

Leila H. Choe
Kelvin H. Lee

Cornell University Chemical
and Biomolecular Engineering,
Ithaca, NY, USA

Quantitative and qualitative measure of intralaboratory two-dimensional protein gel reproducibility and the effects of sample preparation, sample load, and image analysis

We investigate one approach to assess the quantitative variability in two-dimensional gel electrophoresis (2-DE) separations based on gel-to-gel variability, sample preparation variability, sample load differences, and the effect of automation on image analysis. We observe that 95% of spots present in three out of four replicate gels exhibit less than a 0.52 coefficient of variation (CV) in fluorescent stain intensity (% volume) for a single sample run on multiple gels. When four parallel sample preparations are performed, this value increases to 0.57. We do not observe any significant change in quantitative value for an increase or decrease in sample load of 30% when using appropriate image analysis variables. Increasing use of automation, while necessary in modern 2-DE experiments, does change the observed level of quantitative and qualitative variability among replicate gels. The number of spots that change qualitatively for a single sample run in parallel varies from a CV = 0.03 for fully manual analysis to CV = 0.20 for a fully automated analysis. We present a systematic method by which a single laboratory can measure gel-to-gel variability using only three gel runs.

Keywords: *Pseudomonas syringae* / Reproducibility / Two-dimensional protein electrophoresis
DOI 10.1002/elps.200305614

1 Introduction

2-D gel electrophoresis is an important tool in proteomics research. Its power lies in the ability to resolve simultaneously thousands of proteins present in a complex mixture. Using the latest technologies, this separation method is somewhat more reproducible and quantitative than the earlier versions of the technique. Nonetheless, there is well-justified concern about the overall reproducibility of this method when it is used as the sole method to identify qualitative and quantitative differences in protein expression in samples of interest.

Qualitative changes refer to the appearance or the disappearance of individual spots. In theory, if a single sample yields (for example) 2000 spots on one 2-D gel, then this sample should yield 2000 spots in the same relative locations on a parallel gel. However, slight technical variability from gel to gel often yields deviations in the number of spots. Some spots may appear in one gel that are not observed in a parallel gel of the same sample. Thus, when critical experiments are performed, investigators

often choose to run each sample multiple times and assume that any spot that appears in the majority of the resulting 2-DE patterns is 'real.' One approach employed commonly is to study spots present in three out of four gels. The perceived differences from gel to gel can arise from a large number of factors including variability in: sample (biological differences), sample preparation, environmental conditions, sources and lots of reagents, and personnel, among others. Because of the somewhat limited quantitative performance of 2-DE and because investigators are often very interested in 'all or nothing' changes in gene expression, these qualitative differences have become the most important observations resulting from a 2-DE analysis.

Quantitative changes refer to the variations in the observed area, optical density or integrated optical density of individual spots on 2-DE gels. Proteins are detected most commonly using either radioisotopes or total protein stains (e.g., silver, Coomassie blue, SYPRO Ruby). Each of these approaches and their appropriate imaging instrumentation will have some limitations in the quantitative reproducibility and dynamic range. As with qualitative changes, slight technical variability from gel to gel can yield changes in the observed amounts of individual spots (for reasons mentioned above). Taken to an

Correspondence: Dr. Kelvin H. Lee, Cornell University Chemical and Biomolecular Engineering, Ithaca, NY, USA
E-mail: KHL9@cornell.edu
Fax: +607-255-9166

extreme, significant quantitative changes are observed as qualitative differences. Some of the concerns about quantitative variability are addressed by using appropriate image analysis variables. For example, the % volume or normalized volume is an effective means to normalize against small differences in protein load or stain development time. Other image analysis variables such as the scaled volume for silver stain account for differences in background staining [1]. Still, there is significant concern about quantitative reproducibility and most quantitative changes are only considered important if the changes seen are significant (threefold, fourfold, etc.), are validated using other approaches (such as immunoassays, enzyme assays, etc.) and/or are useful in developing a hypothesis about the biology (when the genetic basis for the spot is known). It is very important for the investigator to determine if the observed changes, qualitative or quantitative, arise from technical variability or from biological considerations. Decisions about which spots should be subjected to further analysis (by MS, for example) should be made by considering this information and prioritizing the list of candidates. Thus, each laboratory should set certain standards and measures for their own reproducibility.

