Advances in two-dimensional gel matching technology

S. Church
Nonlinear Dynamics Ltd., Cuthbert House, All Saints, Newcastle upon Tyne NE1 2ET, U.K.

Abstract

For many years, two-dimensional gel electrophoresis has been the method of choice for the investigation of complex mixtures of proteins. Although there are a number of emerging technologies that can be applied to proteomics, none can yet yield routinely the breadth of information available from two-dimensional gels. To be able to obtain instant information regarding molecular mass and pl, as well as to highlight quickly the expression changes or unique proteins across a gel series requires sophisticated and powerful image analysis software. The range of software products offered by Nonlinear Dynamics covers all levels of user application and throughput, from the user-guided Phoretix two-dimensional approach, when working with a small number of gels, to the automatic processing of large numbers of gels with minimal user intervention with Progenesis. Integration of the analysis software with powerful database components allows advanced gel comparisons and data mining to be performed with statistical verification of the results. Spot pick lists can be quickly created and automatically linked to a number of commercially available spot picking robots further increasing the support for proteomics research. The importance of image analysis for accurate, reliable and meaningful results will be discussed. Recent advances in development, with particular attention placed on the impact of noise contamination within gels, are illustrated and how the Progenesis product from Nonlinear Dynamics can be utilized to get the most from two-dimensional gel electrophoresis is shown.

Introduction

There are numerous emerging technologies available to the current crop of proteomics researchers, each attempting to provide improved separation, resolution and automation than is currently available [1]. Column-based separations are becoming even more popular, but after a number of initial separations generally all return to the bedrock of proteomics, gel electrophoresis. An example of this approach is the popular ‘four-dimensional separation’ technique in which column separation (ion-affinity chromatography and reverse-phase HPLC) are sequentially combined with preparative miniprep Isoelectric focusing (Rotofor®) to produce preserparated fragments in preparation for SDS/PAGE and MS [1].

Although extremely useful, this combination of present technologies is still not as powerful as, or universally applicable to, the plethora of sample and experiment types commonly undertaken by proteomics researchers. This therefore maintains the focus of proteomics research firmly in the realm of 2DE (two-dimensional gel electrophoresis) [2].

The realization that the computer is now a vital component in all aspects of research today, means that, to produce results of a high standard, a scientist has to rely on software [3]. Therefore, in addition to the time spent on experimentation, time must also be made available for computer analysis of the laboratory data produced. If the goal is to produce and process routinely a large number of two-dimensional gels, then the reduction in time spent away from the bench is vital. The ability to improve the throughput at both the laboratory and software levels is therefore always high on the agenda.

Recent developments and improvements in software algorithmic design and implementation are discussed with reference to Progenesis Discovery 2DE analysis software produced by Nonlinear Dynamics. A number of articles have described previously the use and assessment of Progenesis in various proteomics laboratories [4–6]. The reader is directed towards these for further details illustrating the power and applicability of Progenesis analysis software to their...
research. This paper concentrates mainly on the most recent improvements to noise reduction techniques.

**Tackling the key issues**

A number of key factors play on the minds of the proteomics scientist. Commonly, the main focus of proteomics research is the identification of expression changes between sample treatments or across a time course. Obvious questions therefore come to mind, ‘which proteins have changed and how big is this change?’ Of course the crucial question, ‘Is the observed change I am seeing statistically significant?’ must also be addressed. On top of this the researcher needs to ask other vital questions, ‘am I seeing only the most abundant spots?’ ‘What is happening at the other end of the spectrum with the expression of the low-abundance spots?’

A further complication can arise when working with limited sample availability, as is often the case in a proteomics study. There may only be enough sample to run a single gel. Therefore the analysis of the gel needs to be objective and the data needs to be both accurate and reliable, no matter what the resultant gel looks like. This is of particular concern when working with sensitive stains such as Sypro Ruby. These stains commonly leave staining artifacts on gels, which can severely hamper accurate and meaningful analysis.

