Proteome Analysis of Intact Proteins in Complex Mixtures*

Julia R. Hayter‡, Duncan H. L. Robertson‡, Simon J. Gaskell§, and Robert J. Beynon‡¶

Analysis of intact protein mixtures by electrospray ionization mass spectrometry requires the resolution of a complex, overlapping set of multiply charged envelopes. To ascertain the ability of a moderate resolution mass spectrometer to resolve such mixtures, we have analyzed the soluble proteins of adult chick skeletal muscle. This is a highly specialized tissue showing a marked bias in expression of glycolytic enzymes in the soluble fraction. SDS-PAGE-resolved proteins were first identified by a combination of matrix-assisted laser desorption ionization time-of-flight (TOF) and electrospray ionization tandem mass spectrometry. Then the mixture of intact proteins was introduced into the electrospray source of a Q-TOF mass spectrometer either by direct infusion or via a C4 desalting trap. In both instances, the complex pattern of peaks could be resolved into true masses, and these masses could in many instances be reconciled with the masses predicted from the known protein sequences when qualified by expected co- and post-translational modifications. These included loss of the N-terminal initiator methionine residue and N-terminal acetylation. The ability to resolve such a complex mixture of proteins with a routine instrument is of considerable value in analyses of protein expression and in the confirmation of post-translational changes in mature proteins. ***Molecular & Cellular Proteomics 2:85–95, 2003.***

A key step in any proteome analysis is the identification of one or more proteins that have been resolved by separation technologies, including one-dimensional and two-dimensional polyacrylamide gel electrophoresis or liquid chromatography. In most instances, the identification begins with complete enzymatic fragmentation of the target proteins with a proteinase of defined specificity such as trypsin. In this way, additional information can be obtained as the properties of the resultant "limit peptides," notably their mass, can be used as a fingerprint to search databases of all known proteins (1, 2). This requires a very good match between the experimental data and the theoretical digest map and usually requires that the target protein sequence already be deposited in the database. Searching with peptide maps derived from organisms for which extensive genome data is not available usually requires the gain of further identification data, specifically composition data (3) or short peptide sequence tags derived by sequencing *de novo* (4, 5).

In peptide mass fingerprinting, the masses of the limit peptides define the positions of the specific amino residues that define the site of cleavage. Proteolysis by trypsin, for example, cleaves the protein into limit peptides typically of ~10–20 amino acids long. The masses of such peptides are readily matched by the high mass accuracy obtained by the mass analyzers in common use for peptide mass fingerprinting, notably a time-of-flight (TOF)[1] mass analyzer, which can achieve a mass accuracy of 10 ppm and a resolution in excess of 10,000 full-width half-maximum.

However, the generation of the set of limit peptides introduces additional steps that can result in sample losses and that diminishes throughput. To circumvent these stages, it has been proposed that the mass of each protein in a proteome can, at an appropriate level of mass accuracy and resolution, be considered to be virtually unique in the data set, which raises the possibility of direct analysis of an intact protein (6). Indeed the superior mass accuracy and resolution of Fourier transform ion cyclotron resonance (FTICR) mass spectrometers might offer a route to highly accurate mass determination of intact proteins that are electrosprayed as multiply charged species (7, 8). With such an instrument, proteins between 10,000 and 100,000 Da can be mass measured to an accuracy of around 1 ppm, which can provide data of sufficiently high quality that identification based only on intact mass might in principle be possible. However, protein identification by mass determination of the intact protein is compromised by post-translational changes, including proteolysis, glycosylation, acetylation, phosphorylation, or even disulfide bond formation (9). All of these modifications elicit a change in the mass of the protein, creating discordance between the measured mass and the mass predicted from the

From the ‡Department of Veterinary Preclinical Sciences, University of Liverpool, Crown Street Liverpool L69 7ZJ, United Kingdom and the §Michael Barber Centre for Mass Spectrometry, Department of Chemistry, University of Manchester Institute of Science and Technology, P.O. Box 88, Manchester M60 1QD, United Kingdom

Received, November 29, 2002

Published, MCP Papers in Press, January 21, 2003, DOI 10.1074/mcp.M200078-MCP200

[1] The abbreviations used are: TOF, time-of-flight; ACN, acetonitrile; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; EST, expressed sequence tag; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; MS/MS, tandem MS; MaxENT 1, maximum entropy software for deconvolution of multiply charged electrospray envelopes; ESI, electrospray ionization; FTICR, Fourier transform ion cyclotron resonance; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol.

(usually) cDNA-inferred protein sequence. Moreover, FTICR instruments are currently expensive and not yet generally available for proteome analysis. Accordingly we embarked on a study of a complex but natural mixture of proteins using an affordable benchtop mass spectrometer. Specifically we used a preparation of the soluble proteins of chicken skeletal muscle. The adult pectoralis muscle of the chicken (*Gallus gallus*) is a highly specialized tissue comprising almost exclusively Type IIB fibers. These fibers rely on glycolysis to provide energy for muscle contraction, and the sarcoplasmic protein fraction is rich in glycolytic enzymes. The marked bias in protein expression of this class of proteins means that an unfractionated high speed supernatant fraction of a homogenate of this tissue provides a complex mixture that is nonetheless sufficiently enriched in relatively few proteins to provide an appropriate test material for this analysis. Specifically we wished to assess the extent to which direct mass measurement of the proteins in this sample could be of value in their identification using a readily available instrument (Q-TOF-micro, Micromass, Manchester, UK).

EXPERIMENTAL PROCEDURES

Experimental Material—Commercial frozen broiler chicken skeletal muscle (5 g) was mechanically homogenized in 25 ml of a buffer consisting of 20 mM BisTris, 50 mM NaCl, pH 6.0 and containing Complete protease inhibitors (Roche Diagnostics). The soluble protein fraction was isolated by centrifugation at 13,000 × *g* for 45 min at 4 °C. The protein concentration of the resulting supernatant fraction was determined using the Coomassie Plus® Protein assay (Perbio Science UK, Tattenhall, UK), and the supernatant fraction was stored at −20 °C in 200-μl aliquots.

Polyacrylamide Gel Electrophoresis—The soluble proteins (4–40 μg) were initially separated by one-dimensional 12.5% (v/v) SDS-PAGE and visualized with Coomassie Brilliant Blue stain. In some instances, the gel was scanned, and the band volumes were analyzed by Phoretix 1-D software (Non-linear Dynamics, Newcastle, UK). The correlation coefficient for the relationship between band volume and protein amount was 0.98. For the two-dimensional electrophoresis, the soluble supernatant fraction containing 100 μg of protein was added to 200 μl of rehydration buffer (8 M urea, 2% (v/v) CHAPS, 50 mM dithiothreitol, 0.2% (v/v) pH 3–10 Biolyte ampholytes). This sample was loaded onto 11-cm linear IPG strips, pH 3–10 (Bio-Rad). Passive rehydration was completed overnight, and the strips were then electrophoresed at 250 V for 15 min followed by 250–8000 V for 35,000 V-h using the Protean IEF system (Bio-Rad). The strips were equilibrated in reducing equilibration buffer (6 M urea, 0.375 M Tris, pH 8.8, 2% (w/v) SDS, 20% (v/v) glycerol, 2% (w/v) dithiothreitol) for 15 min and then in alkylating buffer 2 (6 M urea, 0.375 M Tris, pH 8.8, 2% (w/v) SDS, 20% (v/v) glycerol, 2.5% (w/v) iodoacetamide) for a further 15 min before the second dimension separation was performed in 12% gels using the Hoefer Ettan™ DALT *six* vertical gel system at 600 V for 3 h. The proteins were visualized using Coomassie Brilliant Blue stain.

In-gel Trypsin Digestion—A plug of the protein of interest was excised from the gel with a fine glass pipette and transferred to a microcentrifuge tube. To each tube 25 μl of 50 mM ammonium bicarbonate, 50% (v/v) acetonitrile (ACN) was added and incubated at 37 °C for 20 min. This process was repeated until all the stain had been removed. The plugs were then washed in 50 mM ammonium bicarbonate, which was subsequently discarded. Dithiothreitol (25 μl, 10 mM) was added to each plug and incubated at 37 °C. After 30 min, the supernatant was discarded, iodoacetamide (25 μl, 55 mM) was added to each tube, and incubation continued in the dark for 60 min. The gel was dehydrated using 25 μl of ACN, and incubation at 37 °C was resumed for 15 min. The supernatant was removed from the dehydrated plug, which was allowed to air dry. Once dry, the gel was rehydrated in 50 mM ammonium bicarbonate (9 μl) containing trypsin (1 μl of 100 ng/μl trypsin stock reconstituted in 50 mM acetic acid). After 30 min, 50 mM ammonium bicarbonate (10 μl) was added to each tube, and digestion was allowed to continue overnight at 37 °C; the reaction was halted by the addition of 2 μl of formic acid. In-gel digestion of proteins was performed using the MassPREP automated work station (Micromass).

