Comparison of Three Commercially Available DIGE Analysis Software Packages: Minimal User Intervention in Gel-Based Proteomics

Yunyi Kang, Tanasit Techanukul, Anthanasios Mantalaris, and Judit M. Nagy

*J. Proteome Res.*, 2009, 8 (2), 1077-1084 • DOI: 10.1021/pr800588f • Publication Date (Web): 09 January 2009

Downloaded from http://pubs.acs.org on March 19, 2009
Comparison of Three Commercially Available DIGE Analysis Software Packages: Minimal User Intervention in Gel-Based Proteomics

Yunyi Kang,†,§ Tanasit Techanukul,†,§ Anthanasios Mantalaris,† and Judit M. Nagy*,‡

Department of Chemical Engineering and Chemical Technology, Imperial College London, London, SW7 2AZ, United Kingdom, and Institute of Biomedical Engineering, Imperial College London, London, SW7 2AZ, United Kingdom

Received August 1, 2008

Abstract: The success of high-performance differential gel electrophoresis using fluorescent dyes (DIGE) depends on the quality of the digital image captured after electrophoresis, the DIGE enabled image analysis software tool chosen for highlighting the differences, and the statistical analysis. This study compares three commonly available DIGE enabled software packages for the first time: DeCyder V6.5 (GE-Healthcare), Progenesis SameSpots V3.0 (Nonlinear Dynamics), and Dymension 3 (Syngene). DIGE gel images of cell culture media samples conditioned by HepG2 and END2 cell lines were used to evaluate the software packages both quantitatively and subjectively considering ease of use with minimal user intervention. Consistency of spot matching across the three software packages was compared, focusing on the top fifty spots ranked statistically by each package. In summary, Progenesis SameSpots outperformed the other two software packages in matching accuracy, possibly being benefited by its new approach: that is, identical spot outline across all the gels. Interestingly, the statistical analysis of the software packages was not consistent on account of differences in workflow, algorithms, and default settings. Results obtained for protein fold changes were substantially different in each package, which indicates that in spite of using internal standards, quantification is software dependent. A future research goal must be to reduce or eliminate user controlled settings, either by automatic sample-to-sample optimization by intelligent software, or by alternative parameter-free segmentation methods.

Keywords: Progenesis SameSpots • DeCyder • Dymension • proteomics • gel image analysis

Introduction

Two dimensional difference gel electrophoresis (2D-DIGE) is a powerful tool for the study of differential protein expression. Since samples prelabeled with fluorescent cyanine (Cy)-dyes are simultaneously run on the same gel using a pooled sample as a third internal standard,¹,² DIGE gives higher confidence when the differential protein profile is analyzed, minimizing gel-to-gel experimental variation. The inclusion of an internal standard pooled from aliquots of all biological samples to be studied in an experiment, usually labeled with Cy2, improves the matching of intra- and intergel images and allows normalization across all gels.¹,² The benefit of DIGE can only be realized by analyzing gel images with software tools specifically written for DIGE analysis. The evaluation of software packages dedicated to DIGE analysis is necessary so that users do not compromise valuable DIGE data by using a poor quality evaluation tool.

Several reports have been published addressing comparative evaluation between commercially available 2-DE software packages.³–¹⁰ Raman et al.⁶ provided guidelines by comparing two software packages (Z3 and Melanie) according to three criteria: spot detection, gel matching, and spot quantitation. They also listed the features of 10 available 2-DE analysis software packages. According to their study, Z3 performed better in spot detection while Melanie was able to predict much better in spot quantitation. Recently, Karp et al.¹⁰ compared the quantitative aspect of pre- and postelectrophoresis stained gels using two different methods of image analysis, traditional (DeCyder) and SameSpots (Progenesis) approaches. In their study, it was deduced that the SameSpots is particularly good for the analysis of poststained gels by increasing the proportion of low volume spots.