Because of these concerns, a number of groups have studied the reproducibility of 2-DE. Blomberg *et al.* [2] did a comprehensive interlaboratory reproducibility study involving S35-labelled yeast proteins. IPG isoelectric focusing was performed and three measures of reproducibility were made: *x*-positional reproducibility, *y*-positional reproducibility and quantitative reproducibility as measured by PDQuest (Bio-Rad). The investigators presented their quantitative reproducibility after a log-normalization of the data and after applying a Gaussian fit to detected spots. They matched 470 spots using PDQuest and for these spots, the intralaboratory standard deviation was 20–28%. It is not clear which of the spots from the gels were matched but the study did employ a cut-off factor of threefold to determine qualitative reproducibility. In the present study we measure variability by coefficient of variation and do not use Gaussian fits of spot shapes because some spots, especially those involving post-translational modifications, may not fit accurately with a Gaussian model.

Lopez and Patton [3] measured the reproducibility of a 2-DE separation based on carrier ampholytes and silver stain. This study was founded on the measurement of positional reproducibility from gel to gel and did not emphasize quantitative changes. Voss and Haber [4] employed Melanie II (Bio-Rad) in the image analysis of cell lysates from mononuclear cells. Here they found that matching efficiency dramatically decreased when large

numbers of gels were matched. They also found that 20% of all spots differed by a factor of 2- to 5-fold and 2.3% of spots by 5- to 10-fold when using silver staining for parallel gels. Our study will build on this prior work by measuring reproducibility based on fluorescence stains and laser fluorescence scanning.

Mahon and Dupree [5] performed a thorough investigation of the quantitative reproducibility of *Arabidopsis* proteins separated by 2-DE and analyzed using the Phoretix software (Nonlinear Dynamics). Coomassie blue staining was employed and 198 spots were matched with an error of 32%. The goal of the study, however, emphasized the relative importance of analysis software in measuring the reproducibility of the gels and of the imaging system. Image analysis approaches are a significant and important concern in addition to sample preparation and other technical variations. Tonge *et al.* [6] reported on the use of 2-D differential gel electrophoresis (DIGE) technology (Amersham). This technology has the potential to significantly aid in reproducibility because multiple samples that are labelled independently, are run on the same 2-DE gel. The authors presented carefully controlled experiments in which they determined the quantitative reproducibility based on the measurement of a 5% false positive rate. That is, they determined a cut-off value for which 95% of the observed data are valid. The data are presented as a series of standard deviations over a range of integrated optical densities rather than as a coefficient of variation. They found that from three replicate gels, an 80% prediction threshold is observed for changes that range from 20% quantitative change to a 4.1-fold difference. Depending on the samples used, approximately 400 Cy-labelled spots were studied. In a recent study to evaluate quantitatively the use of SYPRO Ruby stain (Molecular Probes), Nishihara and Champion [7] found a gel-to-gel reproducibility of CV (coefficient of variation) = 0.03–0.33 for spot quantitation of 20 landmarked spots and 0.04–0.11 for spot matching. The authors used both Z3 (Compugen) and Progenesis (Nonlinear Dynamics) in the analysis of their data. A comparison of the use of fluorescent stains *versus* silver stains revealed that silver stains offer a more limited dynamic range than fluorescent stains [8].

