Francesco L. Brancia[1,4]
Amna Butt[2]
Robert J. Beynon[3]
Simon J. Hubbard[2]
Simon J. Gaskell[1]
Stephen G. Oliver[4]

[1]Michael Barber Centre for Mass Spectrometry,
[2]Department of Biomolecular Sciences, UMIST, Manchester, UK
[3]Department of Veterinary Preclinical Studies, University of Liverpool, Liverpool, UK
[4]School of Biological Sciences, University of Manchester, Manchester, UK

A combination of chemical derivatisation and improved bioinformatic tools optimises protein identification for proteomics

The identification of individual protein species within an organism's proteome has been optimised by increasing the information produced from mass spectral analysis through the chemical derivatisation of tryptic peptides and the development of new software tools. Peptide fragments are subjected to two forms of derivatisation. First, lysine residues are converted to homoarginine moieties by guanidination. This procedure has two advantages, first, it usually identifies the *C*-terminal amino acid of the tryptic peptide and also greatly increases the total information content of the mass spectrum by improving the signal response of *C*-terminal lysine fragments. Second, an Edman-type phenylthiocarbamoyl (PTC) modification is carried out on the *N*-terminal amino acid. The renders the first peptide bond highly susceptible to cleavage during mass spectrometry (MS) analysis and consequently allows the ready identification of the *N*-terminal residue. The utility of the procedure has been demonstrated by developing novel bioinformatic tools to exploit the additional mass spectral data in the identification of proteome proteins from the yeast *Saccharomyces cerevisiae*. With this combination of novel chemistry and bioinformatics, it should be possible to identify unambiguously any yeast protein spot or band from either two-dimensional or one-dimensional electropheretograms.

Keywords: Proteomics / Derivatisation / Bioinformatics / Matrix-assisted laser desorption/ionisation

EL 4228

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 *Bacillus subtilis* 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 develomental condition.

Correspondence: Prof. Stephen G. Oliver, School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK
E-mail: steve.oliver@man.ac.uk
Fax: +44-161-6067360

Abbreviations: IEC, ion-exchange chromatography; **PTC**, phenylthiocarbamoyl derivative; **SEC**, size-exclusion chromatography

Of the four levels of analysis, it is the proteome that should be of greatest utility 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 transcriptome) are functional entities within the cell. The last is that (like mRNAs, but unlike metabolites) there is a direct relationship between proteins and genes. The 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, in the most recent survey of the yeast proteome [5], a total of 400 spots were identified on the two-dimensional electropherogram (2-D gel). These corresponded to just 279 genes within the nuclear genome of the organism. If this is a representative proportion, then it implies that the total repertoire of proteins that yeast can call upon under any given set of physiological or developmental conditions comprises some 8800 protein species. The disparity between the number of genes and the number and diversity of their protein products is likely to be even greater for more complex eukaryotes.

Any technique that makes the process of identification of the protein contained within the spots of 2-D gels easier and quicker would have a major impact on progress in

0173-0835/01/0303-552 $17.50+.50/0

proteomics. Previous work indicates that the unambiguous identification of a yeast protein requires 5–7 peptides and often needs further tandem mass spectrometric analyses [6], depending on the mass spectrometric parameters applied. Moreover, increasing attention is being paid to the proteomic analysis of multisubunit complexes in order to uncover function *via* "guilt-by-association" [3]. This requires mass spectral analysis to identify all the members of a protein mixture. Two main approaches have been taken to protein mixture analysis. One is to exploit standard peptide mass fingerprinting and an integrative database-searching regime [6]. This has the twin disadvantages that a relatively high number of peptide fragments is required to identify each individual protein and that, if trypsin is used to generate the fragments, then a number of peaks are usually missing from the mass spectrum due to the poor signal response of peptides with a *C*-terminal lysine [7]. The second approach, taken by Yates and his colleagues [8], uses multidimensional liquid chromatography and tandem MS to separate and fragment peptides. This is a time-consuming procedure requiring the use of instruments that are not routinely available to genome scientists.

