Molecular dissection of membrane-transport proteins: mass spectrometry and sequence determination of the galactose–H^+ symport protein, GalP, of *Escherichia coli* and quantitative assay of the incorporation of [*ring*-2-^{13}C]histidine and $^{15}NH_3$

Henrietta VENTER[1], Alison E. ASHCROFT[1], Jeffrey N. KEEN, Peter J. F. HENDERSON and Richard B. HERBERT[2]

Astbury Centre for Structural Molecular Biology, School of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, U.K.

The molecular mass of the galactose–H^+ symport protein GalP, as its histidine-tagged derivative GalP(His)$_6$, has been determined by electrospray MS (ESI-MS) with an error of $< 0.02\,\%$. One methionine residue, predicted to be present from the DNA sequence, was deduced to be absent. This is a significant advance on the estimation of the molecular masses of membrane-transport proteins by SDS/PAGE, where there is a consistent underestimation of the true molecular mass due to anomalous electrophoretic migration. Addition of a size-exclusion chromatography step after Ni^{2+}-nitrilotriacetate affinity purification was essential to obtain GalP(His)$_6$ suitable for ESI-MS. Controlled trypsin, trypsin + chymotrypsin and CNBr digestion of the protein yielded peptide fragments suitable for ESI-MS and tandem MS analysis, and accurate mass determination of the derived fragments resulted in identification of $82\,\%$ of the GalP(His)$_6$ protein. Tandem MS analysis of selected peptides then afforded $49\,\%$ of the actual amino acid sequence of the protein; the absence of the N-terminal methionine was confirmed. Matrix-assisted laser-desorption ionization MS allowed identification of one peptide that was not detected by ESI-MS. All the protein/peptide mass and sequence determinations were in accord with the predictions of amino acid sequence deduced from the DNA sequence of the *galP* gene. [*ring*-2-^{13}C]Histidine was incorporated into GalP(His)$_6$ *in vivo*, and ESI-MS analysis enabled the measurement of a high ($80\,\%$) and specific incorporation of label into the histidine residues in the protein. MS could also be used to confirm the labelling of the protein by $^{15}NH_3$ ($93\,\%$ enrichment) and [^{19}F] tryptophan ($83\,\%$ enrichment). Such MS measurements will serve in the future analysis of the structures of membrane-transport proteins by NMR, and of their topology by indirect techniques.

Key words: CNBr proteolysis, glucose transporter, NMR, stable isotope labelling, trypsin proteolysis.

INTRODUCTION

Membrane-transport proteins play a crucial metabolic role in living cells by mediating the influx of nutrients and the efflux of toxins and other waste products [1]. A notably high proportion ($5–12\,\%$) of the open reading frames in known bacterial genomes encode for transport proteins [2], of which many are members of the evolutionarily related multi-facilitator superfamily (MFS) [3], found in organisms from cyanobacteria to humans. The general importance of these proteins is thus manifest. The galactose–H^+ symport protein, GalP, of *Escherichia coli* is one such transport protein, which serves as an ideal paradigm for the MFS for the following reasons: GalP shows $22–64\,\%$ amino acid sequence identity with numerous MFS proteins found in, amongst others, bacteria, fungi, protozoan parasites, algae, plants and mammals; in particular, GalP has very similar properties of substrate specificity and inhibitor susceptibility to the human glucose transport protein, GLUT1 [4]; also, very substantial quantities of GalP are available through the facility of overexpression and purification of its active hexa-histidine-tagged derivative, GalP(His)$_6$, in *Escherichia coli* [5].

MFS transport proteins are generally deduced to have, as a dominant motif, twelve membrane-spanning α-helices made up of hydrophobic amino acid residues [6]. However, with a few exceptions, e.g. [7], little three-dimensional structural information has been obtained. Although substantial quantities of bacterial membrane proteins may be produced by cloning and overexpression (GalP can be overexpressed in *Escherichia coli* to the extent of $50–60\,\%$ of total membrane protein [5]), these proteins have so far generally resisted attempts to grow crystals suitable for X-ray analysis or two-dimensional electron crystallography (but see [7–9]). Therefore, driven by the widespread occurrence and metabolic importance of these proteins, other approaches to structure determination are being pioneered, including solid-state NMR spectroscopy [10–13], MS [14–16] and computerized prediction methods [17].

The amino acid sequences predicted from gene sequencing may be inaccurate because of sequencing error, RNA editing and a number of modifications that may arise post-translationally. Also, the estimation of molecular mass by SDS/PAGE is in error because of the anomalous migration of MFS proteins [18]. Thus an accurate determination of the molecular mass of the protein and a confirmation of the amino acid sequence is an essential adjunct to further structural work. We have developed a protein-purification procedure and have established, by means of electrospray ionization (ESI)-MS, the accurate molecular mass of GalP and detected the absence of a single residue. In addition, by measuring the accurate masses of protein fragments by MS following the cleavage of GalP(His)$_6$ with trypsin, trypsin + chymotrypsin and CNBr, we have identified $82\,\%$ of this protein,

Abbreviations used: ESI, electrospray ionization; GalP, galactose–H^+ symport protein; MALDI, matrix-assisted laser-desorption ionization; MS/MS, tandem MS; SEC, size-exclusion chromatography; MFS, multi-facilitator superfamily; DDM, n-dodecyl-β-D-maltoside; TCA, trichloroacetic acid.

[1] These authors share equal seniority as first author.

[2] To whom correspondence should be addressed (e-mail r.b.herbert@chem.leeds.ac.uk).

and by use of tandem MS (MS/MS) we have confirmed 49 % of the specific amino acid sequence in the protein and identified the loss of the N-terminal methionine. Finally, we have used MS to establish the high and specific incorporation (80 % enriched) of [*ring*-2-^{13}C]histidine (99 % ^{13}C), in *E. coli*, into the histidine residues of GalP(His)$_6$. In a complementary way the essentially complete enrichment ($\approx$ 93 %) by $^{15}NH_3$ (98 % enriched) was observed of the nitrogen atoms in the protein, excluding those in the histidine residues.

MATERIALS AND METHODS

Chemicals

[*ring*-2-^{13}C]-L-Histidine, $^{15}NH_4Cl$ and [U-^{13}C]glucose were purchased from Cambridge Isotope Laboratories (Andover, MA, U.S.A.), and n-dodecyl-β-D-maltoside (DDM) was from Melford Laboratories (Ipswich, Suffolk, U.K.). All other chemicals were from Sigma (Poole, Dorset, U.K.).

Growth of organism

Overexpression of GalP(His)$_6$ was achieved by growing *E. coli* strain JM1100 [genotype *HfrC* Δ(*his-gnd*) *thyA galK galP mglP ptsF ptsG ptsM HfrC* Δ(*his-gnd*) containing plasmid pPER3-His] on 25 l of minimal salts medium [19], supplemented with glucose (27 mM), L-histidine (80 mg·l^{-1}), thymine (20 mg·l^{-1}) and 2TY (10 g/l bacto-tryptone, 10 g/l yeast extract and 5 g/l NaCl, pH 7.0; 30 ml·l^{-1}) and containing tetracycline (15 μg·ml^{-1}). *E. coli* strain JM1100 is an auxotroph that requires histidine and thymine for growth [19], so histidine labelling was effected by substitution of the unlabelled compound in the medium with [*ring*-2-^{13}C]L-histidine (10 mg·l^{-1}). The 2TY was then omitted from the medium to minimize dilution of the labelled amino acid. [U-^{15}N]GalP(His)$_6$ was prepared by growing *E. coli* strain JM1100 (pPER3-His) on 10 l of the minimal salts medium as for unlabelled protein. Again the 2TY was omitted from the growth medium and in this case the 42 mM unlabelled NH_4Cl was substituted by 12.5 mM $^{15}NH_4Cl$.

Membrane preparation and protein purification

Harvested cells were disrupted by explosive decompression using a French Press and the inner membrane vesicles were obtained by separation on a sucrose density gradient [18]. Membranes were solubilized in DDM and the GalP(His)$_6$ protein was purified on Ni^{2+}-nitrilotriacetate–agarose (Qiagen, Chatsworth, CA, U.S.A.) as described previously [18]. The purified protein was concentrated and washed after purification to remove imidazole, NaCl, etc., by application to a 2 ml Centricon filtration unit (Amicon) with a cut-off of 100 kDa and washing several times with 10 mM Hepes buffer, pH 7.9. Filtration was aided by centrifugation (180 $\boldsymbol{g}_{av}$, 25 min, 10 °C). Although the molecular mass of GalP(His)$_6$ is $\approx$ 51 kDa, the size of the GalP(His)$_6$–detergent micellar complex was above 100 kDa and no GalP-(His)$_6$ was lost in the filtrate. The use of a 100 kDa cut-off unit was preferred to a cut-off of 50 kDa, because empty DDM micelles also concentrate in the latter, which may lead to protein precipitation. Protein concentrations were determined by a published method [20].

