the CellProfiler pipeline and Melanie software.

similar or very different compartment (potentially useful to select zones for subsequent analysis, see discussion). The HC results are shown with the respective gels in Figure 5C.

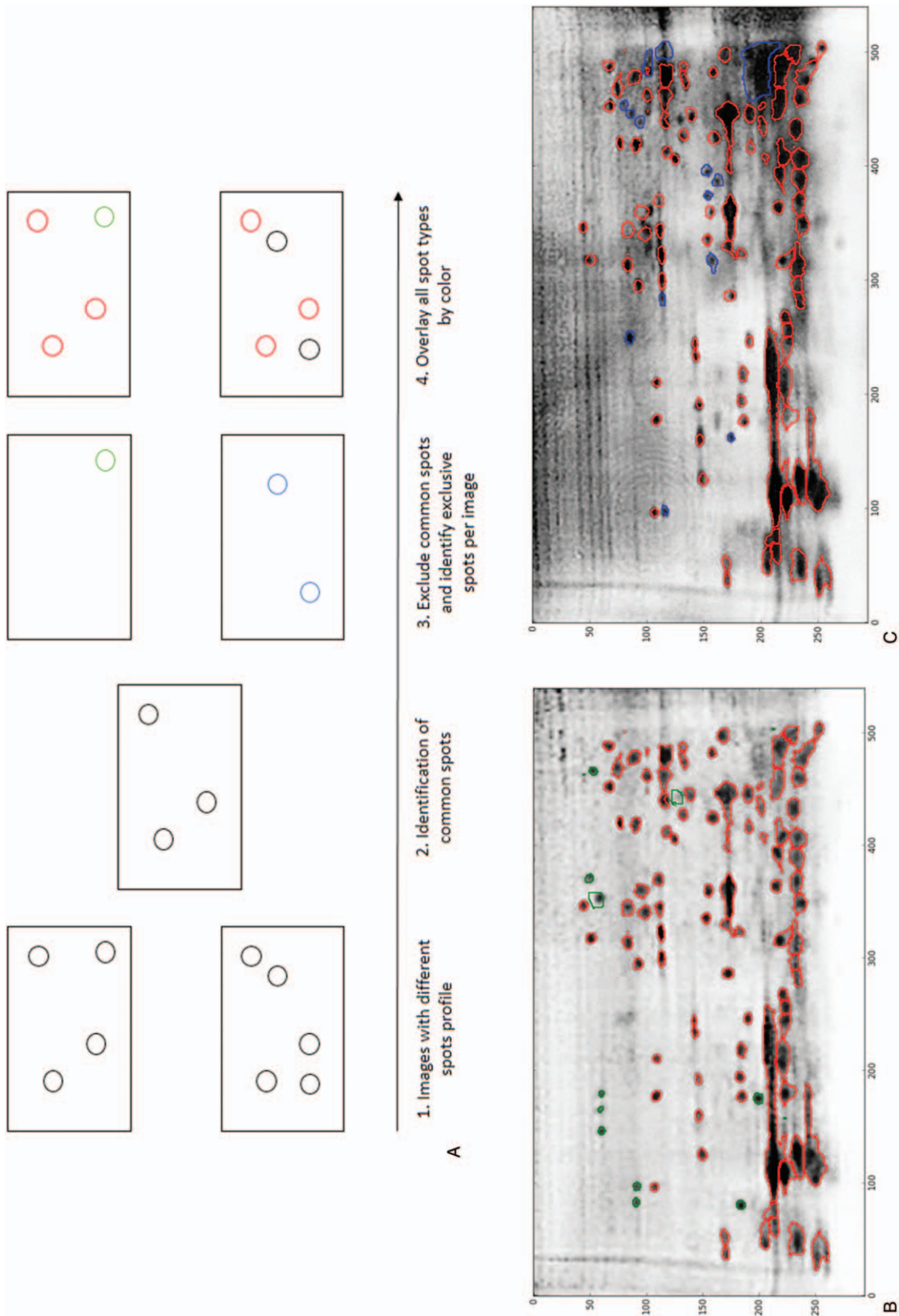
#### 4. Discussion

Proteomics is considered an essential field for the systematic analysis of biological systems, an assessment of changes in the abundance of proteins that occur in living organisms and that can be studied at various levels.<sup>[3]</sup> The two-dimensional gel electrophoresis 2D-GE, separating the proteins according to their isoelectric point and molecular weight, is still used in proteomics laboratories due to the relative ease of implementation in terms of execution and cost, the capacity of solve and visualize miles of proteins in a single run and it is compatible with other high-performance protein techniques, such as mass spectrometry.<sup>[1]</sup> 2D-GE and subsequent strategies have been implemented in recent studies using bacterial models, including application of protein phosphorylation (phosphoproteomics) in *Bacillus anthracis*<sup>[16,17]</sup> or biotechnological applications in *Xanthomonas campestris*.<sup>[18]</sup>

