Amna Butt[1,4], Matthew D. Davison[2], Graeme J. Smith[2], Janice A. Young[2], Simon J. Gaskell[3], Stephen G. Oliver[4], Robert J. Beynon[5]

[1]Department of Biomolecular Sciences, UMIST, Manchester, UK
[2]Proteomics Group, Astra-Zeneca, Cheshire, UK
[3]Michael Barber Centre for Mass Spectrometry, UMIST, Manchester, UK
[4]School of Biological Sciences, University of Manchester, Manchester, UK
[5]Department of Veterinary Preclinical Sciences, University of Liverpool, Liverpool, UK

Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis

Current methods of proteome analysis rely almost solely on two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) followed by the excision of individual spots and protein identification using mass spectrometry (MS) and database searching. 2-D PAGE is denaturing in both dimensions and, thus, cannot indicate functional associations between individual proteins. Moreover, less abundant proteins are difficult to identify. To simplify the proteome, and explore functional associations, nondenaturing anion exchange column chromatography was used to separate a soluble protein extract from *Escherichia coli.* Successive fractions were then analysed using 2-D PAGE and selected spots from both the gels for the start material and the fractionated material were quantified and identified by peptide mass fingerprinting using a MALDI-TOF mass spectrometer. Enrichments of up to 13-fold were attained for individual protein spots and peptide mass fingerprints were of significantly higher quality after chromatographic separation. The marked anomalies between predicted p*I* and column elution position contrasted with the almost perfect correlation with migration distance on isoelectric focusing (IEF) and were explored further for basic proteins.

Keywords: Ion exchange chromatography / Proteome / *Escherichia coli* / MALDI-TOF

PRO 0008

1 Introduction

Functional genomics [1] aims at a comprehensive analysis of gene action and interaction in species with completely sequenced genomes. Four levels of analysis are commonly exploited: genome, transcriptome, proteome, and metabolome [2, 3]. The genome can be taken as a constant during the time frame of most experiments. However, for certain species (*e.g.* the bacterium *E. coli* or the unicellular eukaryote, *Saccharomyces cerevisiae*), the experimenter has the ability to exquisitely manipulate the genome in order to facilitate the systematic analysis of gene function [4]. The last three levels of functional analysis are distinguished from that of the genome because they are context dependent. The entire complement of mRNA molecules, proteins, or metabolites in a tissue, organ or organism varies with its physiological, pathological, or developmental condition.

Of the four levels of analysis, it is the proteome that should be most useful in making functional assignments. There are three main reasons for this. The first is the context dependency of the proteome. The second is that proteins (like the low molecular weight intermediates of the metabolome, but unlike the mRNAs of the transciptome) are functional entities within the cell. The last is that (like mRNAs, but unlike metabolites) there is a direct relationship between proteins and genes. This relationship is not of the one-to-one kind, however. A single gene may have multiple protein products due to the differential splicing of its primary transcript, proteolytic cleavages, or the covalent modification of the mature protein product. Thus, given these multipliers, the complete assignment of all proteins resolved by 2-D PAGE of cellular extracts (the usual approach employed for proteome analysis [5]) is a daunting task. For instance, the first steps towards identifying the total protein complement of *E. coli* were made soon after the development of 2-D PAGE [6] when work to establish a gene-protein database for *E. coli* was initiated [7]. The first edition of this database was published in 1983 [8]. The index attempted to cross reference all known genes of *E. coli* with known proteins but it was far from complete even for those segments of the genome that were well studied.

Current methods of proteome analysis still rely almost solely on 2-D PAGE, although the use of two dimensions of liquid chromatography (LC) has been reported recently [9–11]. The original problems with the reproducibility of

Correspondence: Professor Stephen G. Oliver, School of Biological Sciences, Room 2.205 Stopford Building, University of Manchester, Oxford Road, Manchester MI 3 9PT, UK
E-mail: steve.oliver@bso.man.ac.uk
Fax: +44-161-606-7360

Abbreviation: IOD, integrated optical density

1615-9853/01/0101–0042 $17.50+ .50/0

2-D PAGE separations have been considerably reduced by the recent introduction of immobilised pH gradients [12, 13] and standardised protocols [14]. However, the most important advance for proteome research has been the development of new "soft" methods of ionisation and desorption in mass spectrometry (MS) [15]. In particular, the development of matrix assisted laser desorption/ionisation (MALDI), and electrospray ionisation (ESI) MS has greatly facilitated the assignment of protein identities to spots on 2-D PAGE gels.

The most common method of identification of proteins by MS is the analysis of peptide masses following enzymatic digestion of proteins resolved on 2-D gels using MALDI-TOF [16]. The resultant theoretical "peptide fingerprint" is diagnostic of a protein, tolerant to site-specific, post-translational modifications and can be searched against organism-specific theoretical digest databases. Automation of many of the experimental procedures used has made it possible to identify proteins on a large scale [17]. The high-throughput identification of proteins excised from 2-D gels was demonstrated by Shevchenko *et al.* [16], who used this procedure to identify proteins expressed by yeast, following the resolution of a cell lysate by 2-D PAGE.

This method of protein identification is fairly rapid and reliable, but any approach that relies exclusively on 2-D PAGE suffers from a number of intrinsic limitations. First, the limitation of protein loading means that many low abundance proteins are unlikely to be within the limited dynamic range of the staining protocol, or may be obscured by highly abundant proteins. Second, both dimensions of 2-D PAGE are denaturing (urea or another chaotrope is used in the first dimension and the detergent dodecylsulphate is employed in the second dimension). As such, all noncovalent protein-protein interactions are destroyed and, with them, the possibility of assigning the function of novel gene products based on their interactions with proteins of known function. Such relationships are exploited to make functional assignments using the yeast two-hybrid system [18] in an approach described as "guilt-by-association" [3]. However, two-hybrid analysis identifies potential, rather than actual, protein-protein interaction and thus there is considerable scope for the exploitation of classical biochemical approaches to the separation of proteins in non denaturing solutions in order to elucidate the functions of novel gene products. In this paper, we describe the use and consequences of I-D nondenaturing LC to resolve *E. coli* proteins, prior to resolution on 2-D PAGE and peptide fingerprinting.

2 Materials and Methods

Overnight cultures of *E. coli* strain W3110 were used to inoculate 1 500 ml (2 x 750 mL) of L-broth and this culture was incubated with shaking at 37°C until an OD_{600} of 0.8 was reached. Cells were harvested by centrifugation for 5 min at 10 000 *g*. The supernatant was discarded and the cells were resuspended in 2 mL of 50 mM HEPES buffer, pH 7.5, and transferred to microfuge tubes. Following centrifugation for 1 min, cells were snap-frozen in an ethanol/dry ice bath.

