METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, UK

For further volumes: http://www.springer.com/series/7651

2–D PAGE Map Analysis

Methods and Protocols

Edited by

Emilio Marengo and Elisa Robotti

Department of Sciences and Technological Innovation
Università del Piemonte Orientale
Alessandria, Italy

Humana Press

Editors
Emilio Marengo
Department of Sciences and Technological Innovation
Università del Piemonte Orientale
Alessandria, Italy

Elisa Robotti
Department of Sciences and Technological Innovation
Università del Piemonte Orientale
Alessandria, Italy

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-3254-2 ISBN 978-1-4939-3255-9 (eBook)
DOI 10.1007/978-1-4939-3255-9

Library of Congress Control Number: 2015954566

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Humana Press is a brand of Springer
Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Dedication

To my father who was for me an example of rectitude, honesty and energy.
To my mother for teaching me to love life in all its forms.
To Elena, my future.
To all who love me, for their sustain.

Emilio Marengo

To my mother and my father for their love and sustain and for teaching me to never give up.
To Marco, my present.
To Matteo, my future.

Elisa Robotti

Foreword

It seems only yesterday when, in 1975, a bright star rose suddenly above the horizon of Separation Science. Three labs reported a two-dimensional (2-D, charge coupled orthogonally to mass separation) technique, for analysis of complex protein systems, simultaneously and independently, although most of the credit went just to O'Farrell [1–3]. Perhaps because his system was the most elaborate: in fact, he was able to resolve and detect about 1100 different proteins from lysed *E. coli* cells on a single 2-D map and suggested that the maximum resolution capability might have been as high as 5000 different proteins. Apart from the meticulous attention to detail, major reasons for the advance in resolution obtained by O'Farrell, compared to earlier workers, included the use of samples labeled with ^{14}C or ^{35}S to high specific activity, and the use of thin (0.8 mm) gel slabs for the second dimension, which could then be dried down easily for autoradiography. This detection method was able to reveal protein zones corresponding to one part in 10^7 of the sample (usually 1–20 μg was applied initially, since higher loads caused zone spreading, although up to 100 μg could be loaded). Coomassie blue, in comparison, was about 3 orders of magnitude less sensitive and could reveal only about 400 spots. For the first dimension, O'Farrell adopted gel rods of 13 cm in length and 2.5 mm in diameter. The idea was to run samples fully denatured, in what became known as the "O'Farrell lysis buffer" (9 M urea, 2 % Nonidet P-40, 2 % β-mercaptoethanol, and 2 % carrier ampholytes, in any desired pH interval). For the second SDS-PAGE dimension, O'Farrell [1] used the discontinuous buffer system of Laemmli [4] and, for improved resolution, a concave exponential gradient of polyacrylamide gel (usually in the intervals 9–15 or 10–14 %T, although wide porosity gradients, e.g., 5–22.5 %T, were also suggested). It is thus seen that, since its very inception, O'Farrell carefully selected all the best conditions available at the time; it is no wonder that his system was adopted as such in the avalanche of reports that soon followed (as for this writing, his paper has received about 20,000 citations!). O'Farrell 2-D mapping protocol became the basic methodology for what we would call today "proteomics" of any tissue or biological fluid, where thousands of components were suspected to be present. It remained the gold standard for such investigations at least for the following 25 years, up to the third millennium.

When I stated that the O'Farrell 2-D mapping introduced in 1975 had been a bright star I affirmed a widely held opinion, but it was not a polar star for us navigators in the starry sky represented by the polypeptides display in the 2-D gel. It was wishful thinking, at best. There was indeed a major impediment in this methodology, namely the erratic spot profile obtained by performing the first dimension in conventional IEF with soluble carrier ampholytes (CA), a la Svensson-Vesterberg, if you like [5]. There were no fixed stars in the firmament of 2-D maps: the apparent pI values kept changing, from batch to batch of CAs and, of course, from brand to brand, as manufactured by different companies (a chaotic synthesis, as you might remember) [6]. The situation was so frustrating that the Andersons recommended carbamylation train standards for mapping the pH gradient course and even preparing large volumes of stock solutions of CAs, obtained by carefully blending the various commercial products. The help was soon at hand, since in 1982 Bjellqvist et al. [7] launched another supernova in the sky of bioanalysis: immobilized pH gradients (IPGs), which were soon demonstrated to be able to overcome all these problems, while affording

