Andrew M. Woodward[1]
Naheed Kaderbhai[1]
Mustak Kaderbhai[1]
Adrian Shaw[1]
Jem Rowland[2]
Douglas B. Kell[1]

[1]Institute of Biological Sciences, University of Wales, Aberystwyth, UK
[2]Department of Computer Sciences, University of Wales, Aberystwyth, UK

Histometrics: Improvement of the dynamic range of fluorescently stained proteins resolved in electrophoretic gels using hyperspectral imaging

Most image-based analyses, using absorbance or fluorescence of the spatial distribution of identifiable structures in complex biological systems, use only a very small number of dimensions of possible spectral data for the generation and interpretation of the image. We here extend the concepts of hyperspectral imaging, being developed in remote sensing, into analytical biotechnology. The massive volume of information contained in hyperspectral spectroscopic images requires multivariate analysis in order to extract the chemical and spatial information contained within the data. We here describe the use of multivariate statistical methods to map and quantify common protein staining fluorophores (SYPRO Red, Orange and Tangerine) in electrophoretic gels. Specifically, we find (a) that the 'background' underpinning limits of detection is due more to proteins that have not migrated properly than to impurities or to ineffective destaining, (b) the detailed mechanisms of staining of SYPRO red and orange are apparently not identical, and in particular (c) that these methods can provide two orders of magnitude improvement in the detection limit *per* pixel, to levels well below the limit observable optically.

Keywords: SYPRO stains / Fluorescent staining / Gel electrophoresis / Hyperspectral imaging / Protein quantification PRO 0108

1 Introduction

The post-genomic era has led to the requirement and development of new systems for detection and rapid data analysis, and for their processing into meaningful information. Typically, the wider and more extensive applications of the powerful two-dimensional electrophoresis approaches to analyse complex biological systems at the level of the proteome have greatly increased our understanding of the workings of the living cell [1–10]. Nevertheless, a major limiting factor is the poor dynamic range of the detection of proteins, many of which are present in biological systems in relatively miniscule amounts [11, 12].

Leaving aside radioisotopic detection of proteins in gels, the commonest or most popular means for detection of proteins employ stains that are assayed optically, either *via* absorbance, for example colloidal Coomassie Blue [13] and silver [14], or fluorimetrically for modern stains such as the Cy dyes [15] and SYPRO Orange, Red, Tangerine and Ruby [16–24]. In all cases, the number of channels used to quantify a given protein from a particular cell extract (even if extracts are differentially prelabelled [25]) is one, *i.e.* absorbance or fluorescence are measured at just one wavelength or waveband. A similar strategy is used in flow cytometric detection of particular proteins [26]. However, prototype automatic analysis of polyacrylamide slab gels imaged through a novel optical detection system has recently been developed [27]. The design of this prototype sequencer allows direct optical coupling over the entire read area of the gel and spectrographic separation and detection of the fluorescence emission.

The remote sensing community [28] has recently been developing techniques of *hyperspectral* imaging as a means of mapping otherwise indistinguishable spatial features by means of their spectral content [29–41], and we have begun to apply these ideas to problems in fermentation technology and microbial speciation [42–45]. Since the availability of spectral properties at each spatial location allows a far more precise spatial discrimination, it occurred to us that we could exploit these methods in histology and other cases where biological materials are to be stained (see also [46]). We refer to this combination of histology and chemometrics as *histometrics,* and here apply it to the problem of staining proteins in electrophoretic gels, with highly encouraging results.

Correspondence: Dr. Douglas B. Kell, Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DD, UK
E-mail: dbk@aber.ac.uk
Fax: +44-01970-622354

Abbreviations: PCA, principal components analysis; **PLS**, partial least squares regression

 1615-9853/01/1111–1351 $17.50+.50/0

2 Materials and methods

2.1 Sample preparation

Protein molecular weight markers (P6649) were obtained from Molecular Probes (Molecular Probes, Eugene, OR, USA). The proteins separated in the gels and their respective molecular weights are myosin (205 kDa), β-galactosidase (116 kDa), phosphorylase b (97 kDa), transferrin (80 kDa), bovine serum albumin (66 kDa), glutamate dehydrogenase (55 kDa), ovalbumin (45 kDa), carbonic anhydrase (30 kDa), and trypsin inhibitor (21 kDa). Immediately prior to electrophoresis, 100 μL of the markers were mixed with 400 μL SDS sample buffer (5% w/v SDS, 45% glycerol, 50 mM DTT, 0.05% bromophenol blue, 25 mM Tris-HCl pH 6.8) and denatured by boiling for 5 min. The samples were then serially diluted two-fold with the sample buffer.

2.2 SDS-PAGE

Electrophoresis was performed using a discontinuous gel system as first described by Laemmli [47]. In order to improve the resolution, a homogeneous polyacrylamide (12%) running gel containing 0.1% SDS was overlaid with 1.5% stacking gels. The gels were cast using a Bio-Rad (Hercules, CA, USA) mini-electrophoresis system allowing 10 loadings *per* gel. Samples were applied in equivalent volumes (20 μL) containing variable amounts of protein loading ranging from 0.9 ng to 2.5 μg. Protein quantities were measured using bicinchoninic acid as described by Smith and coworkers [48]. In each gel, the last two lanes were mock loaded with sample buffer without protein. The electrophoretic run was conducted by applying a current of 20 mA and continued until the leading bromophenol blue frontjust ran through the gel.