In this study, we report on our approach to assess qualitative and quantitative reproducibility of 2-DE using a small set of gels to benchmark, and thus infer, overall reproducibility of the technique in the laboratory. We discuss both the qualitative and the quantitative differences that are observed using a relatively automated image analysis package (Progenesis by Nonlinear Dynamics) and performing a fully manual computer-aided image analysis using Melanie 3 (GeneBio) on the same set of data. This approach requires only three gel runs of four

gels each and attempts to assess the relative importance of sample preparation, protein load and running parallel gels as well as the use of different levels of automation in the image analysis tools on quantitative reproducibility. We have attempted to use the most reliable methods including IPGs, fluorescence staining and laser fluorescence scanning, and using fresh reagents at the start of the experiment. These data are useful in the assessment of the importance of an observed quantitative or qualitative change in protein expression as measured by 2-DE. We thus codify a means by which individual laboratories can run a short series of experiments to determine their own reproducibility and thus better determine which changes in detected spots are meaningful.

2 Materials and methods

2.1 Sample preparation

Pseudomonas syringae samples were provided by Dr. Alan Collmer (Cornell University). Protein from cell lysates of *P. syringae* overexpressing HrpL [9] grown to OD₆₀₀ 1.5 were collected for 2-D gel analysis as described previously [10] and stored at -75°C until use. However, we employed three different experiments to test (i) variation from gel to gel for a single sample preparation, (ii) variation across sample preparations, and (iii) variation caused by loading different amounts of protein (Fig. 1). To perform these experiments we took the cell lysate and aliquoted it into five tubes after sonication. One tube was prepared for 2-DE by taking enough sample for four gels and mixing with enough sample buffer (8 M urea, 2% CHAPS, 65 mM DTT, 67 mM Tris, pH 8.0, and a trace of bromophenol blue) for four gels and with enough rehydration solution (8 M urea, 2% CHAPS, 0.3% DTT, 1.33% 3–10 ampholytes, 0.67% 5–7 ampholytes, and a trace of bromophenol blue) for four gels. This mixture was separated into four equal fractions and loaded onto four 18 cm, pH 3–10 nonlinear Immobiline DryStrip IPG strips (Amersham Biosciences, Piscataway, NJ, USA) for overnight rehydration at room temperature. This experiment is designated “1 Prep”. In a separate experiment, the same tube of cell lysate was prepared for 2-DE, but different amounts of this same preparation were loaded onto four IPG strips. Two of these gels had 30% more protein loaded and two of the gels had 30% less protein loaded than in the previous experiment. This experiment is designated “Var Load.” In another experiment, the four other tubes of cell lysate were prepared in parallel and each was individually mixed with sample buffer and rehydration solution and loaded onto IPG strips. This experiment is designated “4 Prep”. Four gels were run for each of the three conditions for twelve total gels. One hundred micrograms of

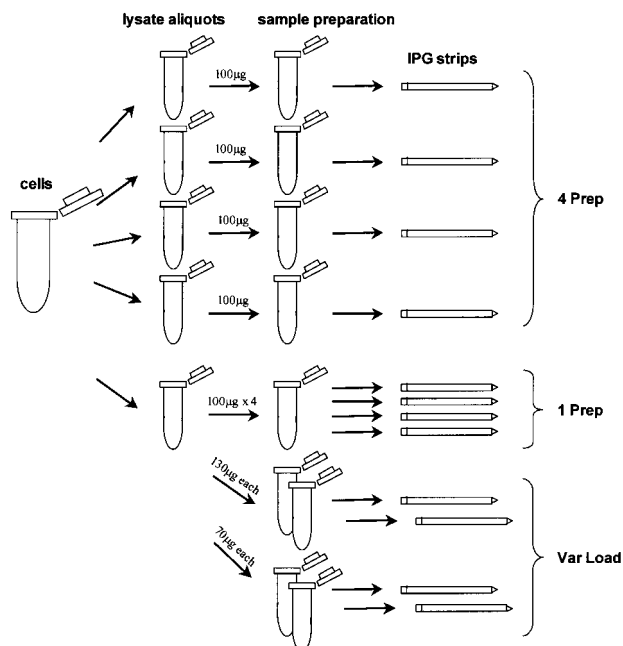


Figure 1. Process used in the 1 Prep, 4 Prep, and Var Load experiments. In the 1 Prep experiment, one sample was prepared and run on four parallel gels. In the 4 Prep experiment, four parallel sample preparations were made and run in parallel. In the Var Load experiment, the amount of protein loaded was 30% greater or 30% less than in the other experiments.

protein was loaded onto IPG strips in the 1 Prep and 4 Prep experiments while either 70 or 130 µg of protein was loaded onto IPG strips in the Var Load experiment.