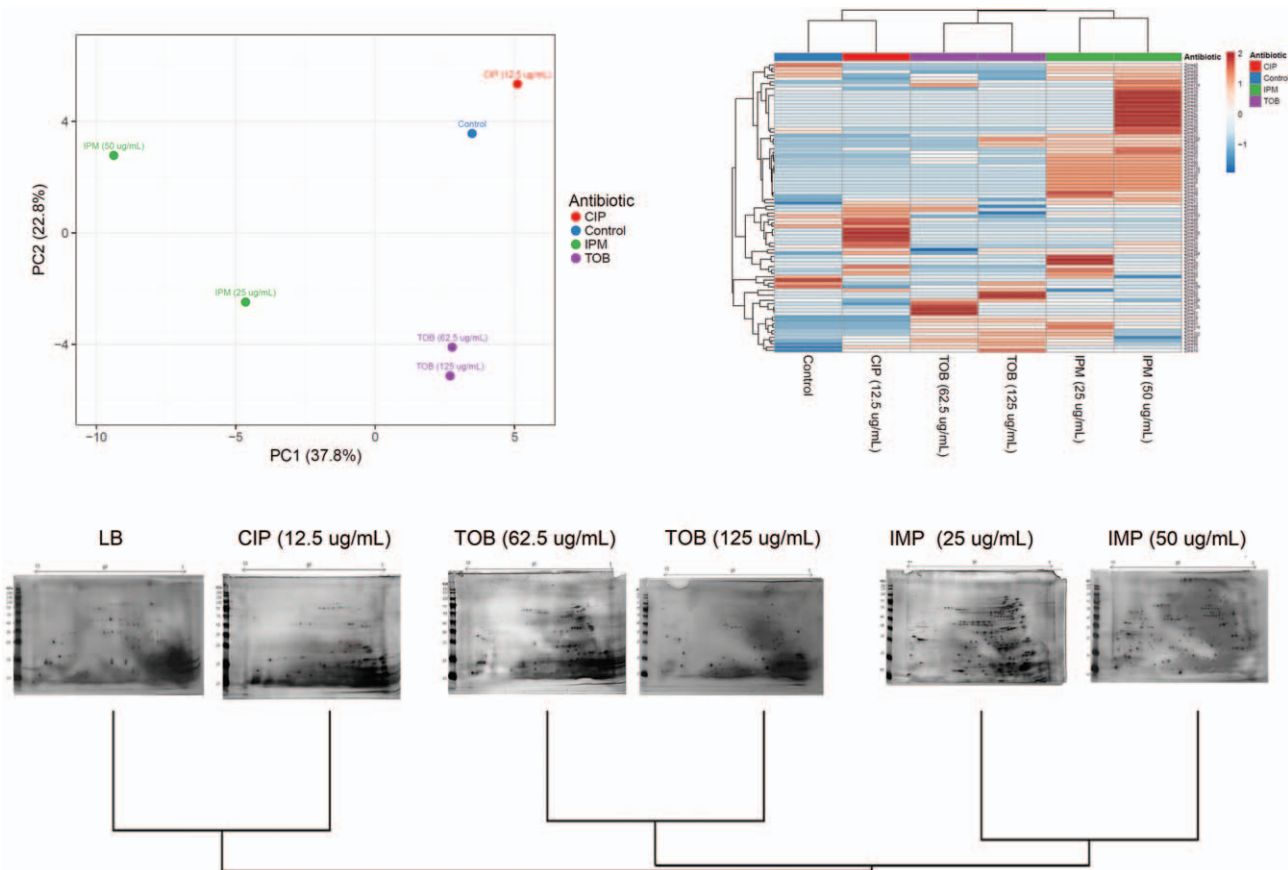
After the experimental phase, the visualization of the proteins is done with the particular stains and gel images are captured, which must be analyzed qualitatively and quantitatively for the extraction of biologically relevant protein information. Of the

existing implementations, although there are some investigations in methods of analysis of gels 2D-GE work directly at the level of pixels, most focus on recognizing spots on gel to describe the abundance in each condition.<sup>[3]</sup> Despite this, there is no protocol for universal or consensus analysis, and multiple limitations are reported in various processing steps.<sup>[4]</sup> At commercial field, the available programs have additional drawbacks of having a high cost, in addition to many of them are for sale with hardware equipment, which restricts the possibilities of use. In addition, due to its nature, the private code of the implementations is not available, which prevents knowing the details of strategy at the level of algorithms and makes the modification impossible for specific applications. In addition, some limitations of commercial or open access programs include the limited number of images to analyze.

