

---

---

# Observation of Heterogeneous Gene Products by FT-ICR MS

Duncan H. L. Robertson,<sup>a</sup> Stephen C. C. Wong,<sup>b</sup> Robert J. Beynon,<sup>a</sup> Jane L. Hurst,<sup>c</sup> and Simon J. Gaskell<sup>b</sup>

<sup>a</sup> Protein Function Group, Faculty of Veterinary Science, University of Liverpool, Liverpool, United Kingdom

<sup>b</sup> Michael Barber Centre for Mass Spectrometry, School of Chemistry and Manchester Interdisciplinary Biocentre, University of Manchester, Manchester, United Kingdom

<sup>c</sup> Mammalian Behaviour and Evolution Group, Faculty of Veterinary Science, University of Liverpool, Leahurst Veterinary Field Station, Neston, United Kingdom

---

The ability of FT-ICR MS to resolve isotopic variants of intact proteins for each of the charge states formed by electrospray ionization offers a sensitive, rapid method for detecting "low mass" heterogeneity, where this is defined as the presence of structural variants differing in mass by 2 Da or less. Such heterogeneity may reflect biological or chemical modifications of structure or may result from the coexpression of related proteins from a multi-gene family. In the analytical approach described here, comparisons are made between observed isotopic distributions and those expected for predicted protein sequences. Close agreement is demonstrated for a homogeneous model protein, and the utility of the method has been evaluated in the study of mouse major urinary proteins (MUPs), a group of closely related sequences. Divergence of the experimental isotopic distribution from distributions predicted for known MUP sequences can be explained, in quantitative terms, by the coexpression of closely related sequences. This approach provides a facile method for the assessment of protein homogeneity and for the detection of structural variants, without recourse to proteolytic digestion and analysis of the resulting products. (J Am Soc Mass Spectrom 2008, 19, 103–110) © 2008 American Society for Mass Spectrometry

---

**A**ccurate measurement of intact protein mass, particularly using electrospray ionization mass spectrometry, has proven to be a valuable analytical tool. Knowledge of intact mass has been used to study protein structure at several levels: to recognize gene expression products, to monitor protein processing, and to characterize many post-translational modifications [1, 2]. An underlying assumption in all these investigations is that the measured mass of a protein arises from a homogeneous molecular population or that any heterogeneity will be resolved by the mass spectrometer. Whilst this is often the case, biological and artefactual circumstances can give rise to protein populations showing low mass heterogeneity (LMH), here defined as distinct molecular species differing in mass by 2 Da or less. Deamidation of glutamine and asparagine residues (+1 Da) and reduction of the disulfide bridge (+2 Da) are examples of biological LMH, both of which are associated with regulation of protein activity [3, 4]. Partial modification in these instances will result in a heterogeneous protein population comprising constituents with overlapping natural isotope distributions. Similarly, proteins coexpressed

by a multi-gene family can vary by a single amino acid, e.g., lysine to glutamate, with an associated mass shift of less than 1 Da [5]. Nonbiological LMH can arise as a result of experimental manipulation and storage [6]. Whilst such subtle heterogeneity is difficult to resolve, it has the potential to indicate important biological processes and sample integrity.

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) can offer distinct advantages over lower resolution techniques; mass accuracy in the sub-part per million (ppm) range and resolution in excess of 1,000,000 may be obtained [7]. Separation of natural isotopic variants (differing in the number of <sup>13</sup>C or other heavy variants of constituent atoms) of proteins may readily be achieved [8] but recognition of the monoisotopic (all light isotopes) variant is problematic as its relative abundance decreases with increasing protein mass. Thus, the difficulty in exploiting FT-ICR MS data for protein characterization derives not from the accuracy and precision of the determination of fractional mass but on the assignment of nominal mass to the monoisotopic variant [9]. Computational approaches have been used to overcome this problem [10], and whilst they have met with some success [11], the observed mass accuracy obtained (10 ppm) is one 100-fold less than the capability of the instrument.

---

Address reprint requests to Dr. D. Robertson, Veterinary Preclinical Sciences, University of Liverpool, Liverpool, L69 7ZJ, United Kingdom. E-mail: dhlr@liverpool.ac.uk

Here we explore the utility of measured protein isotope distributions to provide evidence for LMH in proteins in biological extracts. Where LMH exists, the isotopomer distribution is the sum of two species with different elemental compositions and will thus differ from that of the homogenous molecule. This has been exploited in a recent study of protein deamidation [12] (published during the preparation of this manuscript). In the present study we focus on a systematic evaluation of this approach for the study of heterogeneous gene products.

Major Urinary Proteins (MUPs) are a group of small (~18 kDa) proteins found abundantly in mouse urine [13, 14]. Urinary MUPs are encoded as a multi-gene family and expressed in the liver from where they are rapidly exported to the urine via the blood and the kidney [15]. Individual mice express and excrete a unique pattern of MUPs that provides an olfactory ownership signal to a deposited scent mark [16]. MUPs have many features which make them ideal model proteins for assessing the rigor of analytical techniques. Individual MUP genes show a high level of homology; for example, of the 162 amino acids comprising the polypeptide chain, variations exist at only four sites in the four most abundant MUPs in BALB/c mouse urine [5]. The mass differences between these four proteins range from 1 Da to 59 Da and in previous studies these proteins have been utilized as a model system for assessing mass spectrometric deconvolution techniques [17]. Whilst the mass differences between MUPs are relatively small, the amino acid substitutions often involve charged residues, notably lysine, glutamine and glutamate. This allows ready fractionation of MUPs by anion exchange chromatography, which adds to their utility as model compounds. More generally, purification of MUPs from urine is a simple, one step procedure and the high level of MUP gene expression in liver cells has resulted in a large number of MUP cDNA sequences in the public domain [18].