The software packages that can be used for the analysis of 2-DE gel images are: ImageMaster2D, DeCyder (GE Healthcare), Progenesis (Nonlinear Dynamics), MELANIE (Geneva Bioinformatics), AlphaMatch 2D (Alpha Innotech), PDQuest (Bio-Rad Laboratories), Delta2D (Decodon) and Dymension (Syngene). Some of these software packages have been further developed for DIGE-based experiments. The three commercially available DIGE analysis software tools we compared, using stem cell culture media samples conditioned by HepG2 and END2 cell lines, were DeCyder V6.5 (GE Healthcare), Progenesis SameSpots V3.0 (Nonlinear Dynamics), and Dymension 3 (Syngene).

The purpose of the experiment was to identify the commonly expressed proteins by two cell lines: HepG2 (human hepatocellular carcinoma cell line) and END2 cells (mouse endodermal-like P19 embryonal carcinoma subcell line).¹¹,¹² In our previous
work, we identified factors present in HepG2 and END2 conditioned media that had a great effect on embryonic stem cell development. Identifying these factors is necessary for the understanding of embryonic stem cell differentiation as well as the signaling pathways of embryogenesis. When analyzing two proteomes, the use of DIGE, together with the most suitable software tool, is essential. First, we used three different packages to analyze these difficult images to establish how each package handled the difficult aspects of this experiment, and second, we tested the user friendliness of each package. We report here the results we obtained from the perspective of a proteomics user. The user intervention was kept to a minimum for two reasons, first, because our future aim is to establish a standardized 2D based proteomics workflow and second to make sure that the user-friendly aspect of each software package was compared objectively. Our study includes quantitative tests as well as a general evaluation of the three software packages. Quantitative assessment was performed by comparing the number of detected spots, the cross-matched spots, and the expression differences of selected spots. A second study was also performed using breast cancer cell lines to confirm the performance of the three software packages on a different set of images.

**Materials and Methods**

**Cell Culture and Sample Preparation.** All chemicals, unless stated, were purchased from Sigma, UK. HepG2 cells (ATCC HB-8605, UK) were cultured in T-75 flasks (Fisher Scientific, UK) using Dulbecco’s modified eagle’s medium (DMEM; In-vitrogen, UK) supplemented with insulin-transferrin-selenium supplement (ITSS; Roche, Germany), 0.1 mM β-mercaptoethanol, HEPES, 1 mM t-glucose and 1% Penicillin/streptomycin at a seeding density of 5 × 10^4/cm^2. END2 cells (gift from Prof. Mummery, Hubrecht Laboratory, Holland) were also cultured in the medium described above. Four batches were prepared from consecutive passage for each cell line to create biological replicates. After culturing for four days 50 mL of each conditioned media (CM) was collected and filter-sterilized through a 0.22 µm filtration unit and stored at −80 °C. Thawed CM were 1000-fold concentrated to the final volume of 50 µL and buffer-exchanged into a solution (20 mM Tris, 1 mM EDTA, 1 mM PMSF) using Centricon plus-70 (Millipore, UK; 5000 MWCO) and Microcon centrifugal filter device (Millipore, UK; 3000 MWCO). The protein concentrate was resuspended into a sample buffer containing 2 M Thiourea, 7 M Urea, 4% w/v CHAPS, 1% w/v ASB14, 1% w/v DTT, 1% v/v IPG buffer pH 3–10 NL. The amount of protein was then quantified using 2-D Quant Kit (GE Healthcare, UK).

**Minimal Cy-dye Labeling.** The proteins were labeled with Cy-dyes (CyDye DIGE Fluor, minimal labeling kit, 25–8010–65; GE Healthcare, UK) according to the manufacturer’s protocol. Briefly, the pH of the sample was adjusted to pH 8.5 for Cy dye labeling, 50 µg of each batch of HepG2-CM and END2-CM protein samples was labeled with 400 pmol of Cy3 or Cy5 dyes respectively in a reciprocal manner. The internal standard pooled from the aliquots of all the samples was labeled with Cy2. The labeling reaction was carried out by incubation on ice for 30 min and quenched by the addition of 10 nmol of lysine. Subsequently, equal volume of 2 × sample buffer containing 2% w/v DTT, 2% v/v IPG buffer pH 3–10 NL was added to each sample. Cy2, Cy3 and Cy5 labeled samples were then pooled and applied by cup loading.