Approximately 1 mL of the frozen cells were thawed and resuspended in 5 mL of extraction buffer (50 mM HEPES, pH 7.5) containing 500 units/mL of DNAse I, 50 units/mL of RNase A (Boehringer Mannheim, Indianapolis, IN, USA) and Complete™ Mini EDTA-free protease inhibitor cocktail (Boehringer). In some experiments, the nuclease treatment was omitted. Cells were disrupted by sonication in 10 s bursts interspersed with 30 s intervals on ice. The extract was centrifuged twice at 14 000 *g* for 20 min and the supernatant (clarified cell lysate) collected. The final extract was then filtered through a 0.2 μm PVDF filter (Whatman International, Maidstone, UK) and the protein content was determined using a Coomassie Plus protein assay reagent (Pierce Chemical, Rockford, IL, USA) prior to chromatography.

The total protein extract was fractionated by anion exchange chromatography, on a Mono-Q™ column (bed volume, 1 mL) using a fast protein liquid chromatography (FPLC) system (APBiotech, London, UK). The column was pre-equilibrated with 50 mM HEPES, pH 7.5 before 2 mL of the extract was applied. The bound protein was eluted using a two stage linear gradient of NaCl in the same buffer, firstly from 0–400 mM NaCl in 40 mL and from 400–1 000 mM in 15 mL. The eluate was collected in 1 mL fractions. Following chromatography, fractions were analysed by 1-D SDS-PAGE or 2-D PAGE. Proteins in the fractions were precipitated by 20% w/v TCA precipitation at 4°C for 60 min. The pellet was then sedimented in a microcentrifuge (14 000 *g* for 20 min at 4°C) and washed twice with diethyl ether to remove the TCA. The 1-D SDS–PAGE used a Mini PROTEAN™ Dual Slab cell (Bio-Rad Laboratories, Hercules, CA, USA) using a Tris hydrochloride/Tris-glycine discontinuous buffer system [19]. Proteins were separated on 12.5% SDS–PAGE gels under reducing conditions and stained with Coomassie Brilliant Blue.

For 2-D PAGE of the fractionated material, the whole 1 mL fraction was precipitated as described above and the precipitate resuspended in 400 μL of sample buffer (9 M urea; 0,4% w/v DTT, 0.8% v/v Pharmalytes (APBiotech); 4% w/v CHAPS). Electrophoresis of the whole cell lysate was carried out by precipitating 300 μL of the clarified cell lysate which, again, was resuspended in 400 μL of the same sample buffer. The samples were agitated on a

rotary shaker for at least 30 min to ensure all the precipitate was redissolved before centrifugation at 14 000 *g* for 2 min. Bromophenol blue (5 μL of a 4% w/v solution) was then added to 375 μL of the supernatant and this solution was used to swell Immobiline DryStrips (18 cm, pH range 4–7 or 3–10; APBiotech) overnight. The first dimension gels were run at 20°C under oil as described by the manufacturer, on a Multiphor II (APBiotech). As soon as electrophoresis was completed, the strips were drained of excess oil and used immediately, or stored at –80°C for subsequent loading onto the second dimension. IPG strips were placed in 15 mL of equilibration buffer (50 mM Tris-HCl (pH 8.8), 6 M urea, 30% glycerol, 1% w/v DTT, 1% w/v SDS) and left on a shaker for 10 min. This buffer was discarded and the equilibration step repeated a second time before running the second dimension.

For the second dimension, 12.5% polyarcylamide vertical slab gels (22 x 22 cm) were cast onto Gelbond PAG backing film (FMC BioProducts, Rockland, MA, USA). Equilibration buffer was removed from the IPG strip which was then laid on top of the second dimension gel. The gels were run using the ESA apparatus (Genomic Solutions Ltd., Chelmsford, MA, USA) according to the manufacturer's recommendations, at a constant 300 V and 20°C. Proteins were visualised with Coomassie Brilliant Blue.

In-gel digestions were a modification of that of Shevchenko *et al.* [16]. Excised gel pieces (approximately 1–2 mm cubes) were placed in microtitre plate wells (Costar, Cambridge, MA, USA) and then washed for 1 h with 100 μL of water, followed by a similar wash step in 100 μL of 20 mM Tris HCl, pH 8. This supernatant was discarded and replaced by 100 μL of 50% v/v Tris HCl (pH 8), 50% v/v ACN ((sequencing grade), Perking-Elmer, Norwalk, CT, USA). Gel pieces were left in this solution for 30 min before replacing it with 50 μL of ACN and the gel pieces were left for a further 10 min incubation. After discarding this supernatant the gel pieces were thoroughly dried in a vacuum dryer (Savant, Halbrook, NY, USA) for 15 min. Trypsin stock solution (1 μg/μL of 0.01% v/v TFA) was then diluted 80-fold in 20 mM Tris HCl, pH 8.0 to give a final enzyme concentration of 12.5 ng/μL. The enzyme was added in 10 μL aliquots to each of the dry gel pieces. If all the solution was absorbed in the first addition, a further 10 μL was added. The microtitre plate was sealed and incubated overnight at 37°C. Peptides were extracted by adding 25 μL of 0.1% v/v TFA, 60% v/v ACN to each well and shaking the plate for 30 min. The supernatants were removed to a fresh 96-well plate and the extraction procedure repeated. The combined extracts were then reduced to approcimately 15 μL by vacuum concentration.

The protein digests were analysed by MALDI-TOF MS, using the Voyager De-STR instrument (Perseptive Biosystems, Framingham, MA, USA) set up in autosample mode. The matrix (α-cyano-4-hgydroxycinnamic acid) Aldrich Chemical, Milwaukee, USA) was made up to a saturated solution using 0.1% v/v TFA, 60% ACN. Matrix and sample 0.5 μL of each) were spotted onto a stainless steel, 100-sample, target plate (PerSeptive Biosystems) using a robotic sample loader (Symbiot, PerSeptive Biosystems). Calibration standards were a mixture of two obtained from Perceptive Biosystems and prepared as described in the manual. The monoisotopic mass range covered by the standards was between 904.4681 and 5730.6087 Da (Sequazyme™ instruction manual, PerSeptive Biosystems).

Coomassie stained 2-D gels were scanned using an 8-bit Umax scanner (150 dpi) (Freemont, CA, USA). Images of the fractionated material were overlaid onto the gel for the starting material and proteins considered to have migrated to the same position on both gels were assumed to be identical. These protein spots were excised from both gels and were submitted for MS analysis to confirm their identity. Once the identities of the protein spots, which comigrated on the fractionated and starting material, were shown to be the same, spots from the original images were quantified using HDG Bio-image spot detection software (Genomics Solutions). The intensity of an individual protein spot was calculated using this software and then the integrated intensity of all the spots on the gel was calculated. The scanner/software combination was linear in response ($r = 0.99$).

Database searching was performed using ZIPS (AstraZeneca), a software tool which is mechanistically similar to MS-Fit from Protein Prospector. This tool catalogues all known and putative proteins encoded by the *E. coli* genome, allowing for the production of a theoretical peptide fingerprint for each of these proteins upon digestion with some of the common proteases. ZIPS also takes into account any possible occurrence of mis-cleavages and common modifications of peptide such as oxidised methionine residues. Experimental data from MALDI mass analysis of peptides was compared with the theoretical digest generated by ZIPS using the enzyme trypsin. Identical results were obtained when a public domain search engine (MASCOT, http://www.matrixscience.com) was used.