2.3 Gel staining and destaining

One gel was stained with Coomassie Brilliant Blue R250 to provide a visual comparison; staining and destaining was carried out as described by Chen and colleagues [13]. The remaining three gels were individually stained as described in the manufacturer's protocol using 10 μL of either SYPRO Tangerine, Red or Orange stains (S-12012 SYPRO starter kit: Molecular probes) each diluted with 50 mL of 7.5% v/v acetic acid solution. These gels were left overnight to stain and during this process any background staining was eliminated. These stains are considered to allow detection of nano- to micrograms of protein *per* band in gels with little protein-to-protein signal variability, minimal stain fluorescence in aqueous solutions and bright fluorescence upon binding to detergent-coated proteins in gels.

2.4 Protein detection of SYPRO stained gels

The SYPRO Red (ex300/550, em630), Orange (ex300/470, em570) and Tangerine (ex300/490, em 640) dyes have excitation and emission bands in the spectral range from 400–700 nm and can be visualized with most CCD camera-based imagers. These SYPRO fluorescent stained protein gels were viewed individually using the model 2920 Arthur multiwavelength fluorimager (Perkin Elmer (Wallac), Norwalk, CT, USA; http://www.wallac.com/catalog/art.htm) [24, 49]).

2.5 Hyperspectral imaging

The required hyperspectral images were also taken using the Arthur (now known as ProXpress) multiwavelength fluorimager [24, 49]. This CCD camera-based device uses excitation and emission filters and a white light source to allow six different illumination wavelengths across the visible region from violet (420 nm) through to red (635 nm), and six receiver frequencies over a range slightly longer in wavelength than the illumination range (450–680 nm), allowing imaging of fluorescence over 20 separate and valid excitation/emission combinations. The images thus produced are stacked into a hyperspectral datacube such that the depth dimension of the datacube denotes the relevant excitation/emission frequency pair.

In the present study, the excitations were at 420, 480, 510, 540, 580 and 635 rim, while the emissions were at 450, 480, 530, 575, 630 and 680 nm (each ± 15 nm). The machine was run under a macro written in-house to allow this entire range of valid excitation/emission combinations to be recorded automatically with no further user input after initialisation of the machine. Further details are given in Section 3.

3 Results and discussion

3.1 Improvement of raw images, from the Arthur fluoroimager

The Arthur constructs a large area image (27 cm × 23 cm or 1456 pixels × 1210 pixels) from 8 × 10 tiles of the smaller area of the CCD camera's field of view (182 pixels × 121 pixels) [49]. Due to the non-uniform illumination by the light source, these tiles have a background 'footprint' of strongly varying intensity and consequently do not match up at the edges, leading to sharp lines of different shading criss-crossing the image. This intensity variation and these edge discontinuities must be modelled away before any numerical analysis can be carried out

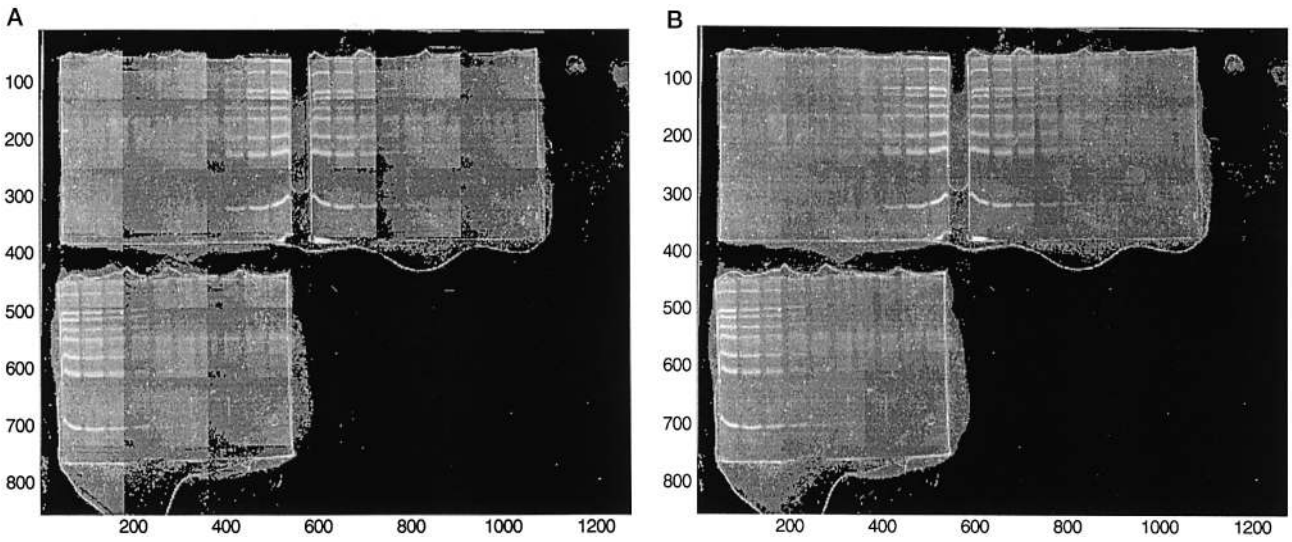

Figure 1. Images of the 480/575 slice of the hyperspectral stack for the gels dyed with SYPRO Tangerine (top left), Red (top right), Orange (bottom left) and nonfluorescent Coomassie Blue (bottom right). (a) Original gel image; (b) image after defootprinting.

on the hyperspectral images, free from artefacts. There is a 'defootprinting' facility provided on the device, but this was very far from being perfect at the time these experiments were performed, so algorithms were developed inhouse to improve this before further analysis. (The latest versions of this instrument have a better camera and shade correction system, with fewer tiles, but the method we describe is of general utility for systems of this type.)

