Computational Biology and Chemistry 61 (2016) 1–7

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research article

Deceptive responsive genes in gel-based proteomics

Sara Hamzelou[a], Hossein Askari[a,*], Nona Abolfathi Nobari[b]

[a] Biotechnology department, Faculty of New Technologies and Energy Engineering, Shahid Beheshti University, G.C. Evin, Tehran, Iran
[b] Mechanical Engineering Department, McGill University, 845 Sherbrooke, Montreal, Quebec, Canada

ARTICLE INFO

Article history:
Received 15 April 2015
Received in revised form 10 November 2015
Accepted 1 December 2015
Available online 3 December 2015

Keywords:
Two-dimensional electrophoresis (2DE)
Proteomics
Mathematical equation

ABSTRACT

The standard method of the global quantitative analysis of gene expression at the protein level combines high-resolution two-dimensional gel electrophoresis (2DE) with mass spectrometric identification of protein spots. One of the major concerns with the application of gel-based proteomics is the need for the analytical and biological accuracy of the datasets. We mathematically and empirically simulated the possibility of the technical regulations of gene expression using 2DE. Our developed equation predicted a detectable alteration in the quantity of protein spots in response to a new protein added in, with various amounts. Testing the predictability of the developed equation, we observed that a new protein could form deceptive expression profiles, classified using prevalent tools for the analysis of 2DE results. In spite of the theoretically predicted overall reduction of proteins that resulted from adding the new protein, the empirical data revealed differential amount of proteins when various quantities of the new protein were added to the protein sample. The present work emphasize that employment of 2DE would not be a reliable approach for biological samples with extensive proteome alterations such as the developmental and differentiation stages of cells without depletion of high abundant proteins.

1. Introduction

Two-dimensional gel electrophoresis (2DE) is a central high-throughput technique in proteomics that has served in many studies for investigating gene expression. This technique sorts proteins extracted from biological samples of interest in two steps, first by isoelectric point and then by molecular weight. The spots displayed on polyacrylamide gels represent the quantification of expressed genes at the protein level. Generally, a reliable extraction and quantification method is employed, depending on the sample. The protein concentration applied to a gel can be varied between several micrograms to one milligram. The first dimension is performed by rehydrated IPG strips in a solution containing the necessary additives and, optionally, the desired protein concentration. In the isoelectric focusing stage, the pH gradient of IPG strips helps the proteins to separate according to their isoelectric point under the electric field strength. In the second dimension, the separation of proteins is performed using SDS-PAGE.

Multiple levels of gene regulation mediate the differentiation of cells during organ morphogenesis. In this way, gel-based proteome analysis has been frequently employed to reach greater overall proteome coverage in uncovering proteomes that are responsible for different developmental stages. Analysis of proteome dynamics, during light-induced development of rice (*Oryza sativa*) chloroplasts from etioplasts using quantitative 2DE, showed the exclusive identification of 237 etioplast proteins in comparison with 118 proteins which were exclusively identified using the shotgun approach reported by Zychlinski et al. (2005) (Kleffmann et al., 2007). Hence, 2DE analysis was preferred to identify molecular processes that may be involved in the regulation of plastids during the early illumination phase. A study on the proteome expression pattern of maize plastids during photo morphogenesis processes at five time points (0, 2, 4, 12 and 48 h) showed that one third of the reproducible spots first appeared after two hours from greening initiation (Lonosky et al., 2004). A dynamic pattern of gene expression can be concluded from the great alteration in 2D gel patterns during different seed developmental stages. Interestingly, around 50% of the most abundant proteins at the final stage of barley seed development

* Corresponding author. Fax: +98 21 22431964.
E-mail addresses: sara_hamzelou@yahoo.com (S. Hamzelou), askarihossein@yahoo.com (H. Askari), nobari_nona@yahoo.com (N.A. Nobari).

http://dx.doi.org/10.1016/j.compbiolchem.2015.12.001
1476-9271/

appeared, whereas these were not detectable in the first stages (Finnie et al., 2002). Proteome analysis of soybean seeds revealed the protein expression profile to be approximately the same at the first developmental stages while around 60% of the total spots were seed storage proteins that were only detected at the final stage (Hajduch et al., 2005). A superficial judgment about proteome maps obtained from the skin tissue of grapes in three stages of ripening (Deytieux et al., 2007) showed a large increase in quantity of some protein spots at the maturity stage compared with those spots in two other stages.

During protein expression analysis using 2DE, all gels have to experience the same protein amount, which is both the prerequisite and the challenge for proteome analysis. This fact, along with a large increase or decrease in the amount of some proteins during the developmental stages, results in the elimination or presence of some less abundant ones. Here, we have indicated a few examples of studies on developmental phases that investigate proteome analysis by utilizing two-dimensional gel electrophoresis regardless of depleting the high abundant proteins. Although depletion of high abundant protein in human, bovine (Faulkner et al., 2011; Hannan et al., 2009) and plant organs (Kim et al., 2013; Gupta et al., 2015) are investigated in several researches, there are several studies carried out without considering this fact.