2.2 Isoelectric focusing and electrophoresis

Isoelectric focusing was performed under the same conditions for all three sets of gels, at 64 200 Vh [10]. IPG strips were equilibrated with DTT and iodoacetamide prior to the second-dimensional electrophoresis on 12%T, 2.6%C polyacrylamide slab gels measuring 18 cm × 16 cm × 1.5 mm. 2-D electrophoresis was performed as previously described [10] with the exception of the agarose overlay for the second-dimensional electrophoresis which was not performed on any of these gels.

2.3 Staining and imaging

Gels were fixed, stained, and destained according to SYPRO Ruby Protein Gel Stain instructions (Molecular Probes, Eugene, OR, USA). Gels were imaged on a Fuji FLA-3000 Fluorescent Image Analyzer (Fuji Photo Film Company, Tokyo, Japan), with 50 µm resolution, 16 bits per pixel and a photomultiplier tube (PMT) gain of 1000.

2.4 Computer-assisted 2-D gel analysis

Gel images were exported as TIFF files from the FLA-3000 and a subset of the gel images (Fig. 2) was imported into gel analysis software. Each gel image set corresponding to one of the three experiments above (1 Prep, 4 Prep, Var Load) was subjected to three separate analyses employing different degrees of automation. In one analysis, designated as fully automatic, Progenesis Workstation Version 2002.01 (Nonlinear Dynamics, Newcastle upon Tyne, UK) was employed utilizing fully the automatic spot detection and matching capabilities. Using Progenesis, images were spot detected under default conditions for each set of gels and a reference gel was automatically created based on the gel with the most number of spots, then background subtraction was performed, and images were matched using gel warping. Unmatched spots were added to the reference gel, and spot volumes were normalized. The parameters were set using the Analysis Wizard and the images were processed overnight. Data from this analysis set was collected without further manipulation and involved minimal user input. A second analysis set, designated as semi-automatic, employed automatic default spot detection, followed by manual correction and editing of spot features. Subsequent analysis steps were carried out with the same parameters as the first set using Progenesis software. This analysis set

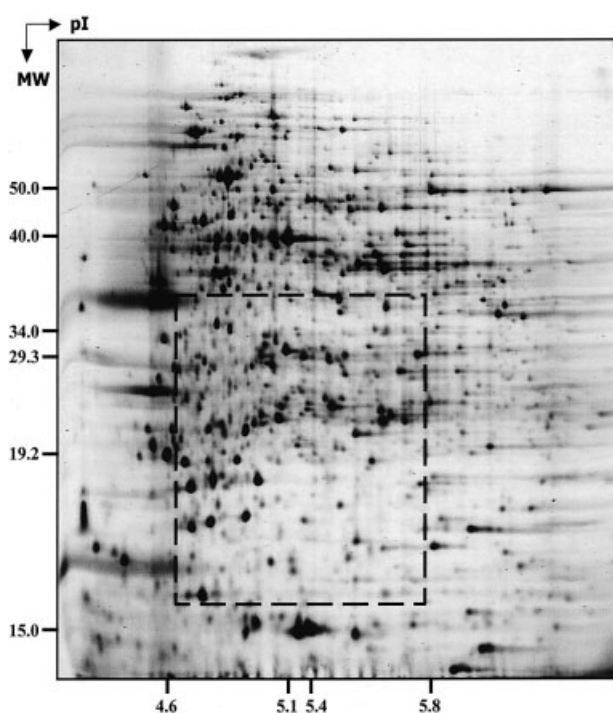


Figure 2. Image of the *P. syringae* proteome used in this study. The proteins have been stained with SYPRO Ruby and the boxed area is a subset corresponding to the region of interest.