Several approaches were investigated, including linear and nonlinear time-domain lowpass filtering methods, 2-D curve fitting, and 2-D Fourier-filtering of a separately recorded reference tile. The Fourier-filtering was found to give the best combination of accuracy, speed of calculation and freedom from requirement for user specification of parameters.

It was also noted that the instrument response could vary with time so that the reference footprint often did not accurately reflect the footprints of the tiles in the main image. This problem was solved by the pixelwise construction of a 'median footprint' derived from the background footprints of all the tiles in the main image itself, which is then used to remove the tiling effect, hence eliminating time variation of instrumental response. (This assumes that most pixels are not spots or bands, and this is true here.) If this footprint is then smoothed using a Fourier-filter as above, then an almost perfect defootprinting can be achieved. Since the median footprinting method also requires no user input, the entire defootprinting procedure can be (and was) automated.

Construction of a median footprint also has the advantage of automatically and completely eliminating blemishes in the footprint. Linear image processing techniques such as Fourier-filtering merely smooth blemishes in the reference footprint, thus reducing and widening them without eliminating them. The median footprint is applicable to all cases where an image is (or is considered) sparse enough that any given pixel in a tile represents background in more than half the total number of tiles. If this is not the case, then a separate reference footprint must be used and imperfect defootprinting tolerated. There was also a small positional irreproducibility between tiles in the same image and the median footprint approach minimised this problem, although it could not completely eliminate it. To do this a local heuristic based on minimising change at the known edge positions of the tiles and interpolating the background level across the tiles from these corrected values is required. An indication of the benefits of this approach is given in Fig. 1 which shows typical pictures of the 480/575 slice of the datacube for the gels dyed with SYPRO Tangerine, Red, Orange and the nonfluorescent Coomassie Blue.

3.2 Identification of residual protein in background areas of fluorescent gels

The human eye has a very poor dynamic range for detection of grey scales, the number that can typically be distinguished in images lacking sharp edges being about 30 [50]. This may be contrasted with the 65 536 levels of grey

available in a 16-bit camera as used herein. Direct observation of protein gels is thus normally a very poor strategy for assessing them! The first question we wished to ask was whether the limit of detection was due to (i) inadequate staining, (ii) a background caused by inadequate destaining, (iii) the presence of numerous proteins at low levels which had migrated to their 'proper' positions, or (iv) to poorly resolved proteins which had been 'left behind' en route to doing so. The latter in particular is a very realistic but under-recognised possibility as it is well known from nucleic acid electrophoresis that the uneven pore size of gels and the presence of charged impurities causes molecules to get 'stuck' and thus be lost from the discrete migrating zone (strategies such as field inversion being required to allow migration past the blockage [51]). We refer to this as the 'slime trail' effect. The above hyperspectral gel images were first rotated and aligned using the Matlab Image Processing Toolbox (MathWorks, Cambridge, UK) so that the tracks were parallel to the sides of the image matrices. They were then defootprinted as described in Section 3.1. Finally each separate image in the datacube was normalized to the brightest band in that image, but no further normalization (*e.g.* to unit variance between equivalent pixels in a given image) was carried out.

Each separate image in the datacube was then unfolded into a linear column vector with each element representing a single pixel. The datacube was thus reshaped into rectangular matrix $\mathbf{X}_{\mathbf{rect}}$ with the rows representing the spectral frequencies and the columns representing the individual pixels, so that each row contained a spectrum at a particular pixel.

Principal components analysis (PCA) [52, 53] involves projecting the original data-matrix *(N* samples; *p* variables) onto a *d*-dimensional subspace using a projection (or *loading*) matrix, thus creating object coordinates (a *scores* matrix) in a new coordinate system. This is achieved by the method known as singular value decomposition (SVD) of X:

$$\mathbf{X}_{\mathbf{N}\times\mathbf{p}} = \mathbf{U}_{\mathbf{N}\times\mathbf{d}}\Lambda_{\mathbf{d}\times\mathbf{d}}\mathbf{L}^{\mathbf{T}}_{\mathbf{p}\times\mathbf{d}} = \mathbf{T}_{\mathbf{N}\times\mathbf{d}}\mathbf{L}^{\mathbf{T}}_{\mathbf{p}\times\mathbf{d}}$$

where, **U** is the unweighted (normalized) score matrix and **T** is the weighted (or biased) score matrix. **L** is the loading matrix where, the columns of **L** are known as eigenvectors or loading-PCs. A is a diagonal matrix (*i.e.* all of the off diagonal elements are equal to zero) containing the square roots of the first *d* eigenvalues of the covariance matrix ($\mathbf{X}^{\mathbf{T}}\mathbf{X}$) where, $d < N$ and $d < p$.