Previous work [9–11] has demonstrated that limited sequence information is of great value in making protein assignments and this is particularly true if a complete genome sequence is available for the database search. In this paper, we report a new approach to the identification of proteome proteins that involves chemical derivatisation of peptides using well-established techniques. First, the *C*-terminal lysine peptides within the tryptic digest mixtures are converted by *O*-methylisourea into homoarginine terminal peptides. Modification of the ε-amino group enhances detection of lysine residues, increasing the absolute number of fragments displayed in the MALDI spectrum, and allowing *C*-terminal lysine and arginine peptides to be distinguished from one other. Second, conventional Edman chemistry allows reaction of the amino group of the peptides with phenylisothiocyanate to form the phenylthiocarbamoyl (PTC) derivative. In the presence of a strong anhydrous acid, the *N*-terminal residue is cleaved to produce a truncated peptide. The close analogy to Edman degradation chemistry in the gas phase has recently been established [12]. The tryptic peptides, as PTC derivatives, are subjected to tandem MS analysis yielding b_1 ions that allow the unambiguous identification of the *N*-terminal residue, with the exception of leucine and isoleucine. The derivatised peptides are analysed using standard matrix assisted laser desorption/ionisation-mass spectrometry (MALDI-MS) coupled with post-source decay analysis (PSD). The protein assignments are then made using a suite of new software tools, PepMAPPER, that are freely available *via* the World Wide Web (http://wolf.bi.umist.ac.uk/mapper). This utility of this approach is demonstrated by experiments with *Saccharomyces cerevisiae*, a eukaryotic organism for which the complete genome sequence is available [13].

2 Materials and methods

2.1 Preparation of yeast protein extracts

2.1.1 Chromatography procedures

Yeast strain S288C was grown in minimal glucose medium (1 L) and harvested in the exponential phase. Cells were lysed in 0.5 mL of extraction buffer (0.1 M HEPES, pH 7.5, containing Complete™ Mini EDTA-free protease inhibitor cocktail; Boehringer, Mannheim, Germany), and then vortexed with an equal volume of acid-washed glass beads (Sigma, St. Louis, MO, USA). Released protein was solubilised in 5 mL of extraction buffer, and left on ice for 2 h. RNase A and DNase I (50 and 500 U, respectively) were added, and the solution left on ice for 30 min. The supernatant was clarified by centrifugation and an aliquot (2 mL) was filtered through a 0.22 μm PVDF syringe filter prior to fractionation using two orthogonal separation steps, size-exclusion (SEC) and ion-exchange (IEC) chromatography. Chromatography was performed using a fast protein liquid chromatography (FPLC) system. SEC was carried out on a Superdex™ 200 prep grade 16/60 size-exclusion column (Amersham-Pharmacia, Uppsala, Sweden), pre-equilibrated with two column volumes of 0.1 M HEPES, pH 7.5 (flow rate, 1 mL/min). The applied protein was eluted using the same buffer and fractions collected at 2 mL intervals. Further simplification of the protein mixture was achieved using IEC. The FPLC system was fitted with a Mono-Q™ column (V_t=1 mL) pre-equilibrated with 10 mL of 50 mM HEPES (pH 7.5) pumped through at a flow rate of 1 mL/min. Aliquots (1 mL) from four consecutive SEC fractions were pooled. The pool (4 mL) was then mixed with an equal volume of Milli-Q water and the sample applied to the Mono-Q™ column. Bound protein was washed using five column volumes of the same buffer (50 mM HEPES, pH 7.5) before eluting with a linear salt gradient of 0–1 M NaCl in 40 mL, using 50 mM HEPES, 1 M NaCl (pH 7.5) Selected fractions from IEC were precipitated using TCA and analysed by SDS-PAGE.

2.1.2 In-gel digestion

In-gel digestions were performed using a modification of the procedure of Shevchenko *et al.* [14]. Excised gel pieces were placed in microtitre plate wells (Costar) and 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. The super-

natant was discarded and replaced by 100 μL of 50% v/v Tris-HCl (pH 8) and 50% v/v acetonitrile (sequencing grade; Perkin-Elmer, Norwalk, CT, USA). Gel pieces were left in this solution for 30 min before replacing it with 50 μL of acetonitrile 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, Holbrook, NY, USA) for 15 min. A trypsin stock solution (1 μL of 0.01% v/v TFA) was then diluted 80-fold in 20 mM Tris-HCl, pH 8, to give a final enzyme concentration of 1.25 ng/μL. The enzyme was added in 10 μL aliquots to each of the dry gel pieces. 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 acetonitrile 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 approximately 15 μL by vacuum concentration.

2.2 Guanidination of lysine residues

The reaction was carried out as described in [15] with some modifications. *O*-methylisourea was obtained from Lancaster (Morecombe, UK). An aqueous solution of the tryptic fragments was mixed with an equal volume of 0.5 M *O*-methylisourea, adjusted to pH 10.5 with NaOH. The reaction was allowed to proceed overnight at room temperature and was stopped by an equal volume of 1% v/v aqueous TFA.