The aim of this study was to use MUPs as model compounds for assessing the utility of FT-ICR MS in detecting LMH in intact proteins. The isotopomer distributions from an FT-ICR spectrum of unfractionated BALB/c MUPs were compared with theoretical distributions derived from cDNA sequences. In each case, the level of heterogeneity was assessed by nonlinear curve fitting. Secondly, the same sample was fractionated by anion exchange chromatography and individual fractions were analyzed by FT-ICR MS. Isotopomer distributions that had indicated heterogeneity in the unfractionated sample were reassessed and in all cases matched the model data to a much higher degree. The data illustrate that FT-ICR MS and isotope distribution modeling provide a rapid, sensitive system for detecting low mass heterogeneity.

## Materials and Methods

### General Reagents

Horse myoglobin and (Glu<sup>1</sup>)fibrinopeptide B were purchased from the Sigma Chemical Co. (Poole, UK). Acetonitrile, water, formic acid, sodium chloride and N-morpholinoethanesulfonic acid (MES) were purchased from VWR Ltd., (Lutterworth, UK). All were AnalaR grade reagents.

### Mouse Urine

Mouse urine was obtained from adult male BALB/c by mice bladder massage. Mice were housed in groups of 6 to 8 with a 12 h light/dark cycle and unlimited food and water. Samples from individual mice were pooled into 1 mL aliquots and immediately stored at -20 °C.

### MUP Preparation

MUPs were purified from urine by spun-column gel permeation chromatography. Columns (5 mL) were packed with pre-swollen Sephadex-G25, which was subsequently equilibrated in deionized water. Excess water was removed from the columns with a 200 × g spin for 2 min. An aliquot (200 μL) of urine was then loaded onto each column which was spun for a further 2 min at 200 g. The eluent from the column was captured in 1.5 mL polypropylene test tubes and either submitted immediately for analysis or stored at -20 °C. Desalted, unfractionated MUPs were diluted 1:500 with 50% (vol/vol) acetonitrile/0.1% (vol/vol) formic acid, and desalted ion-exchange fractions diluted 1:100 with the same diluent before mass spectrometry.

### FT-ICR MS

High-resolution mass spectrometry was performed on a 9.4T Apex III FT-ICR mass spectrometer, fitted with an Apollo electrospray ionization (Bruker Daltonics, Billerica, MA). Desalted unfractionated MUPs, desalted ion-exchange fractions, or 1 pmol/μl myoglobin solution were introduced as a continuous infusion at 2 μL/min. Spectra were acquired using 512 k data points over a *m/z* range of 500 to 5000. The hexapole accumulation time was set such that an accumulated spectrum of 150 scans had total base peak ion intensity in the region of 2 × 10<sup>6</sup>; a typical time was 0.1 s for unfractionated MUP samples. Ions entering the ICR cell were trapped using a timed deflection voltage (SIDEKICK; Bruker Daltonics). Calibration was achieved externally with a 1 pmol/μl solution of horse heart myoglobin run immediately following each sample. Resolution at *m/z* 1100 was typically 100,000 and isotopomer mass accuracy (compared with theoretical values) was between 1 and 3ppm.

### Low-Resolution MS

MUPs purified by ion-exchange chromatography were identified using a Q-TOF Micromass spectrometer (Waters, Manchester, UK), fitted with a nanoflow electrospray ionization source. The quadrupole analyzer was set to rf-only mode to allow unrestricted passage of ions to the time of flight mass spectrometer. Solutions were introduced at a rate of 0.5  $\mu\text{L}/\text{min}$  as a continuous infusion. Spectra were gathered between  $m/z$  700 and 1600 with a scan/interscan time of 2.4/0.1 s. The instrument was calibrated from product ion spectrum of (Glu<sup>1</sup>)fibrinopeptide B ( $M + 2H$ )<sup>2+</sup> ions, acquired before MUP analyses. Raw spectra consisted of between 10 and 30 summed scans and these were subsequently refined and transformed to a true mass scale using the MaxEnt 1 feature within the MassLynx software suite (Waters).

### Construction of Mean Isotopomer Distributions

Mean isotopomer distributions for myoglobin were constructed from eight isotopomer envelopes with charge states from +13 to +20 inclusive. Similar distributions for unfractionated and anion exchange purified MUPs were constructed from six isotopomer envelopes with charge states from +13 to +18 inclusive. In each case, tabulated  $m/z$  and relative intensity values for the isotopomer distributions at each charge state were created and exported to a spreadsheet. For each charge state, the  $m/z$  values of the nine most abundant isotopomers were converted to masses of the corresponding neutrals and mean values obtained using all charge states. To normalize the relative intensities of isotopomer distributions, the most intense ion signal from the most abundant charge state was given a value of 1 and the corresponding ion in each of the remaining charge state distributions was similarly assigned a relative intensity of 1. Mean relative intensity values for each isotopomer in a distribution were constructed from the normalized values at all charge states. These values were then plotted as a function of mean neutral mass within the isotopomer distribution. Unfractionated MUPs were analyzed in triplicate and mean isotopomer distributions were constructed from each replicate. These were subsequently combined to form a further mean distribution.