MALDI-TOF Analysis—All analyses were performed on a reflectron-equipped MALDI-TOF instrument (M@LDI, Micromass), and peptide mass fingerprints were searched using the MASCOT search engine (10). Samples were mixed in a 1:1 ratio with a saturated solution of α-cyano-4-hydroxycinnamic acid in ACN:water:trifluoroacetic acid (50:49:1, v/v/v). Monoisotopic peptide masses in the mass range of 800–4000 Da were collected and used in the database search. The initial search parameters allowed an error of ±250 ppm, a single trypsin missed cleavage, carbamidomethyl modification of cysteine residues, and oxidation of methionine. The taxonomic search space was restricted to Chordata, and there was no restriction regarding mass or pI.

Intact Protein Mass Determination—All analyses were performed using a Q-TOF-micro tandem mass spectrometer (Micromass). In-line desalting was achieved using a C4 reversed phase trap, loading and elution from which were controlled by an LC Packings Ultimate/Switchos chromatographic system. Proteins were bound to the trap in 5% (v/v) aqueous ACN, 0.2% (v/v) formic acid and were eluted in 90% (v/v) aqueous ACN, 0.2% (v/v) formic acid via a capillary directly into the Q-TOF-micro at 200 nl/min. Data were acquired between *m/z* 500 and 2000 with a cycle time of 2.4 s. Typically eight to ten acquisitions were combined into a single spectrum, which was subsequently deconvoluted using MaxENT 1 maximum entropy software (Micromass). Spectra were processed between 5000 and 100,000 Da at 1 Da/channel. The peak width parameter used to construct the damage model was 0.75 Da in all cases. Proteins were identified by comparison of the observed mass with the predicted mass obtained from the Swiss-Prot entries for those proteins that had been identified by MALDI-TOF peptide mass fingerprinting.

Tandem Mass Spectrometry—All analyses were performed using the Q-TOF-micro (Micromass). Samples were introduced by static nanospray from metal-coated capillaries held at a typical potential of 1000 V. Peptides generated from an in-solution digest were first assessed by MALDI to ensure digestion had occurred. An initial mass spectrum was collected between *m/z* 300 and *m/z* 2000 from which multiply charged ions were identified as candidates for MS/MS from their natural isotope envelope. The resultant fragmentation spectra were processed using MaxENT 3 software, and the peptide sequence was determined using the PepSeq software in the Masslynx package. Peptide sequences were searched against the protein database using National Center for Biotechnology Information (NCBI) *protein-protein* BLAST. The taxonomy was limited to *G. gallus*, "expect" was set to 1000, and the word size was set to 2; no other parameters were limited. Identified proteins were cross-referenced to both the intact protein and the MALDI data already obtained. The chicken EST database (www.chick.umist.ac.uk/, Ref. 11) was used to identify peptide sequences that did not match any of the NCBI database entries. The algorithm was tblastn using the PAM70 matrix and either used the set of assembled sequences or a subset of EST sequences specifically derived from chicken skeletal muscle. These were then cross-referenced to any unidentified peptide fingerprints obtained by MALDI-TOF mass spectrometry.

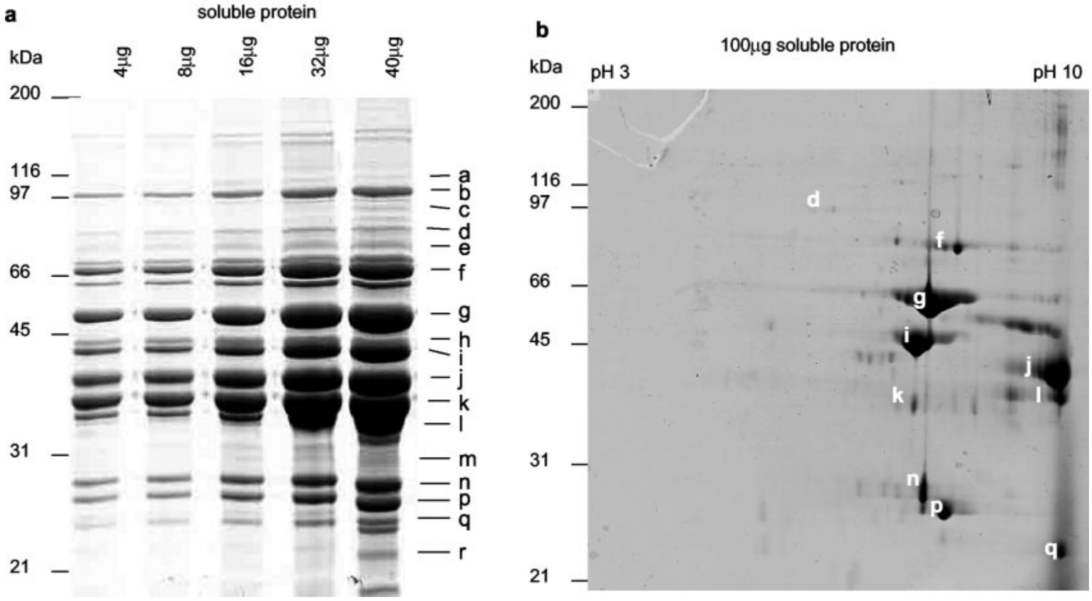


FIG. 1. **The complexity of soluble proteins from chicken skeletal muscle.** Soluble proteins (between 4 and 40 μg) were separated on 12% (v/v) linear SDS-polyacrylamide gels (*panel a*). A further 100 μg of protein was separated by two-dimensional gel electrophoresis using an 11-cm linear immobilized pH gradient strip pH 3–10 for the first dimension and a linear 12% polyacrylamide gel in the second dimension (*panel b*). Both gels were stained with Coomassie Brilliant Blue. Proteins (***a–r***, excluding "**o**," which was omitted for clarity) were identified by in-gel digestion using trypsin and are detailed in Table I.

RESULTS AND DISCUSSION

Characterization of the Protein Mixture—The pattern of proteins in pectoralis muscle of skeletal muscle is relatively simple and reflects the dramatic bias in levels of protein expression that is a feature of this tissue. On one-dimensional SDS-PAGE, ~20 proteins can be readily discerned after Coomassie Blue staining. (Fig. 1*a*). The additional resolution of a two-dimensional separation failed to separate the mixture to any greater extent, although there is some evidence for "charge trains" for the more abundant spots on the gel, implying charge-modifying post-translational changes such as deamidation (Fig. 1*b*). Thus, the one-dimensional separation is a reasonable approximation of the true complexity of the mixture at least in the context of the most abundant proteins.

Before we embarked on an analysis of this mixture in the form of intact proteins, we sought to identify the constituent proteins by MALDI-TOF mass spectrometry of tryptic peptides following "in-gel" digestion of the bands on the one-dimensional SDS-polyacrylamide gel. This approach, recently described as the "bottom up" approach to identification of proteins (12, 13), was an essential prerequisite to our study. The peptide mass fingerprints were analyzed using the MASCOT search engine (10) and were searched against proteins from Chordata or from a local database containing 3554 chicken proteins. The proteins that were analyzed in this way are marked on the one-dimensional gel with the letters **a** to **r**, and of the 17 proteins examined, unequivocal assignments could be made for 14 by matching to known chicken proteins with sequence coverage ranging from 17% for serum albumin to 53% for triose-phosphate isomerase (Table I).

Two proteins, bands **n** and **b**, were tentatively assigned as phosphoglycerate mutase and glycogen phosphorylase ,respectively, on the basis of cross-species hits when searched against a database of sequences from Chordata. In both these instances, the top 20 hits returned from the peptide mass fingerprint search were homologues from other species including sheep, cow, rat, rabbit, fruit fly, and human. Band **j**, a particularly abundant protein, could not be identified with high confidence by peptide mass fingerprinting.

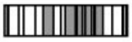
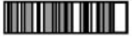
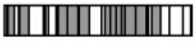
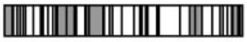
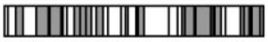
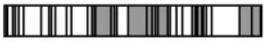
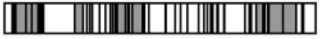
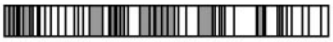
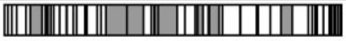
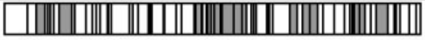
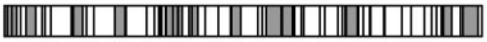
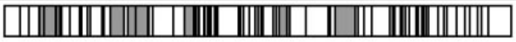
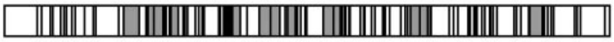
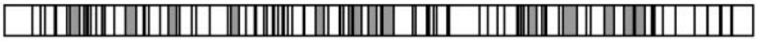
These three protein bands (**n**, **b**, and **j**) required further analysis for an unequivocal assignment. Tryptic fragments of these proteins were sequenced *de novo* by MS/MS; the mass spectrum from a MALDI-TOF analysis of the same sample was used to select peptides that had also been present in the digest of the same band on the one-dimensional separation.

