

Models for teardrop spots in 2-DE gels

Sean M. Pivek¹, Susan T. Weintraub², Carola Wenk¹

¹ University of Texas at San Antonio, Dept. of Computer Science

² University of Texas Health Science Center at San Antonio, Dept. of Biochemistry

Purpose

Develop and compare parametric spot models for teardrop 2D electrophoresis gels, and integrate them into a spot detection algorithm in order to enhance spot detection quality.

Background

Spot detection and matching across multiple sets of gel images are essential first steps for proteomics investigations based on two-dimensional gel electrophoresis (2-DE). A 2-DE gel image potentially contains thousands of so-called spots. Each spot ideally corresponds to one protein. A high-quality spot detection algorithm is of fundamental importance to determine the location and intensity of each of the protein spots because accurate spot detection is the basis for all subsequent analyses.

A parametric spot detection algorithm assumes a given spot shape model, such as a bivariate Gaussian distribution or a saturated Gaussian distribution [2], and fits this model using parameter estimation techniques. Compared to non-parametric algorithms (such as [7]), parametric algorithms allow spots to overlap and provide more accurate spot quantification. If the correct spot shape model is used, parametric algorithms have the potential for higher-quality spot detection.

Although a bivariate Gaussian distribution is a widely accepted modeling assumption for protein spots [1,2,5], it is clear from our 2-D gels and from gels in public databases that many spots exhibit reproducible intensity patterns resembling a “teardrop” shape that do not fit a Gaussian model. Up to now, most spot detection algorithms assume a Gaussian spot shape and have problems modeling teardrop spots.

Methods

We investigated different models for teardrop spots and developed a parametric spot detection algorithm that uses multiple models for spot intensity distribution. The spot detection algorithm identifies a “seed” set of possible spot locations and then fits one or multiple spot models to each location. We used a free implementation [8] of the Levenberg-Marquardt algorithm to perform nonlinear least-squares fitting of the spot models.

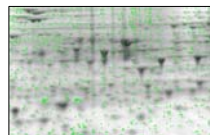


Figure 3. Seed locations in a gel

Methods (continued)

The (x_1, x_2) -position of a protein spot in a 2-DE gel represents the protein’s experimental isoelectric point ($pI \equiv x_1$) and molecular weight ($MW \equiv x_2$). We concentrated on investigating spot models that offer explanations for the cause of the teardrop shape. In particular, we considered the following spot model distributions:

Norm/norm: Bivariate normal (Gaussian) distribution

$$G_N(x) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

for $N=2$, with $x=[x_1, x_2, \dots, x_N]^T$
 $\mu=[\mu_1, \mu_2, \dots, \mu_N]^T$
 $N \times N$ covariance matrix Σ

Norm/lognorm: Normal & modified log-normal distribution

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) \quad f_1(x_1) = G_1(x_1)$$

$$f_2(x_2) = \frac{1}{(2\pi)^{1/2} \sigma(x_2 - \mu)} \exp\left(-\frac{1}{2\sigma^2} (\ln(x_2 - \mu))^2\right)$$

with additional x_2 -offset σ

Norm/ftipp: Normal & modified Fisher-Tippett distribution

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) \quad f_1(x_1) = G_1(x_1)$$

$$f_2(x_2) = \frac{1}{\beta} \exp\left(-\frac{x_2 - \mu_2}{\beta}\right) \exp\left(-\exp\left(-\frac{x_2 - \mu_2}{\beta}\right)\right)$$

with modified $\mu_2 = \mu_2 - \frac{|x_1 - \mu_1|^2}{\sigma_1}$

The **norm/norm** bivariate normal distribution is the widely accepted model for protein spots [1,2,5] and assumes independent normal distributions on the x_1 -axis and on the x_2 -axis. Since the molecular weight on the x_2 -axis is often assumed to have a logarithmic scale, we investigated the **norm/lognorm** distribution which combines a normal distribution on the x_1 -axis with an independent logarithmic normal distribution on the x_2 -axis (which we generalized by adding an x_2 -offset parameter). Furthermore we investigated the **norm/ftipp** distribution, which combines a normal distribution on the x_1 -axis with a modified Fisher-Tippett distribution on the x_2 -axis which is not independent of the x_1 -axis.

We fitted the different spot models on 2-DE gel images from our lab. In addition, we conducted a series of experiments using a pooled stock of protein extract in which we systematically varied parameters associated with running 2-DE gels.

Results

Our teardrop spot models (**norm/lognorm** and **norm/ftipp**) are generally superior to the Gaussian model (**norm/norm**) in fitting teardrop spots. However, the fitting algorithm generally had problems fitting the **norm/lognorm** distribution, which we believe may be due to the nature of the non-linear fitting algorithm. In addition, the **lognorm** distribution is not defined for negative values and therefore does not model an adequate tail, whereas the **ftipp** distribution is defined for negative values. We, therefore, concentrated on comparing the **norm/norm** and the **norm/ftipp** spot models.

Results (continued)

Figures 4-6 show fitting results for representative spots, using the **norm/norm** model and the **norm/ftipp** model. The **norm/ftipp** model generally fits teardrop spots better, based on the error measured as the L_2 -norm of the pixel difference, as well as by visual inspection. Due to the dependency of the x_2 -axis (MW) on the x_1 -axis (pI), the **norm/ftipp** model successfully captures the “smiley” shape of some of the teardrop spots. In general, however, this feature seems somewhat too dominating. In contrast to the bivariate **norm/norm** model, it appears that better teardrop models can be developed (c.f. **norm/ftipp**) when the pI-axis and the molecular weight axis are not assumed to be independent.