The bulk of detergent was removed and the sample concentrated by a solvent-three-phase-partitioning method as described in [21,22]. Briefly, HPLC-grade methanol (300 μl) was added to a 1–2 mg·ml^{-1} protein solution (100 μl). HPLC-grade chloroform (100 μl) was added and the solution mixed to give a single phase. Phase separation was initiated by addition of water (200 μl) and vigorous mixing. After centrifugation (10000 $\boldsymbol{g}$, 2 min), the phases separated out with the precipitated protein at the interphase. The aqueous methanol phase was removed and more methanol (300 μl) was added. After mixing, the protein remained insoluble in this single phase. The protein was obtained by centrifugation (10000 $\boldsymbol{g}$, 2 min) and the pellet was dried under a stream of nitrogen. The protein was dissolved in 90 % formic acid (50 μl) and subjected immediately to size-exclusion chromatography (SEC) to minimize formylation using Sephadex LH-20 (Pharmacia Biotech) prepared by hydroxypropylation of Sephadex G-25 gel-filtration resin. The resulting gel has both hydrophilic and hydrophobic properties and is thus stable in all solvents that are not strongly acidic (i.e. below pH 2) or strong oxidizing agents (from the manufacturer's technical manual). The parent Sephadex LH-20 has an exclusion limit of about 5000 Da. This limit is somewhat reduced by the hydroxyalkylation but varies with the swelling solvent. The large membrane proteins used in this study should elute in the void volume of the column.

Resin ($\approx$ 0.5 g/purification) was swollen overnight in the chromatography solvent (methanol/chloroform/1 % formic acid, 4:4:1, by vol.). A final volume of 2.5 ml of resin (wet volume) was usually used in a home-made glass column (4 mm $\times$ 100 mm). The protein sample (50 μl) was pipetted on to the resin and eluted subsequently with the chromatography solvent when fractions of 0.25 ml were collected. These fractions were concentrated under a stream of nitrogen to a volume of 50 μl. Small volumes of each fraction (2 $\times$ 1 μl) were spotted on 0.45 μm HAWP filters (Millipore) and allowed to dry and assayed for protein [20]. After qualitative inspection of the filters, formic acid (5 μl) was added to a 5 μl volume of each fraction containing protein before ESI-MS was performed.

Enzymic digestion of GalP(His)$_6$

It is generally necessary to cleave a protein into small fragments of about $\leqslant$ 2500 Da in order to obtain sequence information with MS/MS. Both enzymic [23] and chemical cleavage [24] of GalP(His)$_6$ were used in this study.

Digestions were performed on 100–200 μl solutions containing 2 mg·ml^{-1} purified protein in 10 mM Hepes buffer at the optimum pH for the protease used. For digestion with trypsin, the protein was diluted in 10 mM Hepes buffer, pH 7.9/2 mM $CaCl_2$. The $CaCl_2$ stabilized the trypsin over the long incubation periods used. Trypsin digests were carried out in the presence of either 1 or 5 M urea (1 M urea gave less complete digestion, but the resulting peptides were easier to sequence than those resulting from digestion in 5 M urea). SDS (0.1 %) was added to the protein solution prior to digestion with chymotrypsin. Enzymes were added from 1 mg·ml^{-1} stock solutions to give a final enzyme/protein ratio of 1:50. The enzymic cleavage was allowed to proceed at 37 °C (65 °C for thermolysin). Samples were taken at regular intervals and the progress of the reaction was monitored by SDS/PAGE. Digestion was terminated after 18 h, and the resulting peptides were freed of urea and other contaminating material (see below) before ESI-MS was performed.

In order to obtain even smaller peptides, double digests were performed. Several fragments generated by trypsin in the C-terminal half of GalP(His)$_6$, specifically from helices 9 to 12, were undetected. This part of the protein is of particular importance since it contains the predicted substrate/inhibitor-binding site [25–29]. Double digestion with chymotrypsin and trypsin would give peptides smaller than 3000 Da for this part of the protein. Purified GalP(His)$_6$ (200 μl of 5 mg·ml^{-1} solution) in 0.1 % SDS was first digested with chymotrypsin for 18 h at 37 °C. After that,

urea was added to the digestion mixture to a final concentration of 5 M. Trypsin was added and the digestion was allowed to proceed for another 8 h. Trypsin can act in the presence of 0.1 % SDS, whereas chymotrypsin is not active in such a high urea concentration [30], thus preventing the chymotrypsin from hydrolysing the trypsin. Hence, it was not necessary to remove the chymotrypsin from the mixture before trypsin was added.

Purification and preparation of peptide mixtures for MS

The peptides from single enzyme digestion mixtures were purified essentially as described in [31]. After digestion, the peptides were precipitated by adding 1.9 ml of a mixture of 10 ml of acetone, 1 ml of aqueous ammonia (25 %) and 100 mg of trichloroacetic acid (TCA; predissolved in 100 μl of water) to 100 μl of the digested protein solution. The sample was vortexed and then incubated on ice for 30 min. The peptide pellet was obtained by centrifugation (7000 ***g***, 3 min, 4 °C) and resuspended in ice-cold acetone. The peptides were again obtained by centrifugation. The resulting pellet was washed with hexane to remove residual detergent and lipids, and dried under a stream of nitrogen.

Double digestion resulted in a large number of small peptides. Smaller peptides were lost with the TCA/acetone precipitation procedure described above, and peptide yields were unacceptably low. Other means of purification were investigated, of which only the most successful one will be described here. Formic acid (90 %; 20 μl) was added to 20 μl of the digestion mixture. Solvent mixture (ethanol/formic acid, 7:3, v/v; 10 μl) was added. Peptides treated in this way were now completely soluble in the solvent mixture. The digestion mixture was applied to a Sephadex LH-20 column (4 mm × 100 mm) and SEC was performed as described for the intact protein, using ethanol/formic acid (7:3, v/v), as elution buffer. Peptide-containing fractions were subjected to ESI-MS.

Chemical cleavage

CNBr cleaves at the C-terminal side of methionine residues, converting the methionine into a mixture of homoserine and homoserine lactone [24]. Acidic conditions favour the lactone, whereas basic conditions favour the free acid [32].

CNBr (23 mg) was dissolved in 70 % formic acid (50 μl). Detergents and lipids were removed from the purified protein by the solvent-three-phase-partitioning method as described for the preparation of proteins before SEC. The precipitated protein was dissolved in 90 % formic acid (78 μl), and MilliQ water (22 μl) was added to give a final concentration of 2 mg·ml^{-1} protein in 70 % formic acid. CNBr was added to provide a 200-fold molar excess over the methionines. The vial was flushed with nitrogen, sealed and incubated in the dark at room temperature for 18 h. MilliQ water (20 vol.) was added and the water and volatile substances were removed by freeze-drying. The peptides were washed with MilliQ water (10 vol.) twice more before they were analysed by MS.

ESI-MS

Samples were analysed on a Q-TOF (Micromass UK, Manchester, U.K.) orthogonal acceleration quadrupole time-of-flight mass spectrometer equipped with nanoflow ESI. An aliquot (2–3 μl) of the protein- or peptide-containing fractions was used to fill a gold-plated, borosilicate nanospray capillary that was placed inside the ionization source of the mass spectrometer. Positive ionization was used for the sample analyses, with a capillary voltage of 900 V and a sampling cone voltage of 40 V. Nitrogen was employed as the drying gas. An external calibration was made using horse heart myoglobin (molecular mass, 16951.5 Da). For standard MS experiments, the quadrupole analyser was used in wide band-pass mode and the microchannel plate detector was set at 2700 V for protein analysis and at 2850 V for peptide analysis. Data were acquired over the appropriate m/z range and spectra processed using the MassLynx software supplied with the mass spectrometer. The m/z protein spectra were transposed on to a true molecular mass profile using Maximum Entropy processing [33] supplied in the MassLynx suite of programs.