required moderate user interaction for the editing of spot features. However, each gel image was inspected and edited as an individual gel without regard to the features of its matched spots in other gels. For a third analysis, manual input through Melanie 3 Version 3.09 (Genebio, Geneva, Switzerland) was used to approximate a fully manual interpretation of the data for spot detection and matching (*i.e.*, the software is used to measure percent volume of features and for convenience in reporting of data). Gel image sets were inspected as a set, where all four gels of a set were displayed together and features were edited with consideration of their sibling gels. No spot features were added to a gel where no protein stain was visualized by inspection to preserve qualitative changes, and spot features were edited to remain as true to their size and shape, as indicated by their grayscale pixels, as possible. In addition, 40–50 user-defined landmarks were used in each set of images as reference points to improve automatic matching. However, the computer-determined matches were inspected manually within each set and mismatches that were found were corrected.

Because Progenesis creates a reference gel, based on the image and data of one of the gels in an analysis set, and includes an option to add any unmatched spots to the reference gel; it thus creates a reference gel that can include all spots in the data set. When all gels of an analysis set are matched to the reference gel, all matched and unmatched spots are contained in a data set of the reference gel spots. In Melanie 3, however, a separate reference gel is not created; the user defines one gel of the set as the reference gel, to which others are matched. In choosing one gel as the reference gel, without the option of adding unmatched spots, no data is obtained on the matches that may exist among the other three or two gels within a set. To ensure a complete analysis, data was collected separately for each experiment using each of the four gels as reference gels to ensure all matches within a set were recorded. Percent volumes from Melanie 3-derived data and normalized volumes from Progenesis Workstation-derived data were exported to Microsoft Excel and tabulated to generate the results given below. The averages, standard deviations and coefficients of variation were calculated using the Excel-provided formulas. All of the experiments and all of the image analyses were performed by one individual.

3 Results and discussion

As discussed in the introduction, there have been several important studies of 2-DE gel reproducibility already undertaken. However, many of these studies do not

include detailed coefficient of variation data on a large number of spots from the gels. Here we present data on a relatively large number of spots from a region of a 2-DE gel. We use a prokaryotic organism because we are most familiar with this sample and because of a reduced amount of post-translational modifications which results in cleaner spot resolution. We selected a region of the *P. syringae* 2-DE image for analysis and include data on the spots detected in this region. The gel and the region selected are depicted in Fig. 2 and include, depending on the spot detection method used, approximately 725 spots. We have selected this region to avoid known problem areas in 2-DE gels because the goal is to generate the most reproducible data possible as a benchmark for further studies.

An Excel workbook was generated that permits the user to input the % volumes (or normalized volumes) for each of the spots matched across the three series of four gels. Some of the spots will match two out of four gels, some in three out of four and some in all four gels. The variable number of matches is indicative of qualitative differences among gels. This qualitative difference is presented as

Qual CV in Table 1 and measures the CV of the number of spots across the experiment. Using a fully automated matching algorithm, it is possible that some qualitative differences are mismatches, but this is minimized with a manual matching. As expected, the 1 Prep data show better qualitative reproducibility than the 4 Prep experiment. Quantitative variability is presented as the CV for each of the matched spots and the user specifies if the CV is calculated for spots present in two, three or all matched gels. Furthermore, the workbook permits the user to determine the number of matched spots (present in either two, three or four out of four gels) that fall within a specified CV as well as a percent of the spots that meet this criteria. At the outset, one might expect that the 1 Prep experiment would demonstrate better reproducibility (qualitative and quantitative) than the 4 Prep or the Var Load experiments. Any variability in the 1 Prep experiment may be attributed to differences in gel polymerization or other minor technical differences that arise during the procedure. The variability seen among the 4 Prep data can be attributed to the effect of sample preparation (after considering the 1 Prep experiment as a control). Finally, the Var Load experiment tests the effect of small changes in protein load. A signifi-

Table 1. Quantitative and qualitative reproducibility of manual spot detection and matching