**Two Dimensional Gel Electrophoresis.** Eleven cm IPG strips (GE Healthcare, UK) with a nonlinear gradient over the pH range of 3–11 were passively rehydrated overnight with 200 µL of Rehydration solution (2 M Thiourea, 7 M Urea, 4% w/v CHAPS, 1% w/v ASB14, 0.3% v/v DTT, 0.5% v/v IPG buffer pH 3–10 NL). Cy-dye labeled protein samples were loaded using cup-loading application and focused as follows: (1) 500 V, 1 s, gradient; (2) 500 V, 4 h, Step and hold; (3) 1000 V, 1 h, gradient; (4) 1000 V, 1 h, Step and hold; (5) 3500 V, 4 h, Step and hold; (6) 8000 V, 5 h, Step and hold, for a total of 70 kVhr. Prior to second dimension separation the focused strips were equilibrated in the equilibration buffer (50 mM Tris-HCl pH 8.8, 6 M urea, 30% v/v glycerin, 2% w/v SDS, 20 mM DTT, 0.01% w/v bromophenol blue) for 15 min. Second equilibration was carried out in the equilibration buffer containing 4% w/v iodoacetamide instead of DTT for 15 min. Subsequently, equilibrated strips were overlaid onto a precast Bis-Tris 4–12% polyacrylamide gel (Bio-Rad, UK) and ran at 150 V till the bromophenol blue dye migrated to the bottom of the gels.

**Image Acquisition.** DIGE gels were imaged using the Diversity (Syngene, Cambridge, UK) imaging system immediately after the second dimension SDS-PAGE to prevent diffusion of protein spots and subsequent signal loss. Gels were first placed on a low fluorescent glass plate and the thick edges of the gels were removed with a scalpel blade to enable LED side illumination. The exposure time was optimized to achieve maximum dynamic range. Images were then cropped together by applying the same region of interest to all the gels making sure that the regions analyzed in the different software packages were identical. Images were then exported in tagged image file format (TIFF) which is the required format for the software analysis packages. For DeCyder analysis images were saved with a specific file name format, such as Gel01 Cy2 Standard-gel, that allows for DeCyder software to recognize the files.

**Image Analysis.** Four biological replicate gels from each group, namely HepG2-CM and END2-CM, and four internal standard gels were aligned and analyzed using three different software packages; DeCyder V6.5 (GE-Healthcare, Buckinghamshire, UK), Progenesis SameSpots V3.0 (Nonlinear Dynamiscs, Newcastle, UK), and Dymension 3 (Syngene, Cambridge, UK). In each software package the statistically ranked top fifty spots were selected based on p-value of ANOVA (p < 0.05) and then maximum fold change. Among the selected fifty spots, mis-matched spots were counted by manual screening under two conditions: (1) aligned without landmark (definition of landmark: spots determined by users to represent same proteins across replicate gels) and (2) aligned with five landmarks.

**Spot Detection and Analysis by DeCyder.** Spots were encoded in each gel-set using the Differential In-Gel Analysis (DIA) mode of DeCyder. The area of interest was set to the whole image since images were already cropped for uniform analysis to enable unbiased comparison of the software packages. Spots were eliminated for the BVA (Biological Variation Analysis) by applying the exclude filter based on the following threshold: Slope > 1.1, Area < 100, Volume < 100 000, Peak Height < 100 and > 64 000. All the threshold values were default except for the volume, which was chosen following manual assessment of the exclusion of artifacts. The estimated number of spots for each gel was set to default, 2500. Background subtraction and normalization was also performed on the images using default settings. Each sample was grouped in BVA mode and gel images were automatically aligned in two ways,
with no user intervention or with minimal user intervention based on five landmarks set by the user on each gel. Gels were then matched and the spots were compared.