In this paper, we constructed a mathematical method which states that the accuracy of 2DE-based proteome analysis needs to be approved through the determination of a certain threshold level. Along with the mathematical method, an experimental one was simulated by adding a reference protein (BSA) as high abundant protein as well, in order to demonstrate the deceptive gene expression.

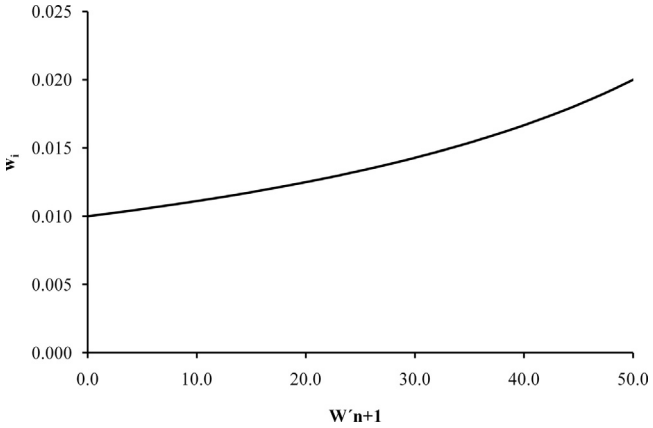

Fig. 1. The quantity of the first eliminated protein (w_i) in addition of new protein (W'_{n+1}). We computed the quantity of the unknown variable (w_i) by putting the known variables including ε, W and W'_{n+1} into the equation *x*. The amount of total protein is represented by total percent volume of the 2-DE gel, thus W constantly accounts for 100. we considered the quantity of W'_{n+1} from 1 to 80 μg or 1.2 to 50%vol. ε is determined 10-4 with taking into account the capability of the software in detection of protein spots percent volume.

2. Materials and methods

2.1. Plant preparation and growth conditions

Rice seeds (*O. sativa* cv. IR651) were obtained from International Rice Research Institute (IRRI). Seeds were germinated in plates at 28 °C. Three-days-old seedlings were transplanted into full strength Yoshida solution and grown in growth chamber at 30 °C/26 °C (day/night), photoperiod 16 h, relative humidity 70%, and a photon flux density 250 $\mu Mm^{-2} S^{-1}$. The root of 75 days old plants were harvested for protein analysis.

2.2. Protein extraction

Total protein was isolated from roots by TRIzol reagent according to manufacturer's guidelines (Molecular Research Center, Inc., Cincinnati, OH). The extracted protein was dissolved in lysis buffer consisted of 8 M Urea, 4% CHAPS, 1% DTT, 1% pH 3–10 ampholytes, 35 mM Tris base and stored in aliquots at −80 °C until further use. Protein concentration was determined by Bradford assay kit (BioRad, Hercules, CA, USA) using BSA as the standard.

2.3. Serial addition of BSA to the root protein

Different amounts of bovine serum albumin fraction *V* (5, 10, 20 and 40 μg) were separately added to 80 μg of root protein, and then 80 μg of the resulting protein solution was subtracted and loaded onto the analytical gels. One of the samples was left without the addition of BSA and nominated as a standard sample.

2.4. Image and data analysis

The gels were scanned using ImageScanner™ III (GE Healthcare Bio Science AB) at a resolution of 600 dots per in. The data were analysed using ImageMaster 2D Platinum 4 software. Protein spots were quantified based on their relative volumes (%vol), computed by the software itself as below, where Vol_S is the volume of spot '*s*' in a gel containing '*n*' spots.

$$\%Vol = \frac{Vol}{\sum_{s=1}^{n} Vol_s}$$

2.5. Classification of co-expressed genes

Hierarchical clustering was carried out using Cluster software version 2.11 (http://rana.lbl.gov/EisenSoftware.htm). Clustering of protein expression patterns was performed as described by the software user manual. Input data for pre-processing were the percentage volume of each protein spot at the BSA treatments by percentage volume of the same protein spot at control conditions (no BSA).

3. Theory/calculation

3.1. Equation generation

The alteration rate of the protein quantity as a result of BSA addition can be modeled using the following equations. In a protein mixture with *n* different types of proteins with various quantities of w_1 to w_n, the total protein quantity (TPQ), designated by *W*, can be found as below:

$$w_1 + w_2 + \cdots + w_i + \cdots + w_n = W \quad (1)$$

where w_i is the quantity of the *i*th protein. It is also assumed that the *i*th protein has the lowest quantity in the mixture. Adding a new protein type indexed by $n+1$ with the quantity of w_{n+1} to the mixture raises the TPQ of the solution to:

$$W_{new} = W + w_{n+1} \quad (2)$$

The appropriate TPQ for the experiment is *W*. Therefore, the quantity of *W* is obtained from the new quantity of W_{new}. In this regard, the new sample of the solution will be comprised of w'_1 to w'_{n+1}, where

$$w'_1 + w'_2 + \cdots + w'_i + \cdots + w'_n + w'_{n+1} = W \quad (3)$$