The purified peptides, obtained from the various cleavage procedures, were dissolved in a minimum amount (20–50 μl) of 90 % formic acid, which was diluted with equal volumes of HPLC-grade water and methanol, before being subjected to ESI-MS. For MS/MS experiments the quadrupole (first) analyser was used to select sequentially the peaks of interest in the m/z spectrum, thus allowing only one particular precursor (or parent) mass ($\pm \approx 2$ amu) to proceed through the collision cell into the time-of-flight (second) analyser. Argon gas was admitted into the collision cell so that the pressure in that region increased by a factor of 10 and the collision energy was optimized for every precursor ion, generally varying from 10–80 eV. The resulting fragment ions were analysed and MS/MS spectra acquired over the appropriate m/z range. A 2–3 μl aliquot of sample was generally sufficient to last $\approx$ 40 min and thus to permit acquisition of spectra from all the components of interest. Data processing was achieved manually with the aid of the BioLynx programs of MassLynx.

Matrix-assisted laser-desorption ionization (MALDI)-MS

Samples were analysed on a TofSpec E (Micromass UK) MALDI–time-of-flight mass spectrometer. The GalP(His)$_6$ peptides obtained from CNBr cleavage were dissolved in 0.1 % DDM (100 μl), sonicated for 5 min to enhance solubilization and diluted between 2 and 25 times in 0.1 % trifluoroacetic acid. An aliquot (2 μl) was added to an equal volume of the matrix α-cyano-4-hydroxycinnamic acid (10 mg·ml^{-1} in ethanol/acetonitrile, 1:1, v/v) and applied to a sample disc. A nitrogen laser at 337 nm was used to desorb solute molecules from the sample disc; the coarse laser energy was set to 20 % with fine adjustment for each sample and 20–30 shots were accumulated for each spectrum. A voltage of 22.5 kV was established in the source region and the microchannel plate detector was set at 1800 V. The peptides were analysed in reflectron mode and spectra were calibrated externally with either corticotropin ('ACTH') or insulin. For protein samples 3,5-dimethoxy-4-hydroxycinnamic acid was used as a matrix.

RESULTS AND DISCUSSION

Determination of the accurate molecular mass for GalP(His)$_6$

Very few accurate molecular mass determinations of membrane proteins have been achieved so far, which is indicative of the difficulty in obtaining membrane-protein samples free of detergents and lipids, purified and solubilized to a level suitable for ESI-MS. Membrane proteins and their cleavage peptides are very hydrophobic and tend to stick almost irreversibly to C_{18} reversed-phase columns that are classically used in the separation of peptides by HPLC. Although smaller membrane proteins have been purified successfully on C_{18} resin [34], our experience with GalP(His)$_6$ was that it bound so strongly to this C_{18} resin that it could only be removed with 90 % formic acid. Recently, accurate molecular mass determination of larger membrane proteins has

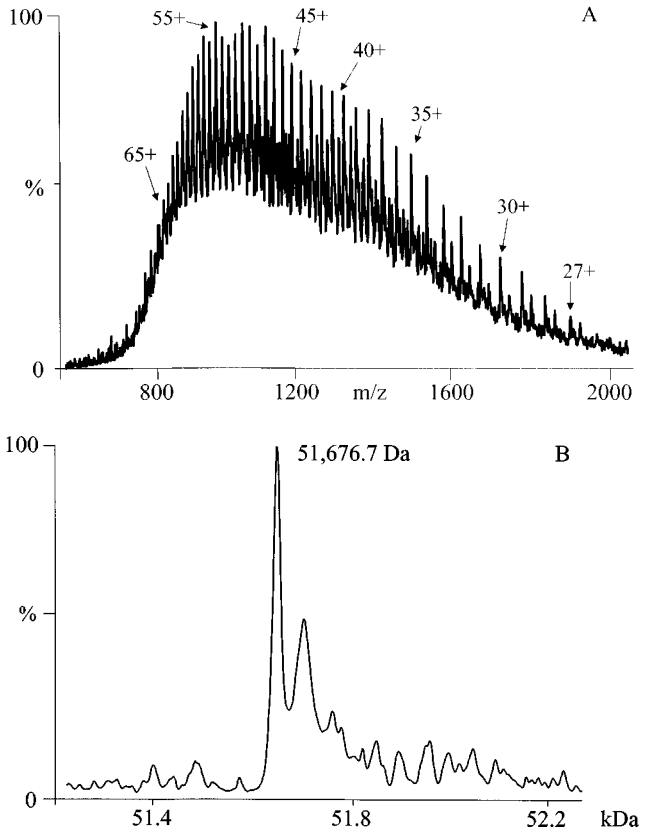

Figure 1 ESI-MS of purified GalP(His)$_6$

(**A**) GalP(His)$_6$, purified by Ni^{2+}-nitrilotriacetate chromatography and SEC on Sephadex LH-20 resin, displayed a series of multiply charged ions extending over an *m/z* range of 700–2000 carrying from 27 to 72 positive charges. (**B**) The molecular mass profile generated by Maximum Entropy processing [33] of the *m/z* spectrum shows essentially one major component of molecular mass 51 676.7 Da. This is in good agreement ($< 0.01\%$ error) with the calculated average molecular mass of 51 674.3 Da for GalP(His)$_6$ without the initiating methionine residue. The major component is accompanied by a less intense species at 51 724.6 Da, $\approx$ 48 Da higher in mass. It is possible that this species may be an adduct of the protein with a molecule of the co-solvent, formic acid (46 Da).

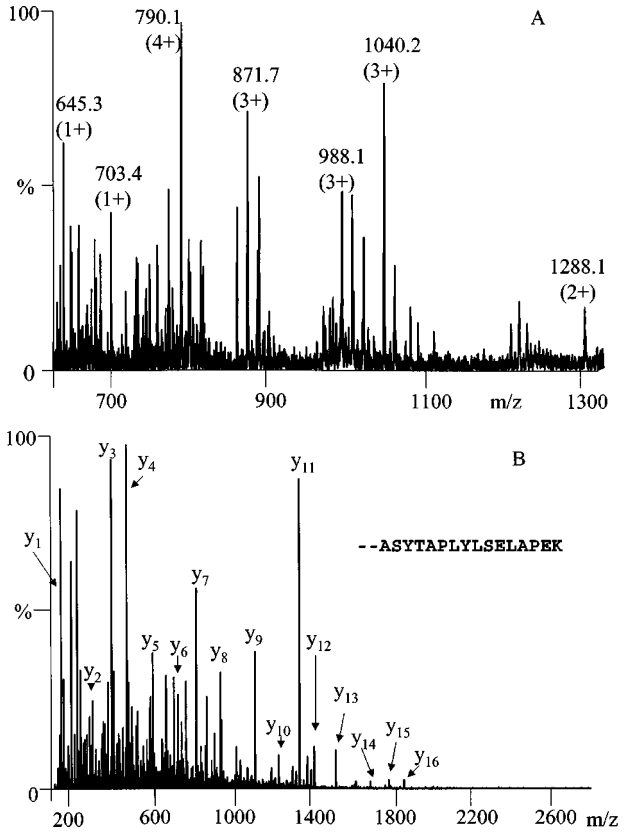

Figure 2 The ESI-MS *m/z* spectrum of the peptide mixture resulting from tryptic digestion of GalP(His)$_6$ in the presence of 5 M urea

(**A**) Peptides were obtained by TCA/acetone precipitation, and treated as described under the Materials and methods section for ESI-MS of peptides. The most significant components, the molecular masses of which have been calculated from the singly, and/or multiply charged ions in the *m/z* spectrum, are shown in Table 1; for clarity only a few have been labelled on the *m/z* spectrum. Many of the peptides were identified further by MS/MS sequencing. (**B**) As an example, the ESI-MS/MS product ion spectrum depicting the fragment ions arising directly from the collisionally activated decomposition of the $(M+3H)^{3+}$ ions centred at *m/z* 859.3 (measured mono-isotopic molecular mass 2573.34 Da; calculated mono-isotopic mass, 2573.44 Da; residues 113–137) is shown. A series of y′′ ions [35] was identified from which the partial sequence ASYTAPLYLSELAPEK was deduced. Low energy MS/MS fragmentation such as this does not distinguish between the isobaric residues leucine and isoleucine; hence L is used to represent either.

been reported that involves the use of SEC in an online HPLC ESI-MS set-up [14].