**Spot Detection and Analysis by Progenesis SameSpots.** Gel images were aligned by automated calculation of alignment vectors after assigning five landmark vectors. To make a fair comparison with other software packages, gels were also processed without setting any landmarks. Automatic analysis was performed on all the aligned images using the analysis wizard. The aligned images were grouped to reflect the biological grouping and the statistically ranked list of spots were evaluated in the review stage of the software package. In addition, separate analysis for each gel was performed in order to calculate total spot number of individual gel. All the settings were left as defaults throughout the analysis.

**Spot Detection and Analysis by Dymension.** Cropped gel images were loaded and grouped using the multiplexed experimental wizard. Region of interest was selected as the whole image. Spots of the whole gels (those from internal standard as well as both groups) were detected automatically with default spot detection settings: Blur radius: 1.2, Peak limit threshold: 0.01, Splitting threshold: 0.001. Detection confidence ratio was set at 20 and separation confidence ratio was set at 20 according to the company’s recommendation. Noise filtering and background correction was performed automatically as default using the following parameters: Raw Volume/Height >10, Height >5. Gels were then automatically warped or ‘manual warping’ function was applied to each gel, which was the only possibility to improve the matching. Pool matching was carried out by automatic gel alignment. Statistical analysis was automatically done when pool matching was carried out.

**Spot Detection and Analysis by DeCyder.** Spot detection settings: Blur radius: 1.2, Peak limit threshold: 0.01, Splitting threshold: 0.001. Detection confidence ratio was set at 20 and separation confidence ratio was set at 20 according to the company’s recommendation. Noise filtering and background correction was performed automatically as default using the following parameters: Raw Volume/Height >10, Height >5. Gels were then automatically warped or ‘manual warping’ function was applied to each gel, which was the only possibility to improve the matching. Pool matching was carried out by automatic gel alignment. Statistical analysis was automatically done when pool matching was carried out.

**Results**

**General Comparison.** The general evaluation was carried out on the three software packages in terms of visualization, normalization method and workflow. All packages gave good graphical user interfaces, which make analysis easy and efficient to perform. Normalization is used for the accurate comparison of spot measurements. In DIGE experiments normalization should be calculated by using the internal standard since it should be the same on each gel. This ratiometric normalization was provided in all of the three software packages as a default setting. All three packages provided univariate and multivariate statistical analysis tools which enable users to compare experimental groups easily. Student’s t-test and ANOVA were used as basic statistical analysis tools in all three software packages. Each software company also markets their own statistical packages separately and these provide additional functionality. Minimal user intervention was needed in all three software tools. In the previous version of Progenesis SameSpots user intervention was only required or possible in the alignment stage of setting landmark vectors while the latest version provides prefilter function allowing users to exclude artifacts. DeCyder and Dymension required more manual intervention at the alignment stage as well as at spot detection, although it can be kept to the minimum using settings recommended by the companies.

The general workflow was compared among the three software packages. The analysis in the three software packages had three main steps: spot detection, alignment, and statistical analysis of spots. Progenesis SameSpots has incorporated a new approach leading to a different workflow from the other two software packages. While spots of each gel were detected before the matching in DeCyder and Dymension, Progenesis SameSpots performed alignment at the beginning and spots were subsequently detected with same spot outlines. In Progenesis SameSpots, aligned spots were automatically detected across gels and full matching was performed based on the identical spot outlines without the need of user intervention. In DeCyder, nonprotein spots were screened using the “exclusion filter” in DIA mode, whereas Progenesis SameSpots carried out this task as default. In our experience not enough artifacts were removed in DeCyder, therefore strict threshold in the “exclude filter” must be applied by the user because the default setting is too liberal. Full matching in DeCyder was carried out similarly to Progenesis SameSpots. Dymension had a different approach from the other two software packages: the spots were accepted and matched only when they were present in all the internal standard gels.