The first group of 2-D gels were run using pH 4–7 IEF strips. The gel number corresponds to the fraction number from the elution profile where the unfractionated material is numbered zero. Protein spot numbers were prefixed with the number of the gel followed by the spot

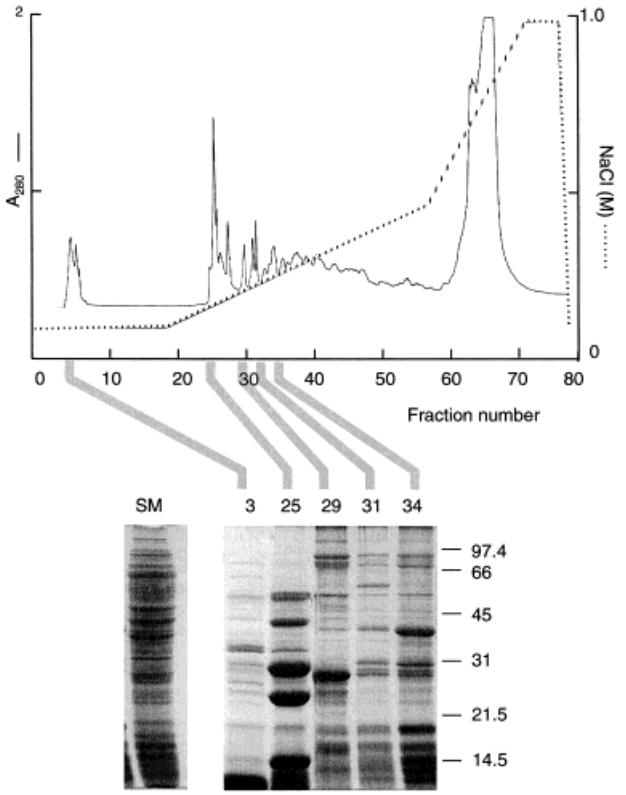

Figure 1. Resolution of *E. coli* soluble proteins by single dimension ion exchange chromatography. Soluble proteins from an *E. coli* cell lysate (16 mg protein) were separated by a linear salt gradient on MonoQ anion exchange chromatography and the absorbance was monitored at 280 nm. Selected fractions were analysed by 1-D SDS-PAGE and stained with Coomassie Blue.

number. The second series of gels were run using full range pH 3–10 IEF strips and the gels were numbered in the same way but prefixed with 'B'.

3 Results and Discussion

The protein concentration of the centrifuge-clarified whole-cell lysate from *E. coli* was approximately 16 mg/mL. There were no significant differences in the elution profile on MonoQ chromatography when loadings of between 8 and 32 mg of protein were applied to the column (results not shown). Routinely, 2 mL of whole cell lysate (starting material) was used as the standard loading level. When fractions were recovered and screened on 1–D SDS-PAGE, the marked enhancement in enrichment of individual proteins is highly evident (Fig. 1). There was no marked preponderance of any one protein in the starting material, but each of the representative fractions were highly enriched in one or more proteins.

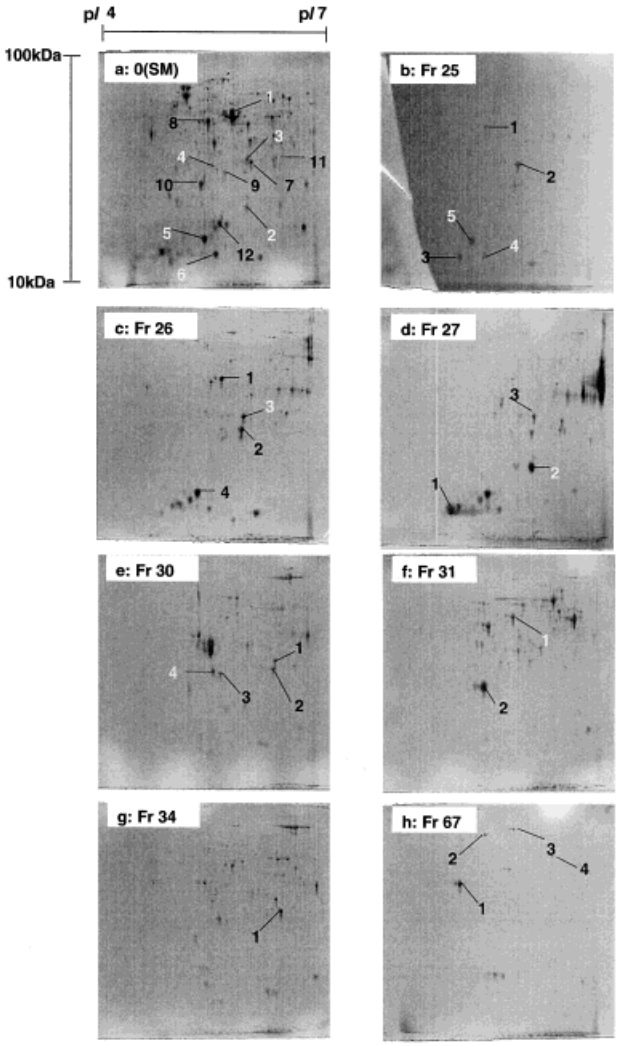

Figure 2. 2-D gel analysis of *E. coli* proteins separated by ion exchange chromatography. Fractions (the entire fraction) from the ion exchange separation such as shown in Fig. 1 were resolved on 2-D gel electrophoresis. For comparison, the starting material (300 μL, 6 mg protein) was resolved in the same way. The 2-D PAGE used linear IPG strips (pH 4–7) for the first dimension, and linear 12.5% gels in the second dimension. Each gel was stained with Coomassie Blue. Spots labelled in black were identified on a single gel. Spots labelled in white were identified both in the starting material and one of the fractions.

The whole cell cleared lysate (Fig. 2a) and selected fractions from the ion exchange column (Figs. 2b–h) were subjected to 2-D PAGE using IEF between pH 4 and pH 7. Since the ion exchange column separates proteins on the basis of charge, proteins co-eluting in the same fraction should have similar apparent isoelectric points, and proteins eluting later on the column should be more