Melanie 3 Analysis									
Spots matched in 2 out of 4 gels									
Expt	# Spots	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Prep	815	76%	88%	94%	97%	99%	99%	100%	100%
4 Prep	842	65%	79%	90%	96%	98%	99%	100%	100%
Var Load	919	72%	85%	92%	97%	98%	99%	100%	100%
Spots matched in 3 out of 4 gels									
Expt	# Spots	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Prep	793	76%	88%	94%	97%	99%	99%	99%	100%
4 Prep	757	65%	79%	90%	96%	99%	99%	100%	100%
Var Load	879	73%	85%	93%	97%	99%	99%	100%	100%
Spots matched in 4 out of 4 gels									
Expt	# Spots	CV = 0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1Prep	771	76%	88%	94%	97%	99%	99%	99%	100%
4 Prep	641	65%	79%	90%	96%	99%	100%	100%	100%
Var Load	824	74%	86%	93%	97%	99%	99%	100%	100%
Expt	Qual CV								
1 Prep	0.03								
4 Prep	0.14								
Var Load	0.05								

Values in the table represent the percentage of spots which fall within a certain CV cutoff value ranging from 0.3 to 1.0. The rows represent the different experiments. Qual CV refers to the number of spots which are qualitatively different.

cant increase or decrease in sample load will certainly affect 2-DE migration patterns and overall reproducibility. However, we are interested in testing whether smaller changes in protein load (from 70 to 130 μg for this sample), that may be representative of different sample preparations or week to week variability in gel runs, may lead to some inherent variability. If such changes do not have a significant impact on gel reproducibility then we expect to observe a Var Load variability similar to the 1 Prep experiment; whereas if small changes in protein load are an important consideration in reproducibility, then one expects poorer reproducibility in the Var Load experiment. We expect that the use of the normalized volume and the % volume image analysis variables for this study that scale the observed volume by the total stain observed should eliminate the effect of sample load.

3.1 Manual spot detection and matching

The analysis performed using Melanie 3 offers the highest degree of manual input because spot detection and spot matches are corrected manually. Thus, the primary role of the software is to aid the user in calculating the number of spots matched and the % volume of each feature. If a fully manual matching process offers the most reliable (albeit slowest and most laborious) approach to 2-DE gel analysis, then these data may provide the best measures of qualitative and quantitative variability in this study. Table 1 presents data for the fully manual image analysis and data are presented as a percentage of observed spots which fall within a given CV (range from 0.3 to 1.0). The number of matched spots is also presented for each experiment as is the qualitative variability for the three experiments. In the 1 Prep experiment we find 793 matched spots present in three out of four gels, and that 76% of these spots have a CV less than or equal to 0.3. Furthermore, 95% of the matched spots have a CV less than 0.52. In the case of the 4 Prep experiment for which we expect a greater CV, only 65% of the 793 spots have a CV less than or equal to 0.3 and 95% of these spots have a CV less than 0.57. The trend in these results, that multiple preparations of a single sample have greater variability than a single preparation of the same sample run in parallel, is consistent for cases in which we study spots present in at least two of four gels, three of four gels or all replicate gels. Furthermore, this trend is consistent with the expectation that doing parallel sample preparations of a given sample will introduce some variability in the observed quantities. Stated in other terms, fewer of the spots from the 4 Prep experiment than the 1 Prep or the Var Load experiments fall within a given cutoff CV value. However, all of the observed spots fall within a two-fold difference. The contribution of sample prepara-

tion variability may be estimated by noting that 95% of spots in the 1 Prep have CV less than or equal to 0.52 and in the 4 Prep experiment this value increases to 0.57. With this perspective, the contribution of sample preparation is less, for these conditions, than gel-to-gel variability. On the other hand, the variability introduced is greater than that introduced by a 30% increase or decrease in sample load.