**Spot Detection and Alignment.** All the gel images were cropped together prior to loading into the software packages to make sure that the region of interest analyzed was identical. The aligned DIGE gel images in each software package were shown in Figure 1. Table 1 shows the result of spot detection performance. Overall, DeCyder detected more spots than the other two software packages. The spot numbers detected by DeCyder were highly dependent on the detection settings using 2000 spots as the initial default setting. When the ‘exclude filter’ was applied (see methods section for details), almost 1000 spots were excluded in all gels apart from gel 3 which had more spots filtered. The higher spot numbers observed in DeCyder results from the choice of inflection points for spot splitting. A spot, which was detected as a single spot in Progenesis SameSpots, was often interpreted as multiple spots in DeCyder as shown in Figure 2. In general, DeCyder is more liberal at spot splitting at inflection points, which often results in oversplits, whereas Progenesis SameSpots and Dymension are more conservative at spot splitting which results in the amalgamation of potentially different protein spots (Figure 2).

For the fair comparison of alignment, we carried out two experiments. In the first, we aligned gels without setting any landmark vectors, and in the second, we aligned with the five landmarks selected manually. We then analyzed the matching performance in each case. Since each software package provides different matching algorithms, the results were not directly comparable. In Progenesis SameSpots the first step of matching was the creation of an averaged gel which contained all the spots from each gel set including both the matched and the unmatched spots. Unmatched spots are matched to the virtual spots of the same boundary in Progenesis SameSpots. Dymension only accepted those spots which appeared in all the internal standard gels as mentioned above. This explains the relatively low matched numbers, 73, found by Dymension as compared to those by Progenesis SameSpots and DeCyder (Table 1). In DeCyder a master gel with the greatest number of spots was chosen and subsequently each gel was matched to this gel.

**Cross-Matching of the Most Highly Ranked Spots.** The statistical analysis tools of the three software packages ranked all spots according to the greatest fold changes across groups at 95% confidence level ($p < 0.05$), and we selected the top fifty of each for comparison. Dymension automatically highlighted the spots which appeared in all the internal standard gels passing only 73 spots across four gel sets. Among those 73 spots, only 3 spots were significantly different at the $p < 0.05$ level while Progenesis SameSpots and DeCyder passed ap-
proximately 150 spots (Table 2). In Dymension, therefore, we chose the top 50 spots, independent of the p value, to compare against the other two packages.

Cross-matched spots were confirmed manually, one by one, and multiple spots, which were detected as a single spot on the other package, were counted as one. Among those fifty spots we counted spots up-regulated in HepG2-CM and those up-regulated in END2-CM. Interestingly, in DeCyder 48 out of 50 spots were up-regulated in HepG2-CM, whereas only 32 and 34 spots were found to be HepG2-CM up-regulated by Progenesis SameSpots and Dymension respectively (Table 2). We found that DeCyder are more liberal in spot splitting, which might result in the generation of multiple spots. Only 5 spots were common to the three sets of 50 top ranked spots selected by the three software packages (Figure 3). Progenesis SameSpots and DeCyder shared the greatest number of cross-matched spots.

The accuracy of matching by each software package was also confirmed manually by judging if the top 50 spots were well-matched or mis-matched. When landmark vectors were not used, all the software packages gave poor matching performance (Table 2). DeCyder and Progenesis SameSpots correctly matched 29 and 33 spots out of the top 50, while Dymension matched 12 without landmarks. By using five landmarks, DeCyder and Progenesis SameSpots improved their matching performance, and Progenesis SameSpots outperformed the other two software packages by correctly matching 92% of spots (Table 2). Dymension only matched 22 spots correctly, even when the ‘manual warping’ function was used (this is the closest function to landmark setting).