The principal components (PCs) can be considered as a basis set used to project the original data matrix, **X**, onto the scores, **T**. In other words the new coordinates are linear combinations of the original variables. The influence of each of the original variables on the new principal components (*i.e.* the contents of the loading matrix) is determined on the basis of the maximum variance criterion. The first PC is considered to lie in the direction describing maximum variance in the original data. Each subsequent PC lies in an orthogonal direction of maximum variance that has not been considered by the former components. The number of PCs computed for a given data set is up to the analyst. However, usually as many PCs are calculated as are needed to explain a pre-set percentage of the total variance in the original data (the number of PCs is always less than or equal to the number of original variables).

PCA was then carried out on this $\mathbf{X}_{\mathbf{rect}}$ matrix to produce a pixel-wise multivariate decomposition. The scores of each PCA factor were then examined. The first principal component in un-shade corrected images from this instrument (*i.e.* those not subjected to defootprinting) often reflects the variance in intensity due to the camera footprint tiling, nominally provides another way of shade correction (not exploited here) and is not of present interest; our attention is thus focused on the second two principal components of the variance.

For the SYPRO Tangerine and Orange, score plots [53] showed that pixels representing background areas of gel lanes containing proteins clustered in a different place from pixels representing blank gel tracks as shown in Fig. 2, indicating that the background areas of active tracks are not merely empty gel, but contain trace concentrations of protein. This strongly supports the 'slime trail' model.

The SYPRO Red dye did not cluster as well, there being only a shift of the centroid of the clusters rather than the clear separation seen in the other two dyes. It is possible that the reduced number of active fluoresent excitation/emission frequency combinations in the datacube for this dye did not give sufficient data to the multivariate model to enable clear separation. Alternatively, and more likely, the exact mechanism of staining of SYPRO Red differs from that of SYPRO Orange (it is of course known to differ form that of SYPRO Tangerine). If the matrices were then transposed and reanalyzed such that the pixels were now represented by PCA loadings [53] rather than by scores, the loadings plots also showed very similar clustering (not shown).

3.3 Prediction of protein concentration in fluorescent gel bands

A set of four identical gels was run, similar to Section 2.3 but with all tracks included in the concentration series, with no blank tracks. Three of the gels were stained with