3.2 Manual spot detection and automated matching

We employed the algorithms of Progenesis Workstation to simplify automated matching. However, the matching was performed on spots that were detected based on manual inspection of individual spots. Here, we expect to observe the same general trends as in the manual analysis in terms of increasing variability for the 1 Prep, Var Load and 4 Prep experiments. We further expect that even the most sophisticated matching algorithms which are employed in the state-of-the-art image analysis software packages (such as Progenesis) are still not as effective as total manual interpretation of spot matches. There is a trade-off in accuracy for speed. Our results with this semi-automated analysis are presented in Table 2. Again, consistent with prior observations and our expectations, we find that the 1 Prep experiment offers the best reproducibility when compared to the 4 Prep and the Var Load experiments. A comparison of the data from the semi-automated analysis and the manual analysis suggests that a manual interpretation identifies a greater fraction of the spots that fall within a given CV cutoff than a semi-automated analysis. In the case of the 1 Prep experiment, for example, 76% of spots matched in three out of four gels have a CV less than 0.3; whereas only 51% of similar spots are identified in the semi-automated analysis. However, 99% of all spots do fall within a twofold quantitative difference for this entire analysis. These observations allude to the importance that should be placed on image analysis in considering quantitative variability in spots. The type of image analysis and degree of automation will impact the level of significance that can be attributed to a specific observed change. As before, there is no significant effect of sample load on the quantitative results for the system tested.

3.3 Automated spot detection and automated matching

In Table 3 we present data using fully automated spot detection and matching. One expects the level of observed variability to further increase. Again, the results were consistent with this expectation. For spots present

Table 2. Quantitative and qualitative reproducibility of manual spot detection and automated matching

Manual detection and automated matching analysis									
Spots matched in 2 out of 4 gels									
Expt	# Spots	CV = 0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Prep	843	50%	65%	77%	87%	93%	95%	97%	99%
4 Prep	823	45%	64%	77%	86%	92%	97%	99%	99%
Var Load	898	51%	68%	80%	87%	93%	95%	97%	99%
Spots matched in 3 out of 4 gels									
Expt	# Spots	CV = 0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Prep	739	51%	66%	79%	89%	93%	96%	98%	99%
4 Prep	683	46%	65%	79%	87%	93%	97%	99%	99%
Var Load	810	50%	67%	79%	87%	92%	95%	97%	99%
Spots matched in 4 out of 4 gels									
Expt	# Spots	CV = 0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Prep	582	54%	70%	82%	91%	95%	97%	98%	99%
4 Prep	500	50%	72%	84%	91%	96%	99%	99%	100%
Var Load	655	53%	70%	82%	89%	94%	96%	98%	99%
Expt	Qual CV								
1 Prep	0.18								
4 Prep	0.24								
Var Load	0.16								

Values in the table represent the percentage of spots which fall within a certain CV cutoff value ranging from 0.3 to 1.0. The rows represent the different experiments. Qual CV refers to the number of spots which are qualitatively different.

Table 3. Quantitative and qualitative reproducibility of automated spot detection and matching

Fully automated analysis									
Spots matched in 2 out of 4 gels									
Expt	# Spots	CV = 0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Prep	720	41%	56%	72%	81%	87%	92%	95%	97%
4 Prep	780	39%	54%	66%	76%	84%	90%	94%	97%
Var Load	798	42%	58%	73%	82%	88%	93%	95%	97%
Spots matched in 3 out of 4 gels									
Expt	# Spots	CV = 0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Prep	614	43%	58%	74%	83%	88%	92%	95%	97%
4 Prep	630	40%	56%	69%	79%	87%	92%	96%	97%
Var Load	681	42%	58%	74%	83%	89%	93%	95%	97%
Spots matched in 4 out of 4 gels									
Expt	# Spots	CV = 0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Prep	482	48%	65%	78%	86%	91%	95%	97%	98%
4 Prep	490	44%	61%	75%	84%	91%	94%	97%	98%
Var Load	538	44%	62%	79%	86%	90%	94%	96%	98%
Expt	Qual CV								
1 Prep	0.20								
4 Prep	0.23								
Var Load	0.19								

Values in the table represent the percentage of spots which fall within a certain CV cutoff value ranging from 0.3 to 1.0. The rows represent the different experiments. Qual CV refers to the number of spots which are qualitatively different.