Quantitative Analysis. Out of the 5 cross-matched spots we selected 3 (two of the cross-matched spots were discarded after the manual check of mis-matched spots) for quantitative analysis and compared the fold changes of the spots calculated by each software package. Surprisingly, whereas it was expected that the estimates of fold change between the three cross-matched spots would be the same, the three packages actually reported significant discrepancies in the estimates of fold changes. DeCyder estimated a much higher differential expression ratio ranging from 60 to 90 fold as compared to Progenesis SameSpots which gave a range from 15 to 40 (Figure 4a). In both cases these values are reflecting on/off differences rather than on/off changes (Figure 5b). Nevertheless, fold changes calculated by each software package remained significantly different as shown in

---

Table 1. Detected and Matched Spot Numbers in Each Gel Set

<table>
<thead>
<tr>
<th>Spot No. in gel</th>
<th>DeCyder</th>
<th>Progenesis SameSpots</th>
<th>Dymension</th>
</tr>
</thead>
<tbody>
<tr>
<td>set1</td>
<td>979</td>
<td>610</td>
<td>479</td>
</tr>
<tr>
<td>set2</td>
<td>984</td>
<td>593</td>
<td>524</td>
</tr>
<tr>
<td>set3</td>
<td>661</td>
<td>654</td>
<td>482</td>
</tr>
<tr>
<td>set4</td>
<td>1019</td>
<td>648</td>
<td>383</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matched spot No.</th>
<th>w/o.l</th>
<th>w.l.</th>
<th>w/o.l</th>
<th>w.l.</th>
<th>w/o.l</th>
<th>w.l.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gel set1</td>
<td>443</td>
<td>495</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gel set2</td>
<td>489</td>
<td>544</td>
<td>620</td>
<td>548</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Gel set3</td>
<td>405</td>
<td>415</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gel set4</td>
<td>1019</td>
<td>1019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* w/o.l., without landmark; w.l, with landmarks. *b* Separate experiment was created for each gel in order to calculate total spot number per gel. *c* “Spot No. in gel set” represents spots from each image that includes all three Cy-dyes.
Figure 5a. In general, the fold changes calculated by Progenesis SameSpots were less than those found by DeCyder and Dymension. The most obvious explanation of the observed differences is the way spot boundaries are defined by each software package. For example, the spot boundary for one of the well-resolved spots, spot 4, varied according to the software package as shown in Figure 5(c-e).

Discussion

In our experience spot detection is a critical step necessary for achieving high quality results at the end of data processing workflow. The advantage of using 2-DE image analysis software is that it performs objective spot detection and quantitation which cannot be done manually. Since none of the software packages are perfect, manual editing cannot be completely eliminated on account of incorrectly detected spots, namely oversplit or undersplit spots. Moreover, software may fail to detect some significant spots while incorrectly identifying artifacts as protein spots. Manual evaluation of detected spots is essential if minimal user intervention is used as a strategy for alignment and matching. Since Progenesis SameSpots does not detect spots from single gels, we loaded gels separately in new experiments in order to compare detected spot numbers from each individual gel. In Progenesis SameSpots the spot detection parameters are already set which leaves no option...
for user intervention. In Dymension users can set the detection parameters such as detection confidence or separation confidence ratios. DeCyder also allows users to set the initial detected spot number although it is recommended to be set at 2500. Spot detection and segmentation are two controversial areas since the number of real spots is not known. The spot segmentation approach is different in each software package. Some use spot boundaries (edge detection algorithm) while others collect continuous pixel groupings (pixel value collection). In particular the split points of partially overlapping spots are judged differently. Our sample contained abundant proteins that appeared as overlapping spots highlighting the segmentation problem (Figure 2). Our results show that each software package interprets spot detection and segmentation in a different manner, which results in different spot numbers. Dymension estimated less spots than the other two software packages possibly due to the conservative settings of spot detection. DeCyder estimated more spots than the other two software packages due to oversplitting as shown in this study. If the segmentation settings are too conservative, spots are amalgamated and post-translationally modified proteins might be missed. Spot detection algorithms require a number of parameters to be set by the operator and these must be tailored subjectively to the data set under consideration in order to achieve optimal segmentation; this takes a significant proportion of the operator’s time. For this reason full automation can only be achieved if the user-set parameters are reduced or eliminated. This may be done by automatic optimization with domain models which are themselves created by intelligent software; software which takes into consideration the sample type, equipment used and image quality. Alternatively, full automation may be achieved by parameter-free segmentation methods. In Progenesis SameSpots, the spot detection parameters cannot be edited in the main workflow, which results in the user not being able to add spots ignored by the software; however spot segmentation or mergence can be corrected by users after spot detection.