### Predicted Protein Sequences and Theoretical Isotopomer Distributions

A list of 15 MUP sequences was compiled from the National Center for Biotechnology Information (NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) protein database. Where sequences contained the intracellular signal peptide, the N-terminal pentapeptide of the mature, urinary MUPs (EEASS...) was identified and a predicted, average molecular mass generated. This mass was adjusted by  $-2$  Da to account for a disulfide bridge formation, and

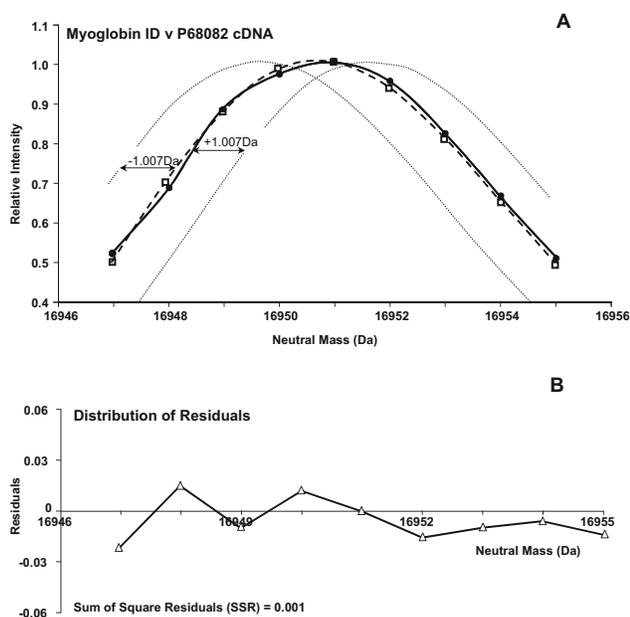
compared with masses previously obtained for proteins from BALB/c urine [5]. Predicted masses that were within 2 Da of observed values were considered as model sequences against which FT-ICR MS isotopomer distributions could be compared. Elemental compositions were generated from sequences that satisfied this criterion using MassLynx software (Waters). Each elemental composition was corrected for the disulfide bridge formation by subtraction of two hydrogen atoms, and this was then used to generate a theoretical isotopomer distribution within the Bruker Xmass software suite. Mean isotopomer distributions derived from FT-ICR MS spectral data were compared to these distributions.

### Anion Exchange Chromatography

Anion exchange chromatography of desalted MUPs was performed on a Resource Q column, ( $V_t = 6$  mL; Amersham Biosciences, Uppsala, Sweden) mounted on a DuoFlow chromatography platform (BioRad, Hemel Hempstead, UK). The column was initially equilibrated in a minimum of five column volumes of 50 mM N-morpholinoethane sulfonic acid (MES) buffer, pH 5.0 at 3 mL/min. A 200  $\mu\text{L}$  aliquot of a solution of desalted BALB/c MUPs (prepared as described above) was then loaded onto the column at the same flow rate, and the bound protein was washed with a further five column volumes of 50 mM MES buffer. The column was then developed with a linear NaCl gradient of 0 to 300 mM in 25 column volumes. The protein content of the column eluent was estimated by continuous monitoring of UV absorption at 280 nm ( $A_{280}$ ). Fractions (1 mL) were collected across the gradient and those corresponding to  $A_{280}$  peaks were pooled. Buffer exchange of pooled fractions into 50% (vol/vol) acetonitrile/0.1% (vol/vol) formic acid was achieved off-line with a C4 reversed-phase cartridge (Presearch, Hitchin, UK).

## Results and Discussion

Before application of FT-ICR MS and isotopomer modeling to MUPs, it was necessary to validate the technique with a known homogeneous protein; horse heart myoglobin was used for this purpose. A mean isotopomer distribution was constructed from ESI/FT-ICR MS data using eight charge states (+13 to +20) and this was modeled to a theoretical isotopomer distribution from a myoglobin amino acid sequence retrieved from the NCBI database (Figure 1a). The mean isotopomer distribution constructed from FT-ICR MS modeled closely to the theoretical distribution. Analysis of the residuals showed no discernible pattern and the sum of squares of the residuals (SSR) value was 0.001 (Figure 1b). The level of agreement between the two distributions is further graphically illustrated in Figure 1a by the inclusion of theoretical myoglobin distributions which were modeled at  $\pm 1$  Da. In both cases, these are well separated from the native myoglobin distributions.



**Figure 1.** Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS) analysis of intact myoglobin. A mean neutral mass isotopomer distribution (top panel) was derived from individual envelopes at charge states +13 to +20 of a myoglobin FT-ICR mass spectrum and compared to a model myoglobin isotopomer distribution derived from a myoglobin cDNA sequence retrieved from the National Council for Biotechnology Information (NCBI) database (accession number P68082). For illustrative purposes, two further model distributions are included corresponding to myoglobin  $\pm 1.007$ Da. These are derived by addition or subtraction of a single hydrogen atom to or from the original sequence. The distribution of residuals (bottom panel) and sum of square residuals value (SSR) are an indication of the error between the observed and theoretical distributions. Solid line: Mean observed isotopomer distribution; dashed line: Model isotopomer distribution.