2.3 Edman derivatisation

Following digestion with trypsin, the mixture was dried under vacuum centrifugation and then converted into the PTC derivative by dissolving in a mixture of ethanol:water:triethylamine:phenyl isothiocyanate (77/11/11/1 by volume). The reaction was allowed to proceed for 10 min at 50°C followed by vacuum centrifugation. The dried, derivatised product was first dissolved in heptane:ethyl acetate (10:1 v/v; 50–200 μL) and then an equal volume of water was added. This mixture was shaken vigorously and then centrifuged to break the dispersion. The organic (top) layer was discarded and extraction of the aqueous layer was repeated twice before performing the MALDI analysis.

2.4 Mass spectrometry

Experiments were conducted in a Voyager DE-Elite reflector MALDI-time of flight-mass spectrometer (PE Biosystems, Framingham, MA, USA). All spectra were collected in reflector mode with each mass determination being an average of 256 spectra, using four peptides of known mass as external calibration standards. The matrix was α-cyano-4-hydroxycinnamic acid (CHCA) which was prepared as a saturated solution, at room temperature, in acetonitrile:water (1:1 v/v) containing 0.1% v/v TFA acid. The peptide samples were mixed with 50 μL of the matrix solution and then deposited directly onto the target and allowed to dry. PSD spectra were acquired with accelerating voltages of 20 000 V whose grid voltage (defining the two-stage ion acceleration) was held at 73.5% with decrement ratio of 0.75 (1.000, 0.75, 0.5625, 0.4219, 0.3164, 0.2373, 0.1780 , 0.1335, 0.1001, 0.0751) using, as precursor ions, the two PTC-derivatives at 930 and at 1887 Da.

3 Results

3.1 Search efficiencies

In order to explore the value of limited sequence information in making protein assignments from peptide mass fingerprints of the yeast proteome, we have exploited two derivatisation procedures. PTC derivatisation provides information about the identity of the *N*-terminal amino acid of a peptide [9], while guanidination [15] converts all the lysine residues in a peptide into homoarginines. Guanidination will, in most instances, permit the discrimination between *C*-terminal arginine and *C*-terminal lysine peptides following trypsin digestion. This is because there is a mass increase of 42.02 Da for every lysine residue in a peptide that is converted to homoarginine as a result of the derivatisation. Furthermore, the information content of the mass spectrum of a tryptic digest is significantly increased by improving the MALDI response of *C*-terminal lysine-containing fragments.

In order to assess the increase in database search efficiency using these derivatisation techniques, a series of bioinformatic simulations were conducted with typical search parameters using our in-house tools (PepMAPPER) on the yeast proteome. The search “gain” was assessed in two ways. When searching a given peptide mass against a database of all potential tryptic peptide masses generated from the yeast proteome, we calculate the number of peptides in the database which match, given a defined mass accuracy. This is repeated using the additional criteria that the sequence of the matching peptide is consistent with the extra sequence information about the search peptide that can be identified from derivatisation. In other words, for the guanidination experiment, the number of lysine residues in the search and database peptides must also match; for PTC derivatisation the *N*-terminal amino acids must be the same.