During the matching process of DIGE gels, the software packages started with the alignment of the internal standards before spot matching. Since the internal standard represents all the spots among the groups and replicates, the alignment is more accurate when using internal standard gels. In spite of the inclusion of the internal standard there were several discrepancies in matched spots because each software package provided different alignment and subsequent matching algorithm. Since Dymension’s matching algorithm is based on the perfectly matched spots across all the replicates it could be favored when stricter criteria of experimental variation are
required. However, this could result in missing spots when users need to take into consideration the interinternal standard experimental variation. The experiment discussed in this assessment was a good example for this problem. Dymension only included 73 spots for the statistical analysis, missing out some spots of interest. Since our sample is a complex mixture of proteins with biological replicates the matching became more challenging in the analysis. When ‘landmark’ function was used the accuracy of matching was dramatically increased in all the software packages. Progenesis SameSpots improved the matching accuracy up to 92% with landmarks. In Progenesis SameSpots the “landmark” function was integrated in the alignment step as default before spots were detected. Since spot boundaries were determined after alignment in Progenesis SameSpots, the spots of the same boundary on each gel were matched up maybe improving spot-matching efficiency. As a result of this if alignment was corrected Progenesis SameSpots had to recalculate the whole data set with the new corrections. Nonetheless none of the packages gave perfect matching suggesting that without user intervention this cannot be achieved for the types of samples analyzed in this study. The conditioned media samples are particularly difficult to analyze due to the presence of highly abundant proteins resulting in unresolved protein spot chains. Samples with higher resolution might provide different matching results. Therefore, the analysis was also carried out with gels that include more discrete spots and Progenesis SameSpots also outperformed the other two software packages in terms of matching accuracy (See supplementary data).

When we looked at the cross-matched spots among the top 50 ranked proteins to assess the consistency of analysis, only 5 proteins were completely matched across the three software applications. Since alignments and subsequent matching were performed in a different manner in each software package, inconsistencies of the ranked spots were expected.

Results of this study show that the software packages exhibited different expression levels when analyzed on several randomly selected spots as well as the completely cross-matched spots. DeCyder and Dymension showed wider range of responses to the investigated spots in terms of differential spot volumes. Quantitation, one of the most important information from the biologist’s point of view, depends on the results of the detection and matching steps. Researchers have studied the performance of different software packages on gel-based protein quantitation.5–6 These studies evaluated spot quantitation using ideal-Gaussian-distributed spots in artificially created gels only.6 In reality, spots may not be elliptical and gels usually have streaks, artifacts and other noises and the inclusion of these affects background subtraction and quantitation subsequently. The huge discrepancy observed in quantitation of spots with large fold changes between software packages must be due to different background estimation. Although each software package gave a relatively linear response to the wide dynamic ranges of spot intensity, different spot boundaries and matches resulted in the discrepancy in quantitation. Our study also showed that the spot boundaries of well-resolved spots were differently drawn as well in each software package, which affected the final data processing, quantitation of spots and statistical analysis (Figure 5).

All three packages offer the standard statistical tools of Student’s t-test and ANOVA. Recent papers have highlighted the pitfalls surrounding conventional application of these techniques, especially when concerned with the small sample sizes employed.6 It is particularly important to adjust the p-values to avoid the dangers of multiple hypothesis testing. Bonferroni correction is a widely used technique, where the p-value is divided by the number of tests to be performed, but this has been found to be too conservative. More advanced techniques involve performing a power analysis. For example, Student’s t-tests can be modified with a data permutation technique that estimates the false discovery rate.17 Experimental power is typically improved by increasing the number of replicates under study, either using the same sample (technical replicates) or with different samples from the same treatment group (biological replicates). Karp et al.18 showed that mixing the two types in the same statistical test, greatly increases the false positive rate when conventional tests are applied. They advocate the use of nested-ANOVA, where the test statistic is corrected by the ratio of biological variance to technical variance. In the three packages under test such nesting is not possible, so a correct analysis of mixed replicate types is inconvenient. One must also note that it has now been formally proved that technical replicates should not be run at the expense of biological replicates unless there is no choice e.g. unavailability of samples.16