With the aim of implementing and evaluating an image analysis protocol for the recognition of spots in 2D-GE gels images, using open-source software, parallelizable, and applicable to hundreds of images, we obtained experimental data of protein profiles of *P. aeruginosa* AG1 under standardized conditions with or without antibiotics. The general protocol was presented in Figure 1. Although it is possible to find variations between runs for the same sample, in our work, we used data from different runs but the same run. Comparison of other protein concentrations, experimental conditions, or



**Figure 4.** Spots differential identification and comparison of 2D-GE gel images from two experimental conditions (clones from same strain). (A) General pipeline for identifying common (red, 124 spots) and exclusive spots (blue or green), which was applied to two different proteomic profiles, Clone 1 with 11 exclusive spots (B) or Clone 2 with 14 exclusive spots (C), respectively.



**Figure 5.** Machine learning approach of clustering analysis for comparing 2D-GE gel images from multiple experimental conditions. (A) PCA algorithm, (B) HC analysis showing zones and spots count, and (C) HC showing images.

replicates are known to produce changes in the proteomic profile, and further analyses are required to study these effects and the performance of our pipeline considering this.

Images were aligned with the bUnwarpl package in the ImageJ program. This step is required as preprocessing of data since the final performance depends to a great extent on the quality of the images to be processed. This processing includes the alignment of images to match the corresponding protein points of different conditions.<sup>[1]</sup> In our case, the larger image was adjusted to the smaller one and as an example the case of two protein profiles of two clones of the bacteria was presented in the control condition with LB culture medium (Fig. 2). Although in our final implementation we use 6 images when aligning, the alignment of hundreds or thousands of images is possible using a single reference, as we did in another application with data of cell cultures followed over time, aligning 600 images to the initial image (unpublished data), showing the potential of using this package for the analysis of multiple gel images. Other applications with other types of images show this fact.<sup>[19,19,20]</sup>

After the preprocessing, we carried out the implementation of two protocols with CellProfiler software. Particular features of this software are discussed below. In a first approach, we recognized total spots (Fig. 3), allowing the counting of spots and the quantification of the area and intensity integrated by each one. Additionally, we compared the performance of this protocol

with a pipeline of the commercial software Melanie, showing an equivalent performance when comparing the intensity obtained per object. Due to the fact that in both protocols a module of manual editing of the identification is implemented, the count of elements was intentionally controlled according to expert criteria, for a total of 124 spots. Similar results in performance have been previously reported when an analysis with ImageJ was compared with Melanie,<sup>[14]</sup> but as mentioned before, with limited number of images to be processed. In the case of CellProfiler, automation is an essential component from its design, as well as the option to parallelize in computer clusters.<sup>[14]</sup>

In a second protocol (Fig. 4), we implemented a procedure to differentially recognize the expression of proteins in pairs of experimental conditions, allowing us to identify common and exclusive spots of experimental conditions. To do this, our strategy was based on the construction of a new image using the minimum value of pixels of the two images aligned and inverted, using the MathImage function of the program. In this way common spots were preserved. The recognition by segmentation based on intensity allowed the identification of objects, which were later excluded in each. In a second phase, the remaining spots were recognized in each image, to then differentially represent the edges of the shared and exclusive spots.

To the best of our knowledge, there are no approximations that allow the display of common and exclusive spots



automatically, given that it is regularly done manually. This information is used to identify proteins differentially expressed in the conditions studied. However, our approach is very robust considering only the presence or absence of spots, and true cases of differential expression with significant changes in intensity are not contemplated, so we consider that this protocol allows the differential identification of spots, but not properly the differential protein expression analysis. This last type of analysis is carried out by commercially available packages, but they are mainly based on the intensity and area, and due to the preprocessing of the image in terms of image contrast, dimensions and other modifications, the normalization and transformation of data it remains a challenge.<sup>[1]</sup>

Regarding the CellProfiler program and its convenience for this implementation, this software offers the management of hundreds of thousands of images, freely available and with an open and flexible code platform to share, test, and develop new methods by experts in image analysis. In addition, it offers an easy-to-use interface and the possibility of implementing in computational clusters.<sup>[8]</sup> In addition, due to its nature of automation, the program is capable of handling hundreds of thousands of images, which high performance infrastructure is required for massive analyzes, such as those implemented at the omics level. Although many of the applications of the CellProfiler program are formulated for cells, other applications have been implemented at the level of recognition of complete organism in images, such as the parasite *Caenorhabditis elegans*,<sup>[21]</sup> or complete tumors, colonies of yeast or bacteria, and other images of biological origin, as evidenced in the Educational Modules section of the web page (<https://CellProfiler.org/outreach/>).