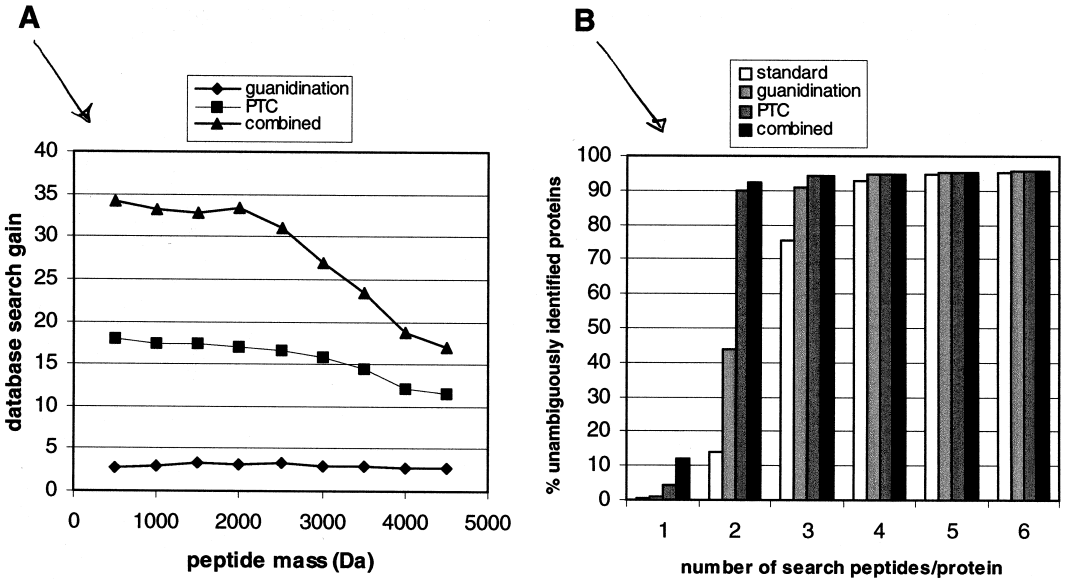


Figure 1. Derivatisation effects on database searching of the yeast proteome. (A) The average gain in database searching efficiency was calculated for 500 Da steps across the mass range for the entire yeast proteome, assuming a measured mass accuracy of 500 ppm, a fixed charge state of +1 and up to 1 missed tryptic cleavage. The gain was assessed as the ratio of the number of valid peptide matches in the database at a certain mass to the reduced number of matches obtained when using the additional search information implied from the respective hypothetical derivatisation experiment. In the case guanidination, the number of lysines in the peptide was assumed to be known, whilst the identity of the *N*-terminal amino acid was known for PTC derivatisation (excepting Ile/Leu, which are indistinguishable). The effect of using both types of extra information was assessed as a combined approach. (B) Each simulation was carried out over 500 000 steps, picking a fixed number of randomly selected peptides from a randomly selected protein at each step. Using these peptides as search criteria, the number of matching yeast proteins was calculated, with or without the derivatisation-derived extra information. The searches assumed a measured mass accuracy of 500 ppm, and singly charged monoisotopic species, with up to 1 missed cleavage. For each of the 24 simulations, the number of times that only one protein was consistent with the search data (an unambiguous identification) was calculated and is expressed as a percentage of the number of simulation trials. All experiments used a 6462 yeast protein database derived from the PIR databank.

The database search "gain" is the ratio of these two numbers of matching database peptides, using the reduced number of matches found with the extra sequence information as the denominator to give values greater or equal to 1 (see Fig. 1 A). The fewer the number of matches to the database, the greater the certainty of identification of a protein. The second approach used computer simulations to determine the number of search peptides required for the unambiguous identification of any protein. Figure 1 B shows the percentage of all yeast proteins that can be unambiguously identified by searching the database with a given number of peptides, both with and without derivatisation data. These simulations underestimate the impact of guanidination since they take no account of the increased number of peptide fragments that can be detected in the mass spectrometer as a result of the conversion of *C*-terminal lysines to homoarginines. Note that the upper limit on unambiguous identification is *ca.* 95% of all yeast proteins due to the redundancy in the genome [16]. We will go on to demonstrate that the potential gains in the efficiency of protein identification that these derivatisation procedures offer can be realised in the laboratory.

3.2 The impact of guanidination

The derivatisation involves treatment of tryptic digest of a protein fraction overnight with *O*-methylisourea (see Section 2.2 for a full description of the technique), following which the material can be loaded straight onto the MALDI target. Standard proteins were digested from the gel and subsequently derivatised with *O*-methylisourea in order to test the sensitivity in MALDI-MS for the analysis of guanidinated peptides. Typically this sensitivity reached the femtomole level (200 fmol for a tryptic digest, 50 fmol for a solution of a single peptide). Figure 2 B provides an example of the efficacy of guanidination in protein derivatisation. A yeast total protein extract was subjected to SEC and IEC and one fraction from the second column