Isotopomer distributions were compared for myoglobin spectra acquired on different days to assess the variability between experiments. The second myoglobin spectrum (not shown) gave an SSR value of 0.002 when compared with the same model distribution as used previously. A similarity index (SI) of 0.028 was calculated for the measured relative intensities of the isotope distributions using eq 1 [19], where  $i$  and  $i_0$  are the relative intensities of the mean myoglobin isotope distributions. The number of isotope signals used for the calculation ( $N$ ) was 9. Using this index, indistinguishable spectra will give SI values equal to 0, within the precision of the method. (It is worth noting that each FT-ICR mass spectrum is itself the accumulation of 150 scans, which will greatly reduce the effect of variability in ion signals in the individual scans).

$$SI = \sqrt{\frac{\sum_i \{(i - i_0/i + i_0) \times 100\}^2 / N}{N}} \quad (1)$$

The duplicate myoglobin spectra acquired for this study are insufficient to accurately define the variability in SSR measurements between spectra. By comparing

composite theoretical isotope distributions (derived from native myoglobin and myoglobin + 1Da) with the theoretical distribution from native myoglobin, the level of heterogeneity can be related to SSR values. A 10:1 mixture of native myoglobin and myoglobin + 1 Da, produced an SSR value of 0.0019, similar to the values observed in the two myoglobin spectra used to calculate the SI (0.001 and 0.002). This indicates that, within the precision of this method, it is unlikely to distinguish heterogeneity of proportions greater than 10:1. In the absence of more replicate myoglobin spectra, an SSR value  $\leq 0.002$  is considered to be homogeneous with the caveat that the technique will not discriminate LMH of lower than 10% of the native molecule.

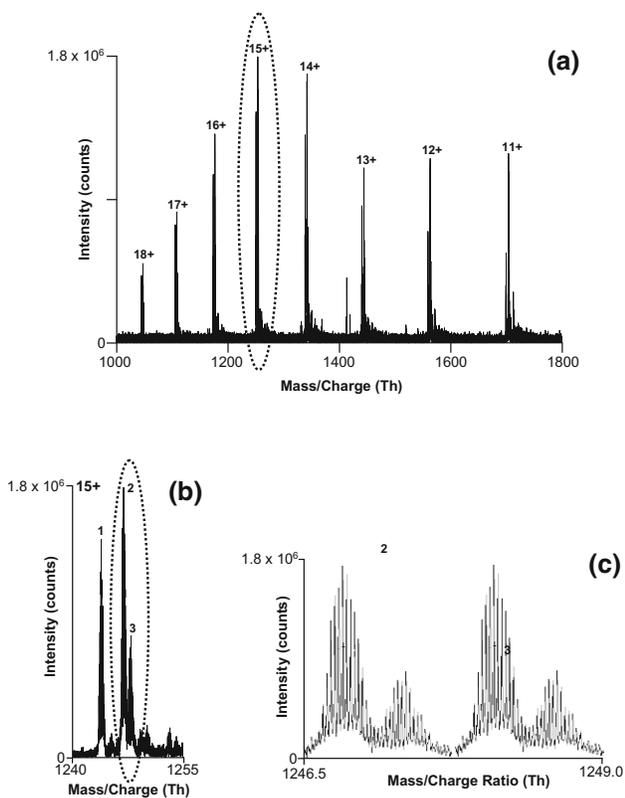
The high levels of MUP gene expression and ready isolation of the proteins from urine allowed a list of candidate MUP amino acid sequences to be assembled, from which isotopomer models could be constructed. A search of the NCBI database returned 15 complete, non-redundant MUP sequences. Of these, five were identified as candidate sequences for urinary MUPs on the basis of their predicted average masses and the observed masses of urinary MUPs [5] (Table 1). Unprocessed ESI/FT-ICR MS data from unfractionated BALB/c MUPs consisted of ions at charge states 11+ to 18+ (Figure 2a). Closer inspection of the +15 charge state (Figure 2b) revealed three distinct isotopomer distributions (labeled ID1, 2 and 3), two of which are shown in greater detail (Figure 2c). Mean isotopomer distributions were constructed in all three cases from a minimum of five charge states and these were compared with theoretical distributions generated from the MUP amino acid sequences. In each case, the theoretical distributions were corrected by a factor of  $-2.014$  Da, to account for the single disulfide bridge known to be present in urinary MUPs [20].

Modeling of the mean isotopomer distributions from unfractionated MUPs gave three distinct results. The first distribution (ID1) modeled closely to a theoretical distribution from MUP sequence AAH91744 (predicted average mass 18,645.846 Da) (Figure 3a). The distribu-

**Table 1.** MUP gene sequences used to construct theoretical isotopomer distributions

NCBI accession number	Predicted average mass (Da)	Nearest observed average mass (Da)
CAM13626	18692.8903	18693
AAH91744	18645.8463	18644
NP_032673	18692.8469	18693
AAH99597	18707.9019	18708
P02762	18693.8316	18693

An initial list of 15 MUP sequences was retrieved from the NCBI nucleic acid databases. MUP sequences CAM13626, AAH91744, NP\_032673, AAH99597, and P02762 were identified as candidate model sequences for isotopomer distribution modeling by comparison of their predicted average masses with published experimental masses for urinary MUPs derived from a low resolution mass spectrometer [5].