exquisite resolution when run in narrow and ultra-narrow pH ranges. IPGs went largely unnoticed for about a decade, even though they brought about some out-of-(terrestrial) space results in bio-separations, including a resolution limit of $\Delta pI = 0.001$ for IPGs, vs. a maximum resolving power of conventional IEF in soluble carrier ampholytes of $\Delta pI = 0.01$, one order of magnitude less. Together with that, IPGs brought "democracy" for the first time in electrokinetic processes. Up to their introduction, 2-D maps had been conducted only in linear pH gradients, which penalize acidic proteins, jammed in the overcrowded zone of the pH 4–6 region, where >60 % of all proteins focus. Already in 1985, we were able to describe a broad range, nonlinear IPG, strongly flattened in the crowded region, with a sharp upward turn at alkaline pH values [8]; these ranges are by far the most popular in today 2-D map analyses. Needless to say, IPGs proved to offer a loading ability much superior to that of conventional IEF. Gels could be massively overloaded without isoelectric precipitations or smears. This unique property could be exploited in 2-D maps for detecting low-abundance proteins; whereas the typical protein load in 18 × 20 cm gel slabs was up to 0.5 mg, with IPGs the load could be incremented up to 10 mg per gel [9]!

Two-dimensional maps represent only one half of the proteomics panorama of the present day (excluding from the count mass spectrometry, which of course has an enormous significance in this field). The other half is a chromatographic approach developed by Yates III and coworkers [10, 11], consisting of an online 2-D ion-exchange column coupled to RP-HPLC, termed MudPit, for separating tryptic digests of 80S ribosomes from yeast. The acidified peptide mixture is loaded onto a strong cation exchanger (SCX) column; discrete eluate fractions are then fed onto a RP (reversed-phase) column, whose effluent is injected directly into a mass spectrometer. This iterative process is repeated 12 times by using increasing salt gradient elution from the SCX bed and an increasing organic solvent concentration from the RP beads (typically a C_{18} phase). In a total yeast lysate, the MudPit strategy allowed the identification of almost 1500 proteins [11]. There are major differences, though, between these two methods, in that the first one (2-D gel mapping) consists in separating intact, proteins, as found in the original tissue in which they were expressed, whereas in the second, chromatographic approach, only digested species are analyzed, which means that subtle differences in expression (e.g., deamidation, proteolytic cleavage products originating in vivo) are usually lost. Ideally, though, a lab should utilize both approaches, since it is claimed that the advantage of MudPit (an unfortunate acronym for such a powerful technique, since it literally means "hole full of mud") would be the ability of detecting also scarcely soluble membrane proteins and very acidic or basic species whose pI values would be outside the range of IEF/IPG. It turns out, though, that a kind of dichotomy developed, by which in the USA MudPit is mostly adopted (on the grounds that it is mostly an instrumental approach, involving little labor from lab technicians) whereas on this side of the Atlantic 2-D gel approaches are still much in vogue.

What should I state about the present book? In principle, there are so many books already published describing in detail all methodological approaches, tips and hints on 2-D gel slabs, that an additional one would seem to be pleonastic. Yet, by looking at the table of contents and at the list of chapters (no less than 15), it is easy to note that this book is a very special one in the panorama of manuals published up to the present. Whereas all the others are "cookbooks" giving just about only recipes, this one, on the contrary, gives mostly and perhaps only theory, a field too much ignored in all treatises on 2-D gel maps. So, I believe that this is a most useful and unique approach, in that it would help users to avoid common pitfalls due to ignorance of the basic theoretical mechanisms underlying the technique, including data handling and proper tools for spot analysis. Of course potential users had

better have a minimum of mathematical background in order to be able to understand all the theories here proposed.