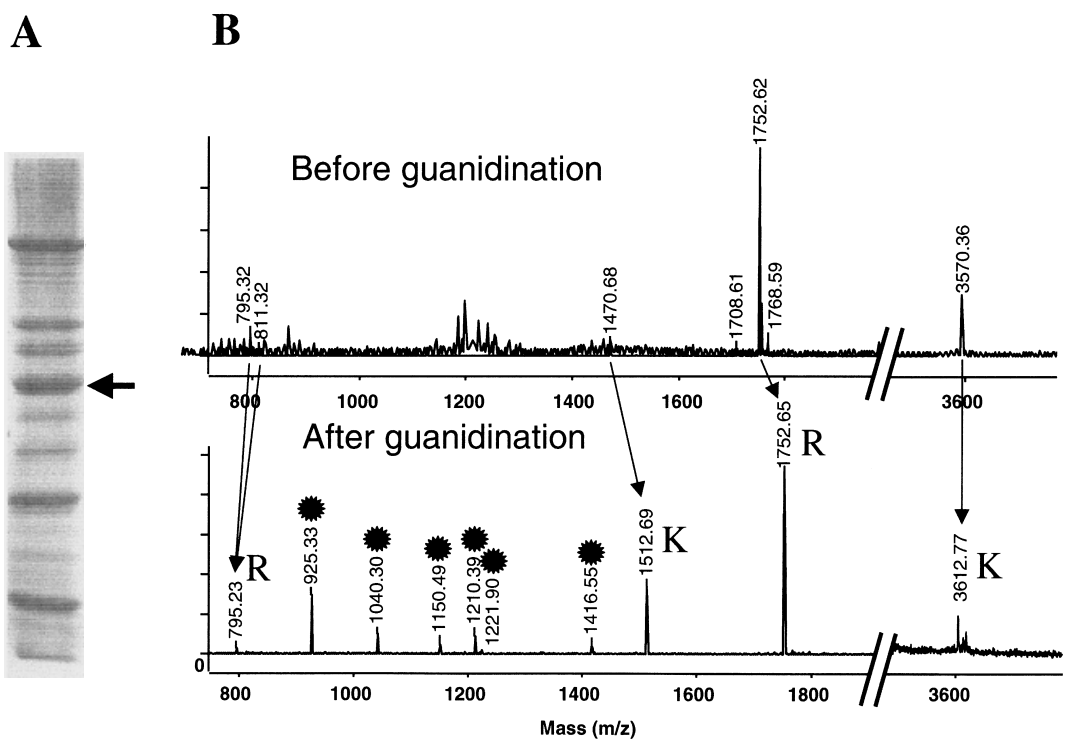
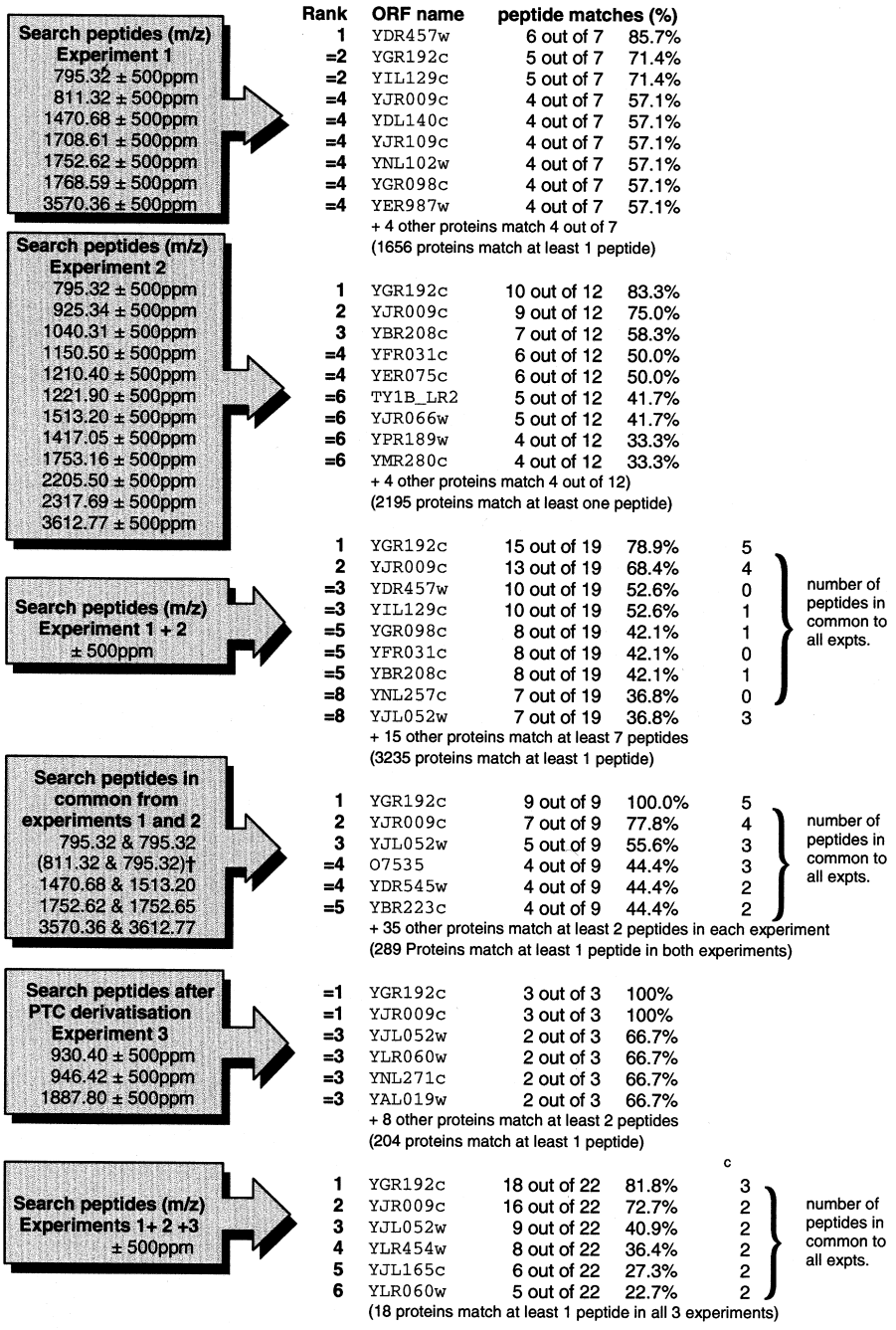