In our experiments aimed at the purification of membrane-transport proteins for MS measurements, simple SEC (Sephadex LH-20) using methanol/chloroform/1% aqueous formic acid (4:4:1, by vol.) as the chromatography solvent was found, in a pivotal final purification step, to yield pure GalP(His)$_6$ for which a molecular mass of 51676.7 Da was measured by ESI-MS (Figure 1). A high level of purity for the samples was confirmed by the quality of the spectra obtained (Figure 1). Following this we are successfully applying the use of Sephadex LH-20 in the purification of other, related proteins for MS. The presence of lipids/detergent and other possible adducts, e.g. Na^+, imidazole and buffer, in a sample of GalP(His)$_6$ that had not been purified by the size-exclusion step leads to an inaccurate mass determination obscured by multiple peaks.

The molecular mass predicted from the DNA sequence for GalP(His)$_6$ was 51805.5 Da, allowing for natural abundance of ^{13}C. The difference from the experimental value of 51676.7 Da corresponds most closely to the absence of a methionine residue (predicted mass, 51674.3 Da). This was confirmed when the N-terminal methionine was shown to be missing both by conventional Edman degradation and by ESI-MS of the CNBr-cleaved protein (see below). An average of 51680.01 ± 3.35 Da ($\pm$S.D.) was obtained for the molecular mass from five estimations on at least three separate purifications of different protein samples, so the difference between the experimental and predicted molecular masses is less than 0.02%. The value of using MS to detect even small changes from the predicted sequence of these proteins is thus demonstrated clearly.

Proteolytic digestion of GalP(His)$_6$

Not unexpectedly, because of the general intractability of membrane-transport proteins, it proved difficult to achieve satisfactory hydrolysis of GalP to obtain peptidic fragments suitable for MS analysis.

The heat-stable protease thermolysin [23] initially appeared to be a suitable candidate for hydrolysis of GalP, because it retains activity at 65 °C [30], a temperature at which membrane-transport

Table 1 Summary of ESI-MS data for trypsin-digested GalP(His)$_6$

Observed mass (Da)	Calculated mono-isotopic mass (Da)	Residues	Verified by partial MS/MS sequence
416.2*	416.24	227–229	Yes
489.21*	489.22	206–209	Yes
504.2	504.25	218–222	Yes
531.24*	531.27	226–229	Yes
553.27*	553.3	446–449	No
588.28*	588.29	205–209	Yes
644.32*†	644.35	225–229	Yes
682.3*	682.34	445–449	Yes
702.36	702.39	230–235	Yes
720.29	720.33	216–222	Yes
735.34	735.36	204–209	No
765.32	765.34	244–249	No
773.36	773.39	224–229	Yes
882.42*	882.46	443–449	No
929.48	929.49	223–229	No
1118.56	1118.58	441–449	Yes
1161.56*	1161.63	128–137	No
1243.52*	1243.72	447–456	No
1462.58	1462.64	459–470	No
2238.88*	2238.15	231–249	No
2573.34	2573.44	113–137	Yes
2610.72*	2610.54	294–317	Yes
2958.39	2960.54	251–276	Yes
3036.36*	3036.55	226–250	No
3116.43	3116.64	250–276	Yes
3154.12*	3155.82	108–137	Yes

* The N-termini of these peptides arose from digestion at residues other than R, K and H, i.e. they did not accord with the expected specificity of trypsin, although the C-termini ended with R or K, as expected.

† This peptide resulted from non-specific cleavage by trypsin at the C-terminal side of a glutamate residue, and its identity was determined by MS/MS sequencing.

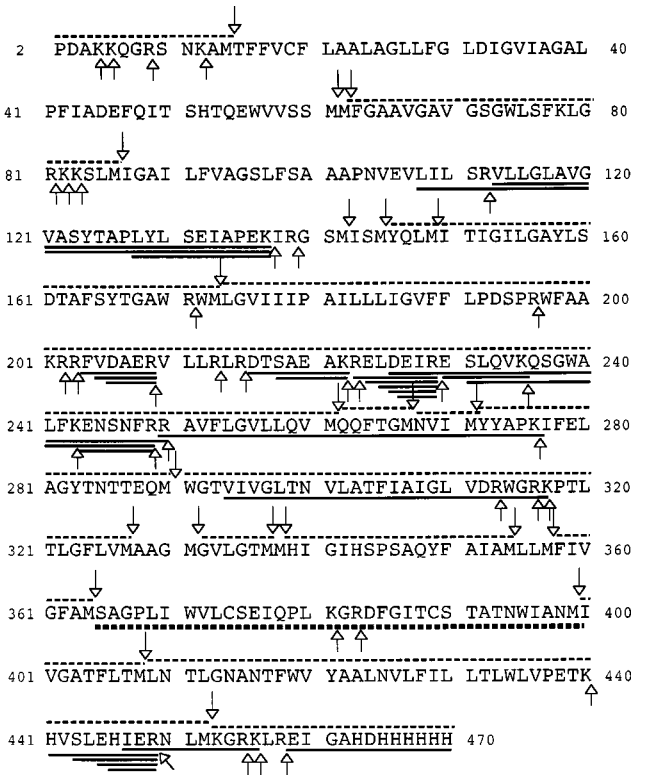

Figure 3 Amino acid sequence of GalP(His)$_6$

The absent initiating methionine is omitted. The amino acids that were identified by ESI-MS/MS sequencing of tryptic peptides are indicated by solid underlining. Potential CNBr cleavage sites are indicated by a downward arrow over the sequence, and potential trypsin cleavage sites by an upward arrow below the sequence. The peptides from CNBr cleavage that could be identified from their masses in ESI-MS are indicated by a dashed overline. The peptide identified from MALDI-MS is indicated with a dashed bold underline.

proteins, including GalP(His)$_6$, lose a significant amount of their secondary structure [18]. However, no GalP(His)$_6$ peptides suitable for MS analysis were obtained either with this enzyme or with protease K [23]. The broad amino acid specificity of chymotrypsin resulted in a profusion of small peptides from which very little sequence information could be gained.

Trypsin alone proved to be the most effective of all the proteases investigated (Figure 2). Its high amino acid specificity (lysine and arginine) and its tolerance of denaturing reagents such as 5 M urea [30] make it a uniquely suitable protease for protein-sequencing studies by ESI-MS. In addition, all tryptic peptides have basic arginine and lysine residues at their C-termini that are amenable to protonation and hence make them good MS candidates. Charge retention at the C-terminal fragments is thus favoured, the y″ ion species tend to dominate the MS/MS spectra and the spectral interpretation is simplified [35,36] (see these references also for an explanation of this nomenclature). The main peptides observed in the trypsin-catalysed cleavage of GalP(His)$_6$ are listed in Table 1. The results of both ESI-MS and an example of the subsequent MS/MS sequencing of a GalP(His)$_6$ tryptic peptide, where a sequence of 16 out of 25 residues was detected, are presented in Figure 2 and the locations of tryptic peptides in the GalP sequence are given in Figure 3.

Since the tryptic digestion did not give complete coverage of the GalP(His)$_6$ sequence (Figure 3), double digestions were performed, with chymotrypsin first followed by trypsin. This successfully yielded smaller peptides that were identified with sequences by their molecular mass and MS/MS sequencing. However, the additional information served only to confirm the already established identities of trypsin-only digests (none of the results are shown). Nevertheless, such double digests may prove of more value with other proteins.

Chemical degradation of GalP(His)$_6$ using CNBr

Chemical degradation of membrane proteins with CNBr usually results in the production of large hydrophobic peptides that are difficult to separate and purify on the traditional reversed-phase HPLC systems [30]. However, in this study, using MS, the best success with the degradation of GalP(His)$_6$ was actually achieved using CNBr in 70 % aqueous formic acid. The first gains were practical in the improved solubility of the protein, the avoidance of denaturing agents, and the ease of removal of reagent and by-products by evaporation. The specificity of the reagent for methionine residues was an important aid in interpreting the mass spectra. Under the acid conditions used the formation of homoserine lactone at the C-terminus after cleavage at methionine residues is favoured over homoserine [32] and only the homoserine lactone derivatives were observed (Table 2).

The peptides generated by CNBr treatment of labelled or unlabelled GalP(His)$_6$ were detected as one or both of two forms. The molecular mass of one form corresponded to the calculated mass of the peptide after cleavage of the peptide backbone at the

Table 2 Summary of ESI-MS and MALDI-MS data for CNBr-digested GalP(His)$_6$

In the second column, for peptides of mass < 10000 Da, the mono-isotopic mass applies; for peptides of mass > 10000 Da, the average mass (ave) applies.