Milano, 30th April 2015

Pier Giorgio Righetti
Honorary Professor, Politecnico di Milano
Milano, Italy

References

1. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021
2. Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals Humangenetik 26:231–243
3. Scheele GA (1975) Two-dimensional gel analysis of soluble proteins. Characterization of guinea pig exocrine pancreatic proteins J Biol Chem 250:5375–5385
4. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–681
5. Righetti PG (1983) Isoelectric Focusing: Theory. Methodology and Applications. Elsevier, Amsterdam, pp 1–386
6. Righetti PG, Simò C, Sebastiano R, Citterio A (2007) Carrier ampholytes for IEF, on their fortieth anniversary (1967-2007), brought to trial in court: The verdict. Electrophoresis 28:3799–3810
7. Bjellqvist B, Ek K, Righetti PG, Gianazza E, Goerg A, Postel W, Westermeier R (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. J Biochem Biophys Methods 6:317–339
8. Gianazza E, Giacon P, Sahlin B, Righetti PG (1985) Non-linear pH courses with immobilized pH gradients. Electrophoresis 6:53–56
9. Righetti PG, Gelfi C (1984) Immobilized pH gradients for isoelectric focusing. III Preparative separations in highly diluted gels J Biochem Biophys Methods 9:103–111
10. McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR 3rd (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. Anal Chem 69:767–776
11. Washburn MP, Wolters D, Yates JR 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotech 19:242–247

Preface

As commonly acknowledged, two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is one of the most widespread techniques for the separation of complex protein extracts, above all in the research field, for the identification of candidate biomarkers of different biological effects (pathologies, drug effects, ripening effects, etc.). Notwithstanding its being quite time-consuming and laboratory intensive, it is still one of the most exploited techniques for the separation of protein extracts due to its low cost as compared to high-throughput high-resolution methods, and its versatility. SDS 2-D PAGE can also be considered nowadays as a phase of sample pre-treatment to enhance or facilitate the subsequent analysis based on mass spectrometry, and remove the masking effect of the most abundant proteins.

The achievement of the final 2-D map is based on a multistep experimental procedure involving: sample preparation and pretreatment, isoelectric focusing, interfacing of the first and second dimension, separation according to the molecular mass, and staining and destaining of the final 2-D map. Once the final maps are obtained, they have to undergo a final step usually named "differential analysis," providing sets of candidate biomarkers by comparison of maps from different classes of samples (e.g., control vs. pathological, control vs. drug treated etc.). Actually, this final step is in turn based on a multistep complex procedure based on the exploitation of image analysis tools to provide information about the differences existing between the groups of investigated samples, i.e., the candidate biomarkers. In most cases the comparison of different 2-D maps is accomplished by dedicated software packages guiding the operator through a wizard-like procedure for noise removal (background, artifacts, etc.) and image warping, with the final aim of aligning gel images and matching protein spots across gels, to open the way toward the final quantification of spot intensities across all gel images. Once protein spots are matched and quantified, statistical methods can be applied to identify the relevant upregulation and downregulation.

The present volume is focused on deepening the analysis of 2-D maps by bioinformatics tools for what regards both the image analysis process to detect and quantify protein spots and the statistical analysis carried out to identify candidate biomarkers (i.e., spots upregulated or downregulated across samples).

Two main approaches to the analysis of 2-D maps images are available:

- The first involves a step of spot detection on each gel image, to provide a final list of spots present in all the investigated gels, each characterized by its volume: the final differential analysis is then performed on the spot volume dataset obtained.
- The second is instead based on the direct differential analysis of the 2-D maps images following a pixel-based procedure.

In the traditional approach spot boundaries are identified on each gel and spots are matched across multiple gels using a reference or master gel. However, this approach suffers from an important drawback: when the matching method fails, missing values are introduced in the spot volume data table; they can be due to the absence of a spot on the gel (in this case the missing value is a true zero) or to a failure in matching (in this case instead the missing value cannot be substituted by zero). Another problem regards the

definition of spot boundaries, which is particularly challenging in the case of overlapping spots. The pixel-based approach can overcome these problems, but both approaches rely on proper gel alignment; moreover, the pixel-based method is computationally intensive.

The volume is structured in four parts. The first part is devoted to deepening the problem of 2-D maps reproducibility and maps modeling. The validity of the results obtained by the final differential analysis deeply depends on the choices made during the experimental planning. This aspect is addressed in the first part of the book, to provide general good practices for a correct experimental design. In this part, the problem of spot overlapping is also addressed, and the main software packages available for 2-D maps analysis are presented.