in three out of four gels for the 1 Prep experiment, we observed 76% of spots in the manual method, 51% in the semi-automated analysis and 43% in the fully automated analysis to fall within a CV cutoff of 0.3. These trends are consistent across all of the observed data. Here, 97% or better of all matched spots fall within a twofold quantitative change. A fully manual interpretation of 2-DE data is not practical for more than a very small number of gel images. Thus, there is a need for automated spot detection and matching. While we have selected only one commercially available software package for this part of the study, our observations with these data are that automated interpretation of data (and using other packages) performs very well. A twofold cutoff for determining the significance of a quantitative change is reasonable when using these packages for these experiments and is commonly used in the field. However, for those able to invest more resources into image analysis, one can achieve a 95% true positive rate that is closer to a 52% observed change in % volume under “best” conditions.

4 Concluding remarks

We tested the effects of gel-to-gel variability (for a given sample preparation), sample preparation, and sample load on overall 2-DE variability. We further studied the use of automated spot detection and matching algorithms for spots present in two out of four, three out of four and four out of four matched gels. All of these factors should be considered when determining the significance of observed changes in 2-DE patterns and it is difficult to discuss a single number that can be used to quantify reproducibility of the technique. In particular, 95% true positive rate can be assigned to a given observed quantitative change, but only in the context of understanding what image analysis method was employed and the number of replicate gels that are included or excluded in the analysis. We can state that for a spot present in three out of four replicate gels, we observe that 95% of spots fall within a CV of 0.52. This analysis is, to the best of our knowledge, the most comprehensive study of

observed quantitative variability (in terms of numbers of spots) to employ state-of-the-art techniques including SYPRO Ruby staining and laser fluorescence scanning. Quantitative reproducibility is certainly laboratory specific. This work suggests a short series of experiments that can be employed by any laboratory to help quantify intra-laboratory reproducibility and thus streamline the identification of significant spots for further analysis and characterization. The selection of the degree of automation for the analysis of gel images should be based on available resources and can have an important impact on the resulting data.

This work was supported in part by the USDA ARS SCA #58-1907-1-146, NIH R01 MH59926 and NSF BES-0120315. The authors would like to thank Zsofia Franck for very important discussions. The authors would like to thank Nonlinear Dynamics for the complimentary demonstration license of the Progenesis Workstation software.

Received December 18, 2002

5 References

- [1] Dutt, M., Lee, K. H., *Electrophoresis* 2001, 22, 1627–1632.
- [2] Blomberg, A., Blomberg, L., Norbeck, J., Fey, S. J., Larsen, P. M., Larsen, M., Roepstorff, P., Degand, H., Boutry, M., Posch, A., Goerg, A., *Electrophoresis* 1995, 16, 1935–1945.
- [3] Lopez, M. F., Patton, W. F., *Electrophoresis* 1997, 18, 338–343.
- [4] Voss, T., Haberl, P., *Electrophoresis* 2000, 21, 3345–3350.
- [5] Mahon, P., Dupree, P., *Electrophoresis* 2001, 22, 2075–2085.
- [6] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., Davison, M., *Proteomics* 2001, 1, 377–396.
- [7] Nishihara, J. C., Champion, K. M., *Electrophoresis* 2002, 23, 2203–2215.
- [8] Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., Patton, W. F., *Electrophoresis* 2000, 21, 3673–3683.
- [9] Fouts, D. E., Abramovitch, R. B., Alfano, J. R., Baldo, A. M., Buell, C. R., Cartinhour, S., Chatterjee, A. K., D’Ascenzo, M., Gwinn, M. L., Lazarowitz, S. G., Lin, N. C., Martin, G. B., Rehm, A. H., van Dijk, K., Tang, X., Collmer, A., *Proc. Natl. Acad. Sci USA* 2002, 99, 2275–2280.
- [10] Choe L., Lee, K. H., *Electrophoresis* 1999, 20, 798–805.