Observed mass by ESI-MS (Da)	Calculated mono-isotopic mass (Da)	Residues	Verified by partial MS/MS sequence
427.3	427.25	268–271	Yes
505.3	505.28	146–149	Yes
528.33	528.29	332–337	Yes
662.37	662.31	262–267	Yes
735.44	735.4	358–364	Yes
903.56, 931.56	903.51, 931.5	400–408, CO + 400–408	Yes, Yes
1409.62*	1409.74	CO + 2–14	No
1722.0*	1721.86	CO + 339–354	No
2191.26*, 2219.26	2191.05, 2219.04	272–290, CO + 272–290	No, Yes
2201.15*	2201.13	453–470	Yes
2418.54*	2418.34	63–86	No
2644.56*, 2672.5, 2700.48	2644.34, 2672.33, 2700.33	150–173, CO + 150–173, CO + CO + 150–173	No, No, No
3979.68*, 4007.6, 4035.64	3979.28, 4007.27, 4035.27	291–327, CO + 291–327, CO + CO + 291–327	No, Yes, Yes
5058.2*, 5086.4*	5057.76, 5085.75	409–452, CO + 409–452	No, Yes
10108.4, 10136.0, 10164.0	10107.03 (ave), 10135.04 (ave), 10163.05 (ave)	174–261, CO + 174–261, CO + CO + 174–261	No, No, No
Peptide only observed by MALDI-MS			
3744.50*	3746.37 (ave)	365–399	Not applicable

* Peptides observed also by MALDI-MS; for such peptides, the average mass applies.

C-terminal side of a methionine residue, followed by condensation to produce the homoserine lactone residue (designated hsl) at the C-terminus of the cleaved peptide, as expected. The other form was 28 Da higher in mass and was concluded to be a formylated derivative, accounted for by the fact that the chemical degradation was carried out in 70% formic acid, in conditions under which formylation of the N-terminus and lysine have been observed [37].

Overall, with CNBr much larger tracts of sequence could be identified using ESI-MS mass mapping than with protease digestion (above; Figure 3 and Table 2); MS/MS also yielded sequence.

MALDI-MS is more tolerant of detergent than ESI-MS [38] and this was exploited in order to attempt the detection of the most hydrophobic peptides arising from the α-helices of GalP that were considered too difficult to solubilize for ESI-MS. In the event, it was found that solubilization of the lysate in the detergent DDM (0.1%) allowed nearly all of the ESI-MS-identified fragments to be observed also with MALDI-MS, but one peptide not seen with ESI-MS was of average molecular mass 3744.50 Da, corresponding to the amino acid sequence 365–399 (Table 2). This identification is valuable for further research because it corresponds to the putative binding sites for cytochalasin B and forskolin, which are inhibitors of GalP [25,26], and includes the cysteine residue, 374, that reacts, uniquely, with thiol reagents [29]. Detection of this peptide by MALDI-MS affirms the idea that lack of solubility of the protein/peptides can be a practical handicap for ESI-MS.

Detection and quantification by MS of the incorporation of [*ring*-2-^{13}C]histidine into GalP(His)$_6$ in *E. coli*

Specific labelling of particular amino acid residues in proteins with NMR-active isotopic labels is a fundamental requirement for the determination of three-dimensional protein structure by NMR spectroscopy. This labelling may be achieved by incorporation into the protein *in vivo* of an amino acid bearing one or more isotopic labels, e.g. ^{13}C. Accurate analysis of this incorporation is manifestly essential and it provides, incidentally, for membrane proteins such as GalP, a severe test of the methods discussed above.

[*ring*-2-^{13}C]-L-Histidine was chosen to be examined for incorporation into GalP in *E. coli* for the following practical reasons. (i) The *E. coli* strain used for the overexpression of GalP is a histidine auxotroph, thus promising, from the beginning, a high and specific incorporation of externally added labelled amino acid. (ii) Histidine is not used as an *E. coli* carbon source (glucose or glycerol are employed) and thus dissipation of label into other amino acid residues should not occur [39] and the relatively small number, six, of histidines in GalP [apart from those added as the (His)$_6$ tag] facilitates later assignment of NMR signals to individual histidines; and all except one histidine are in the C-terminal half of the protein, where a binding site for galactose/cytochalasin B/forskolin has already been identified [27,29].

GalP(His)$_6$ was labelled with [*ring*-2-^{13}C]histidine *in vivo* (see the Materials and methods section), and the protein was isolated, purified, cleaved with CNBr and subjected to ESI-MS analysis. Peptidic fragments containing histidine residues include those with measured molecular masses of 2201.15 Da (residues 453–470) and 5058.2/5086.4 Da (residues 409–452; see Table 2). Two histidine residues are present in the latter peptide, which was detected by ESI-MS mainly as formylated [37] pentuply charged ions centred at m/z 1019. An appropriate increase, over unlabelled material, in average mass of 1.85 amu was indicated for this labelled fragment (Figure 4A).

Comparison of the molecular masses measured for the non-histidine-containing peptide fragments, which were in excellent agreement with the calculated values, did not vary between the labelled and unlabelled peptide mixtures. Examples of these include the peptide fragment comprising residues 358–364 (FIVGFA.hsl; measured mass, 735.44 Da; calculated mass, 735.4 Da) and the peptide fragment comprising residues 400–408 (IVGATFLT.hsl; measured mass, 903.56 Da; calculated mass, 903.51 Da; Figure 4B and Table 2). The latter component co-existed with a formylated derivative some 28 Da higher in mass.

When the peptide of mass 5086.4 Da (residues 409–452) was analysed by MS/MS, 18 out of the 43 amino acid residues in the sequence could be assigned from a mixture of singly and doubly charged b, y$''$ and x series ions, i.e. WVYAALNVLFILLTLWLV (Figure 5; for an explanation of this nomenclature see [35,36]). The masses of the b series ions for this peptide include those with

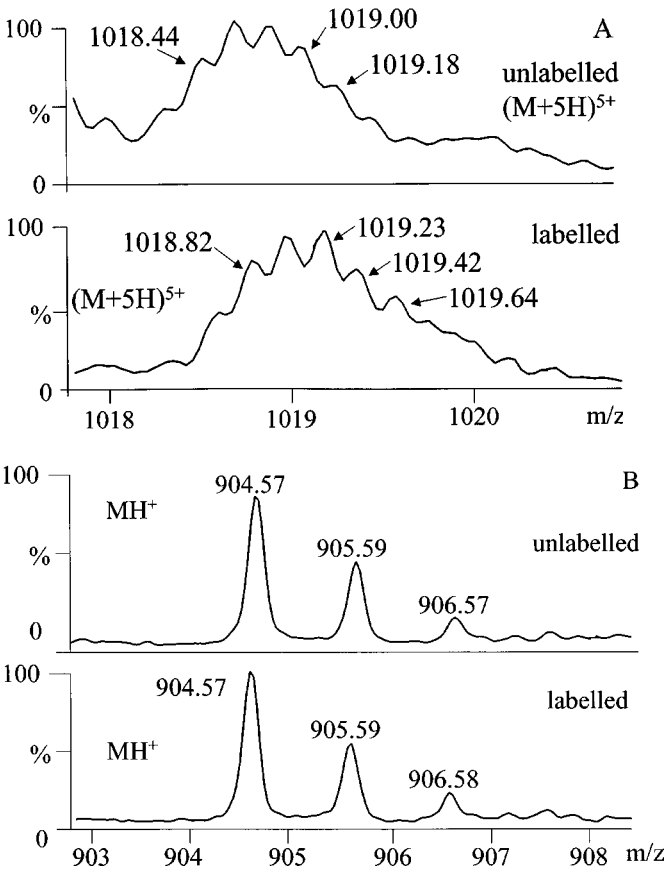

Figure 4 ESI-MS detection of [*ring*-2-^{13}C]His incorporated into peptide 409–452 and absent from peptide 400–408 of GalP(His)$_6$

(**A**) A comparison is shown of the ESI-MS *m*/*z* spectra of the peptide mixtures arising from chemical degradation of [*ring*-2-^{13}C]His-labelled (lower trace) and unlabelled (upper trace) GalP(His)$_6$ with CNBr, highlighting the region of *m*/*z* ≈ 1019, which shows the $(M+5H)^{5+}$ ions of this component (measured mono-isotopic mass, 5086.4 Da; calculated mono-isotopic mass including N-terminal formylation, 5085.75 Da; residues 409–452). The trace corresponding to the labelled protein shows a shift to higher *m*/*z* values, equivalent to an overall shift of 1.85 Da, thus indicating significant incorporation of the [*ring*-2-^{13}C]histidine label in both of the histidine residues occurring in this part of the amino acid sequence. (**B**) A similar comparison made of peptides corresponding to residues 400–408 (measured mono-isotopic mass, 903.56 Da; calculated mono-isotopic mass, 903.51 Da) from the labelled (lower trace) and unlabelled (upper trace) samples; this region is predicted not to contain any histidine residues. The coincidence of the two spectra demonstrates that there is no instrumental drift during the course of the measurements.