**High throughput proteomics, a reality**

The availability of sophisticated image analysis and data mining software for the fast and accurate analysis of two-dimensional gels has been the ultimate goal for Nonlinear Dynamics since its conception in 1989. The rapid advances introduced into this field have allowed two-dimensional gel analysis software to be produced, which increases objectivity, crucially removes subjectivity and allows high throughput proteomics to become a reality from 2DE.

Full automation of gel analysis can be achieved using Nonlinear Dynamics 2DE gel analysis software. In addition to the advanced noise reduction, detection and matching algorithms employed, a wizard approach (Figure 1) is taken to allow users to set up and define quickly any number of experiments, which can all be automatically processed in...
The power of INCA and whole image background subtraction on successful spot detection

Progenesis Discovery was utilized in a study where the sensitive fluorescent stain Sypro Ruby was used. (A, C, E) The gel image, the presence of noise spikes on the gel and the background levels associated with the image before INCA processing. (B, D, F) The same regions after INCA processing. The influence on spot detection can be seen in (G, H) which show how only the real spot material has been detected.

sequence. This approach allows many separate gel images, in many experiments, to be analysed with no user intervention. Thus valuable time can be focused in the laboratory and not in front of a computer screen [7].

Advanced noise reduction and data QC (quality control) techniques

The cornerstone of this reliable automated approach to image analysis is the INCA (Intelligent Noise Correction Algorithm) that intelligently cleans and processes the two-dimensional gel images, and Data QC that allows the quality of the data to be scrutinized. INCA allows the recognition of noise in the image and therefore the accurate definition of actual spot material (Figure 2). Figure 2 shows data, part of a larger data set obtained from a multinational pharmaceutical company, which exhibits the commonly observed Sypro Ruby stain crystals throughout the gel. In the past, it would have been impossible to obtain meaningful quantitative data from such gels. Figures 2(A), (C), (E) and (G) show the
image and associated noise before noise removal using INCA. Figures 2(B), (D), (F) and (H) show the same region of the gel post-INCA processing. It can be seen that the noise will heavily skew the quantitative data obtainable from this image and make any fold differences observed totally unreliable. After INCA application all the noise gets removed without altering the actual spot measurements, allowing accurate spot volume quantification to take place and hence reliable protein expression changes to be investigated. Panels E and F illustrate the level of background intensity associated with this image and show how Progenesis Discovery delivers extremely accurate background subtraction based on features across the whole image. This method provides a more stringent and reliable method of subtraction than is obtainable using spot-based methods. Working in this manner the sensitivity of detection is extremely high, increasing the detection of the valuable low-abundance protein spots. INCA improves the reliability of all observed expression differences, since the influence of the noise is removed from the analysis. Therefore changes observed are real fold differences based on the actual raw data. Therefore the statistical significance of the change can be assessed easily, decreasing the number of false positives and negatives significantly.

Progenesis Discovery was also utilized in a recent study to examine 11 2DE gels stained with the common fluorescent protein stain Sypro Ruby. The analysis wizard was used to create two ‘averaged groups’, representing sets of replicate gels, and a fully automatic analysis was performed. Those spots showing a 2-fold or greater expression change between the two replicate groups were highlighted and the influence of noise spikes within the gel were studied (Figure 3). The example shown indicates that when not using the INCA-processed information, protein spot 1201 on the image (arrowed) would not have been flagged as exhibiting a significant expression change (1.9-fold change). However, it can be seen that this spot is contaminated with many noise spikes. When this noise is INCA corrected, the actual volume change of this protein between the samples is nearly 3-fold, and therefore the spot is clearly worth investigating further.

Progenesis Discovery allows accurate differentiation between what are real expression changes and those that are incorrectly highlighted. INCA prevents the presence of noise within spots from contributing to the protein volume and hence allows the identification of false negatives and positives. This maintains the high throughput requirement of proteomics, since time is not being wasted on any further study of meaningless proteins.

**Hunting for expression differences**

The ability to pinpoint confidently true expression differences between gels with minimal user intervention hinges on the ability of the software to recognize common proteins across any number of gels. Progenesis achieves this using ‘whole image warping’ before ‘matching’ (Figure 4). This warping automatically takes into account any running distortions within the gels and produces sets of overlaid images where common spots are superimposed. These common proteins across the gel images can then be confidently and accurately matched. Even when gels exhibit severe distortions, simple manual intervention allows highly accurate matching to be obtained from gels that in the past would have had to be discarded.