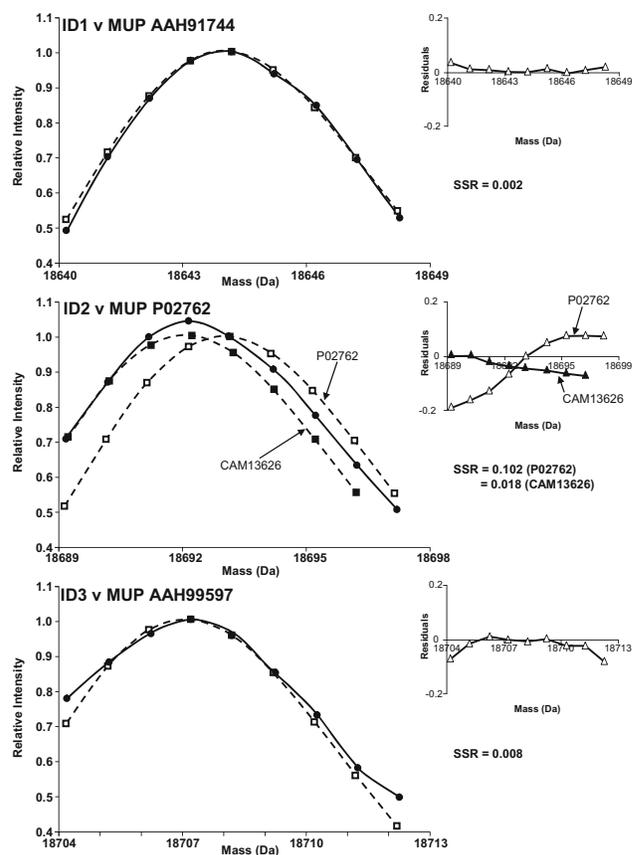
**Figure 2.** Electrospray ionization Fourier transform ion cyclotron resonance (ESI-FTICR) mass spectra of unfractionated BALB/c MUPs. The FT-ICR mass spectrum of unfractionated BALB/c MUPs (a), the sum of 150 scans, consisted of eight groups of isotopomer envelopes with charge states between 11+ and 18+. Closer examination of a single group of isotopomer envelopes (15+) revealed three distinct isotopomer distributions (b). An expanded view of two of these isotopomer distributions is also shown (c).

tion of residuals and the SSR (0.002) were comparable to those from myoglobin (Figure 1b). The second distribution (ID2) was significantly distorted (Figure 3b). The two nearest theoretical MUP isotopomer distributions were from sequences CAM13626 (18,692.890 Da) and P02762 (18,693.832 Da), of which P02762 showed the lowest SSR (0.102). Another theoretical isotopomer distribution, derived from sequence NP\_032673 (18,692.847 Da), which differed in average mass from CAM13626 by  $-0.043$  Da, was omitted from Figure 3b to enhance clarity and on the basis that it has not been previously observed as a urinary MUP in inbred mice [5]. Finally, the third MUP isotopomer distribution (ID3, Figure 3c) modeled closely to the theoretical distribution from MUP AAH99597 (18,707.902 Da). In this instance, however, the SSR was significantly higher (0.008) than those observed for myoglobin

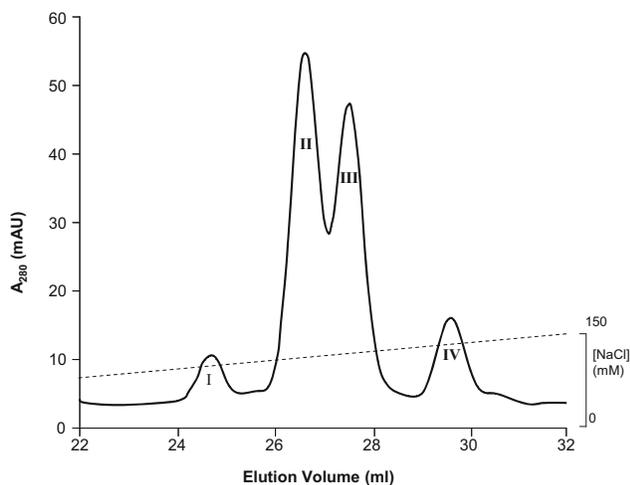
Three distinct circumstances apply to the comparisons of the observed and predicted isotopomer distributions in Figure 3. The close agreement between ID1 and the theoretical distribution from MUP AAH91744 is consistent with expression of this gene resulting in a homogeneous protein at this mass. In contrast, ID2 did

not model well to any predicted gene product. The nearest matching theoretical protein sequences fell into two groups, the first containing CAM13626 and NP\_032673 and the second containing P02762. The two groups are separated in mass by  $\sim 1$  Da, and a possible explanation for the distortion observed in ID2 is that proteins expressed from both groups contribute to the isotopomer distribution.

A degree of distortion was also observed in ID3 but this was both less pronounced and different to ID2. Large portions of ID3 modeled closely to the theoretical distribution from MUP AAH99597 but substantial errors were observed at the extremes of the distribution. This is attributed to the fact that ID3 was generated



**Figure 3.** Comparison of isotopomer distributions from unfractionated MUPs and model distributions derived from MUP DNA sequences. Three mean isotopomer distributions were constructed from the FTICR mass spectrum of unfractionated BALB/c MUPs (Figure 2). The processed data from individual isotopomer distributions (IDs) were compared to model data derived from four MUP DNA sequences; ID1 v MUP AAH91744 (upper panel), ID2 v MUP P02762 and CAM 13,626 (middle panel), and ID3 v MUP AAH99597 (lower panel). In each case, a correction of  $-2.014$  Da was applied to the model data to account for a single disulfide bridge known to be present in MUPs. Assessment of the error between observed and model data were estimated using nonlinear curve fitting. The residual differences between observed and model distributions are shown as a function of isotopomer mass (insets) and the sum of square residuals (SSR) for each distribution is shown. Solid line: Observed mean isotopomer distribution, dashed line: Model distribution.