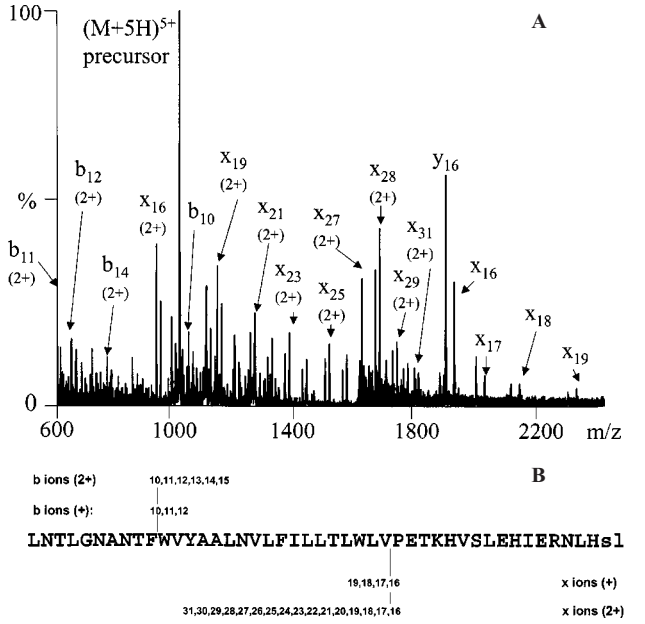

Figure 5 Tandem ESI-MS/MS location of [*ring*-2-^{13}C]His in peptide 409–452 of GalP(His)$_6$

(**A**) ESI-MS/MS product ion spectrum is shown depicting the fragment ions arising directly from the collisionally activated decomposition of the $(M+5H)^{5+}$ ions centred at *m*/*z* 1019 of the peptide 409–452 (unlabelled measured mono-isotopic molecular mass, 5086.4 Da; calculated mono-isotopic mass including N-terminal formylation, 5085.75 Da). (**B**) The sequence of this peptide is displayed with the fragment ions observed in the spectrum. Fragment ions are labelled according to the nomenclature proposed by [35,36].

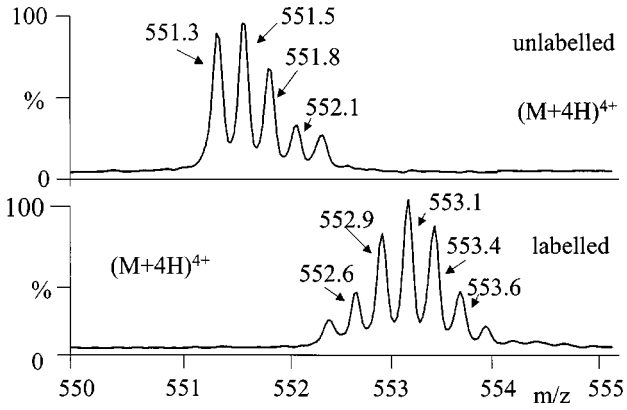

Figure 6 ESI-MS detection of [*ring*-2-^{13}C]His incorporated into peptide 453–470

ESI-MS/MS product ion spectra are shown depicting the $(M+4H)^{4+}$ ions at *m*/*z* 551–554 of peptide 453–470 (unlabelled measured molecular mass, 2201.15 Da; calculated mono-isotopic molecular mass, 2201.13 Da; KGRKLREIGAHDHHHHHH) in Table 2, showing the isotope distributions of the labelled (lower trace) and unlabelled (upper trace) peptides.

an extra 28 Da resulting from formylation [37], indicating that formylation has taken place towards the N-terminus of the peptide, the N-terminal amino group being a primary candidate. In these data, where the C-terminus ends with a homoserine lactone, it is of interest to note that the x series of fragment ions appears to dominate the spectrum.

Towards the C-terminus of the peptide there are two histidine residues, but there are none towards the N-terminus (Figure 5B). No mass differences were observed in the b series of fragmentations (N-terminal side) observed for the labelled and unlabelled peptides, indicating that there was no labelling in the N-terminal part of this peptide. But the x and y″ series of ions (singly and doubly charged) both indicated an increase of 2 Da in the labelled peptide fragments, establishing the presence of two labelled histidine residues on the C-terminal side. Thus from the different modes of fragmentation in a single peptide can the histidine labelling be observed rather elegantly.

The presence of several labelled histidine residues in a peptide renders the isotope distribution of the molecular-related $[(M+nH)^{n+}]$ ion complex and an accurate estimation of the percentage label incorporation difficult to achieve, especially if the peptide contains a mixture of labelled and unlabelled residues. Figure 6 shows the $(M+4H)^{4+}$ region of the C-terminal peptide of GalP(His)$_6$, comprising residues 453–470, which includes the lone His-

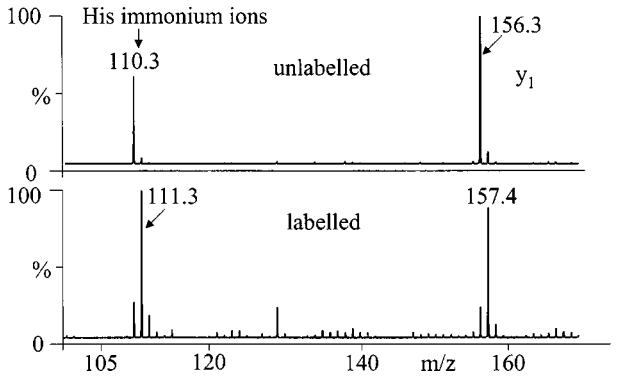

Figure 7 Tandem ESI-MS/MS detection of [*ring*-2-^{13}C]His in peptide 453–470 of GalP(His)$_6$

ESI-MS/MS product ion spectrum depicting the fragment ions in the low m/z region arising directly from the collisionally activated decomposition of the $(M+4H)^{4+}$ ions at $m/z \approx 552$ of peptide 453–470 (unlabelled measured molecular mass, 2201.15 Da; calculated monoisotopic molecular mass, 2201.13 Da; KGRKLREIGAHDHHHHHH) of Figure 6, showing the histidine immonium ions at m/z 110 and also the y_1'' (histidine) ions at m/z 156 for the unlabelled peptide (upper trace) and the corresponding ions for the histidine-labelled peptide (lower trace).

463 residue as well as the hexa-histidine tag. MS/MS collisionally activated decomposition of these quadruply charged, quasi-molecular ions gave rise (*inter alia*) to characteristic, singly charged, histidine immonium ions at m/z 110 (unlabelled species) and the C-terminus y_1'' ions at m/z 156 (Figure 7). Both of these ions contain a single histidine residue and no other amino acid residues. Peak-area integration and comparison of these ions in the spectra acquired from the unlabelled and labelled peptides allowed the ^{13}C enrichment to be calculated at a very satisfactory and practicable 80 %. The 20 % of unlabelled material is attributed to the unlabelled histidine present in the culture inoculum.

Detection and quantification by MS of the incorporation of ^{15}N, ^{13}C or ^{19}F into GalP(His)$_6$ in *E. coli*

Preparation of [^{15}N]GalP(His)$_6$

We considered that GalP(His)$_6$, which was generally (but near completely) labelled with ^{15}N label, would yield indicative information on the structure of this protein by solid-state NMR; initial results are encouraging (H. Venter, P. J. R. Spooner, A. Watts, P. J. F. Henderson and R. B. Herbert, unpublished work). But it was important first to measure the extent of isotopic enrichment by MS. The preparation of [^{15}N]GalP(His)$_6$ was simply achieved by growth of *E. coli* strain JM1100 (pPER3-His) under standard conditions in minimal medium (with addition of thymine and histidine, but no tryptone yeast extract; see the Materials and methods section), but with 12.5 mM $^{15}NH_4Cl$ (98 % enriched) instead of the usual 42 mM unlabelled NH_4Cl. There was no reduction in the level of expression of the GalP(His)$_6$ protein.