Figure 2. SDS-PAGE and comparative mass spectrometric analysis of tryptic digests of yeast proteome proteins. (A) SDS-PAGE analysis of native yeast proteins. A mixture of yeast proteins obtained following native state 2-D chromatography (Ferro *et al.*, in preparation) was precipitated using TCA and visualised on a Coomassie-stained polyacrylamide gel. The indicated band was *in situ* digested, using trypsin (Butt *et al.*, in preparation), and the extracted peptides derivatised. (B) The upper MALDI spectrum was obtained from the tryptic digestion of the protein. The lower spectrum displays the same mixture after derivatisation with *O*-methylisourea. The arrows between the upper and lower peaks indicate the tryptic digest fragments present in both spectra, whilst the stars show the new signal corresponding to lysine-containing peptides converted into homoarginine derivatives.

was resolved by one-dimensional PAGE. A band (indicated in Fig. 2 A) was cut from the gel and subjected to in-gel digestion with trypsin. The tryptic digest was extracted and divided into two aliquots, one of which was subjected to *O*-methylisourea treatment. The MALDI-MS spectra for the derivatised and underivatised proteins are shown in Fig. 2.

Before derivatisation, 7 peaks of sufficient intensity for use in database searching are present in the spectrum, whereas 12 peaks are available for the search after guanidination. The seven-peak search of the yeast database gave YDR457w (a hypothetical protein) as the most likely hit (based on the number of matched peptides). YGR192c and YJR009c (two paralogues of glyceraldehyde-3-phosphate dehydrogenase) were, respectively, one of two equal second-choice and one of nine equal third-choice identifications (see Fig. 3). Searching with the 12 masses provided by the spectrum of the derivatised protein identified YGR192c as the unique top match, and YJR009c as the unique second-choice identification. If all masses from the two experiments are combined, the same two top choices are obtained. However, what distinguishes these two proteins from all the other database matches is that they share a large proportion of corresponding peptides between the two spectra (five peptides for YGR192c and four for YJR009c; see Fig. 3). This ability to distinguish the top choices from all other database hits is highlighted in Fig. 3 when just these corresponding peptides are used in the PepMAPPER search. A final decision between these two options can be made by considering the fact that only YGR192c possesses a lysine-containing peptide with an (underivatised) mass of 3570.36 Da. This unambiguuos identification of the correct paralogues provides a good illustration of the utility of our approach since these two enzymes share 96% amino acid sequence identity and also have the same SDS-PAGE mobility and elution position.

These data demonstrate the two key gains for database searching from this derivatisation approach. First, an increased total number of peptides is available for search-



† (The 811.32 Da peptide is the methione-oxidised peptide at 795.32 Da).

Figure 3. Database search results utilising guanidination and phenylcarbamoyl-derivatisation information. The systematic yeast proteome open reading frame name for matching database proteins are displayed for each search, ranked by the number of peptides they match in each case. Further statistics concerning the number of peptides in common for combined searches are listed, along with statistics indicating the total number of putative matches in each case. In all cases, searches were conducted using a fixed charge state of +1, a 500 ppm mass accuracy, and permitting up to 1 missed cleavage per peptide. The possible oxidation of all methionine residues was also considerd.

ing. Second, a comparison of the spectra before and after guanidination provides extra information as to the lysine content of the different peptide fragments. The information on lysine content in yeast tryptic peptide fragments provides an almost threefold reduction in the number of matching database proteins (Fig. 1 A). In our example, only four peptides were required to make the identification (a fifth peptide, used in the search, is an *in vitro* oxidation product – an artefact that PepMAPPER has been designed to handle).