For band **n**, the closest match by peptide mass fingerprinting was to skeletal muscle phosphoglycerate mutase from rat liver for which six peptides, covering 30% of the sequence, were matched at a tolerance of 250 ppm. The predicted molecular mass of this protein, at 28 kDa, was consistent with the mobility in the one-dimensional gel equivalent to approximately 30 kDa. There is no sequence available for the muscle isoform of the chicken enzyme, but the match to the chicken liver isoform was substantially worse (Fig. 2). The sequence of the rat muscle enzyme was used to search (with tblastn) the chicken skeletal muscle EST database, and five EST-derived sequences were retrieved that aligned to the C terminus of the protein. Seven peptides in the MALDI-TOF spectrum matched

Table I

Proteins identified using peptide maps generated from trypsin in-gel digestion

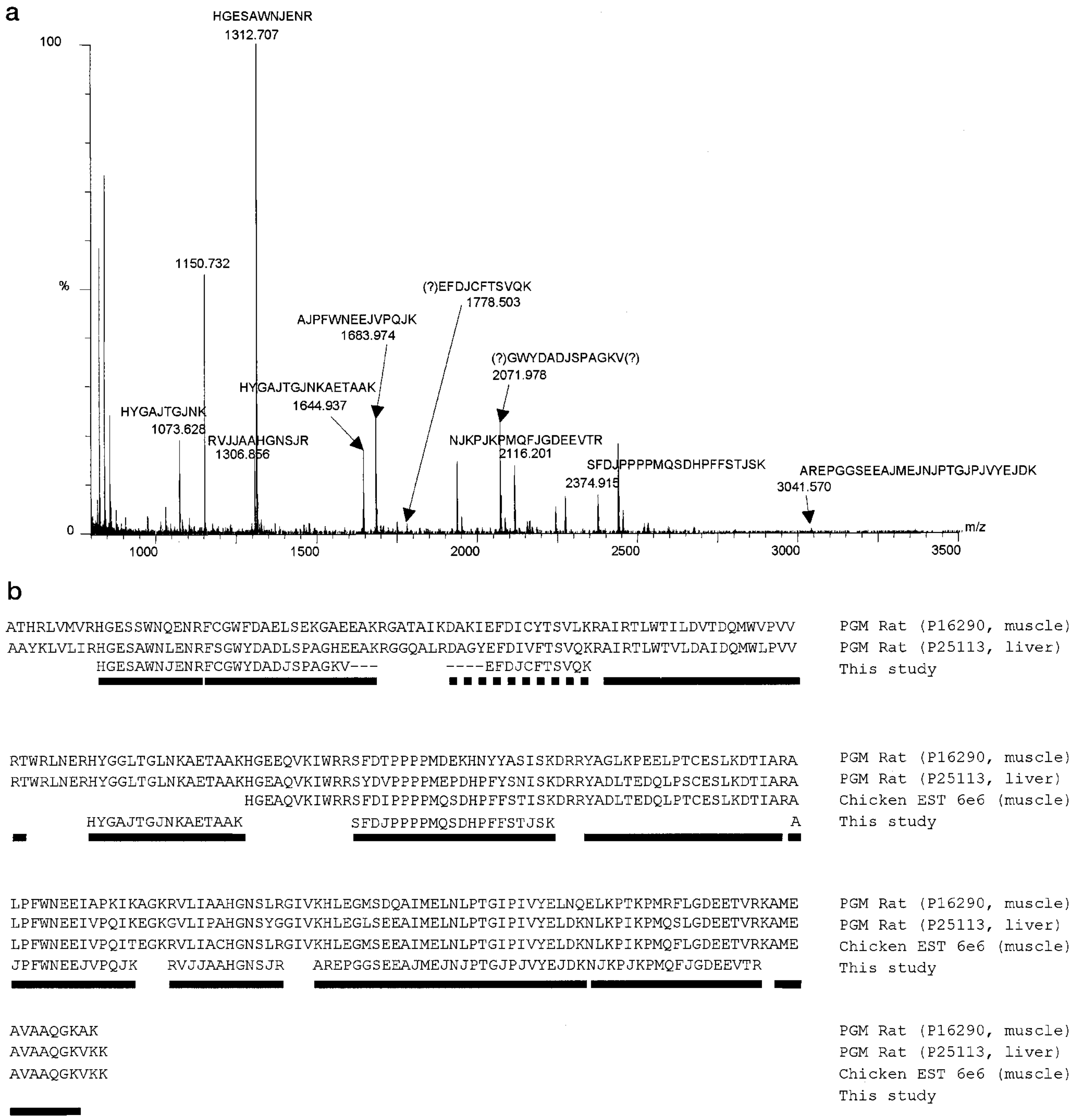
Soluble proteins of chicken skeletal muscle were resolved by gel electrophoresis and analyzed by peptide mass fingerprinting. For all proteins for which an identification was made, a coverage map indicates those peptides that were unambiguously identified. The filled areas indicate peptides that were observed in the MALDI-TOF spectra. Three proteins were identified by cross-species matching. For these three proteins, selected peptides were sequenced by tandem mass spectrometry to confirm the identification (see text).

Peak	Band	MALDI ID	Coverage diagram	Coverage (%)
	r	Actin polymerisation inhibitor (A39644)		29
2	q	Adenylate kinase (A25237)		29
3	p	Triosephosphate Isomerase (P00940)		53
4	n	Phosphoglycerate mutase	Cross species match confirmed by ms/ms data – see text	N/A
	m	Succinate dehydrogenase IP subunit (AAC72372)		27
5	l	L-Lactate dehydrogenase M chain (P00340)		40
6	k	Glyceraldehyde 3-phosphate dehydrogenase (P00356)		36
7	j	Aldolase A	Cross species match confirmed by ms/ms data – see text	N/A
8	i	Creatine kinase (P00565)		37
9	h	Phosphoglycerate kinase (P51903)		32
10	g	β-enolase (P07322)		30
11	f	Pyruvate kinase (P00548)		25
	e	Serum albumin (mature sequence from P19121)		17
	d	Heat shock protein 70kDa (P08106)		19
	c	Heat shock protein 90kDa (P11501)		22
	b	Glycogen phosphorylase	Cross species match confirmed by ms/ms data – see text	N/A
	a	α-actinin-2 (P20111)		21

exactly to the retrieved chick EST sequence. Nine peptides derived from this protein (equivalent to 146 residues) were sequenced by tandem MS covering the region mapped by the EST data but also the N-terminal region of the protein. From these data the assignment of this protein as the muscle isoform of phosphoglycerate mutase was unambiguous.

For band **b**, a detailed MALDI-TOF spectrum was obtained (Fig. 3*a*). Between 10 and 13 peptides above the 10% intensity threshold were matched to the muscle isoform of glycogen phosphorylase from six species. The observed mass on a gel of ~100 kDa also supported this identification as muscle glycogen phosphorylase from several species has a mass of approximately 97 kDa. Final confirmation of the identity of this protein came from tandem mass spectrometry of selected tryptic peptides (data not shown). A total of 16 tryptic peptides were sequenced, yielding 186 amino acid residues of sequenced protein, and confirmed the identity of this protein. The tandem MS data also highlighted a sequencing error in an EST clone where a putative deletion in the EST set was disproved by virtue of identification of that region of the polypeptide sequence.