The protein was purified as described (see the Materials and methods section) and its molecular mass was measured to be 52211.5 Da by ESI-MS (results not shown). Since the 12 histidine residues in GalP(His)$_6$ would not be labelled, because this auxotrophic strain only utilizes externally added (unlabelled) histidine and, furthermore, cannot exchange histidine nitrogen with $^{15}NH_3$, the calculated molecular mass of protein to be labelled is 51674.3 Da. This would increase to 52249.3 Da if ^{15}N was incorporated into all of the 575 non-histidine nitrogen atoms. The observed molecular mass thus indicated a very satisfactory level of enrichment at ≈ 93 %. No other proteins were detectable in the MS spectrum.

Preparation of [^{19}F]GalP(His)$_6$

The natural isotope of fluorine, i.e. ^{19}F, promises a number of advantages over other isotopic labels in structural studies of proteins by solid-state NMR, i.e. sensitivity in detection similar to ^{1}H NMR, large distribution of chemical shift and sensitivity to environment [40]. Using *E. coli* strain JM1100 5-fluorotryptophan was incorporated into GalP(His)$_6$ with no reduction in the level of expression. The subsequent molecular mass of the purified protein measured by ESI-MS (results not shown) of 51854.55 Da indicated satisfactory labelling (≈ 83 % enrichment) of the 12 tryptophan residues in the protein. Preliminary work has been carried out with other fluorotryptophan derivatives as well as with 4-fluorophenylalanine and 3-fluorotyrosine [40].

Conclusions

In summary, we have shown that: (i) ESI-MS can be used to determine accurately the molecular mass of a membrane-transport protein, exemplified by GalP(His)$_6$, by employing a SEC column in the purification procedure, which is both inexpensive and also essential for success; (ii) trypsin, trypsin + chymotrypsin and CNBr fragmentation of the protein followed by the determination of the accurate molecular masses of the peptidic fragments led to the confirmation of a major proportion, i.e. 82 %, of the sequence for GalP(His)$_6$ and MS/MS measurements allowed identification of 49 % of actual amino acid sequence for this protein; both high values are unprecedented for a 12-helix membrane-transport protein; (iii) the specific incorporation of an isotopically labelled amino acid (histidine), resulting in minor mass changes, could be detected, characterized and moreover quantified by MS/MS; and (iv) the incorporation of a general isotope label, e.g. ^{15}N from $^{15}NH_3$, could also be measured accurately. In addition to the incorporation of ^{13}C, ^{15}N and ^{19}F described in this paper, incorporation of Se from derivatized Met has also been measured, for isomorphous replacement in X-ray crystallography (J. Haddon, S. E. V. Phillips and A. E. Ashcroft, unpublished work; [41]). The way is now open for wide exploitation of MS in the study of these important membrane-transport proteins, most notably in support of solid-state NMR, X-ray crystallography and the elucidation of topology (see below).

The specificity of trypsin cleavage and the ease with which the peptides could be sequenced (see above) readily made trypsin the most practically useful enzyme, of those examined, for obtaining GalP sequence; chemical cleavage with CNBr also gave excellent results for this membrane protein. It is noteworthy, however, that about one-third of the peptides identified after cleavage by trypsin derived from hydrolyses that were not preceded by a positively charged residue, so the trypsin may not be as specific as expected, or the preparation may contain traces of other proteases. In three such apparently anomalous cases the result was a consecutive series of peptides, all ending in arginine (Figure 3), which suggested that scission by trypsin was followed by some aminopeptidase activity. This, fortuitously, gave rise to extra indirect sequence information to add to that obtained by MS/MS. Seven more peptides were identified by ESI-MS after trypsin digestion (omitting five components of amu < 1000 that may not be peptides), in addition to those listed in Table 1 that matched predicted peptides from GalP(His)$_6$.

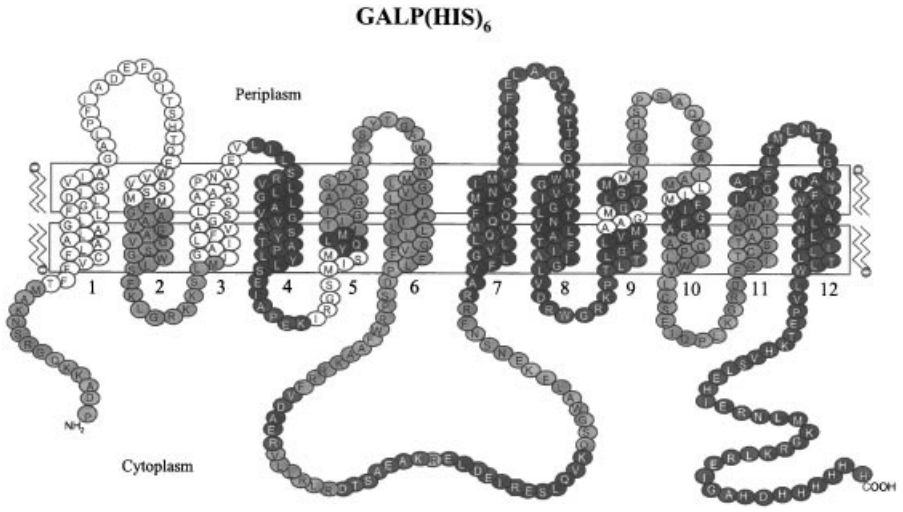

Figure 8 Two-dimensional model of GalP(His)$_6$ showing the combined results of trypsin and CNBr cleavage

The parts of the sequence shaded in grey were identified by mapping the accurate molecular masses of the peptides observed in ESI-MS and MALDI-MS to the predicted trypsin- or CNBr-cleavage maps of GalP(His)$_6$. Sequence-specific individual amino acids (in black) were observed by ESI-MS/MS of these peptides. The parts of the sequence that are so far unobserved are shown in white.

These all had molecular masses that could correspond to more than one sequence of GalP(His)$_6$ and/or trypsin amino acid residues which prevented unequivocal identification.

Nearly all the peptides after reaction with CNBr were from the C-terminal side of a methionine, as expected. Only two out of the 27 CNBr-generated fragments were not identified.

The use of chymotrypsin, followed by trypsin, brought some large peptides into a suitable mass range for analysis, but its relative lack of specificity renders chymotrypsin of limited utility. Importantly, judicious choice of MALDI-MS enabled the mass measurement of a peptide that had been refractory for ESI-MS, most probably due to the fact that this peptide required the presence of detergent for solubilization.

Most of the GalP sequence was observed by ESI-MS combined with MALDI-MS, as summarized diagramatically in Figure 8. It follows that these digestion and MS methods can now be applied for sequence analysis of other MFS transport proteins, and membrane proteins in general.

ESI-MS has also recently been used to establish, with high accuracy, the molecular mass of the 12-helical lactose membrane-transport protein, as well as of derivatives where thiol groups were functionalized [14,15]. By combining ESI-MS and MS/MS sequencing in future, the precise position(s) of covalently bound ligands in peptide digests may be determined in order to localize binding sites. Also, the positions of topological reagents, such as membrane-impermeant maleimides reacting with cysteine mutants [29] can be used to show which residues are outside or inside the cell, or buried in the membrane.

We are grateful to the Wellcome Trust, Biotechnology and Biological Sciences Research Council (BBSRC), SmithKline Beecham Pharmaceuticals and Glaxo Wellcome (now GlaxoSmithKline), the EU, The Royal Society, Leeds University and Micromass UK for financial support of this research. H. V. acknowledges an Overseas Research Studentship of the UK Committee of Vice Chancellors and Principals. This work was carried out as part of the BBSRC-funded North of England Structural Biology Centre. We thank J. O'Reilly for growth of *E. coli* and preparation of membrane fractions, N. G. Rutherford for assistance with tryptic digestions and J. P. Whitelegge for discussions on SEC.