**Figure 4.** Fractionation of BALB/c MUPs by anion exchange chromatography. A 200  $\mu$ L aliquot of desalted BALB/c MUPs was fractionated by anion exchange chromatography. The protein content of the column eluent was estimated by continuous monitoring of UV absorbance at 280 nm and the portion of the elution profile containing MUPs is shown. To determine which MUPs were present in each of the four peaks (I–IV), associated fractions (1 mL) were pooled and desalted with an off-line reversed-phase trap before infusion into a Q-TOF micro mass spectrometer (Waters) fitted with an electrospray ionization source. Anion exchange peaks containing MUPs subsequently analyzed by FT-ICR MS are annotated with bold type.

using the least intense mass spectrometric signals, and at the extremes of the distribution had very low signal to noise ratios. (It is for this reason that the points in the distribution figures are linked by interpolation rather than curve fitting, which would inevitably place undue weight on the extreme points in the distribution.)

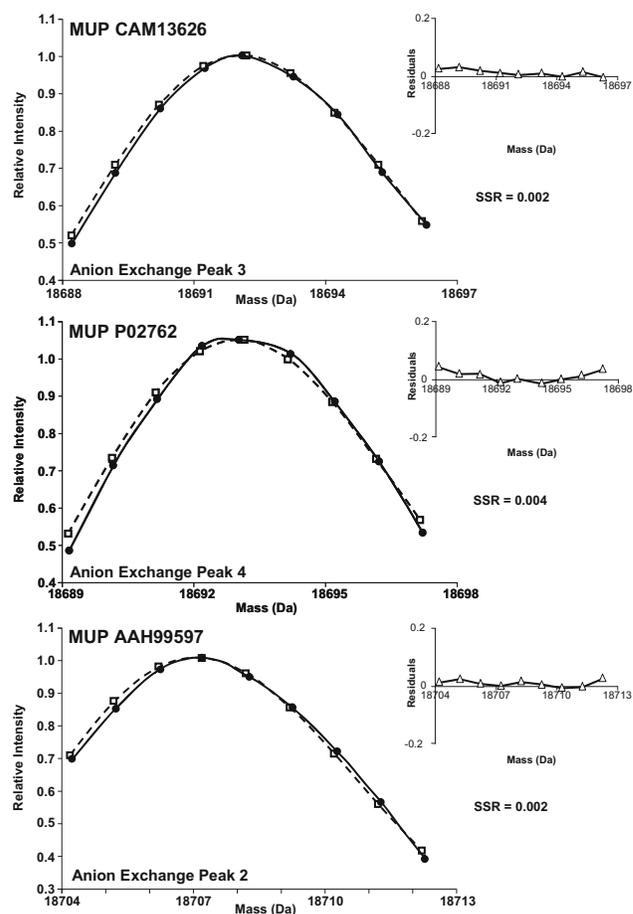
Whilst the individual MUPs in mouse urine share a high degree of sequence homology (>98%), the few variations in the amino acid chain often involve charged residues [5]. As a result, MUPs are readily fractionated by ion-exchange chromatography. This provides a means of preparing MUPs with significantly more homogeneity than observed in the unfractionated sample. As the separation by anion exchange chromatography is charge based, it has the potential to separate MUPs with small mass differences, providing these differences result from substitutions in charged amino acid residues.

The MUPs from BALB/c urine were fractionated into four discrete peaks (Figure 4) and mean isotopomer distributions constructed from FT-ICR MS analyses of MUPs within Peaks II, III, and IV of the anion exchange chromatogram were compared with theoretical distributions from MUPs AAH99597, CAM13626, and P02762, respectively (Figure 5). In each case the error between mean and theoretical distributions was significantly less than that observed in the unfractionated sample. This is reflected in the SSR values (0.002, 0.002, and 0.004, respectively), which were comparable to that observed for myoglobin.

The isotopomer distribution from anion exchange

Peak II (Figure 5, lower panel) matched the theoretical distribution from MUP AAH99597 better than the unfractionated mixture (ID3). The preparation of the anion exchange peak for ESI-FTICR/MS included a reversed-phase concentration and desalting step, which greatly increased the concentration of the final preparation. The increased signal to noise ratio in the FTICR spectrum resulting from this preparation is the most likely explanation for the improved agreement between the experimental and theoretical distributions.