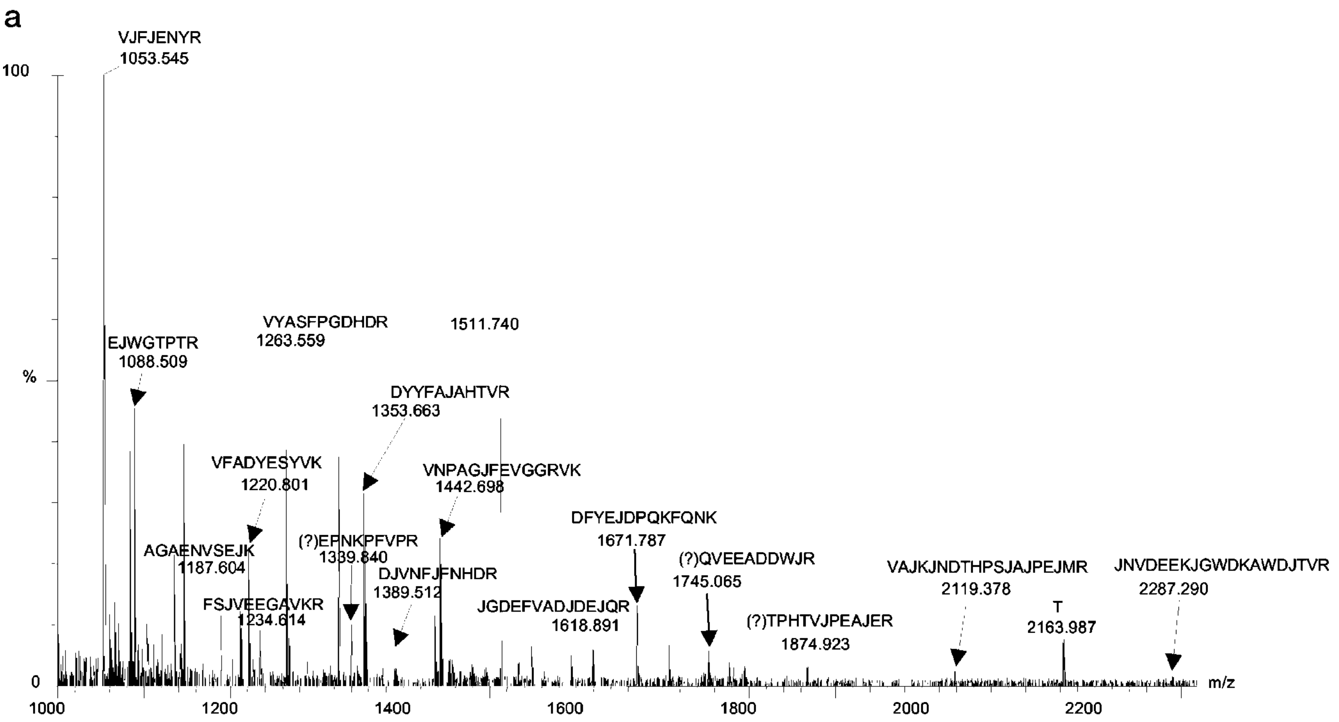
One particularly abundant protein, band **j**, was not readily identifiable by peptide mass fingerprinting. Eight peptides were sequenced by tandem mass spectrometry, yielding 111 amino acids of sequence information (Fig. 4). All of the peptides showed extensive sequence similarity to fructose-1,6-bisphosphate aldolase (aldolase) from several species. The muscle-specific isoform of this protein from chicken (Type A) has not been cloned or sequenced, but the sequences of both the liver (Type B) and brain forms (Type C) of chicken aldolase



```
ATHRLVMVRHGESSWNQENRFCGWFDAELSEKGAEEAKRGATAIKDAKIEFDICYTSVLKRAIRTLWTILDVTDQMWVPVV     PGM Rat (P16290, muscle)
AAYKLVLIRHGESAWNLENRFSGWYDADLSPAGHEEAKRGGQALRDAGYEFDIVFTSVQKRAIRTLWTVLDAIDQMWLPVV     PGM Rat (P25113, liver)
         HGESAWNJENRFCGWYDADJSPAGKV---            ----EFDJCFTSVQK                        This study

RTWRLNERHYGGLTGLNKAETAAKHGEEQVKIWRRSFDTPPPPMDEKHNYYASISKDRRYAGLKPEELPTCESLKDTIARA     PGM Rat (P16290, muscle)
RTWRLNERHYGGLTGLNKAETAAKHGEAQVKIWRRSYDVPPPPMEPDHPFYSNISKDRRYADLTEDQLPSCESLKDTIARA     PGM Rat (P25113, liver)
                        HGEAQVKIWRRSFDIPPPPMQSDHPFFSTISKDRRYADLTEDQLPTCESLKDTIARA     Chicken EST 6e6 (muscle)
        HYGAJTGJNKAETAAK           SFDJPPPPMQSDHPFFSTJSK                           A     This study

LPFWNEEIAPKIKAGKRVLIAAHGNSLRGIVKHLEGMSDQAIMELNLPTGIPIVYELNQELKPTKPMRFLGDEETVRKAME     PGM Rat (P16290, muscle)
LPFWNEEIVPQIKEGKGVLIPAHGNSYGGIVKHLEGLSEEAIMELNLPTGIPIVYELDKNLKPIKPMQSLGDEETVRKAME     PGM Rat (P25113, liver)
LPFWNEEIVPQITEGKRVLIACHGNSLRGIVKHLEGMSEEAIMELNLPTGIPIVYELDKNLKPIKPMQFLGDEETVRKAME     Chicken EST 6e6 (muscle)
JPFWNEEJVPQJK   RVJJAAHGNSJR   AREPGGSEEAJMEJNJPTGJPJVYEJDKNJKPJKPMQFJGDEEVTR        This study

AVAAQGKAK                                                                            PGM Rat (P16290, muscle)
AVAAQGKVKK                                                                           PGM Rat (P25113, liver)
AVAAQGKVKK                                                                           Chicken EST 6e6 (muscle)
                                                                                     This study
```

FIG. 2. **Identification of band n as phosphoglycerate mutase (*PGM*).** The protein in band and spot **n** was subjected to in-gel digestion with trypsin, and the masses of the tryptic peptides were determined by MALDI-TOF mass spectrometry (*panel a*). Some of these peptides were also analyzed by tandem mass spectrometry (ESI-Q-TOF) to derive new sequence information. Four sets of sequence data are aligned: the sequences of the same enzyme from rat liver and muscle, a chicken EST sequence, and the sequences obtained by ESI-Q-TOF mass spectrometry (*panel b*). These sequences are used to label the appropriate peaks in the MALDI-TOF spectrum (*panel a*). The *letter J* in the new peptide sequence data is used to indicate that the amino acid was one of an isomeric leucine/isoleucine pair; – indicates that a residue could not be determined. The *black bars* indicate peptides that were observed in the MALDI-TOF data. A *dotted black bar* indicates that a peptide was seen in both the ESI-MS and MALDI-TOF spectra but that the peptide could not be fully defined because of the incomplete sequence data.

are known. These were used to search the chick EST database to retrieve aldolase EST sequences from skeletal muscle. Two overlapping clones (ChEST 188h15 and ChEST 190j22) were recovered to give an overlapping partial protein sequence 275 amino acids in length. A further two partial sequences of chick aldolase A were recovered from the NCBI