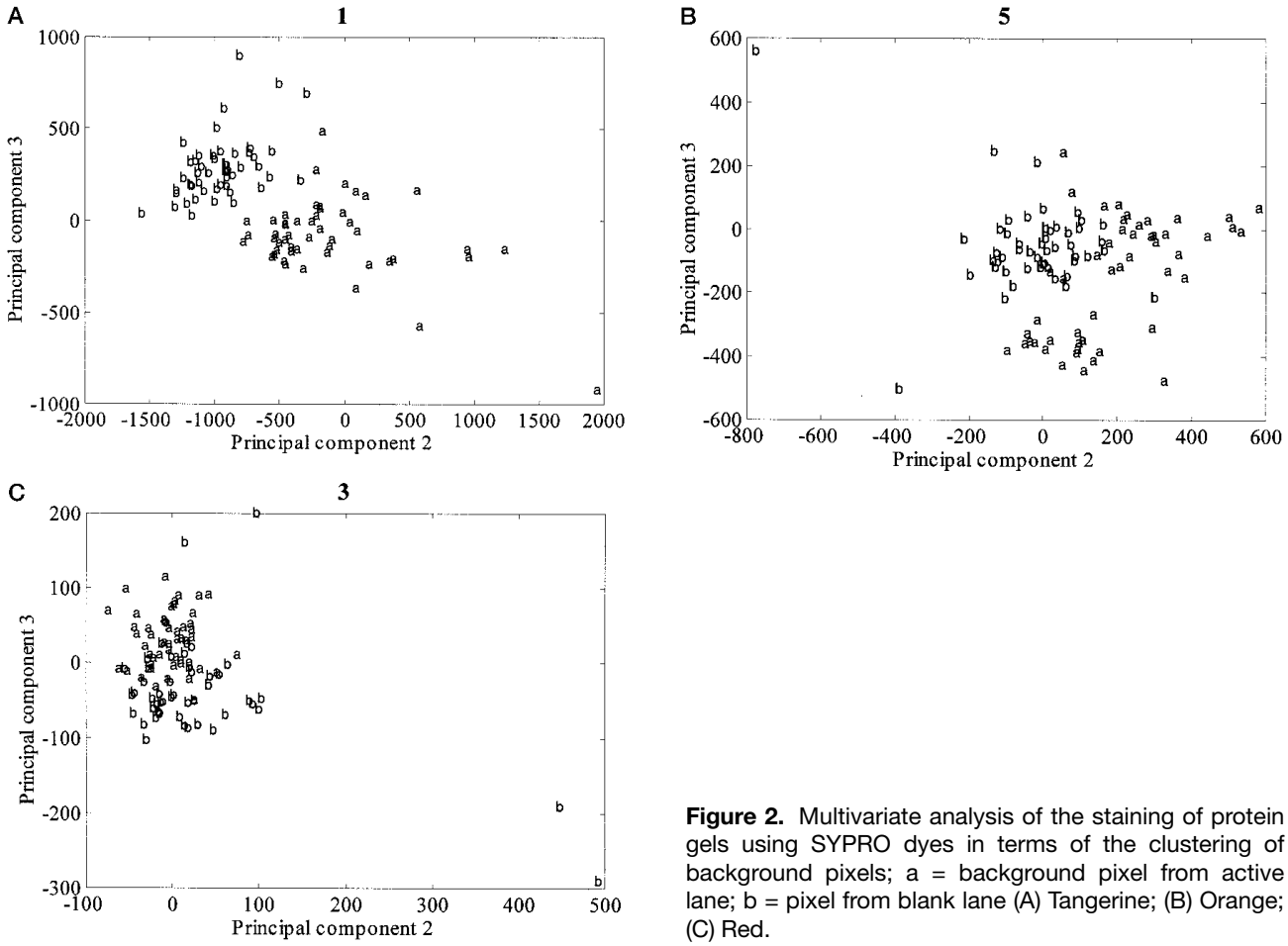

Figure 2. Multivariate analysis of the staining of protein gels using SYPRO dyes in terms of the clustering of background pixels; a = background pixel from active lane; b = pixel from blank lane (A) Tangerine; (B) Orange; (C) Red.

SYPRO Red dye and the remaining gel with the nonfluorescent Coomassie Blue both to act as visual comparison and as a nonfluorescent reference gel for assessing the variation in illumination for defootprinting. The gels were then imaged and the images preprocessed as in Section 2.5. Additionally the images were passed through a 2-D median filter, a program written in-house, to remove blemishes and noise spikes. Each gel image contained 10 tracks which represent a geometric increase in concentration of the protein markers run up the track, over a concentration range 5, 12.5, 25, 50, 125, 250, 500, 1250, 2500, 5000 ng protein/band. Automatically identifying the position of each band and aligning these bands unambiguously throughout all three datacubes is an image-registration problem beyond current technology. A method of unambiguously assigning concentration data to the training of a multivariate model must therefore be found.

Accordingly strips 10 pixels wide were picked from the centre of each track to form truncated hyperspectral images. The resulting truncated images were unfolded such that each row contained the equivalent row of pixels from the images at all frequencies in the datacube, concatenated end to end. An unambiguous concentration figure representing all the data in this row can then be assigned. Continuing this process for all rows in the hyperspectral images produces three large 2-D matrices, one for each datacube, with each row in the matrix associated with a particular concentration. These matrices were then assigned as training, validation and test sets and a partial least squares regression (PLS) model [53–55] formed on the training and validation sets. PLS is related to PCA (see Section 3.2) except that the factors which are extracted are known as latent variables and are extracted (here) on the basis of the variance in both the spectral and the concentration (target) variables. This model was then used to predict the test set. It was found to produce a very precise prediction of the higher concentrations but a poor-to-negligible prediction of the dimmer/invisible gel bands corresponding to the lower concentrations as shown in Fig. 3a.