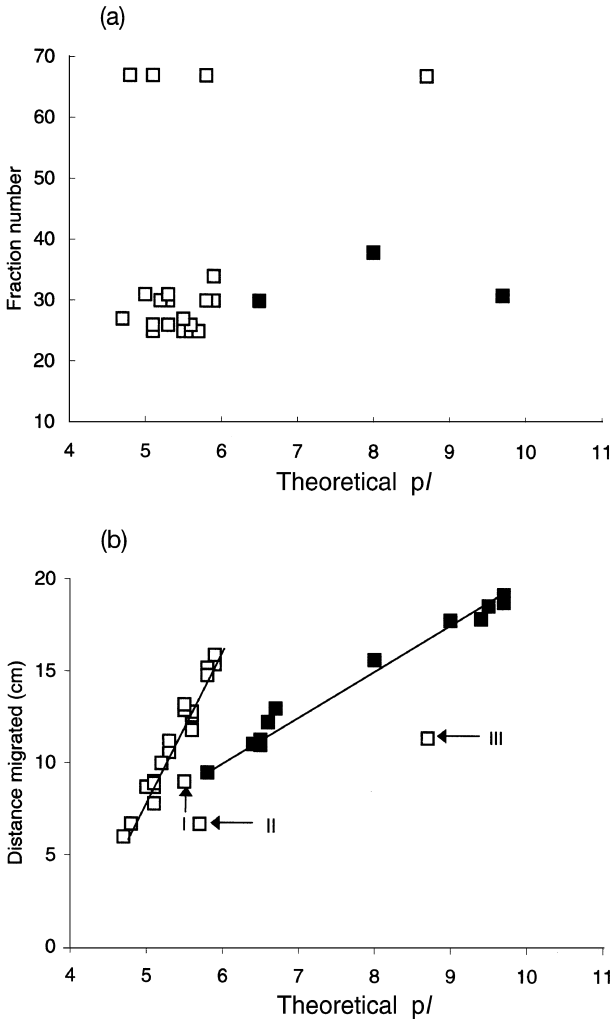

Figure 3. Correlation between the theoretical p*I* of *E. coli* proteins with their elution volume following chromatography and their migration on a 2-D gel. For proteins that were identified by MALDI-TOF, the theoretical p*I* was calculated from the published protein sequence (http://www.expasy.ch/tools/pi_tool.html). The relationship between elution position on the anion exchange column was correlated with theoretical p*I* (a). Additionally, the migration on IEF gels was correlated with elution position for both series of experiments (b). In the second series of experiments (closed squares) a broad range p*I* gel was used, which explains the difference in slope of the correlation. Three proteins that exhibited anomalous migration on 2-D PAGE (I, II and III) are discussed further in the text.

acidic. However, the 2-D gels gave no evidence of protein clustering in particular regions of the gels, and there was no clear indication of a trend in the distribution of the p*I* values of the eluted protein during column development. For proteins that were subsequently identified by MALDI-TOF, we calculated the theoretical p*I* (http://www.expasy.ch/tools/pi_tool.html) and related this parameter to the elution volume on the MonoQ column and the migration distance on IEF (Fig. 3a). There is no correlation between predicted p*I* and elution position on ion exchange chromatography. By contrast, the migration of proteins in the first dimension of the 2-D gel correlates strongly with the theoretical p*I,* irrespective of the pH range over which the proteins are separated (Fig. 3b). There is therefore no evidence for widespread, charge-shifting, post-translational modifications, and the strength of the latter correlation serves to emphasise the lack of correlation between predicted p*I* and elution position in Fig. 3a. Composition derived p*I* has little influence on chromatographic retention which is modified by surface charge features, by formation of protein-protein complexes (in which the combined p*I* is different to the composite p*I* of the constituents) or by non-ionic interactions with the chromatographic matrix. Although this behavior is often appreciated by protein chromatographers, it emphasises that the translation of the gel based techniques to LC will introduce new dimensions of separation, which of course may be advantageous.

Only three proteins (marked I, II and III on Fig. 3b) showed a markedly anomalous relationship between theoretical p*I* and migration position of IEF. Spots I and II were identified to be MalE and HdeB, respectively. MalE has a predicted p*I* of 5.5, but migrated at an apparent p*I* of 5.2. HdeB has a predicted p*I* of 5.7, but migrated at an observed p*I* of 4.8. In both cases, the shift to a more acidic position can be explained by a loss of a basic signal peptide – the theoretical p*I* values of the mature forms of these proteins were calculated as 5.2 and 4.9, respectively. No *N*-terminal peptides for the signal peptide region were present in either mass spectrum. Spot III was DeaD, a ribosomal associated protein and one of the proteins involved in the *E. coli* cold shock response. It probably has an aberrant migrational position through loss of part of the arginine-rich *C*-terminus. Given the strength of the theoretical/observed correlation seen for the rest of the proteins that were identified, such anomalies are of value in identification of charge-shifting, post-translational changes.

For this study, it was not our intention to identify large numbers of proteins. Rather, we aimed to recover spots that we putatively assessed as being the same (by virtue of their 2-D co-ordinates on the gel) in the starting material and the chomatographic fractions to demonstrate chromatographic enrichment. Moreover, the choice of Coomassie Blue staining meant that we were explicitly exploring the behaviour of the more abundant proteins. In total, a sample of 34 proteins were excised and identified by MALDI-TOF, these represented 25 different pro-

teins. Proteins identified from the whole-cell lysate are prefixed with the number 0 and those from the fractionated material are prefixed with their fraction number on ion exchange chromatography. There were instances where two protein spots at the same coordinates on different gels were subsequently identified as different proteins (Pgk, 0–8 and MalE, 25–1; Upp, 0–9 and Apt, 30–3). This serves to emphasise that multiple proteins occupy the same position on a 2-D gel (a key problem, particularly with complex proteomes), and the role that chromatography can play in separation of such proteins.

A further observation (Table 1) is the identification of 4 spots (0–5, 0–6, 25–5 and 25–4; Figs. 2a and 2b, respectively) as the same protein, YfiD, a 127-amino acid theoretical ORF (14.3 kDa, p*I* 5.1). The spectra for all four spots were reliable, with a least 9 peptides accurately matched in each case. The spots have different migrational properties in both dimensions, and the larger of the two proteins has a migration in the second dimension commensurate with the predicted mass of the intact protein. The spectra for spots 0–6 and 25–4 contained no peptides representing the amino acid sequence after residue number 99, whereas the spectra for spots 0–5 and 25–5 contained peptides for the protein sequence up to residue 116. The tryptic fragments representing residues 117–121 and 121–127 would be too small to be detected (573 Da and 697 Da, respectively). Cleavage of the *C*-terminal amino acids (residues 100–127) would yield a theoretical mass of 11.1 kDa and a theoretical p*I* of 5.0. It is likely that the proteolytic cleavage of the *C*-terminal amino acids resulted in the observed mass and p*I* shift of the proteins. At this stage, it is not possible to state whether or not the proteins were fragmented before or after cell lysis, although protease inhibitors were included in the lysis buffer. Although the two spots of YfiD have different size and charge properties associated with proteolysis, they nonetheless elute in the same fraction on the chromatographic trace.