3.3 The impact of combining guanidination and Edman derivatisation

The use of successive PTC derivatisation in sequencing proteins or peptides has been familiar to protein chemists since Edman first introduced this procedure in 1967 [17]. The technique is exploited today in identifying protein spots from 2-D gels, either directly [10], or with tandem MS being used as the analytical procedure [20]. The key to its use in the present context, as with much of func-

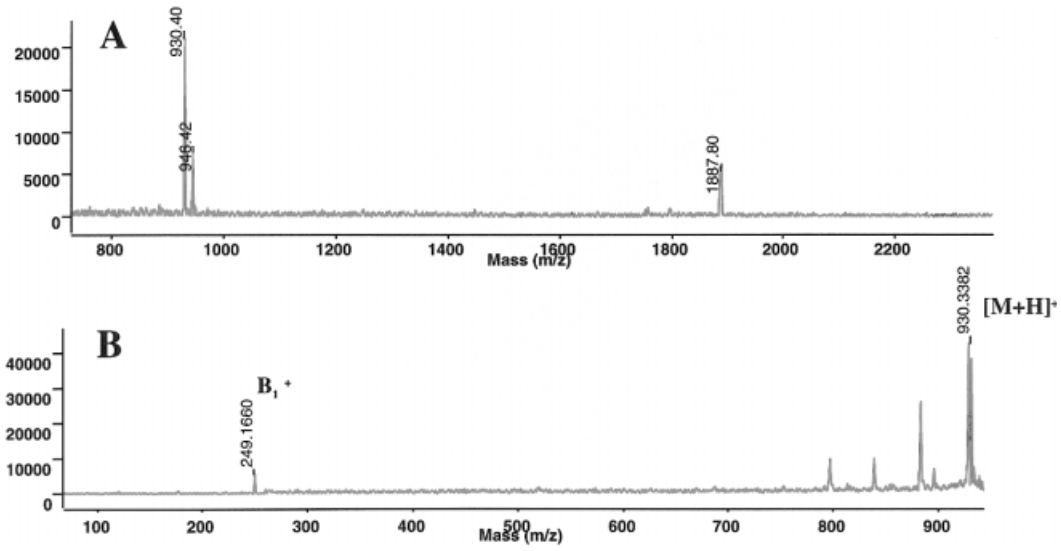


Figure 4. Mass spectra of PTC-derivatised yeast protein tryptic peptide fragments. (A) MALDI spectrum obtained from the derivatisation of the tryptic digest (Fig. 2B, upper spectrum) with phenyl thiocyanate. Three peaks are present in the spectrum due to conversion of the signals at 795.32, 811.32 (methionine oxidation of 795.32), 1752.62. (B) PSD-MALDI spectrum using the derivatised peptide at 930.40, as precursor ion; the b_1 ion appears at 249.7 corresponding to the leucine/isoleucine *N*-terminal fragment.

tional genomics, is to ask: what difference does it make to have a complete genome sequence? The answer is that much less information on protein sequence is required to make an unambiguous identification, and that using PTC derivatisation to identify only the *N*-terminal amino acid [9] is sufficient, especially when it is combined with the guanidination approach.

Figure 1 A shows that PTC derivatisation should provide a nearly 20-fold reduction in the number of database matches in the yeast proteome when searching with a given peptide fragment mass. Furthermore, the number of peptide masses required for a protein identification at the 90% confidence limit is reduced to just two (assuming an undemanding *m/z* accuracy of ± 500 ppm; Fig. 1B). If the information from guanidination is combined with that obtained from PTC derivatisation, then the reduction in available matches is almost 35-fold (see Fig. 1 A) and two peptides yield unambiguous assignment approaching the 95% confidence limit (Fig. 1B). Both these points are illustrated by the data provided in Fig. 4, which shows the MALDI mass spectrum of the tryptic digest of the same protein band as that analysed in Fig. 2, but this time subjected to PTC derivatisation. If the three peaks from this spectrum are used to search the yeast database, PepMAPPER identifies only two proteins containing all three peptide fragments. These are the two glyceraldehyde-3-phosphate dehydrogenases, YGR192c and YJR009c.

If the data from the two derivatisation regimes are combined, then there is a significant increase in search efficiency. In the search using the three peptides from the PTC-derived spectrum (Fig. 4), 204 yeast proteins were found to contain at least one match to this set of three tryptic peptides. When the information about lysine content and the identify of the *C*-terminal residue, which was provided by the guanidination data, is included, then 186 of these 204 proteins may be excluded from the search, an order-of-magnitude gain in search efficiency.