b

```
MSRPLSDQEKRKQISVRGLAGVENVTELKKNFNRHLHFTLVKDRNVATPRDYYFALAHTVRDHLVGRWIRQQHYYEKDPKRIYYLSLEFYMGRTLQNTMV hGPB(P11217)
MSRPLSDHEKRKQISVRGLAGAENVSELKKGFNRHLHFTLVKDRNVATSRDYYMALAHTVRDHLVGRWLRTQQHYYEKDPKRIYYLSLEFMGRTLQNTMI Chicken muscle EST 1
                 --AGAENVSEJK                     DYYFAJAHTVR                                        This study
                 ■ ■ ■ ■ ■ ■                      ━━━━━━━━━━━

NLALENACDEATYQLGLDMEELEEIEEDAGLGNGGLGRLAACFLDSMATLGLAAYGYGIRYEFGIFNQKISGGWQMEEADDWLRYGNPWEKARPEFTLPV hGPB(P11217)
NLGLQSSCDEAVYQLGLDMEELQEIEEDAGLGNGGLGRLAACFLDSMATXGLAAY                                              Chicken muscle EST 1
                                                                     -----QVEEADDWJR                 This study
                                                                     ■ ■ ■ ■ ■ ■ ■ ■

HFYGHVEHTSQGAKWVDTQVVLAMPYDTPVPGYRNNVVNTMRLWSAKAPNDFNLKDFNVGGYIQAVLDRNLAENISRVLYPNDNFFEGKELRLKQEYFVV hGPB(P11217)
                                                                                                     Chicken muscle EST
                                                                                                     This study

AATLQDIIRRFKSSKFGCRDPVRTNFDAFPDKVAIQLNDTHPSLAIPELMRILVDLERMDWDKAWDVTVRTCAYTNHTVLPEALERWPVHLLETLLPRHL hGPB(P11217)
                                                LMRILVDEEKLGWDKAXDITVRTXAYSNHTVLPEAL**************** Chicken muscle EST 2
                                VAJKJNDTHPSJAJPEJMRJNVDEEKJGWDQAWDJTVR    TPHTVJPEAJER               This study
                                ━━━━━━━━━━━━━━━━━━ ━━━━━━━━━━━━━━━━━━━    ━━━━━━━━━━━━

QIIYEINQRFLNRVAAAFPGDVDRLRRMSLVEEGAVKRINMAHLCIAGSHAVNGVARIHSEILKKTIFKDFYELEPHKFQNKTNGITPRRWLVLCNPGLA hGPB(P11217)
*********************************GAVKRINMNHXXIAXAHAVNGVAQIHXD                                        Chicken muscle EST 2
                               EEGAVKRINMAHLCIAGAHAVNGVAQIHSDILKHSVFKDFYELDPQKFQNKTNGITPRRWLLLCNPGLA Chicken muscle EST 3
             VYASFPGDHDR   FSJVEEGAVKR                               DFYEJDPQKFQNK                   This study
             ━━━━━━━━━━━   ━━━━━━━━━━━                               ━━━━━━━━━━━━━

EVIAERIGEDFISDLDQLRKLLSFVDDEAFIRDVAKVKQENKLKFAAYLEREYKVHINPNSLFDIQVKRIHEYKRQLLNCLHVITLYNRIKREPNKFFVP hGPB(P11217)
EVIAERIGDEFVADLDQLQRLKALVDDDAFIRDVAKVKQENKAKFAAQLERDYGLRVNPAGLFVDQVKRIHEYKRQLLNCLHVITLYNRIKKEPNKPFVP Chicken muscle EST 3
                                       ENKAKFAAQLERDYGLRVNPAGLFDVQVKRIHEYKRQLLNCLHVITLYNRIKKEPNKPFVP Chicken muscle EST 4
      JGDEFVADJDEJQR                                    VNPAGJFDVQVK                      --EPNKPFVP This study
      ━━━━━━━━━━━━━━                                    ━━━━━━━━━━━━                      ■ ■ ■ ■ ■

RTVMIGGKAAPGYHMAKMIIRLVTAIGDVVNHDPAVGDRLRVIFLENYRVSLAEKVIPAADLSEQISTAGTEASGTGNMKFMLNGALTIGTMDGANVEMA hGPB(P11217)
RTVMIGGKAAPGYHMAKLIIKLITSIGDVVNHDPAVG                                                                Chicken muscle EST 4
RTVMIGGKAAPGYHMAKLIIKLITSIGDVVNHDPAVGDRLKVIFLENYRVSLAEKVIPAADLSEQISTAGTEASGTGNMKFMLNGALTIGTMDGANVEMA Chicken muscle EST 5
R                                        VJFJENYR                                                    This study
■                                        ━━━━━━━━

EEAGEENFFIFGMRVEDVDKLDQRGYNAQEYYDRIPELRQVIEQLSSGFFSPKQPDLFKDIVNMLMHHDRFKVFADYEDYIKCQEKVSALYKNPREWTRM hGPB(P11217)
EEAGEENLFIFGMRVEDVERLXRQGYCARDYYERLPELRTAVDQLXSGFFSPRQPDX                                            Chicken muscle EST 5
                                                           DJVNFJFNHDR  VFADYESYVK                   This study
                                                           ━━━━━━━━━━━  ━━━━━━━━━━

VIRNIATSGKFSSDRTIAQYAREIWGVEPSRQRLPAPDEAI                                                            hGPB(P11217)
                      EJWGTEPTR                                                                      This study
                      ━━━━━━━━━
```

FIG. 3. **Identification of band b as glycogen phosphorylase (*GPB*).** The protein in band **b** was subjected to in-gel digestion with trypsin, and the masses of the tryptic peptides were determined by MALDI-TOF mass spectrometry (*panel a*). Trypsin autolysis products are designated by the *letter T*. Some of these peptides were also analyzed by tandem mass spectrometry (ESI-Q-TOF) to derive new sequence information.

protein database (accession numbers I51292 and AAA99864). The alignment of the tandem MS data and the matching of specific peptides in the MALDI-TOF spectrum to EST-derived sequences clearly identifies this protein as aldolase A, although there was some evidence of minor sequencing errors or polymorphic variation. Two further peptides aligned with chick aldolases B and C at the N-terminal end of the protein, but there is no information on this region of the chicken aldolase A sequence to confirm these matches.

Analysis of Intact Proteins by ESI-MS—Having identified all of the major components of this mixture, we then focused on the behavior of the intact proteins in the mass spectrometer. In contrast to the bottom up strategy, a "top down" approach focuses on the intact mass of the protein as a key identifying feature, usually supported by confirmatory studies such as fragment ions (13–15). To test the applicability of this approach to a relatively complex mixture of proteins, we electrosprayed the chicken muscle soluble protein fraction directly into the source of our Q-TOF-micro mass spectrometer. Our goal was to determine the intact protein masses of as many components of the mixture as was feasible. In the first experiment, the protein mixture was diluted 50-fold in 5% (v/v) acetonitrile, 0.2% (v/v) formic acid and electrosprayed directly into the source (Fig. 5*a*). As anticipated, the raw spectrum was very complex with multiple charge envelopes overlapping between *m/z* 600 and *m/z* 1800. Manual resolution of this spectrum into a true mass spectrum (displaying the analyzed masses of the proteins) was not feasible, and instead we used the MaxENT 1 maximum entropy deconvolution software to recover the true masses. An initial survey at 10 Da/channel was used to indicate the feasibility of the analysis. This was then refined to 1 Da/channel for more precise mass determination. The time taken to process the data was typically about 1 h on a Pentium 4 desktop computer running at 1.7 GHz. In all instances, the software converged to a solution and generated high quality true mass spectra.

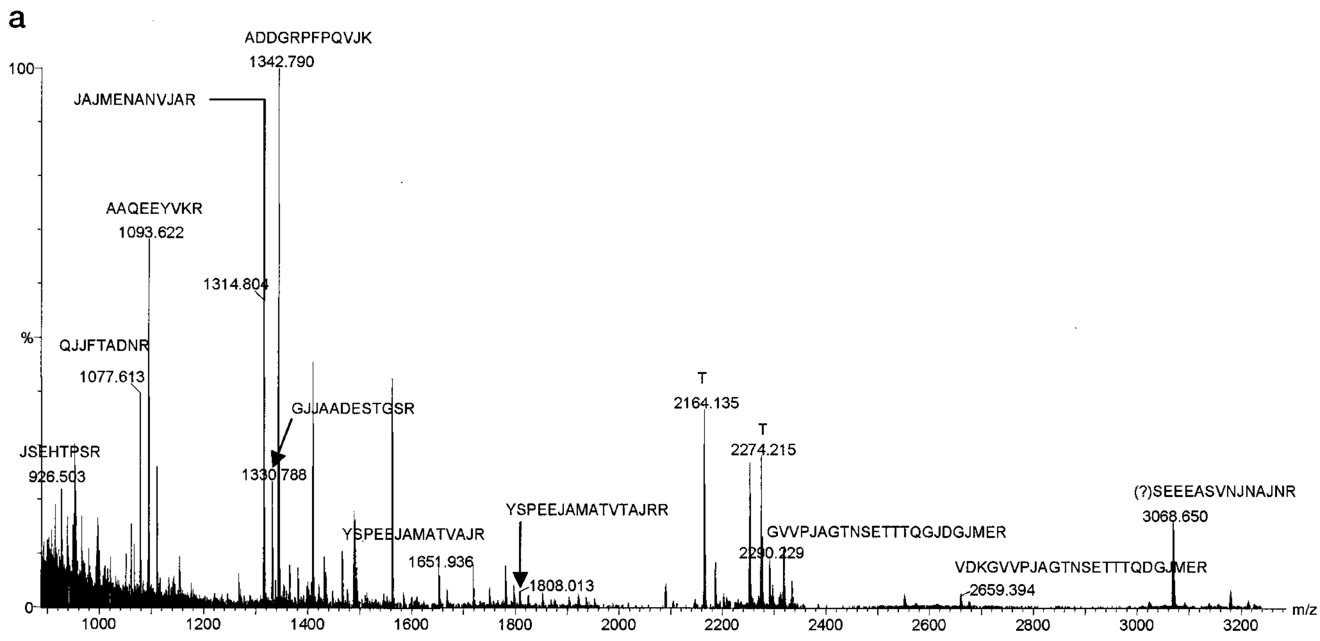
To eliminate the possibility of buffer-mediated suppression effects, the protein mixture was also concentrated and desalted on a 0.3- × 1-mm C4 reversed phase microcolumn. The protein mixture was diluted 50-fold in 5% (v/v) acetonitrile, 0.2% (v/v) formic acid and applied to the column. The column was then washed with the same buffer, and finally the mixture was desorbed from the column with 90% (v/v) acetonitrile, 0.2% (v/v) formic acid. Again a complex raw spectrum was obtained that, after processing, yielded a high quality true mass spectrum. The overall pattern of peaks obtained with the raw sample or with the sample preconcentrated and desalted on a C4 reversed phase trap was very similar (Fig. 5*b*). There was no evidence in either spectrum for harmonic artifacts that can be a feature of this type of analysis (16).