The second part instead is devoted to spot-based methods: algorithms for maps denoising, background removal, and normalization are presented; the problem of image warping and spot detection and matching are presented and the most widespread algorithms available are described in detail.

The third part is mainly devoted to the description of classical and multivariate statistical methods that can be applied to spot volume datasets for the identification of candidate biomarkers.

The last part finally is focused on direct image analysis tools through pixel-based approaches.

The mathematical and statistical procedures are described from a theoretical point of view, to provide the basis for their correct applications, but examples of applications are also provided. The book is in fact thought to be of use for both the insiders of 2-D map image analysis and the researchers exploiting 2-D maps analysis in a wizard-like procedure: to the first ones the book is intended to as a compendium of the most recent applications, while for the second ones as a guide to help in the understanding of all the main steps of image analysis, to avoid errors and misinterpretations during the image analysis process.

Alessandria, Italy *Emilio Marengo*
Elisa Robotti

Contents

Contributors

Massimo Alessio • *Proteome Biochemistry, San Raffaele Scientific Institute, Milan, Italy*
Alfonso Caiazzo • *WIAS Berlin, Berlin, Germany*
Carlo Vittorio Cannistraci • *Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Technische Universität Dresden, Dresden, Germany*
Sebastien C. Carpentier • *Department of Biosystems, Faculty of Bioscience Engineering, K.U. Leuven, Leuven, Belgium; SYBIOMA: Facility for Systems Biology Based Mass Spectrometry, Leuven, Belgium*
Daniela Cecconi • *Mass Spectrometry & Proteomics lab, Department of Biotechnology, University of Verona, Verona, Italy*
Marco Demartini • *Department of Sciences and Technological Innovation, University of Piemonte Orientale, Alessandria, Italy*
Francesco Dondi • *Department of Chemical and Pharmaceutical Sciences, University of Ferrara, Ferrara, Italy*
Elisa Ficarra • *Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy*
Howard B. Gutstein • *Anesthesiology and Perioperative Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA*
Marcello Manfredi • *Department of Sciences and Technological Innovation, University of Piemonte Orientale, Alessandria, Italy; High Resolution Mass Spectrometry Lab, ISALIT SRL, Spin-off of University of Piemonte Orientale, Alessandria, Italy*
Nicola Marchetti • *Department of Chemical and Pharmaceutical Sciences, University of Ferrara, Ferrara, Italy*
Emilio Marengo • *Department of Sciences and Technological Innovation, University of Piemonte Orientale, Alessandria, Italy*
Simona Martinotti • *DiSIT—Dipartimento di Scienze e Innovazione Tecnologica, University of Piemonte Orientale, Alessandria, Italy*
Martin H. Maurer • *Department of Physiology and Pathophysiology, University of Heidelberg, Heidelberg, Germany; Mariaberg Hospital for Child and Adolescent Psychiatry, Gammertingen-Mariaberg, Germany*
Jeffrey S. Morris • *Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA*
Massimo Natale • *Global ICT, Unicredit SpA, Milano, Italy*
Maria Chiara Pietrogrande • *Department of Chemical and Pharmaceutical Sciences, University of Ferrara, Ferrara, Italy*
Fabio Quasso • *Department of Sciences and Technological Innovation, University of Piedmont Orientale, Alessandria, Italy*
Elia Ranzato • *DiSIT—Dipartimento di Scienze e Innovazione Tecnologica, University of Piemonte Orientale, Alessandria, Italy*
Elisa Robotti • *Department of Sciences and Technological Innovation, University of Piedmont Orientale, Alessandria, Italy*

Cristina-Maria Valcu • *Department of Behavioral Ecology & Evolutionary Genetics, Max Planck Institute for Ornithology, Seewiesen, Germany*
Mihai Valcu • *Department of Behavioral Ecology & Evolutionary Genetics, Max Plank Institute for Ornithology, Seewiesen, Germany*
Hua-Mei Xin • *College of Physics & Electronics, Shandong Normal University, Shandong, China; CREATIS, INSA Lyon, University of Lyon, Villeurbanne, France*
Yuemin Zhu • *CREATIS, INSA Lyon, University of Lyon, Villeurbanne, France*