4 Discussion

As an increased number of complete genome sequences becomes available, it should be possible to applly this combination of bioinformatic and chemical derivatisation procedures to analyse the proteomes of many organisms. In the example that we have provided for yeast, the combination of chemical derivatisation techniques progressively limited the search space for the identification of the correct protein by PepMAPPER. Searching with the masses, identified from the underivatised protein band from the 1-D gel, identified 1656 yeast proteins that matched at least one tryptic peptide. Moreover, the correct protein was not ranked first in this list. When tryptic peptides that occurred in both the underivatised and guanidinated protein spectra were used in the search, then 289 proteins match at least one fragment. Finally, when

data from all three experiments (underivatised, guanidination, and PTC derivatisation) were combined, then the search space was limited to just 18 proteins. Only two of these (YGR192c and YJR009c) matched 3/3 peptides at this stage of the analysis, and the persistent presence of the 3570.36 (underivatised) and 3612.77 (guanidinated) peaks at all earlier stages in the analysis permitted the identification of the correct paralogue. We would point out that, in terms of the yield of information from the PTC derivatisation, the example that we have given, represents the worst case. This is because the *N*-terminal amino acid for both paralogues is leucine (which cannot be destinguished from its isomer, isoleucine, by MS). These two isomers are the most abundant *N*-terminal amino acids in yeast tryptic peptides, yielding a search gain of around five (compared to gains of 35 + for rare amino acids such as Cys, Trp and His).

The approach that we have presented does not require sophisticated and expensive mass spectrometers, such as electrospray ionisation (ESI) with tandem MS devices, nor the postprocessing of the tryptic digest by LC. The PepMAPPER bioinformatics tools are freely available and these searches could be easily implemented with other similar, commonly used search tools [18–20]. Thus, the technique is accessible to most protein chemistry laboratories and should improve the speed and efficiency of protein assignment in proteomic studies. Moreover, the simple aqueous reaction involved in guanidination, and the prospect of carrying out the PTC derivatisation in the gas-phase [12], means that our procedure is amenable to automation, with further improvements in throughput available by carrying out the Edman reaction directly on the guanidinated products. We are currently using this combined approach to extend and improve the use of PTC derivatisation in the identification of components of protein mixtures [9].

This work was supported by CASE awards to FB and AB from the BBSRC and, respectively, Amersham-Pharmacia and AstraZeneca, and by a grant (BIF10517) to SJH, SGO and Olaf Wolkenhauer from the BBSRC/EPSRC Bioinformatics Initiative. We thank Dr. Sue Fowler for her advice and encouragement.

Received July 1, 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] Perrot, M., Sagliocco, F., Mini, T., Monribot, C., Schneider, U., Shevchenko, A., Mann, M., Jeno, P., Boucherie, H., *Electrophoresis* 1999, *20*, 2280–2298.

[6] Jensen, O. N., Podtelejnikov, A. V., Mann, M., *Anal. Chem.* 1997, *69*, 4741–4750.

[7] Krause, E., Wenschuh, H., Jungblut, P. R., *Anal. Chem.* 1999, *71*, 4160–4165.

[8] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, M., Yates III, J. R., *Nature Biotechnol* 1999, *17*, 672–682.

[9] Summerfield, S. G., Hubbart, S. J., Beynon, R. J., Gaskell, S. J., *J. Am. Soc. Mass Spectrom. 2000*, in press.

[10] Wilkins, M. R., Gasteiger, E., Tonella, L., Ou, K., Tyler, M., Sanchez, J. C., Gooley, A. A., Walsh, B. J., Bairoch, A., Appelt, R. D., Williams, K. L., Hochstrasser, D. F., *J. Mol. Biol.* 1998, *278*, 599–608.

[11] Perkins, D. N., Pappin, D. C., Creasy, D. M., Cottrell, J. S., *Electrophoresis* 1999, *20*, 3551–3567.

[12] Summerfield, G. S., Bolgar, M. S., Gaskell, S. J., *J. Spectrom.* 1997, *32*, 225–231.

[13] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S. G., *Science* 1996, *274*, 546–567.

[14] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

[15] Bonetto, V., Bergman, A. C., Jornvall, H., Sillard, R., *Anal. Chem.* 1997, *69*, 1315–1319.

[16] Mewes, H. W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S. G., Pfeiffer, F., Zollner, A., *Nature* 1997, *387*, 7–65.

[17] Edman, P., Begg, G., *Eur. J. Biochem.* 1967, *1*, 80–91.

[18] Clauser, K. R., Hall, C., Smith, D. M., Webb, J. W., Andrews, L. E., Tran, H. M., Epstein, L. B., Burlingame, A. L., *Proc. Natl. Acad. Sci. USA*, 1995, *92*, 5072–5076.

[19] Fenyo, D., Qin, J., Chait, B., *Electrophoresis* 1998, *19*, 998–1005.

[20] Mann, M., Wilm, M., *Anal. Chem.* 1994, *66*, 4390–4399.