At least 20 discrete peaks resulted from this analysis. The masses of these proteins were compared with the proteins that had been identified by MALDI-TOF analysis and for which complete protein sequences were available (Table I) and with other chicken skeletal muscle proteins. The singularly most intense peak in the true mass spectrum (peak 6) yielded a mass of 35,573 Da. By comparison, the most abundant protein on a one-dimensional SDS-polyacrylamide gel of this protein mixture is glyceraldehyde-3-phosphate dehydrogenase, migrating at ~35 kDa on the gel. The full sequence of this protein is known, and the mass of this protein predicted from the cDNA-inferred sequence (NCBI accession number P00356) is 35,703 Da. Thus the difference between the observed and the theoretical masses is 131 Da. This is equivalent to the mass of a methionine residue. The database entry includes the N-terminal initiator methionine residue, which is often lost as a co-translational or post-translational event (9, 17). Indeed many of the database entries for this protein omit the N-terminal methionine residue in the sequence file (see for example, Swiss-Prot accession number P00356) in which case the observed mass of protein matches exactly that of the predicted mass. The exact correspondence between the observed mass and the predicted mass further suggests that sequence variants (conflicts) described in the database entry P00356 are either specific to another strain of chicken or are sequencing artifacts.

It was possible to assign a further 10 proteins to mass peaks in the deconvoluted electrospray mass spectrum (Table II). In some instances, confirmation of the assignment was made by a repeat analysis of proteins fractionated further by native state column chromatography (results not shown). Peak 2 corresponds to adenylate kinase, modified by removal of the N-terminal methionyl residue with subsequent N-terminal acetylation. Loss of the initiating methionine is often followed by N-terminal acetylation (17). Acetylation would add 42 Da to the observed mass; hence the loss of methionine followed by addition of an acetyl group would result in a loss of mass of 89 Da. Peak 2 has a measured mass of 21,597 Da, which is exactly equivalent to the mass of (adenylate kinase − methionine + acetyl group).

Peak 3 is exactly coincident with the mass of triose-phosphate isomerase lacking the N-terminal methionine. Peak 5 was assigned to lactate dehydrogenase. In this instance, the difference between the predicted and observed masses was

The sequence alignment includes human muscle glycogen phosphorylase, several chicken EST sequences for the muscle form of the enzyme, and the sequences obtained by ESI-Q-TOF mass spectrometry (*panel b*). These sequences are used to label the appropriate peaks in the MALDI-TOF spectrum (*panel a*). The *letter J* in the new peptide sequence data is used to indicate that the amino acid was one of an isomeric leucine/isoleucine pair; − indicates that a residue could not be determined. The *black bars* indicate peptides that were observed in the MALDI-TOF data. A *dotted back bar* indicates that a peptide was seen in both the ESI-MS and MALDI-TOF spectra but that the peptide could not be fully defined because of the incomplete sequence data (see text).



b

```
MTHQFPALSPEQKKALSDIAQRIVASGKGILAADESVGTMGNRLQRINVENTEENRRAFREILFSSDASISKSIGGVILF  Ch. Aldolase B (P07341)
------ALTAEQKKELSDIALAIVAPGKGILAADESVGSMAKRLNQIGVENTEENRRLYRQILFSADSRVKKCIGGVIFF  Ch. Aldolase C (P53449)
--------------------------------------------------------------------------------  Ch. Aldolase A EST
                            GJJAADESTGSJAK                  QJJFTADNR             This study
                            ━━━━━━━━━━━━━━                  ━━━━━━━━━

HETLYQKDSSGKPFPAIIKEKGMVVGIKLDAGTAPLAGTNGETTIQGLDKLAERCAQYKKDGADFGKWRAVLKISSTTPS  Ch. Aldolase B (P07341)
HETMYQKADDGTPFVQMIKDKGIVVGIKVDKGVVPLAGTDGETTTQGLDGLSERCAQYKKDGADFAKWRCVLKISDNTPS  Ch. Aldolase C (P53449)
---------DGRPFPQVIKAKGGVVGIKVDKGVVPLAGTNGETTTQGLDGLMERCAQYKKDGADFAKWRCVLKISEHTPS  Ch. Aldolase A EST
       ADDGRPFPQVJK         VDKGVVPJAGTNSETTTQGJDGJMER                            This study
       ━━━━━━━━━━━━         ━━━━━━━━━━━━━━━━━━━━━━━━━━                    ━━━━━━━━

QLAIQENANTLARYASICQQNGLVPIVEPEVLPDGDHDLQRCQYVTEKVLAAVYKALNDHHVYLEGTLLKPNMVTAGHSC  Ch. Aldolase B (P07341)
ALAIMENANVLARYASICQQNGIVPIVEPEILPDGDHDLKRCQYVTEKVLAAVYKALSDHHVYLEGTLLKPNMVTPGHSC  Ch. Aldolase C (P53449)
RLAIMENANVLARYASICQQNGIVPIVEPEILPDGDHDLKRCQYVTEKVLAAVYKALSDHHIYLEGTLLKPNMVTAGHSC  Ch. Aldolase A EST
-------------------------------------------------------------------LLKPNMVTPGHSC  Ch. Aldolase A (I51292)
 JAJMENANVJAR                                                                     This study
 ━━━━━━━━━━━━

PKKYTPQDVAVATVTTLLRTVPAAVPGICFLSGGQSEEEASLNLNAMNQSPLPKPWKLTFSYGRALQASALAAWLGKSEN  Ch. Aldolase B (P07341)
PTKYSPEEIAMATVSPLRRTVPPAVPGVTFLSGGQSEEEASINLNAINTCPLFAPWALTFSYGRPLQASALSAWRGQRDN  Ch. Aldolase C (P53449)
TKKYS---------------------------------------------------------------------------  Ch. Aldolase A EST
PTKYSPEEIAMATVTALRRTVPPAVPGVTFLSGGQSEEEASINLNAINTCPLVRPWALTFSYGRALQASALSAWRGQRDN  Ch. Aldolase A (I51292)
   YSPEEJAMATVTAJR ----------------SEEEASVNJNAJNR                                  This study
   ━━━━━━━━━━━━━━━ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■

KKAAQEAFCKRAQINSLACRGQYVTSGKTDTAATQSLFTASYTY                                      Ch. Aldolase B (P07341)
ANAATEEFVKRAEVNWLAALGKYEGSGDDSGAAGQSLYVANHAY                                      Ch. Aldolase C (P53449)
--------------------------------------------                                      Ch. Aldolase A EST
ANAATEEFVKRAEVNGLAALGKYEGSGDDSGAAGQSLYVANHAY                                      Ch. Aldolase A (I51292)
  AAQEEYVKRALANSLACQGKYTPSGHAGAAASESLFISNHAY                                      Ch. Aldolase A (AAA99864)
  AAQEEYVKR                                                                       This study
  ━━━━━━━━━
```

FIG. 4. **Identification of band j as fructose-1,6-bisphosphate aldolase.** The protein in band and spot **j** was subjected to in-gel digestion with trypsin, and the masses of the tryptic peptides were determined by MALDI-TOF mass spectrometry (*panel a*). Trypsin autolysis products are designated by the *letter T*. Some of these peptides were also analyzed by tandem mass spectrometry (ESI-Q-TOF) to derive new sequence