Visual comparison of the gel images from the IEF fractions with that of the starting material (Figs. 1 and 2) indicated that the fractions were not only simpler than the whole cell lysate but were also enriched for particular protein species. The simpler protein patterns meant that the amount of individual proteins represented on a gel could be increased greatly, without exceeding the protein loading of the gel. To assess the enrichment, image analysis software was used to calculate the integrated optical density (IOD) of equivalent proteins (identified by MALDI-TOF) from the starting and the fractionated material. The abundance of the individual protein was then calculated as a percentage of the total protein of the gel. Four proteins identified in the starting material and in specific fractions

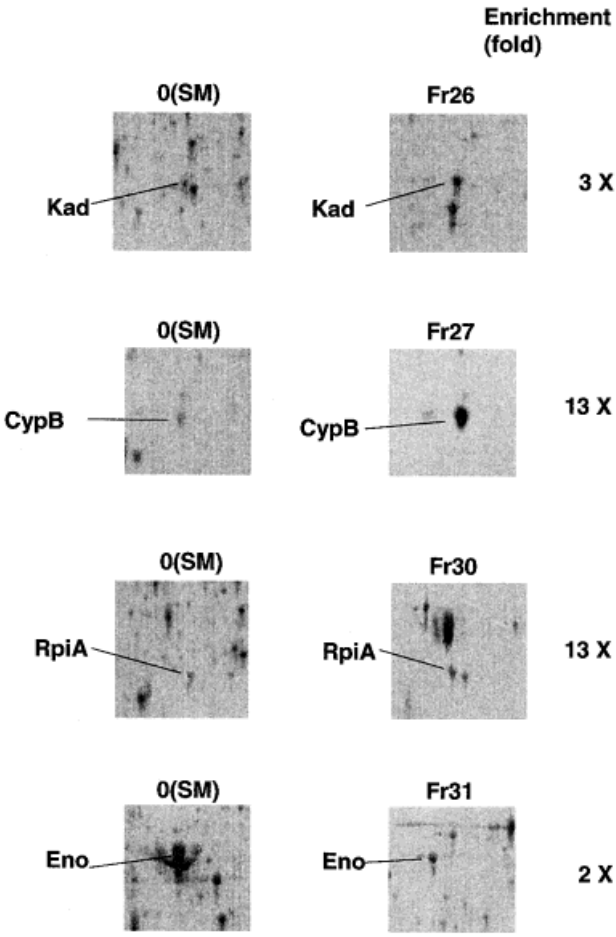

Figure 4. Chromatographic enrichment of proteins by LC. Spots in the 2-D gels of *E. coli* lysate and ion exchange fractions were selected on the basis of identical *x, y* coordinates and analysed by peptide mass fingerprinting. When mass spectral identifications confirmed that the proteins were identical, the IOD of the spot was expressed relative to the summed IOD of all spots on the gel, and the enrichments of the identified proteins were calculated as the increase in relative IOD. The enrichment values are indicated.

were enriched between two-fold and 13-fold (Fig. 4). The variability in enrichment should not be surprising, as any one protein may elute in consecutive fractions (for example, a protein with the same migrational properties as enolase can be seen in the gels for fractions 30 and 34 as well as 31) or a protein may be eluted from the column in different fractions if it is present both as a free species and as part of a multiprotein complex.

Chromatographic enrichment improved the peptide mass fingerprint data, giving mass spectra with increased number of peptide masses detected and improved signals. To

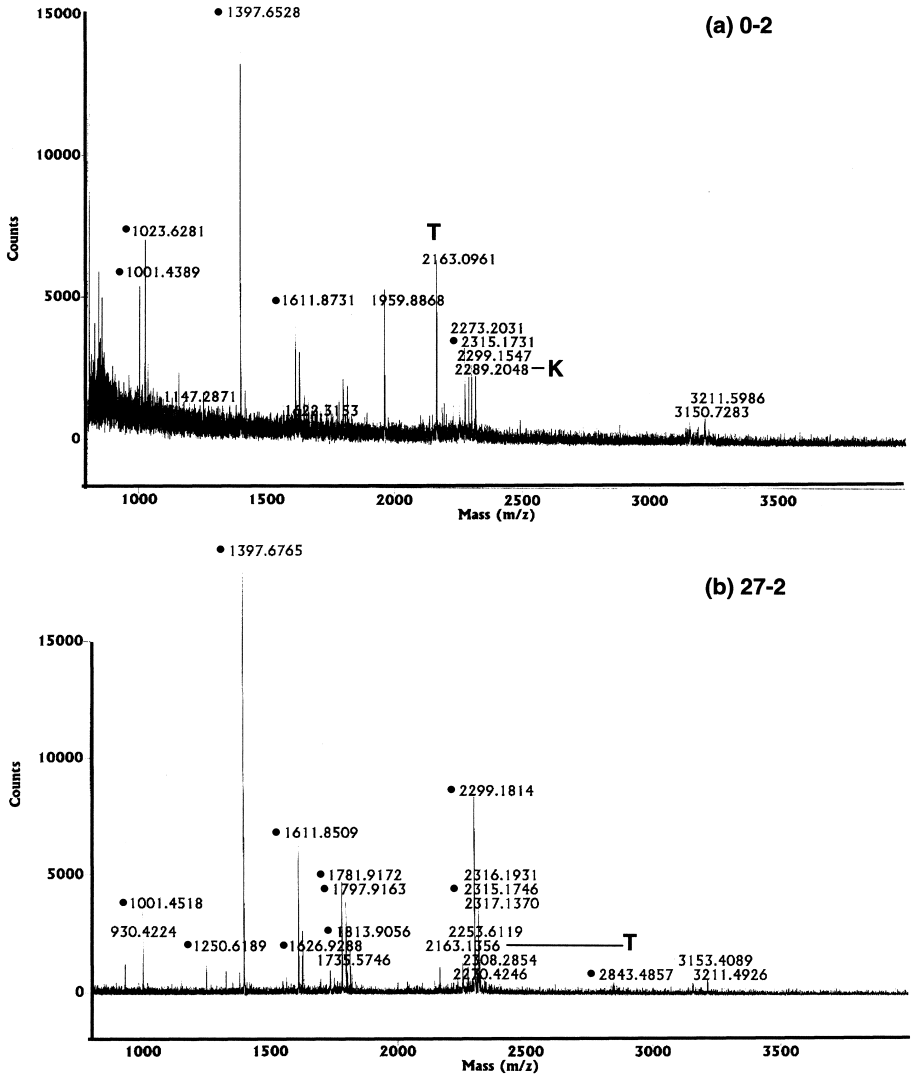

Figure 5. Effect of chromatographic enrichment on spectral quality. A representative protein (CypB), recovered from a 2-D gel of whole soluble protein extract (a) and from a single ion exchange fraction (b) (spots 0–2 and 27–2 respectively), was analysed by in-gel digest and MALDI-TOF MS. Peptide masses indicated with T were derived from trypsin, and with K were derived from keratin. Peptides indicated by a closed circle were present in the target CypB protein.

illustrate, the mass spectra for spots 0–2 and 27–2 (both identified as the protein CypB) were compared. CypB was enriched some 13-fold by the chromatographic step. This is not immediately evident from the signal intensities of the two spectra, but is should be noted that a much higher laser power (2 340 as opposed to 2 100) was needed to generate a signal for digest 0–2. The signal to noise ratio of the spot derived from a chromatographically purfied fraction (27–2) was dramatically improved (Fig. 5) and new peptides were detected which were not observed in the weaker spectrum for 0–2. This is attributable to the increased amount of protein available for digestion relative to the background noise. Eleven out of a possible 16 peptides matched in the spectrum from spot 27–2, these peptides represented 49% of the total sequence for this protein. From spot 0–2 only five peptides were matched (21% of the total protein sequence). All peptides assigned in 0–2 were also matched for 27–2, but the additional peptides identified from 27–2 significantly enhanced the reliability of identification, and may be particularly relevant with less abundant proteins.