In contrast, we finally carried out the implementation of a machine learning analysis to compare protein profiles from gel images using PCA and HC clustering algorithms. This type of strategy has been previously implemented with PCA and heuristic clustering algorithms,<sup>[22–24]</sup> as well as supervised classification algorithms to separate conditions, including Support Vector Machine.<sup>[23]</sup> Other approaches have implemented comparison modules using directly the properties of intensity, brightness, and contrast of images to contrast with databases,<sup>[25]</sup> or, other levels of proteomic analysis, such as mass spectrometry.<sup>[26]</sup> Regarding the methodology used in our case of division by zones and grouping of regions with a similar profile, this strategy can be used to make subsequent decisions of work in proteomics laboratories, where the task after the gels is the selection of spots and continue with identification applications with techniques such as HPLC or mass spectrometry.<sup>[1]</sup>

In the biological aspect according to the results obtained when comparing 6 experimental conditions with *P aeruginosa* AG1 bacteria with or without antibiotic, it was possible to identify the relationships between the total protein expression profiles. Both with the results of the analysis by PCA and by HC, it is concluded that there is greater similarity between the profiles obtained for the same antibiotic at different concentrations, and that they are separated from the conditions of other antibiotics, congruent according to the mechanisms of action of each type of antibiotic. In the case of ciprofloxacin, its profile was separated to a greater degree from the other antibiotics and was grouped with the control with LB medium.

Because the bacterial strain *P aeruginosa* AG1 is resistant to those antibiotics, this information and subsequent analysis at the proteomic level, together with other genomic, transcriptomic, and phenomic analyzes that we are conducting, will allow us to

obtain new findings of the biological relationships to molecular level that provide insights to begin to explain the mechanisms of tolerance to antibiotics and the modulation of biological processes in response to cellular stress.

## 5. Conclusions

In the context of proteomics and its importance for the study of different biological conditions, our implementation of the image analysis of gels 2D-GE offers an opportunity to continue with studies of analysis of protein profiles. Using the open-source software, CellProfiler (and bUnwarpJ for preprocessing), we achieved the alignment of images, the identification of spots and the final comparison of protein profiles. These workflow also allow analyze a large number of images automatically as well as enabling the parallelization in computational clusters to counteract the complexity of processing this type of data. Regarding the biological meaning, exposure to ciprofloxacin in *P aeruginosa* AG1 showed a similar pattern to control without treatment, and other groups were generated according to the antibiotic class. This information will be integrated with other molecular analyses using antibiotics in this multiresistant strain to gain insights regarding the mechanisms of tolerance to antibiotics and the modulation of biological processes in response to cellular stress.

## Acknowledgments

We thank the student Daniel Solano Alvarado for his collaboration as an assistant in various activities of the project.

## Author contributions

**Conceptualization:** Fernando García.

**Formal analysis:** Jose Arturo Molina-Mora, Diana Chinchilla-Montero, Carolina Castro-Peña.

**Methodology:** Jose Arturo Molina-Mora, Diana Chinchilla-Montero, Carolina Castro-Peña, Fernando García.

**Software:** Jose Arturo Molina-Mora.

**Supervision:** Fernando García.

**Visualization:** Jose Arturo Molina-Mora.

**Writing – original draft:** Jose Arturo Molina-Mora.

**Writing – review & editing:** Jose Arturo Molina-Mora, Diana Chinchilla-Montero, Carolina Castro-Peña, Fernando García.