The presence of two MUPs in the anion exchange profile (Figure 4, Peaks III and IV) with a mass differ-



**Figure 5.** Comparison of mean isotopomer distributions from anion exchange purified MUPs to model distributions derived from DNA sequences. Mean isotopomer distributions were constructed from FTICR mass spectra of MUP anion exchange Peaks II, III, and IV (see Figure 4). The processed data from individual anion exchange purified MUPs were compared with model data from MUP DNA sequences; anion exchange peak 3 v MUP CAM13626 (upper panel), anion exchange peak 4 v MUP P02762 (middle panel), and anion exchange peak 2 v MUP AAH99597 (lower panel). In each case a correction of  $-2.014$  Da was applied to the model data to account for a single disulphide bridge known to occur in MUPs. Assessment of the error between observed and model data were estimated using nonlinear curve fitting. The residual differences between observed and model distributions are shown as a function of isotopomer mass (insets) and the sum of square residuals (SSR) for each distribution is annotated. Solid line: Observed mean isotopomer distribution; dashed line: Model isotopomer distribution.

ence of 1 Da (Figure 5, upper and middle panels) is consistent with the hypothesis that a contribution from both causes distortion in the isotopomer distribution for component ID2 in the unfractionated sample (Figure 3, middle panel). This hypothesis was tested further by the construction of a composite theoretical isotopomer distribution, containing contributions from both MUPs at proportions determined from the peak areas of the anion exchange chromatogram, namely 3.3:1 for CAM13626:P02762.

A comparison of the composite isotopomer distribution to ID2 (Figure 6) gave low errors (SSR 0.001), consistent with the proposition that the heterogeneity originally observed in ID2 resulted from expression of both these MUP genes. For guidance, theoretical isotopomer distributions constructed from other CAM13626:P02762 ratios are also included in the figure. This conclusion is consistent with a previous study [5], which used MALDI-TOF/MS to analyze endoproteinase Lys-C peptides from these two MUP anion exchange peaks. This established that the only difference in sequence between the MUPs in Peaks III and IV was a substitution between the charged residues Lys and Glu at position 140 (mature protein sequence numbering).

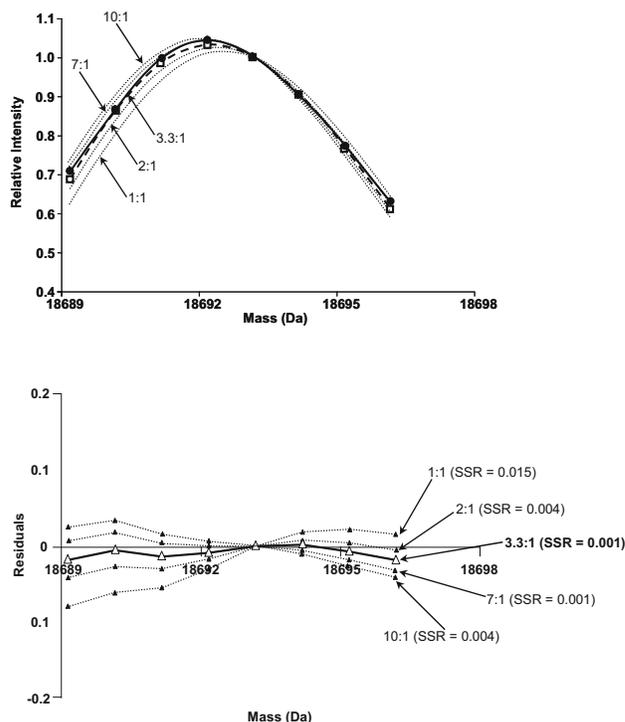
## Conclusions

Heterogeneity in a protein preparation can include the presence of proteins of related sequence with mass differences of 2 Da or less. This can arise for a number of reasons that reflect both biological and nonbiological processes [3–6]. In both cases, characterization of the heterogeneity is crucial. The ability of FT-ICR MS to resolve protein isotopomers for each of the charge states formed by electrospray ionization offers a sensitive, rapid method for detecting heterogeneity.

Pure preparations of MUPs are a useful research tool, which are used to unravel the complex olfactory signals they convey [16]. MUPs have provided an ideal model for assessing the validity of this technique in the context of a real biological system. FT-ICR MS and isotopomer distribution modeling demonstrated heterogeneity in unfractionated MUPs that would not be apparent following analysis by lower resolution mass spectrometric techniques. Previous attempts to characterize the MUPs in BALB/c urine have produced similar findings to this study [5] but required a significantly more complex and lengthy analytical approach. Application of the current approach to anion exchange purified MUPs provided useful evidence of the purity of these preparations, particularly when results are evaluated (as here) alongside data from equivalent analysis of a known homogeneous protein.

The work presented here illustrates the utility of the information contained within a protein isotopomer envelope and exemplifies the benefits of intact protein analysis (rather than analysis of proteolytic fragments). A survey of the protein modification website Deltamass

### I.D.2 v MUP P02762/CAM13626 COMPOSITE DISTRIBUTION



**Figure 6.** Comparison of an isotopomer distribution from unfractionated MUPs to a composite model distribution constructed from two DNA sequences. To test the hypothesis that the distortion previously observed in isotopomer distribution 2 from unfractionated MUPs was attributed to the presence of two closely related proteins, a composite isotopomer distribution was constructed. This contained contributions from sequences CAM13626 and P02762 (both identified as component MUPs by anion exchange chromatography and FT-ICR MS) at a ratio of 3.3:1, respectively. This ratio was derived from the relevant peak areas in the anion exchange elution profile (see Figure 5). The second mean isotopomer distribution constructed from unfractionated MUPs (ID2) was compared with this composite distribution (top panel). For illustrative purposes, theoretical distributions derived from varying compositions of CAM13626 and P02762 are also included. In each case the annotated ratio describes the proportions of CAM13626:P02762. The residual differences between observed and model distributions are shown as a function of isotopomer mass (bottom panel). For illustrative purposes, the residual distributions between the observed data and distributions containing varying compositions of CAM13626 and P02762 are also included. In each case, the ratio of the composite distribution and the SSR value are annotated. Solid line: Observed mean isotopomer distribution; dashed line: Model composite isotopomer distribution.