One explanation for the lack of correlation between theoretical p*I* and elution position is that the individual proteins are assembled into larger, heterogeneous complexes. Ion exchange chromatography at near-neutral pH values, in relatively low salt polar buffers is nondenaturing, and strongly associated proteins should therefore co-elute. This is illustrated by the proteins identified from fraction 67 (Fig. 2h). It is apparent that the strong A280 nm peak seen very late in the salt gradient (Fig. 1, fractions 60–70) was unlikely to represent protein (confirmed by protein analysis and by the 2-D gel of fraction 67, Fig. 2h). This absorbance peak also had a high absorbance at 260nm (A260nm/A280nm) 2.1) and was likely to be nucleic acid that binds strongly to the anion exchange beads. The pro-

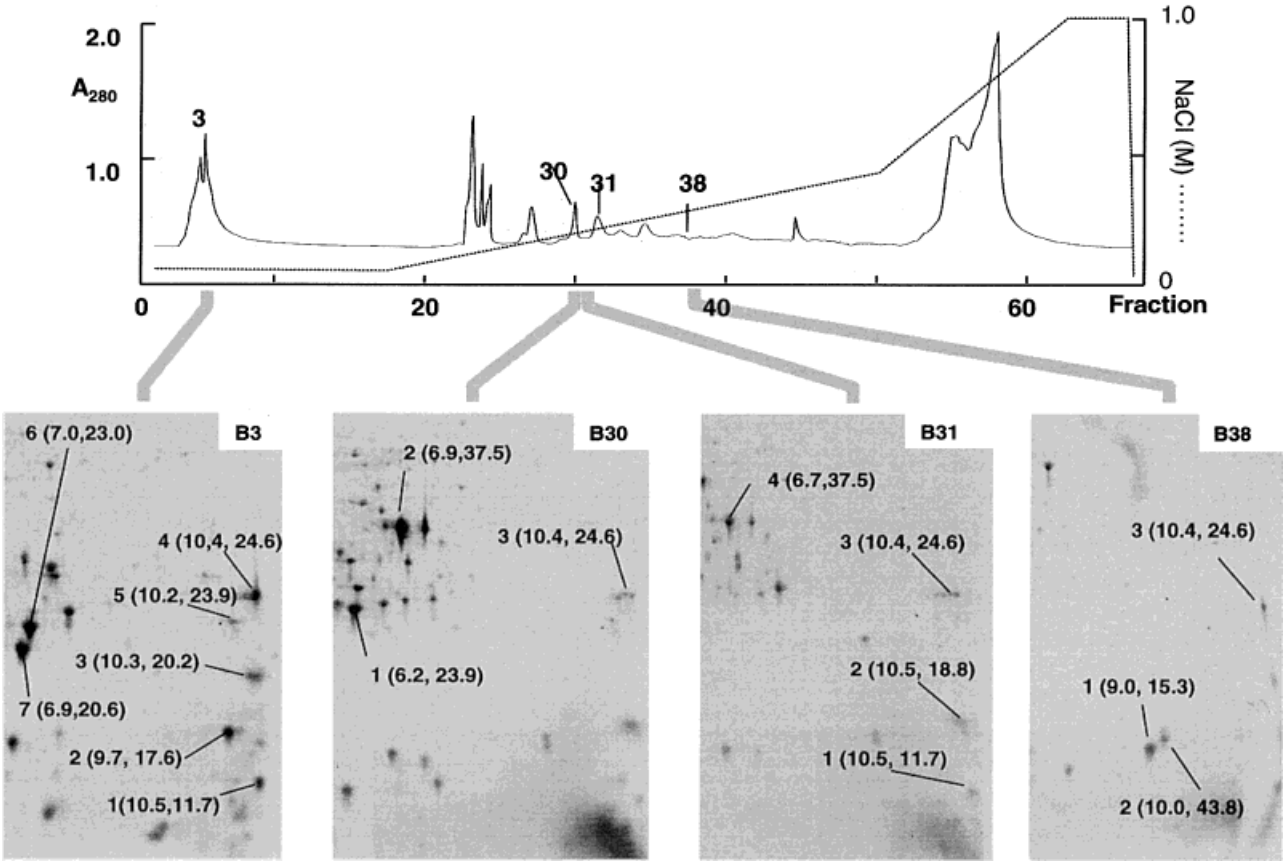

Figure 6. Chromatographic behaviour of basic proteins. Soluble proteins from an *E. coli* cell lysate (16 mg protein) were separated by a linear salt gradient on MonoQ anion exchange chromatography and the absorbance was monitored at 280 nm. Selected fractions were analysed by 2-D SDS-PAGE used linear IPG strips (pH 3–10) for the first dimension, and linear 12.5% gels in the second dimension. Protein spots corresponding to high p*I* values were excised and analysed by peptide mass fingerprinting. A few proteins at the acidic end of the gel were also analysed.

teins in fraction 67 could therefore elute late in the gradients because of ionic association with nucleic acids. Only a few proteins were detected on the 2-D gel of fraction 67. However, none of these proteins possessed the high p*I* values that might have been anticipated given the apparent strength of interaction between the proteins and the column matrix. When the lysate was treated with nucleases prior to chromatography, there was no strong absorbance peak in this region of the salt gradient and protein assays or gel electrophoresis of these late fractions confirmed the absence of proteins (results not shown).

Four spots were identified from fraction 67. Spot 67–1 is RfaD (ADP-L-glycero-D-mannoheptose-6-epimerase), 67–2 is Pnp (polynucleotice nucleotidyltransferase), 67–3 is DeaD (ATP-dependent RNA helicase) and 67–4 is DldH (dihydrolipoamide dehydrogenase). Pnp is part of the RNA degradasome [20–22], a multi-enzyme complex involved in mRNA processing and degradation, and DeaD is a helicase that is also ribosomally associated [23]. It is likely that both of these proteins eluted late in the gradient because of association with nucleic acid. It is less clear why DldH and RfaD should have eluted so late, possibly a previously unknown ability to bind nucleic acids, or an unexpected association with a nucleic acid binding protein. The value of such anomalies lies in their ability to highlight previously unsuspected interactions.

The nature of the MonoQ anion exchange chromatography means that proteins binding to the column and eluting within the salt gradient would be expected to have an apparent p*I* between 4 and 7. Accordingly, total cell extract was resolved on MonoQ and selected fractions were resolved on wide-range (p*I* 3–10) 2-D PAGE. Proteins migrating towards the extreme alkaline pH end of the gel, irrespective of elution position, were then considered to have anomalous behaviour and representative proteins of that subset were identified (Table 2).