## References

- [1] Goetz MM, Torres-Madroñero MC, Röthlisberger S, et al. Preprocessing of 2-dimensional gel electrophoresis images applied to proteomic analysis: a review. *Genomics Proteomics Bioinformatics* 2018;16:63–72.
- [2] O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975;250:4007–21.
- [3] Silva TS, Richard N, Dias JP, et al. Data visualization and feature selection methods in gel-based proteomics. *Curr Protein Pept Sci* 2014; 15:4–22.
- [4] Natale M, Maresca B, Abrescia P, et al. Image analysis workflow for 2-D electrophoresis gels based on imageJ. *Proteomics Insights* 2011;4:37–49.
- [5] Abdallah C, Dumas-Gaudot E, Renaut J, et al. Gel-based and gel-free quantitative proteomics approaches at a glance. *Int J Plant Genomics* 2012;2012.
- [6] Dowsey AW, Morris JS, Gutstein HB, et al. Informatics and statistics for analyzing 2-D gel electrophoresis images. *Methods Mol Biol* 2010; 604:239–55.
- [7] Abramoff MD, Magalhães PJ, Ram SJ. Image processing with Image. *J Biophotonics Int* 2004;11:36–42.
- [8] Lamprecht M, Sabatini D, Carpenter A. CellProfilerTM: free, versatile software for automated biological image analysis. *Biotechniques* 2007; 42:71–5.

- [9] Schindelin J, Rueden CT, Hiner MC, et al. The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol Reprod Dev* 2015;82:518–29.
- [10] Cirz RT, O'Neill BM, Hammond JA, et al. Defining the *Pseudomonas aeruginosa* SOS response and its role in the global response to the antibiotic ciprofloxacin. *J Bacteriol* 2006;188:7101–10.
- [11] Toval F, Guzmán-Martel A, Madriz V, et al. Predominance of carbapenem-resistant *Pseudomonas aeruginosa* isolates carrying blaIMP and blaVIM metallo- $\beta$ -lactamases in a major hospital in Costa Rica. *J Med Microbiol* 2015;64:37–43.
- [12] Arganda-carreras I, Sorzano COS, Kybic J, et al. bUnwarp: Consistent and Elastic Registration in ImageJ. *Methods and Applications. Image (Rochester, NY)*; 2006.
- [13] Collins T. ImageJ for microscopy. *Biotechniques* 2007;43(S1):S25–30.
- [14] Kametsky L, Jones TR, Fraser A, et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* 2011;27:1179–80.
- [15] Ames GF, Prody C, Kustu S. Simple, rapid, and quantitative release of periplasmic proteins by chloroform. *J Bacteriol* 1984;160:1181–3.
- [16] Virmani R, Sajid A, Singhal A, et al. The Ser/Thr protein kinase PrkC imprints phenotypic memory in *Bacillus anthracis* spores by phosphorylating the glycolytic enzyme enolase. *J Biol Chem* 2019;294:8930–41.
- [17] Arora G, Sajid A, Virmani R, et al. Ser/Thr protein kinase PrkC-mediated regulation of GroEL is critical for biofilm formation in *Bacillus anthracis*. *Npj Biofilms Microbiomes* 2017;3:7.
- [18] Schulte F, Hardt M, Niehaus K. A robust protocol for the isolation of cellular proteins from *Xanthomonas campestris* to analyze the methionine effect in 2D-gel experiments. *Electrophoresis* 2017;38:2603–9.
- [19] Kindle LM, Kakadiaris IA, Ju T, et al. A semiautomated approach for artefact removal in serial tissue cryosections. *J Microsc* 2011;241:200–6.
- [20] Seiler C, Reyes M. Displacement Vector Field Regularization for Modelling of Soft Tissue Deformations; 2008. Available at: <https://christofseiler.github.io/MasterThesis.pdf>. [Accessed June 4, 2019].
- [21] Moy TI, Conery AL, Larkins-Ford J, et al. High throughput screen for novel antimicrobials using a whole animal infection model. *ACS Chem Biol* 2009;4:527.
- [22] Appel R, Hochstrasser D, Roch C, et al. Automatic classification of two-dimensional gel electrophoresis pictures by heuristic clustering analysis: a step toward machine learning. *Electrophoresis* 1988;9:136–42.
- [23] Supek F, Peharec P, Krsnik-Rasol M, et al. Enhanced analytical power of SDS-PAGE using machine learning algorithms. *Proteomics* 2008;8:28–31.
- [24] Castillejo MÁ, Fernández-Aparicio M, Rubiales D. Proteomic analysis by two-dimensional differential in gel electrophoresis (2D DIGE) of the early response of *Pisum sativum* to *Orobanche crenata*. *J Exp Bot* 2012;63:107–19.
- [25] Kush A, Raghava GPS. AC2DGel: analysis and comparison of 2D Gels. *J Proteomics Bioinform* 2008;01:043–6.
- [26] Kelchtermans P, Bittremieux W, De Grave K, et al. Machine learning applications in proteomics research: how the past can boost the future. *Proteomics* 2014;14:353–66.