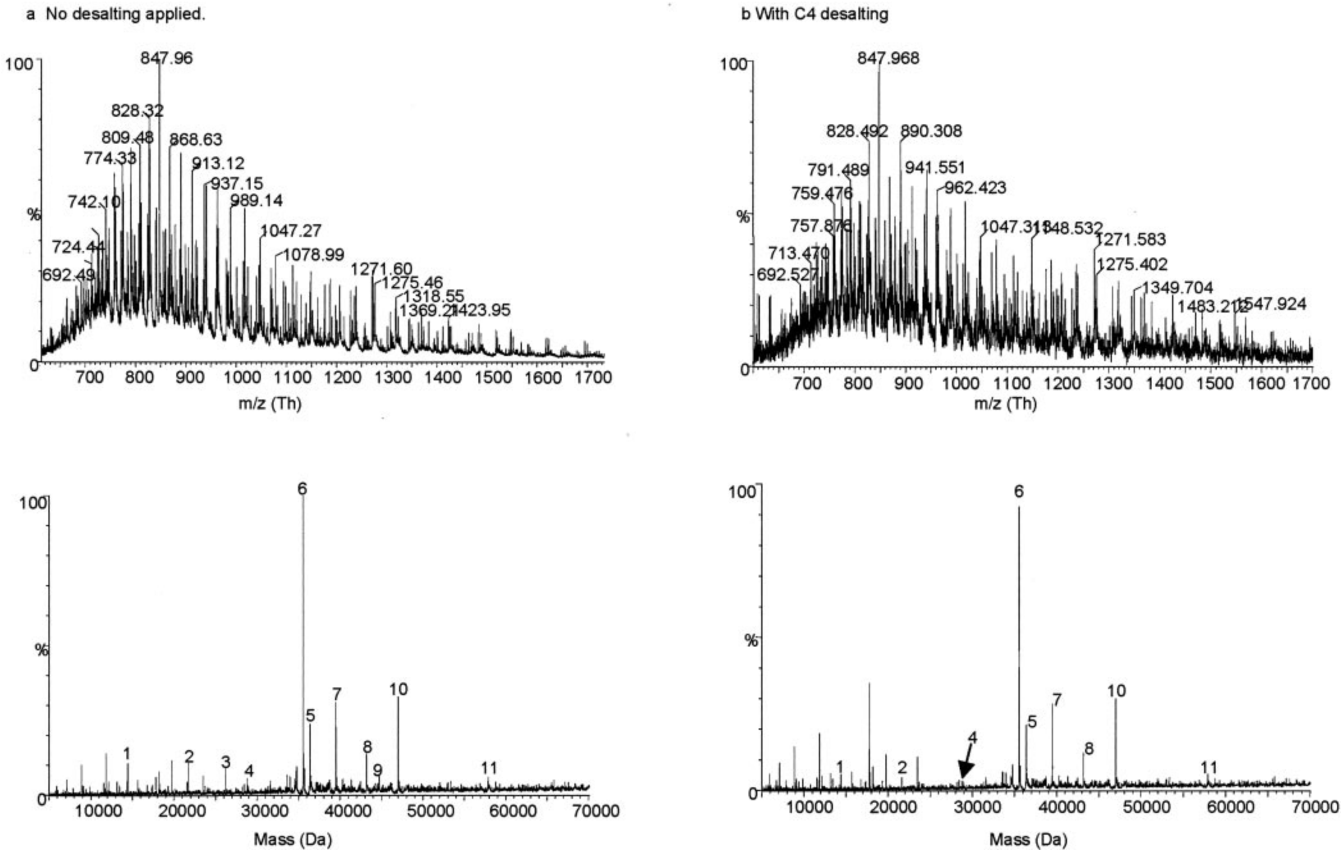


FIG. 5. **Mass determination for soluble skeletal muscle proteins.** The soluble protein extract from chicken skeletal muscle was diluted 50-fold in formic acid/acetonitrile and electrosprayed directly into the source of the mass spectrometry. The *left-hand panels* show the multiple charge state envelope derived from the complex mixture, after spectrum averaging, background subtraction, and smoothing (*top*), and the MaxENT 1-deconvoluted true mass spectrum (*bottom*). The mass peaks marked *1–11* are discussed further in the text. In a separate series of experiments (*right-hand panels*), the protein mixture was fully desalted on an in-line C4 reversed phase peptide trap prior to mass spectrometry. The multiple charge state envelope (*top*) and MaxENT 1-processed data (*bottom*) were treated as described above.

−101 Da. Additionally a single peptide in the MALDI-TOF tryptic map was 28 Da larger than anticipated. These two differences can be reconciled as loss of N-terminal methionine (−131 Da) and a putative mutation of a threonine residue to a glutamine residue (+28 Da), respectively. However, there are two threonine residues in this peptide, and discrimination between the two mutations is not possible in the absence of tandem MS data for this peptide. The mass of the protein in peak 8 is exactly coincident with the mass of creatine kinase lacking the N-terminal methionine. Peaks 9 and 11 cannot be precisely assigned on the basis of mass (although we have demonstrated the co-elution of proteins of these masses with bands on SDS-PAGE that yield MALDI-TOF data for the tryptic digest identifying the same protein). Although in most instances the loss of the methionyl residue was inferred from the mass measurement of the intact protein, in one case (peak 9), the des-Met N-terminal peptide of phosphoglycerate kinase was observed in the MALI-TOF spectrum of the digest. Whereas the N-terminal peptide (T1+T2) would be expected at $[M + H]^+$ of 1248.67 Da with the methionine intact, this peptide was absent from the analyte, but a new peptide was evident at $[M + H]^+ = 1117.63$ (corresponding to T1+T2-Met). Previous studies on the purified proteins confirm the loss of the N-terminal methionine, but the residual mass differences are unexplained. Further work will be required to identify the precise polymorphisms that elicit these mass changes.

Peak 10 has an observed mass of 47,023 Da. The mass of

information. The sequence alignment includes the B and C (non-muscle) isoforms of this enzyme from chicken, several chicken EST sequences for the muscle form (aldolase A), and the sequences obtained by ESI-Q-TOF mass spectrometry (*panel b*). These sequences are used to label the appropriate peaks in the MALDI-TOF spectrum (*panel a*). The *letter J* in the new peptide sequence data is used to indicate that the amino acid was one of an isomeric leucine/isoleucine pair; − indicates that a residue could not be determined. The *black bars* indicate peptides that were observed in the MALDI-TOF data. A *dotted back bar* indicates that a peptide was seen in both the ESI-MS and MALDI-TOF spectra but that the peptide could not be fully defined because of the incomplete sequence data. *Ch.*, chicken.

TABLE II

Proteins identified using intact protein masses determined by ESI

Soluble proteins of chicken skeletal muscle were analyzed by electrospray ionization mass spectrometry and analyzed using the MaxENT 1 software. The masses of these proteins were obtained by a narrow range scan (typically 5000 Da) at 1 Da/channel after an initial low resolution scan (10 Da/channel, 10,000–70,000). For several mass peaks, it was possible to reconcile the measured masses with the masses of known chicken skeletal muscle proteins after adjustment for likely post-translational modifications and polymorphic variation that was confirmed by MALDI-TOF MS. For other proteins, the lack of a full-length sequence precluded prediction of the mass, or there were residual mass differences that are not immediately explained. ND, not determined; N/A, not applicable.

Peak	Band	MALDI ID	Accession no.	Predicted mass	Measured mass		Discrepancy	Comment	Residual mass
					Broad range	Narrow range			
				Da	*Da*		*Da*		*Da*
1	ND	Annexin I	Q92108	14,398	14,440	14,440	+2	No post-translational modifications	+2
2	**q**	Adenylate kinase	P05081	21,683	21,597	21,597	−86	Loss of N-terminal methionine and *N*-acetylation	+3
3	**p**	Triose-phosphate isomerase	P00940	26,620	26,489	26,489	−131	Loss of N-terminal methionine	0
4	**n**	Phosphoglycerate mutase	N/A	N/A	28,809	28,809			
5	**l**	Lactate dehydrogenase	P00340	36,514	36,413	36,413	−101	Loss of N-terminal methionine (−131 Da) T247E (+28 Da) or T261E (+28 Da)	+2
6	**k**	Glyceraldehyde-3-phosphate dehydrogenase	P00356	35,704	35,573	35,572	−132	Loss of N-terminal methionine	−1
7	**j**	Aldolase	N/A	N/A	39,505	39,500			
8	**i**	Creatine kinase	P00565	43,328	43,197	43,198	−131	Loss of N-terminal methionine	0
9	**h**	Phosphoglycerate kinase	P51903	44,716	44,578	44,576	−140	Loss of N-terminal methionine (−131 Da) Loss of 9 Da (unexplained)	−9
10	**g**	Enolase	P07322	47,196	47,024	47,023	−173	Loss of N-terminal methionine (−131 Da) and *N*-acetylation (+42 Da) Loss of 84 Da (unexplained)	−84
11	**f**	Pyruvate kinase	P00548	58,014	57,933	57,922	−92	Loss of N-terminal methionine (−131 Da) and *N*-acetylation (+42 Da)	−3

the protein and native state column chromatography followed by ESI-MS of the intact protein and MALDI-TOF MS of the digest (data not shown) provide convincing evidence that this protein is enolase. However, there are two sequence entries for chicken skeletal muscle enolase in the sequence databases, P07322 and JC4187 (18, 19). The sequence of the former was derived by Edman degradation and protein chemistry approaches, while the sequence of the second was derived by cDNA sequencing. There are significant numbers of sequence conflicts between the two entries. However, our MALDI-TOF data provide clear evidence that enolase in our samples is most similar to the entry P07322. The extent of coverage of the sequence does not allow explanation of the residual unexplained mass difference.