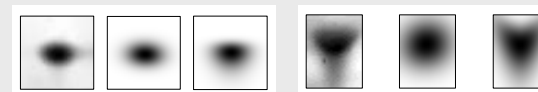


Figure 4. Gaussian-shaped spot (left), fit with **norm/norm** model (center), fit with **norm/ftipp** model (right)

Figure 5. Teardrop-shaped spot (left), fit with **norm/norm** model (center), fit with **norm/ftipp** model (right)

Error	1.34	3.85	2.44	2.61	0.69	5.78	0.78	8.81	28.2	n/a	6.78	11.56
Best fit with norm/ftipp model												
Original 2-DE gel spot												
Best fit with norm/norm model												
Error	3.31	9.11	3.35	6.14	1.55	7.30	n/a	11.18	44.65	n/a	6.39	14.04

Figure 6. Fitting results for multiple representative spots. The error is the (unnormalized) L_2 -norm of the pixel difference between image and model.

We conducted a series of experiments separating a pooled stock of protein extract on acrylamide 2-DE gels purchased from Bio-Rad. We varied the voltage (100V, 200V), the protein load (50µg, 100µg), and the acrylamide percentage (10%, 12%), which are common parameters associated with running 2-DE gels. The results, shown in Figure 7, suggest that teardrop spots are present for all parameter choices.

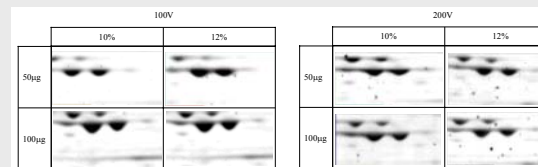


Figure 7. Spot shapes for varied voltage, protein load, and acrylamide percentage in a BioRad 2-DE gel.

Conclusions

Many 2-DE gels exhibit teardrop spots that cannot be modeled by current analysis software packages. Our investigations offer the first attempt at modeling and explaining this phenomenon.

- Teardrop spots were observed in 2DE-gels using variations in voltage, protein load, and acrylamide percentage.
- The standard bivariate Gaussian distribution (**norm/norm**) does not fit teardrop spots well.
- The **norm/ftipp** distribution is a better fit for teardrop spots, and captures “smiley” shapes.
- It appears that better teardrop models can be developed (c.f. **norm/ftipp**) when the pI-axis and the molecular weight axis are not assumed to be independent.

References

- [1] R. Appel, J. Vargas, P. Palagi, D. Walther, and D. Hochstrasser. Melanie II, a third generation software package for analysis of two dimensional electrophoresis images: II. Algorithms. *Electrophoresis*, 18:23735–2748, 1997.
- [2] E. Bettens. Peak characterization using parameter estimation methods. PhD thesis, Universiteit Antwerpen, Faculteit Wetenschappen, 1999.
- [3] A. Efrat, F. Hoffmann, C. Krauer, K. Kriegel, G. Rote, and C. Wenk. Covering with ellipses. *Algorithmica* 38(2):145–160, 2004.
- [4] A. Efrat, F. Hoffmann, K. Kriegel, C. Schultz, and C. Wenk. Geometric algorithms for the analysis of 2D-electrophoresis gels. *Journal of Computational Biology: special issue dedicated to RECOMB 2001*, 9(2):299–316, 2002.
- [5] J. Garrels. The QUEST system for quantitative analysis of 2D gels. *J. Biol. Chem.*, 264:5269–5282, 1989.
- [6] F. Hoffmann, K. Kriegel, and C.Wenk. An applied point pattern matching problem: comparing 2D patterns of protein spots. *Discrete Applied Mathematics*, 93:75–88, 1999.
- [7] K. Kriegel, I. Seefeldt, F. Hoffmann, C. Schultz, C. Wenk, V. Regitz-Zagrosek, H. Oswald, and E. Fleck. An alternative approach to deal with geometric uncertainties in computer analysis of two-dimensional electrophoresis gels. *Electrophoresis*, 21:2637–2640, 2000.
- [8] M.I.A. Lourakis, Ievmar. Levenberg-Marquardt nonlinear least squares algorithms in C/C++, <http://www.ics.forth.gr/~lourakis/levmar/>, 2004.
- [9] K.-P. Pleißner, F. Hoffmann, K. Kriegel, C. Wenk, S. Wegner, A. Sahström, H. Oswald, H. Alt, and E. Fleck. New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases. *Electrophoresis*, 20:755–765, 1999.
- [10] M. Rogers, J. Graham, and R.P. Tonge. Statistical models of shape for the analysis of protein spots in two-dimensional electrophoresis gel images. *Proteomics*, 3:887–896, 2003.

Acknowledgements

Supported in part by the National Science Foundation (NSF) grant NSF CAREER CCF-0643597, by a seed grant part of a UT System grant with PI James Bower (UTHSCSA), and by the National Centers for Research Resources (NCRR) Research Centers in Minority Institutions (RCMI) grant 5G12 RR013646-08 awarded to The University of Texas at San Antonio. We are grateful for the assistance of Kevin Hakala and Curtis Johnson (UTHSCSA) who ran the 2-DE gels.