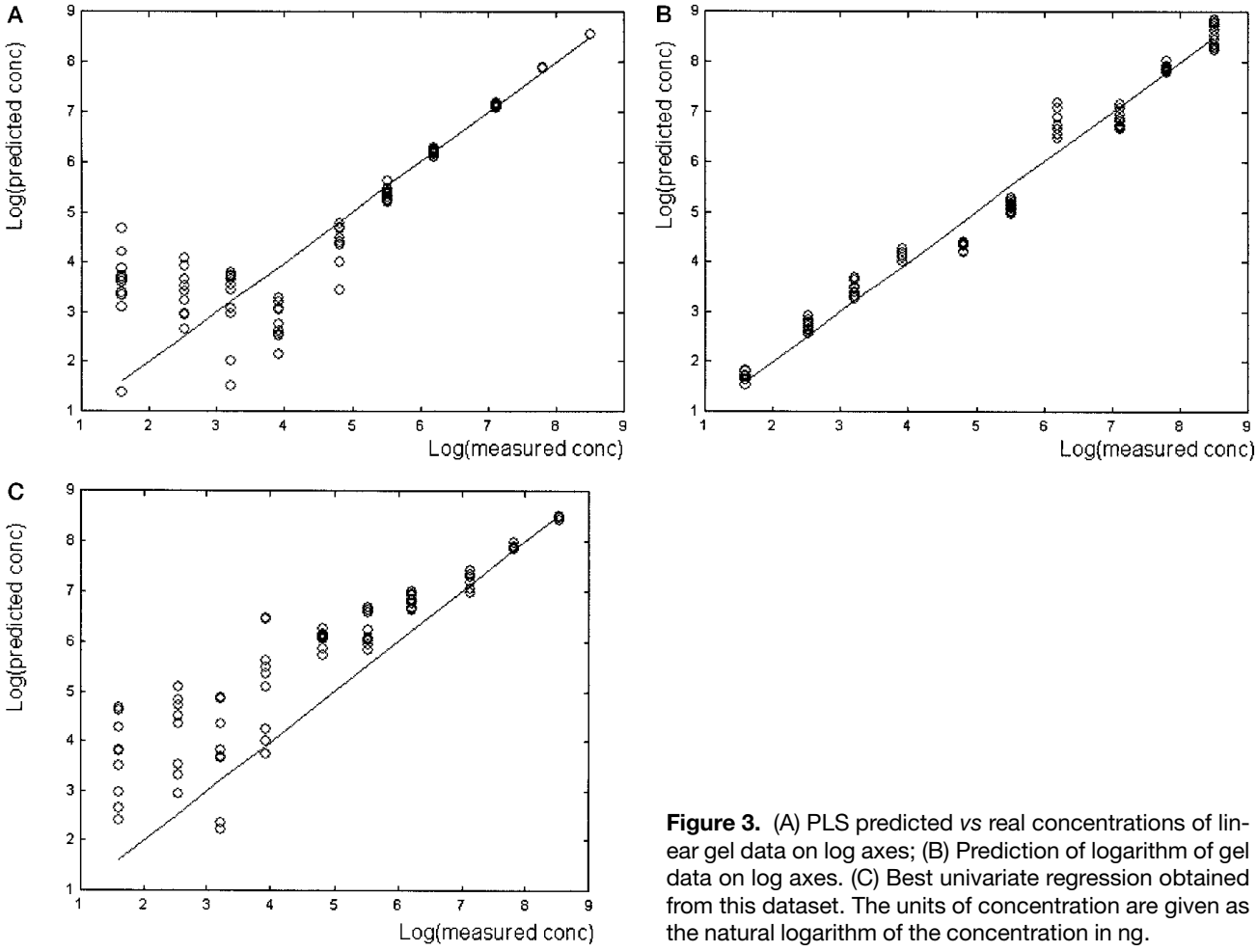

Figure 3. (A) PLS predicted *vs* real concentrations of linear gel data on log axes; (B) Prediction of logarithm of gel data on log axes. (C) Best univariate regression obtained from this dataset. The units of concentration are given as the natural logarithm of the concentration in ng.

When the logarithm of the intensity data was used to linearise the geometric concentration dependence then the PLS model trains to proportional rather than absolute error and greatly improved the prediction of the lowest concentrations (dimmest bands) at the cost of a slight loss in precision at the highest concentrations (brightest bands). This prediction was precise to a root mean square error of prediction (RMSEP) compared to the true concentration values of just 6.4% using only three PLS factors as shown in Fig. 3b. Carrying out a univariate prediction equivalent to the multivariate prediction of Fig. 3b on the logarithm of each individual column of the truncated images produced a null prediction. The best linear prediction equivalent to Fig. 3a had an RMSEP of 23% and is shown in Fig. 3c.

Study of the first factor PLS loadings of the prediction of Fig. 3b, as shown in Fig. 4, clearly shows the model is forming exclusively on those variables corresponding to the pixels in protein bands and ignoring those corres-

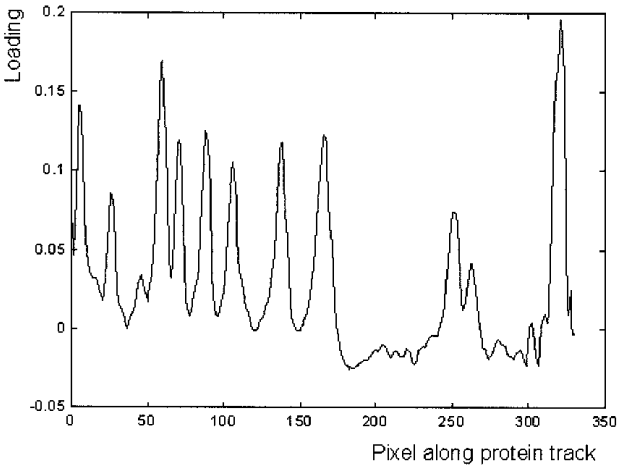

Figure 4. First factor loadings of PLS model on hyperspectral gel tracks. The peaks correspond to gel bands in the tracks, showing that the model forms predominantly on these peaks.

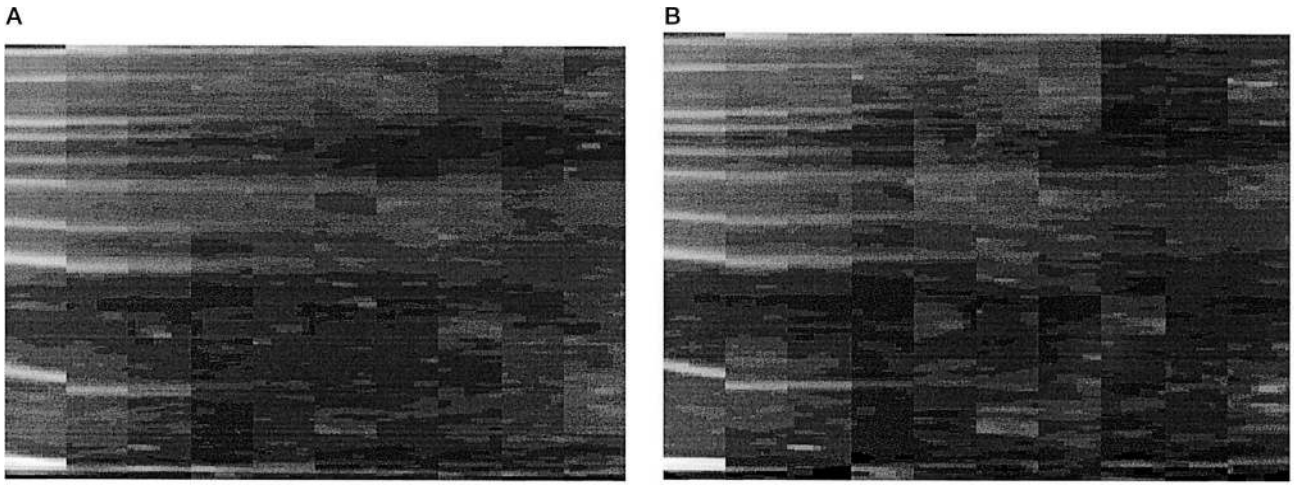

Figure 5. Truncated hyperspectral image: (A) true data and (B) hyperspectral prediction.