Table 1. Identification of spots on narrow range p*I* 2-D gels by MALDI-TOF MS

Spot	Protein (abbreviation, name)	Peptides matched	Sequence covered %	p*I*	Mass (Da)
0-1, 31-1	Eno, Enolase	14, 17	48, 53	5.3	45524
0-2, 27-2	CypB, Peptidyl-prolyl *cis*-trans isomerase b	5, 11	33, 65	5.5	18154
0-3, 26-3	Kad, Adenylate kinase	14, 15	63, 73	5.6	23586
0-4, 30-4	RpiA, Ribose-5-phosphate epimerase	7, 4	15, 23	5.2	22861
0-5, 25-5; 0-6, 25-4	YfiD, Unknown 14.3 kDa protein	9, 10; 11, 11	65, 75; 73, 73	5.1	14284
0-7	TpiS, Triose-phosphate isomerase	8	28	5.9	26972
0-8	PgK, Phospho-glycerate kinase	22	63	4.9	40987
0-9	Upp, Uracil phosphoribosyl transferase	11	54	5.2	22533
0-10	IpyR, Inorganic pyrophosphatase	7	47	4.9	19572
0-11	YgiD, Hypothetical 36 kDa protein	9	26	6.2	36025
0-12	RS6, 30S ribosomal protein 6	8	56	4.9	15704
25-1	MalE, Maltodextrin-binding protein	17	54	5.5	43388
25-2	Kad, Adenylate kinase	19	78	5.6	
25-3	HdeB, Protein HdeB precursor	5	43	5.7	12043
26-1	ManA, Mannose-6-phosphate isomerase	18	73	5.3	42850
26-2	AlkH, 4-hydroxy-2-oxogluterate aldolase	7	45	5.6	22284
26-4	YfiD	9	84		
27-1	Thio, Thioredoxin	6	51	4.7	11675
27-3	DeoC, Deoxyribose-phosphate aldolase	12	51	5.5	27734
30-1	Pmgl, Phospho-glycerate mutase	16	74	5.9	28425
30-2	NfnB, Dihydropteridine-oxidoreductase	20	60	5.8	23905
30-3	Apt, Adine phosphoribosyl transferase	4	37	5.3	19859
31-2	AhpC, Alkyl hydroperoxide reductase c22 protein	14	78	5.0	20630
34-1	Pmgl	15	65		
67-1	RfaD, ADP-L-glycero-D-mannoheptose-6-epimerase	17	53	4.8	34893
67-2	Pnp, Polyribonucleotide nucleotidyltransferase	20	35	5.1	77101
67-3	DeaD, ATP-dependent RNA helicase DeaD	7	16	8.7	72700
67-4	DldH, Dihyldrolipoamide dehydrogenase	10	24	5.8	50558

Each spot is prefixed by the fraction number for which the 2-D gel analysis was carried out (gel series A, Figs. 2b–h). The proteins identified from the gel of the whole cell lysate (Fig. 2a) are prefixed by O.

Table 2. Identification of spots on broad range p*I* 2-D gels by MALDI-TOF MS

Spot	Protein (abbreviation, name)	Peptides matched	Sequence covered %	p*I*	Mass (Da)
B0-1	RS10, 30S ribosomal protein S10	7	55	9.7	11736
B0-2	RL10, 50S ribosomal protein L10	6	38	9.0	17580
B0-3	RL10	7	45		
B0-4	RL1, 50S ribosomal protein L1	10	41	9.7	24598
B0-5	RS2, 30S ribosomal protein S2	13	60	6.7	26613
B0-6	G3P1, Glyceraldehyde 3-phosphate dehydrogenase A	15	66	6.6	35401
B3-1	RS10	8	76		
B3-2	RL10	10	68		
B3-3	RL5, 50S ribosomal protein L5	12	47	9.5	20170
B3-4	RL1	8	44		
B3-5	YrbC, Hypothetical 24 kDa protein	7	34	9.4	23963
B3-6	SodM, Superoxide dismutase	11	67	6.5	22966
B3-7	Rrf, Ribosome recycling factor	13	84	6.4	20639
B30-1	NfnB, Dihydropteridine reductase	12	57	5.8	23905
B30-2	Syw, Tryptophanyl-tRNA synthetase	15	46	6.5	37497
B30-3	RL1	6	23		
B31-1	RS10	11	95		
B31-2	RL6, 50S ribosomal protein L6	8	47	9.7	18773
B31-3	RL1	6	38		
B31-4	Syw	12	34		
B38-1	StpA, DNA binding protein stpA	8	52	8.0	15348
B38-2	YagA, Hypothetical 43.8 kDa protein	4	16	9.3	43769
B38-3	RL1	10	41		

Each spot number is prefixed by the fraction number for which the 2-D gel analysis was carried out (gel series B, Fig. 6).

Many of the proteins observed at the high pH region of the gels were ribosomal proteins that are extremely basic due to their considerable (~20%) lysine and arginine content [24, 25]. The *E. coli* ribosome is a complex structure, containing a total of 52 proteins segregated into the 30S and 50S subunits. Studies of the 30S ribosome *in vitro* indicate that much of the assembly of ribosomal proteins into subunits and then into this complex is sequential and co-operative [26, 27]. As was presumed, a number of the ribosomal proteins were found in the flow-through fraction *i.e.* the material that had not bound to the column. However, the same ribosomal proteins were also identified in fractions recovered from the anion exchange column. At this stage, we do not have an explanation for the anomalous elution of these proteins, but assembly into partial ribosome complexes, or association with nucleic acids are both likely. The ability of native state chromatography to divide proteins into separable entities, possibly through different functional associations, is a strength of this approach that should not be overlooked.

4 Concluding remarks

The proteome, even that derived from organisms with simple genomes, such as *E. coli,* is too complex for analysis by a single 2-D gel, or even a series of narrow range p*I* gels. There are many more spots on a 2-D gel than there are encoded proteins in the genome [28], whether derived from post-translational processing or artefactual modification [29]. Single gene products appear as multiple, post-translationally modified spots, and the background spectral noise of abundant proteins detracts from the peptide fingerprint mass spectra of less abundant proteins. As more and more complex proteomes are analysed, so will the need for prior fractionation and simplification of the proteome before the proteins are identified. The most promising application of native state chromatography is in the resolution of functional complexes. In this regard, we might expect non selective methods such as size exclusion chromatography and ion exchange chromatography to be supplemented by a range of affinity based chromatographic steps.

There have been some notable studies of directly coupled n-dimensional chromatography in proteome analysis [9–11, 30, 31]. Unger *et al.* [10] linked ion exchange columns with RP columns and have proven the utility of this system for resolution of protein mixtures, although in this instance the goal was resolution of individual proteins rather than discovery of functional assemblies. Other applications have focused on the use of chromatographic separations to resolve tryptic peptides derived from protein digests [11, 31], an approach that may be the key to discovery of the members of a functional assembly. In these examples, the separation is deferred until after tryptic fragmentation, inevitably creating a more complex mixture. However, against this increase in complexity should be balanced the improvement in resolution that is characteristic of ion exchange and RP chromatography of peptides compared to proteins.