Although the MaxENT 1 mass spectrum is highly processed, we repeatedly observed a similar distribution of peak areas. We were interested to explore whether or not the peak heights could be correlated with relative abundances of the proteins determined independently. Accordingly we scanned a one-dimensional gel of the protein mixture at different loadings and quantified each protein by densitometry. When the band volumes were compared with the peak areas obtained by mass spectrometry the correlation was not really good enough to justify the use of mass spectrometric data for quantification of expression. Quantification will require the use of an internal standard, possibly by the use of a polymorphic variant of the protein of different mass (we have indeed identified such polymorphic variants here) or by the use of stable isotope-labeled internal standards (corresponding to the intact proteins or tryptic peptides acting as surrogates for quantification purposes[2]). These would need to be prepared by a chemical modification that gave an adequate mass separation or by labeling *in vivo* (20), which might be challenging with animal systems.

Skeletal muscle demonstrates a remarkable pattern of protein expression. Not only does 40% of the muscle protein comprise the contractile protein actin and myosin, but the soluble sarcoplasmic proteins also show a remarkable bias. The soluble proteins are predominantly glycolytic enzymes as might be expected in a muscle (pectoralis) that in this species comprises almost exclusively white, fast twitch fibers. This restricted subset of proteins dominates a one-dimensional or a two-dimensional gel and obscures the less abundant proteins that might be expected to be there in large numbers. However, in any analysis of muscle growth and protein turnover, it is these proteins together with the contractile proteins that comprise the bulk of the products of biosynthetic activity in the tissue. This study has demonstrated that a complex mixture of proteins can be analyzed by electrospray ionization mass spectrometer using a benchtop instrument of relatively

[2] D. Firth, J. Brookman, and S. J. Gaskell, in preparation.

modest performance. Complex mixtures of proteins with heavily interdigitated multiply charged envelopes can be readily deconvoluted using the maximum entropy software, and the masses of the proteins that are identified are measured within a few daltons of the predicted masses.

Previous studies on mass measurements of intact proteins have either concentrated on relatively simple mixtures (21–23) or have required off-line (15) or on-line (9, 24) separation of protein mixtures prior to mass measurement. Here we show that instruments of modest performance can be used to resolve relatively complex mixtures, which might diminish the need for at least one dimension of separation. Unlike bottom up approaches where the masses of tryptic peptides are concentrated over a relatively narrow range, a protein-focused top down approach has the added advantage of distributing the protein masses over a much broader mass range, enhancing identifiability. Such an approach is predicated on resolution of individual peaks in the multiply charged envelope, but as we have demonstrated, an instrument with a nominal resolution of 5000 full-width half-maximum performs well in this context.

The protein masses predicted from sequence databases must be modified to account for likely post-translational changes, notably the loss of the initiator methionine and blocking of the N-terminal amino group by acetylation. A peripheral outcome of this study is that we have defined the extent of these post-translational changes for seven of the proteins in the mixture, information that had not previously been known. We have also predicted the mature masses of two proteins (phosphoglycerate mutase and aldolase) and will await the publication of the full cDNA sequences with interest.

* This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 44-151-794-4312; Fax: 44-151-794-4243; E-mail: r.beynon@liv.ac.uk.

REFERENCES

1. Cottrell, J. S. (1994) Protein identification by peptide mass fingerprinting. *Pept. Res.* **7,** 115–124
2. James, P., Quadroni, M., Carafoli, E. & Gonnet, G. (1994) Protein identification in DNA databases by peptide mass fingerprinting. *Protein Sci.* **3,** 1347–1350
3. Pratt, J. M., Robertson, D. H., Gaskell, S. J., Riba-Garcia, I., Hubbard, S. J., Sidhu, K., Oliver, S. G., Butler, P., Hayes, A., Petty, J. & Beynon, R. J. (2002) Stable isotope labelling in vivo as an aid to protein identification in peptide mass fingerprinting. *Proteomics* **2,** 157–163
4. Mann, M. & Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66,** 4390–4399
5. Mann, M. (1996) A shortcut to interesting human genes: peptide sequence tags, expressed-sequence tags and computers. *Trends Biochem. Sci.* **21,** 494–495
6. Smith, R. D., Pasa-Tolic, L., Lipton, M. S., Jensen, P. K., Anderson, G. A., Shen, Y., Conrads, T. P., Udseth, H. R., Harkewicz, R., Belov, M. E., Masselon, C. & Veenstra, T. D. (2001) Rapid quantitative measurements of proteomes by Fourier transform ion cyclotron resonance mass spectrometry. *Electrophoresis* **22,** 1652–1668
7. Veenstra, T. D., Martinovic, S., Anderson, G. A., Pasa-Tolic, L. & Smith, R. D. (2000) Proteome analysis using selective incorporation of isotopically labeled amino acids. *J. Am. Soc. Mass Spectrom.* **11,** 78–82
8. Martinovic, S., Veenstra, T. D., Anderson, G. A., Pasa-Tolic, L. & Smith, R. D. (2002) Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *J. Mass Spectrom.* **37,** 99–107
9. Lee, S. W., Berger, S. J., Martinovic, S., Pasa-Tolic, L., Anderson, G. A., Shen, Y., Zhao, R. & Smith, R. D. (2002) Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 5942–5947
10. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567
11. Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E., Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. & Hubbard. S. J. (2002) A comprehensive collection of chicken cDNAs. *Curr. Biol.* **12,** 1965–1969
12. Kettman, J. R., Frey, J. R. & Lefkovits, I. (2001) Proteome, transcriptome and genome: top down or bottom up analysis? *Biomol. Eng.* **18,** 207–212
13. Reid, G. E. & McLuckey, S. A. (2002) 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **37,** 663–675
14. Stephenson, J. L., McLuckey, S. A., Reid, G. E., Wells, J. M. & Bundy, J. L. (2002) Ion/ion chemistry as a top-down approach for protein analysis. *Curr. Opin. Biotechnol.* **13,** 57–64
15. VerBerkmoes, N. C., Bundy, J. L., Hauser, L., Asano, K. G., Razumovskaya, J., Larimer, F. Hettich, R. L. & Stephenson, J. L. (2002) Integrating "top-down" and "bottom-up" mass spectrometric approaches for proteomic analysis of *Shewanella oneidensis*. *J. Proteome Res.* **1,** 239–252
16. Ferrige, A. G., Seddon, M. J. & Jarvis, S. (1991) Maximum-entropy deconvolution in electrospray mass-spectrometry. *Rapid Commun. Mass Spectrom.* **5,** 374–377
17. Polevoda, B. & Sherman, F. (2000) N-terminal acetylation of proteins. *J. Biol. Chem.* **275,** 36479–36482
18. Russell, G. A., Dunbar, B. & Fothergill-Gilmore, L. A. (1986) The complete amino acid sequence of chicken skeletal-muscle enolase. *Biochem. J.* **236,** 115–126
19. Tanaka, M., Maeda, K. & Nakashima, K. (1995) Chicken α-enolase but not β-enolase has a Src-dependent tyrosine-phosphorylation site: cDNA cloning and nucleotide sequence analysis. *J. Biochem.* **117,** 554–559
20. Pratt, J. M., Petty, J., Riba-Garcia, I., Robertson, D. H., Gaskell, S. J., Oliver, S. G. & Beynon, R. J. (2002) Dynamics of protein turnover, a missing dimension in proteomics. *Mol. Cell. Proteomics* **1,** 579–591
21. Robertson, D. H., Hurst, J. L., Bolgar, M. S., Gaskell, S. J. & Beynon, R. J. (1997) Molecular heterogeneity of urinary proteins in wild house mouse populations. *Rapid Commun. Mass Spectrom.* **11,** 786–790
22. Beynon, R. J., Veggerby, C., Payne, C. E., Robertson, D. H., Gaskell, S. J., Humphries, R. E. & Hurst, J. L. (2002) Polymorphism in major urinary proteins: molecular heterogeneity in a wild mouse population. *J. Chem. Ecol.* **28,** 1429–1446
23. Reid, G. E., Shang, H., Hogan, J. M., Lee, G. U. & McLuckey, S. A. (2002) Gas-phase concentration, purification, and identification of whole proteins from complex mixtures. *J. Am. Chem. Soc.* **124,** 7353–7362
24. Wang, H., Kachman, M. T., Schwartz, D. R., Cho, K. R. & Lubman, D. M. (2002) A protein molecular weight map of ES2 clear cell ovarian carcinoma cells using a two-dimensional liquid separations/mass mapping technique. *Electrophoresis* **23,** 3168–3181