ponding to background pixels, and shows that the method is able to select and model the protein bands itself without user assistance. Mutual Information (MI) (see [56–59]) is a generalised version of correlation. Whereas correlation assumes linear relationships and Gaussian-distributed data, MI makes no assumptions about the two data series being compared. It is based on calculating the information content of one signal that is also contained in the other. Mathematical details are given in [56–59]. The same variables as described above were also picked out as having the highest MI of the datacube with the known concentration data (not shown). What is particularly important is that the PLS model can form precise predictions of protein bands at concentrations so low that they are not visible by eye on either the images produced from the Arthur multiwavelength fluorimager or visually in the Coomassie Blue stained gel. If the predicted dataset is reconstructed into an image from the datacube, it is almost indistinguishable from the true hyperspectral image. One example image from the actual truncated datacube of the test set is shown in Fig. 5a and the equivalent from hyperspectral prediction is Fig. 5b.

4 Concluding remarks

The separation of principal component scores originating from blank gel tracks with no protein from those scores originating from background areas of tracks which have had protein run up them clearly show that these background areas are not merely empty gel but contain traces of protein residue. The only possible source of this protein is as a result of the migration of bands through these regions ('slime trail' model). Univariate analyses of absorbance or fluorescence signals cannot discriminate signals that come from background from signals due to stain, and only the application of chemometrics to hyperspectral images has indicated this difference. The result shows that backround areas of blank tracks cannot be treated as if they are merely protein-free empty areas of pure gel, and has significant implications for our understanding of, and improvement in, the dynamic range of proteome gels.

PLS modelling on datacubes clearly demonstrates an ability to quantify fluorophores precisely in gels at far lower concentrations and greater sensitivity than are either visible in the images themselves or can be determined *via* the integration of band intensities in univariate images. It is able to produce a precise prediction to 6.4% error over a range of three orders of magnitude of concentration, from 5.10^{-9} g protein/band to 5.10^{-6} g protein/band (above which the gel spots are so bright as to be easily quantifiable by most methods). This represents an improvement in the dynamic range *per* pixel of two orders of magnitude.

We thank the UK BBSRC for financial support.

Received May 20, 2001

5 References

[1] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., *et al.*, *Proc. Natl. Acad. Sci. USA* 1996, *93,* 14440–14445.

[2] Humphery-Smith, I., Cordwell, S. J., Blackstock, W. P., *Electrophoresis* 1997, *18,* 1217–1242.

[3] Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F., (Eds.), *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997.

[4] Yates, J. R., *J. Mass Spectrom.* 1998, *33,* 1–19.

[5] O'Connor, C. D., Farris, M., Hunt, L. G., Wright, J. N., *Methods Microbiol.* 1998, *27,* 191–204.

[6] Hochstrasser, D. F., *Clin. Chem. Lab. Med.* 1998, *36,* 825–836.

[7] Blackstock, W. P., Weir, M. P., *Trends Biotechnol.* 1999, *17,* 121–127.

[8] Link, A. J., (Ed.), *2-D Proteome Analysis Protocols*, Humana, Totowa, NJ 1999.

[9] Page, M. J., Amess, B., Rohlff, C., Stubberfield, C., Parekh, R., *Drug Discov. Today* 1999, *4,* 55–62.

[10] Gygi, S. P., Rist, B., Aebersold, R., *Curr. Opin. Biotechnol.* 2000, *11,* 396–401.

[11] Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., Sanchez, J. C., *Electrophoresis* 2000, *21,* 1104–1115.

[12] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R., *Proc. Natl. Acad. Sci. USA* 2000, *97,* 9390–9395.

[13] Chen, H., Cheng, H., Bjerknes, M., *Anal. Biochem.* 1993, *212,* 295.

[14] Swain, M., Ross, N. W., *Electrophoresis* 1995, *16,* 948.

[15] Gruber, H. J., Hahn, C. D., Kada, G., Riener, C. K., *et al.*, *Bioconj. Chem.* 2000, *11,* 696–704.

[16] Alba, F. J., Daban, J. R., *Electrophoresis* 1998, *19,* 2407–2411.

[17] Berggren, K., Steinberg, T. H., Lauber, W. M., Carroll, J. A., *et al.*, *Anal. Biochem.* 1999, *276,* 129–143.

[18] Steinberg, T. H., Jones, L. J., Haugland, R. P., Singer, V. L., *Anal. Biochem.* 1996, *239,* 223–237.

[19] Steinberg, T. H., Haugland, R. P., Singer, V. L., *Anal. Biochem.* 1996, *239,* 238–245.

[20] Steinberg, T. H., White, H. M., Singer, V. L., *Anal. Biochem.* 1997, *248,* 168–172.

[21] Steinberg, T. H., Chernokalskaya, E., Berggren, K., Lopez, M. F., *et al.*, *Electrophoresis* 2000, *21,* 486–496.