**Statistical validation and QC**

It is crucial that any expression changes observed are real changes and, as outlined above, INCA processing allows the influence of noise within a gel to be accounted for. Statistical validation is achieved using the powerful ‘Bootstrapping’ method of data re-sampling [8]. Nonlinear Dynamics have implemented this ‘Data QC’ methodology into Progenesis Discovery to allow confidence limits for any given spot measurement to be produced. This allows QC of the data, again removing the chance of false results and hence maintaining the high-throughput requirement.

In addition to this ‘Data QC’, both single and paired t tests can be performed on the analysis data. The data can also be validated using the ANOVA test with a 2DE gel set.

**Analysis of multiplexed 2DE gels**

As the requirements and expectations of 2DE increase, new technologies emerge in a bid to more accurately capture the sometimes small, but significant, changes occurring in proteomics experiments. Therefore a proteomics researcher requires software that is extremely sensitive and still maintains his confidence in the analysis. One such technological jump in 2DE has been the introduction of multiplexed gel-staining methods [9]. Samples are visualized using different fluorescent stains (such as the Cy2, Cy3 and Cy5 stains commonly used) and run on the same physical two-dimensional gel. Multiple images are then produced allowing the analysis of each sample to be performed.

Progenesis Discovery contains all the functionality required for this type of analysis and, as always, maintains the high-throughput requirements by analysing every image from every gel automatically. Even when working manually on the analysis the workload is significantly reduced. The advanced spot detection, warping and matching algorithms present in the 2DE analysis software from Nonlinear Dynamics means that all manual procedures need only be performed on one-third of the images in the entire experiment.

**From unknown spots to identified proteins**

Once spots have been identified as being of possible interest for further studies, the identification of the proteins is commonly required. Progenesis Discovery allows quick and simple creation of spot-pick lists, from both gels stained using the standard methods (Coomassie, Silver, Sypro etc.) and those stained using a multiplexed staining approach.
Figure 3 | Identification of false negatives using Progenesis Discovery

In this study, false negatives and false positives were correctly identified. Here two gels are compared for expression differences. The native gel image and INCA processed images are compared. The expression change of protein 1201 would have been incorrectly missed without the use of INCA.

Compatibility is available for transferring the pick lists to a number of commercially available spot-picking robots, making the process as automatic as possible.

Completing the proteomics cycle

Two-dimensional electrophoresis and gel analysis are only the first part of the proteomics cycle. As discussed previously, the results obtained from the analysis must be accurate and reliable. Commonly, spot-pick lists are created for MS to identify the proteins that have been highlighted as being of possible interest for further study. The recent introduction of database software that works in conjunction with Progenesis Discovery allows all the gel analysis data from many experiments to be archived in a relational database. Combining this with the MS data and sample information obtained allows the creation of a vital central store of the proteomics data. Advanced data mining tools can then be
utilized to obtain further insight into the data produced \[10\].

**Conclusions**

The use of gel electrophoresis is still the most useful tool in the armoury of the proteomics researcher today. It quickly provides valuable information on the proteins under study, such as molecular mass and pI, presence and absence and expression changes. With the impact of column-based methods linking into 2DE, proteomics has never been more powerful and allows the in-depth study of the expression changes occurring in the low-abundance proteins.

To be able to achieve high throughput, maintaining at the same time the highest level of accuracy and reliability of gel data, a proteomics scientist must put his trust in dependable analysis software. Progenesis Discovery provides fully automated 2DE image analysis of gels stained with the traditional stains (Coomassie, Silver and Sypro) and also applies the same advanced detection methods to gels stained using a multiplexed approach. Continued development on Nonlinear Dynamics 2DE gel analysis \[11\] and database software allows the production of leading edge software to meet the challenges set by proteomics research today and in the future.

The gel images shown were kindly provided by Dr H. Siitari (VIT Medical Biotechnology, Turku, Finland) and also from a large multinational pharmaceutical company.

**References**


Received 25 November 2003