Another important consideration regarding differential expression analysis is the impact of systematic errors such as channel-specific and spatial biases between gels. A recent study of DIGE experiments showed a loss of sensitivity with three multiplexed dyes compared to two, and different system gains and offsets for each dye.19 It is likely that the three packages under examination are limited to a global linear correction for each channel. Unfortunately, a nonlinear relationship has been demonstrated between measured and actual protein abundances,20 which has also been shown to vary across each gel’s surface.21 Methods from microarray gene analysis can be used to estimate this relationship from stable expression in a pair of experiments. Academic projects have applied these techniques to proteomics, including global quantile normalization17,22 and spatial calibration through loess smoothing (local polynomial regression fitting).21 It is hoped that similar techniques are in future incorporated into commercial gel analysis packages.

Removal of stochastic bias (expression-variance dependence) is particularly important for meaningful statistical testing. Employing a log-transformation of the data to account for a purely multiplicative dependence between the expression level and variance of protein spots is inaccurate at low expression levels due to an underlying additive dependency caused by image capture noise and background variation. A more elaborate two parameters model (multiplicative and additive) has been proposed in the microarray community24 and robust automated techniques developed to estimate the parameters and so variance stabilize on a per-gel basis.25 This arsinh transform is approximately linear for small expression levels and logarithmic for large expression levels. Difference detection on variance stabilized data is clearly a more principled and precise reflection of statistical confidence, and has recently been applied successfully to DIGE spot-list data20,26 and now appears to be incorporated into Progenesis SameSpots.27 However, as demonstrated in this paper, there is still significant ambiguity in the spot detection process, with arbitrary decisions made when delineating spot boundaries. This variance contributes significantly to the overall problem, and while this issue still exists, the advantages of these new statistical and stochastic techniques will be constrained by suboptimal input data.
Conclusions

 DIGE is a high-throughput technology, which overcomes the disadvantages of conventional 2-DE. To obtain the best from DIGE applications it is essential that appropriate software be used which ideally offers a complete and reliable data analysis for the sample type. Our results show that all three software packages provide a reliable, automated image analysis with minimal user intervention. They performed in a generally satisfactory manner, each with strengths and weaknesses. Based on matching accuracy, Progenesis SameSpots performed best. The mis-matches found in DeCyder and Dymension were caused by errors in spot detection and alignment. By optimizing detection settings and making manual vector corrections, the accuracy of analysis can be significantly increased. This was shown by increased matching accuracy after inclusion of landmarks, therefore user intervention is essential in order to achieve the best analysis results, even in fully automated and apparently sound software analysis. Moreover, users should be cautious when drawing inferences based on the magnitude of differences between samples as calculated by software packages; statistical analysis must be validated by alternative methods such as Western blot.

Acknowledgment

The authors are indebted to Dr Andrew Dowsey and Prof. Guang-Zhong Yang for their invaluable advice and their contribution towards the manuscript. We acknowledge the support and training provided by Martin Wells (Nonlinear Dynamics), Paru Oatey (Syngene) and Richard Fowler (GE Healthcare). We gratefully thank Prof. Mummery and for her kind gift of END2 cell line and Prof. Anthony E.G. Cass for his support. TT also thanks Faculty of Medicine, Chulalongkorn University for the support. This study was funded by Department of Trade and Industry (DTI).

Supporting Information Available: Supplementary Tables. This material is available free of charge via the Internet at http://pubs.acs.org.

References


PR80058BF

Kang et al.