[22] Steinberg, T. H., Lauber, W. M., Berggren, K., Kemper, C., *et al.*, *Electrophoresis* 2000, *21,* 497–508.

[23] Berggren, K., Chernokalskaya, E., Steinberg, T. H., Kemper, C., *et al.*, *Electrophoresis* 2000, *21,* 2509–2521.

[24] Yan, J. X., Harry, R. A., Spibey, C., Dunn, M. J., *Electrophoresis* 2000, *21,* 3657–3665.

[25] Unlu, M., Morgan, M. E., Minden, J. S., *Electrophoresis* 1997, *18,* 2071–2077.

[26] Davey, H. M., Kell, D. B., *Microbiol. Rev.* 1996, *60,* 641–696.

[27] O'Brien, K. M., Wren, J., Dave, V. K., Bai, D., *et al.*, *Rev. Sci. Instr.* 1998, *69,* 2141–2146.

[28] Curran, P. J., *Prog. Phys. Geog.* 1994, *18,* 247–266.

[29] Chiou, W. C., *Appl. Optics* 1985, *24,* 2085–2091.

[30] Goetz, A. F. H., Vane, G., Solomon, J., Rock, B. N., *Science* 1985, *228,* 1147–1153.

[31] Birk, R. J., McCord, T. B., *IEEE Aerospace Electronic Systems Magazine* 1994, *9,* 26–33.

[32] Filiberti, D. P., Marsh, S. E., Schowengerdt, R. A., *Optical Eng.* 1994, *33,* 2520–2528.

[33] Abousleman, G. P., Gifford, E., Hunt, B. R., *Optical Eng.* 1994, *33,* 2562–2571.

[34] Lee, Z. P., Carder, K. L., Hawes, S. K., Steward, R. G., *et al.*, *Appl. Optics* 1994, *33,* 5721–5732.

[35] Rogers, S. K., Colombi, J. M., Martin, C. E., Gainey, J. C., *et al.*, *Neural Netw.* 1995, *8,* 1153–1184.

[36] Wilson, T. A., Rogers, S. K., Myers, L. R., *Optical Eng.* 1995, *34,* 3154–3164.

[37] Gaddis, L. R., Soderblom, L. A., Kieffer, H. H., Becker, K. J., *et al.*, *IEEE Trans. Geosci. Remote Sensing* 1996, *34,* 163–178.

[38] Babey, S. K., Anger, C. D., *Proc. SPIE* 1996, *2599,* 302–307.

[39] Kruse, F. A., *Int. J. Remote Sensing* 1996, *17,* 1623–1632.

[40] Roger, R. E., *Int. J. Remote Sensing* 1996, *17,* 589–613.

[41] Willoughby, C. T., Folkman, M. A., Figueroa, M. A., *Proc. SPIE* 1996, *2599,* 264–272.

[42] Goodacre, R., Timmins, É. M., Burton, R., Kaderbhai, N., *et al.*, *Microbiology* 1998, *144,* 1157–1170.

[43] Goodacre, R., Burton, R., Kaderbhai, N., Timmins, É. M., *et al.*, in: Stopa, P. J., Bartoszcze, M. A., (Eds.) *Rapid Methods for Monitoring the Environment for Biological Hazards,* Kluwer, Dordrecht 2000, pp. 111–136.

[44] McGovern, A. C., Ernill, R., Kara, B. V., Kell, D. B., Goodacre, R., *J. Biotechnol.* 1999, *72,* 157–167.

[45] Winson, M. K., Goodacre, R., Timmins, É. M., Jones, A., *et al.*, *Anal. Chim. Acta* 1997, *348,* 273–282.

[46] Schultz, R. A., Nielsen, T., Zavaleta, J. R., Ruch, R., *et al.*, *Cytometry* 2001, *43,* 239–247.

[47] Laemmli, U. K., *Nature* 1970, *227,* 680–685.

[48] Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., *et al.*, *Anal. Biochem.* 1985, *150,* 76–85.

[49] Spibey, C. A., Jackson, P., Herick, K., *Electrophoresis* 2001, *22,* 829–836.

[50] Drury, S. A., *Image Interpretation in Geology, 2nd Edition,* Chapman and Hall, London 1993, p. 23.

[51] Schwartz, D. C., Cantor, C. R., *Cell* 1984, *37,* 67–75.

[52] Jolliffe, I. T., *Principal Component Analysis,* Springer-Verlag, New York 1986.

[53] Martens, H., Næs, T., *Multivariate Calibration,* John Wiley, Chichester 1989.

[54] Martens, M., Martens, H., in: Piggott, J. R. (Ed.), *Statistical Procedures in Food Research,* Elsevier, Amsterdam 1996, pp. 293–359.

[55] Höskuldsson, A., *J. Chemometrics* 1988, *2,* 211–228.

[56] Shannon, C. E., Weaver, W., *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, IL 1949.

[57] Battiti, R., *IEEE Trans. Neural Networks* 1994, *5,* 537–550.

[58] Gilbert, R. J., Goodacre, R., Woodward, A. M., Kell, D. B., *Anal. Chem.* 1997, *69,* 4381–4389.

[59] Alsberg, B. K., Woodward, A. M., Winson, M. K., Rowland, J. J., Kell, D. B., *Anal. Chim. Acta* 1998, *368,* 29–44.