([www.abrf.org](http://www.abrf.org)) lists 13 types of chemical modification that result in a mass change of between +2 Da and –2 Da. Determination of proteins differentiated by such modifications should be readily achievable using the experimental approach described here, comprising FT-ICR MS and isotopomer distribution modeling. Finally, the data processing steps in this study are amenable to automation that would encourage more widespread application.

## Acknowledgments

The authors thank the UK Engineering and Physical Sciences Research Council (EP/D013615/1 to SJG) and the Biotechnology and Biological Sciences Research Council (S19816 to JLH and RJB) for financial support of this work.

## References

- Hayter J. R.; Robertson D. H. L.; Gaskell S. J.; Beynon, R. J., Proteome Analysis of Intact Proteins in Complex Mixtures. *Mol. Cell. Proteom.* **2003**, *2*, 85–95.
- Wang, H.; Hanash, S. Intact Protein Based Sample Preparation Strategies for Proteome Analysis in Combination with Mass Spectrometry. *Mass Spectrom. Rev.* **2005**, *24*, 413–426.
- Shimizu T.; Matsuoka Y.; Shirasawa, T. Biological Significance of Isoaspartate and its Repair System. *Biol. Pharm. Bull.* **2005**, *28*, 1590–1596.
- Amato A.; Connolly C.N.; Moss S.J.; Smart, T. G. Modulation of Neuronal and Recombinant GABAA receptors by Redox Reagents. *J. Physiol.* **1999**, *517*, 35–50.
- Robertson D. H. L.; Evershed R. P.; Cox, K.; Gaskell S.; Beynon, R. J. Characterisation of Molecular Heterogeneity in the Major Urinary Proteins of the House Mouse *Mus musculus*. *Biochem. J.* **1995**, *315*, 461–466.
- Lindner H.; Helliger, W. Age-Dependent Deamidation of Asparagine Residues in Proteins. *Exp. Gerontol.* **2001**, *36*, 1551–1563.
- Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.
- Loo, J. A.; Quinn, J. P.; Ryu, S. I.; Henry, K. D.; Senko, M.W.; McLafferty, F. W. High Resolution Tandem Mass Spectrometry of Large Biomolecules. *Proc. Nat. Acad. Sci. U.S.A.* **1992**, *89*, 286–328.
- Green M. K.; Vestling, M. M.; Johnston, M. V.; Larsen, B. S. Distinguishing Small Molecular Mass Differences of Proteins by Mass Spectrometry. *Anal. Biochem.* **1998**, *260*, 204–211.
- Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320–332.
- Lee, S. W.; Berger, S. J.; Martinovic, S.; Pasa-Tolic, L.; Anderson, G. A.; Shen, Y.; Zhao, R.; Smith, R. D. Direct Mass Spectrometric Analysis of Intact Proteins of the Yeast Large Ribosomal Subunit Using Capillary LC-FTICR. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5942–5947.
- Robinson, N. E.; Zabrouskov, V.; Zhang, J.; Lampi, K. J.; Robinson, A. B. Measurement of Deamidation of Intact Proteins by Isotopic Envelope and Mass Defect with Ion Cyclotron Resonance Fourier Transform Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3535–3541.
- Finlayson, J. S.; Baumann, C. A. Mouse Proteinuria. *Am. J. Physiol.* **1958**, *192*, 69–72.
- Finlayson, J. S.; Potter, M.; Runner, C. R. Electrophoretic Variation and Sex Dimorphism of the Major Urinary Protein Complex in Inbred Mice: A New Genetic Marker. *J. Natl. Cancer Inst.* **1963**, *31*, 91–107.
- Finlayson, J. S.; Asofsky, R.; Potter, R.; Runner, C. C. Major Urinary Protein Complex of Mice: Origin. *Science* **1965**, *149*, 981–982.
- Hurst, J. L.; Payne, C. E.; Nevison, C. M.; Marie, A. D.; Humphries, R. E.; Robertson, D. H. L.; Cavaggioni, A.; Beynon, R. J. Individual Recognition in Mice Mediated by Major Urinary Proteins. *Nature* **2001**, *414*, 631–634.
- Evershed, R. P.; Robertson, D. H.; Beynon, R. J.; Green, B. N. Application of Electrospray Ionization Mass Spectrometry with Maximum Entropy Analysis to Allelic “Fingerprinting” of Major Urinary Proteins. *Rapid Commun. Mass Spectrom.* **1993**, *10*, 882–886.
- Clark, A. J.; Chave-Cox, A.; Ma, X.; Bishop, J. O. Analysis of Mouse Major Urinary Protein Genes: Variation Between the Exonic Sequences of Group 1 Genes and a Comparison with an Active Gene Out with Group 1 Both Suggest that Gene Conversion Has Occurred Between MUP Genes. *EMBO J.* **1985**, *4*, 3167–3171.
- Wan, K. X.; Vidavsky, I.; Gross, M. L. Comparing Similar Spectra: From Similarity Index to Spectral Contrast Angle. *J. Am. Soc. Mass Spectrom.* **2001**, *13*, 85–88.
- Szoka, P. R.; Gallagher, J. F.; Held, W. A. In Vitro Synthesis and Characterization of precursors of the Mouse Major Urinary Proteins. *J. Biol. Chem.* **1980**, *255*, 1367–1373.