We do not believe that one – or two-dimensional chromatography should be projected as an alternative to 2-D PAGE, and in this regard, the requirement for direct coupling of multiple stages of chromatography may not always be necessary. Chromatographic systems that use, for example, a microtitre plate format to pass fractions from the first dimension to the second impose little additional workload, and remove the need to intersperse the second dimension of separation in a narrow window during the first dimension separation. The gain in throughput by automation of spot recovery, in-gel digestion and automatic processing of MALDI-TOF samples may more than compensate for the additional time invested in functionally relevant separations. However, in systems offering high throughput proteome-wide screening, a rapid chromatographic substitute for 2-D PAGE may have appeal.

A relatively simple modification of introducing a single dimension of rather nonselective chromatography, can enhance the quality of mass spectral data from low abundance spots. Moreover, the association of proteins in a chromatographic fraction is of value in identifying putative multiprotein complexes. Recent studies have exposed some of the limitations of 2-D PAGE as the only separation technology in proteome analysis [28, 32] although it remains the highest resolution methodology for protein separation. Liquid phase and gel-based separations should be seen as complementary, providing different perspectives on subsets of the proteome, and providing a conjunction of low-resolution isolation of functional assemblies and high-resolution separation of their components. Such approaches are also complementary to genetic approaches to discovery of functional associations [3, 33, 34], and extend the capability to macromolecular assemblies involving more than two or three proteins. The marked lack of correlation between theoretical p*I* and elution position on ion exchange chromatography serves to illustrate the potential of nondenaturing, liquid phase and mild separation technologies as a third dimension in deconstruction of the proteome.

We would like to thank Mathew Woodwark for his advice and encouragement, and Rachel Rowlinson for her assistance with MALDI-TOF analysis. AB was supported by a studentship from Astra-Zeneca.

Received August 20, 2000

5 References

[1] Hieter, P., Boguski M., *Science* 1997, *278,* 601–602.
[2] Oliver, S. G., Winson, M. K., Kell, D. B., Baganz, F., *Trends Biotechnol.* 1998, *16,* 373–378.
[3] Oliver, S. G., *Nature* 2000, *403,* 601–603.
[4] Oliver, S. G., *Nature* 1996, *379,* 597–600.
[5] Wilkins, M. R., Pasquali C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A. Hughes G., Humphery-Smith I., *BioTechnol.* 1996, *14,* 61–65.
[6] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250,* 4007–4021.
[7] Pedersen, S., Bloch, P. L., Reeh, S., Neidhardt F. C., *Cell* 1978, *14,* 179–190.
[8] Neidhardt, F. C., Vaughn, F., Philips, T. A., Bloch P. L., *Microbiol. Rev.* 1983, *47,* 231–284.
[9] Opiteck, G. J., Ramirez, S. M. Jorgenson, J. W., Moseley, M. A. III., *Anal. Biochem.* 1998, *258,* 349–361.
[10] Unger, K. K., Racaityte, K., Wagner, K., Miliotis, T., Edholm, L. E., Bischoff, R., Marko-Varga G., *J. High. Resol. Chromatogr.* 2000, *23,* 259–265.
[11] Gao, H., Shen, Y., Veenstra, T. D., Harkewicz, R., Anderson, G. A., Bruce, J. E., Pasa-Tolic, L., Smith R. D., *J. Microcolumn Separations* 2000, *12,* 383–390.

[12] Görg, A., Boguth, G., Obermaier, C., Posch, A., Weiss, W., *Electrophoresis* 1995, *16,* 1079–1086.
[13] Sanchez, J. C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R. Hochstrasser, D. F., *Electrophoresis* 1997, *18,* 324–327.
[14] Vanbogelan, R. A., Neidhardt, F. C., *Electrophoresis* 1991, *12,* 955–994.
[15] Suisdak, G., *Mass spectrometry for biotechnology,* Academic Press, New York 1996.
[16] Shevchenko, A., Jensen, O. N., Podtelejnikow, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., *Proc. Acad. Sci. USA* 1996, *93,* 14440–14445.
[17] Jensen, O. N., Mortensen, P., Vorm, O., Mann, M., *Anal. Chem.* 1997, *69,* 1707–1714.
[18] Uetz, O. N., Hughes R. E., *Curr. Opin. Microbiol.* 2000, *3,* 303–308.
[19] Laemmli, U. K., *Nature* 1970, *227,* 680–685.
[20] Py, B., Higgins, C. F., Krisch, H. M., Carpousis, A. J., *Nature* 1996, *381,* 169–172.
[21] Vanzo, N. F., Li, Y. S., Py, B., Blum, E., Higgins, C. F., Raynal, L. C., Krisch, H. M., Carpousis, A. J., *Genes Dev.* 1998, *12,* 2770–2781.
[22] Miczak, A., Kaberdin, V. R., Wie, C. L., Lin-Chao, S., *Proc. Natl. Acad. Sci. USA* 1996, *93,* 3865–3869.
[23] Jones, P. G., Mitta, M., Kim, Y., Jiang, W., Inouye M., *Proc. Natl. Acad. Sci. USA* 1996, *93,* 76–80.
[24] Noller, H. F., Nomura, M., in: Ingraham, J. L., Low, K. B., Magasanik, B., Schaecter, M., Umbarger, H. E. (Eds). *Escherichia coli* and *Salmonella typhimimuriam: cellular and molecular biology,* 2nd Edn, ASM Press, Washington DC 1996, pp. 167–186.
[25] Nomura, M., Traub, P., Guthrie, C., Nashimoto. H., *J. Cell. Physiol.* 1969, *74,* 241–252.
[26] Powers, T., Noller, H. F., *J. Mol. Biol.* 1993, *232,* 362–374.
[27] Rohl, R., Nierhaus, K. H., *Proc. Natl. Acad. Sci. USA* 1982, *79,* 729–733.
[28] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R., *Proc. Natl. Acad. Sci. USA* 2000, *97,* 9390–9395.
[29] Sarioglu, H., Lottspeich, F., Walk, T., Jung, G., Eckerskorn, C., *Electrophoresis* 2000, *21,* 2209–2218.
[30] Link, A. J., Eng, J., Schietz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates, J. R., *Nature Biotechnol.* 2000, *17,* 676–682.
[31] Yu, L., Zeng, R., Shao, X., Wang, N., Xu, Y., Xia, Q., *Electrophoresis* 2000, *21,* 3058–3068.
[32] Celis, J. E., Gromov P., *Curr, Opin. Biotechnol*. 1999, *10,* 16–21.
[33] Legrain, P., Jestin, J.-L., Schachter, V., *Curr. Opin. Biotechnol.* 2000, *11,* 402–407.
[34] Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., Seraphin, B., *Nature Biotechnol.* 1999, *17,* 1030–1032.