REFERENCES

1 Henderson, P. J. F. (1997) Function and structure of membrane transport proteins. In The Transporter Factsbook (Griffith, J. K. and Sansom, C., eds.), pp. 3–29, Academic Press, London
2 Paulsen, I. T., Chen, J., Nelson, K. E. and Saier, M. H. (2001) Comparative genomics of microbial drug efflux systems. J. Mol. Microbiol. Biotechnol. **3**, 145–150
3 Pao, S. S., Paulsen, I. T. and Saier, Jr, M. H. (1998) The major facilitator superfamily. Microbiol. Mol. Biol. Rev. **62**, 1–32
4 Maiden, M. C. J., Davies, E. O., Baldwin, S. A., Moore, D. C. M. and Henderson, P. J. F. (1987) Mammalian and bacterial sugar-transport proteins are homologous. Nature (London) **325**, 641–643
5 Ward, A., O'Reilly, J., Rutherford, N. G., Ferguson, S. M., Hoyle, C. K., Palmer, S. A., Clough, J. L., Venter, H., Xie, H., Litherland, G. J. et al. (1999) Expression of prokaryotic membrane transport systems in *Escherichia coli*. Biochem. Soc. Trans. **27**, 893–899
6 Henderson, P. J. F. (1993) The 12-transmembrane helix transporters. Curr. Opin. Cell Biol. **5**, 708–721
7 Williams, K. A. (2000) Three-dimensional structure of the ion-coupled transport protein NhaA. Nature (London) **403**, 112–115
8 Zhuang, J. P., Prive, G. G., Werner, G. E., Ringler, P., Kaback, H. R. and Engel, A. (1999) Two-dimensional crystallisation of *Escherichia coli* lactose permease. J. Struct. Biol. **125**, 63–75
9 Yin, C.-C., Aldema-Ramos, M. L., Borges-Walmsley, A. R., Taylor, R. W., Walmsley, A. R., Levy, S. B. and Bullough, P. A. (2000) The quaternary molecular architecture of TetA, a secondary tetracycline transporter from *Escherichia coli*. Mol. Microbiol. **38**, 482–492
10 Spooner, P. J. R., Rutherford, N. G., Watts, A. and Henderson, P. J. F. (1994) NMR observation of substrate in the binding site of an active sugar-H^+ symport protein in native membranes. Proc. Natl. Acad. Sci. U.S.A. **91**, 3877–3881
11 Watts, A., Ulrich, A. S. and Middleton, D. A. (1995) Membrane protein structure: the contribution and potential of novel solid state NMR approaches. Mol. Membr. Biol. **12**, 233–246
12 Spooner, P. J. R., O'Reilly, W. J., Homans, S. W., Rutherford, N. G., Henderson, P. J. F. and Watts, A. (1998) Weak substrate binding to transport proteins studied by NMR. Biophys. J. **75**, 2794–2800
13 Appleyard, A., Herbert, R. B., Henderson, P. J. F., Watts, A. and Spooner, P. J. R. (2000) Selective NMR observation of inhibitor and sugar binding to the galactose-H^+ symport protein GalP, of *Escherichia coli*. Biochim. Biophys. Acta **1509**, 55–64
14 Whitelegge, J. P., Le Coutre, J., Lee, J. C., Engel, C. K., Prive, G. G., Faull, K. F. and Kaback, H. R. (1999) Toward the bilayer proteonome, electrospray ionization-mass spectrometry of large, intact transmembrane proteins. Proc. Natl. Acad. Sci. U.S.A. **96**, 10695–10698

15 Le Coutre, J., Whitelegge, J. P., Gross, A., Turk, E., Wright, E. M., Kaback, H. R. and Faull, K. F. (2000) Proteonomics on full-length membrane proteins using mass spectrometry. Biochemistry **39**, 4237–4242
16 Griffiths, W. J., Jonsson, A. P., Liu, S., Rai, D. K. and Wang, Y. (2001) Electrospray and tandem mass spectrometry in biochemistry. Biochem. J. **355**, 545–561
17 Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305**, 567–580
18 Ward, A., Sanderson, N. M., O'Reilly, J., Rutherford, N. G., Poolman, B. and Henderson, P. J. F. (2000) The amplified expression, identification, purification, assay and properties of histidine-tagged bacterial membrane transport proteins. In Membrane Transport – a Practical Approach (Baldwin, S. A., ed.), pp. 141–166, Blackwell Press, Oxford
19 Henderson, P. J. F., Giddens, R. A. and Jones-Mortimer, M. C. (1977) Transport of galactose, glucose and their molecular analogues by *Escherichia coli* K12. Biochem. J. **162**, 309–320
20 Schaffner, W. and Weissmann, C. (1973) A rapid, sensitive, and specific method for the determination of protein in dilute solution. Anal. Biochem. **56**, 502–514
21 Wessel, D. and Flügge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal. Biochem. **138**, 141–143
22 Whitelegge, J. P., Gundersen, C. B. and Faull, K. F. (1998) Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. Protein Sci. **7**, 1423–1430
23 Bond, J. S. (1989) Commercially available proteases. In Proteolytic Enzymes, a Practical Approach (Beynon, R. J. and Bond, J. S., eds.), pp. 232–240, IRL Press, Oxford
24 Witkop, B. (1961) Nonenzymatic methods for the preferential and selective cleavage and modification of proteins. Adv. Protein Chem. **16**, 221–321
25 Cairns, M. T., McDonald, T. P., Horne, P., Henderson, P. J. F. and Baldwin, S. A. (1991) Cytochalasin B as a probe of protein structure and substrate recognition by the galactose-H^+ transporter of *Escherichia coli*. J. Biol. Chem. **266**, 8176–8183
26 Martin, G. E. M., Seamon, K. B., Brown, F. M., Shanahan, M. F., Roberts, P. E. and Henderson, P. J. F. (1994) Forskolin specifically inhibits the bacterial galactose-H^+ symport protein, GalP, of *Escherichia coli*. J. Biol. Chem. **269**, 24870–24877
27 McDonald, T. P., Walmsley, A. R., Martin, G. E. M. and Henderson, P. J. F. (1995) The role of tryptophans 371 and 395 in the binding of antibiotics and the transport of sugars by the D-galactose-H^+ symport protein (GalP) from *Escherichia coli*. J. Biol. Chem. **270**, 30359–30370

Received 9 October 2001/6 December 2001; accepted 25 January 2002

28 McDonald, T. P., Walmsley, A. R. and Henderson, P. J. F. (1997) Asparagine 394 in putative helix 11 of the galactose-H^+ symport protein (GalP) from *Escherichia coli* is associated with the internal binding site for cytochalasin B and sugar. J. Biol. Chem. **272**, 15189–15199
29 McDonald, T. P. and Henderson, P. J. F. (2001) Cysteine residues in the D-galactose-H^+ symport protein of *Escherichia coli*: effects of mutagenesis on transport, reaction with *N*-ethylmaleimide and antibiotic binding. Biochem. J. **353**, 709–717
30 Carrey, E. A. (1990) Peptide mapping. In Protein Structure, a Practical Approach (Creighton, T. E., ed.), pp. 117–144, IRL Press, Oxford
31 Hufnagel, P., Schweiger, U., Eckerskorn, C. and Oesterhelt, D. (1996) Electrospray ionization mass spectrometry of genetically and chemically modified bacteriorhodopsins. Anal. Biochem. **243**, 46–54
32 Lee, T. D. and Shively, J. E. (1990) Enzymatic and chemical digestion of proteins for mass spectrometry. Methods Enzymol. **193**, 361–374
33 Ferrige, A. G., Seddon, M. J., Green, B. N., Jarvis, S. A. and Skilling, J. (1992) Disentangling electrospray spectra with maximum-entropy. Rapid Commun. Mass Spectrom. **6**, 707–711
34 Whitelegge, J. P., Jewess, P., Peckering, M. G., Gerrish, C., Camilleri, P. and Bowyer, J. R. (1992) Sequence analysis of photoaffinity-labelled peptides derived by proteolysis of photosystem-2 reaction centres from thylakoid membranes treated with [^{14}C]-azidoatrazine. Eur. J. Biochem. **207**, 1077–1084
35 Roepstorff, P. and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom. **11**, 601
36 Johnson, R. S., Martin, S. A. and Biemann, K. (1988) Collision-induced fragmentation of $(M+H)^+$ ions of peptides – side chain specific sequence ions. Int. J. Mass Spectrom. Ion Processes **86**, 137–154
37 Zhang, X., Dillen, L., Vanhoutte, K., Van Dongen, W., Esmans, E. and Claeys, M. (1996) Characterization of unstable intermediates and oxidized products formed during cyanogen bromide cleavage of peptides and proteins by electrospray mass spectrometry. Anal. Chem. **68**, 3422–3430
38 Cadene, M. and Chait, B. T. (2000) A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. Anal. Chem. **72**, 5655–5658
39 McFall, E. and Newman, E. B. (1996) Amino acids as carbon sources. In *Escherichia coli* and *Salmonella* (Neidhardt, F. C., ed.), pp. 358–379, ASM Press, Washington DC
40 Venter, H. (2001) Applications of Biophysical Techniques to the Analysis of Membrane Transport Proteins, Ph.D. Thesis, University of Leeds
41 Hadden, J. M., Covery, M. A., Déclais, A.-C., Lilley, D. M. J. and Phillips, S. E. V. (2001) Crystal structure of the Holliday junction resolving enzyme T7 endonuclease 1. Nat. Struct. Biol. **8**, 62–67