Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 852 (2007) 188–194
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

A general, robust method for the quality control of intact proteins using LC–ESI-MS

Gustav Sundqvist [a], Maria Stenvall [a], Helena Berglund [b], Jenny Ottosson [a], Harry Brumer [a,*]

[a] School of Biotechnology, Royal Institute of Technology, Albanova University Centre, 10691 Stockholm, Sweden
[b] Structural Genomics Consortium, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden

Received 24 October 2006; accepted 10 January 2007
Available online 18 January 2007

Abstract

A simple and robust method for the routine quality control of intact proteins based on liquid chromatography coupled to electrospray ionization mass spectrometry (LC–ESI-MS) is presented. A wide range of prokaryotic and eukaryotic proteins expressed recombinantly in *Escherichia coli* or *Pichia pastoris* has been analyzed with medium- to high-throughput with on-line desalting from multi-well sample plates. Particular advantages of the method include fast chromatography and short cycle times, the use of inexpensive trapping/desalting columns, low sample carryover, and the ability to analyze proteins with masses ranging from 5 to 100 kDa with greater than 50 ppm accuracy. Moreover, the method can be readily coupled with optimized chemical reduction and alkylation steps to facilitate the analysis of denatured or incorrectly folded proteins (e.g., recombinant proteins sequestered in *E. coli* inclusion bodies) bearing cysteine residues, which otherwise form intractable multimers and non-specific adducts by disulfide bond formation.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Electrospray ionization; Liquid chromatography; Mass spectrometry; Protein analysis; Quality control; Reduction; Alkylation; Column regeneration; Glycoprotein; Xyloglucan endo-transglycosylase (XET)

1. Introduction

Verification of the identity of recombinant proteins and engineered variants thereof is a mandatory step in all biochemical analysis, including antibody generation, enzyme kinetic studies, and structure determination. Specifically, the presence of side products, degradation products, or unwanted protein variants which can confound future utilization or analysis must be revealed. Protein expression in high-throughput, multiwell plate format presents a specific risk of unintentional cross-contamination of proteins during sample handling, due to the close proximity of the individual sample chambers. Quality control methods for the high-throughput production of proteins must likewise be free from sample-to-sample carry over, and should possess sufficient mass accuracy and resolution to reveal cross contamination or protein heterogeneity. The ability to identify single amino acid variants and low-mass post-translational modifications across a large mass range, e.g., 5–100 kDa, is desirable. Whereas denaturing polyacrylamide gel analysis has become the *de facto* standard to assess protein purity, this technique suffers from limited mass accuracy and resolution. Mass spectrometry (MS), on the other hand, is an excellent tool allowing highly accurate protein mass determination.

Two ionization techniques are widely available for the production of intact, gas phase protein ions: matrix assisted laser desorption ionization (MALDI) [1] and electrospray ionization (ESI) [2]. As samples are typically introduced into the mass spectrometer in an array format, MALDI–MS is particularly well-suited to high-throughput applications, and has therefore become a workhorse for the rapid analysis of peptide digests (<4 kDa). However, the use of MALDI–MS in the analysis of larger proteins is hampered by decreasing ionization efficiency with increasing mass and the production of predominantly single- or double-charged molecular ions ($[M+H]^+$ or $[M+2H]^{2+}$, respectively). Linear time-of-flight (TOF) mass analyzers used with MALDI for intact protein analysis thus have high mass-to-charge (*m/z*) ranges at the expense of resolution and mass accuracy, which are often insufficient to unambiguously confirm protein sequences. In contrast, ESI generates a distribution of multiply charged molecular ions ($[M+nH]^{n+}$)

* Corresponding author. Tel.: +46 8 5537 8367; fax: +46 8 5537 8468.
E-mail address: harry@biotech.kth.se (H. Brumer).

1570-0232/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2007.01.011

that, in the case of monomeric proteins, typically lies in the range 500–2000 *m/z*. A high degree of data redundancy, coupled with the ability to accurately calibrate high resolution analyzers in this range (e.g., reflectron TOF), yields highly accurate protein masses after peak deconvolution. For example, ESI-orthogonal acceleration TOF MS has been used to achieve <5 ppm accuracy in the analysis of an intact 30 kDa glycoprotein in the presence of a 17 kDa internal protein standard [3].

The continuous nature of ESI allows for straightforward interfacing with liquid chromatography (LC), thus providing the opportunity for on-line sample concentration, desalting, and separation. Both single stream and parallel stream LC–ESI-MS configurations have been described for the high-throughput analysis of small molecules [4] and proteins up to 9 kDa [5]. A limited number of reports have been presented describing the LC–ESI-MS analysis of larger proteins, but these are limited in scope by the need for long chromatography cycles, protein-specific eluants, and costly chromatography media. For example, a LC–ESI-MS system was used for the analysis of the intact intrinsic membrane protein bacteriorhodopsin, which had an elution time of over 40 min [6]. Furthermore, no generally applicable LC–ESI-MS methods have been described for the routine mass determination of intact proteins above 10 kDa in a multi-well plate format. Motivated by demands to characterize and perform quality control analysis on a diversity of proteins expressed in microbial hosts, we have developed a rapid, robust system for MS analysis of proteins in the mass range 10–100 kDa from liquid samples, including on-line desalting. Demonstrative examples include the mass analysis of functional enzymes, heterogeneously *N*-glycosylated proteins, a range of native (i.e., folded and soluble) human protein targets for crystallography, and urea-solubilized human protein fragments for antibody generation. For the analysis of denatured protein samples, such as those produced from *E. coli* inclusion bodies, it was further shown that chemical reduction and alkylation was essential to cleave non-specific disulfide bonds to cysteine residues, improve signal quality, and simplify MS analysis.

2. Experimental

2.1. Chemicals

HPLC gradient acetonitrile (ACN), was from Carlo Erba (Peypin, France). Formic acid (FA) was from Fluka Chemie Gmbh (Buchs, Germany). Dithiothreitol (DTT), iodoacetamide (IAA), and horse heart myoglobin (HHM) were from Sigma–Aldrich Chemie Gmbh (Steinheim, Germany).

2.2. Equipment

The CapLC System^{TM} and Q-Tof ^{TM} II quadrupole/orthogonal acceleration time-of-flight mass spectrometer were from Waters Corporation, Micromass MS Technologies (Manchester, UK). Protein trap cartridges (300 μm × 5 mm, filled with C4 pepMap300 matrix) were from LC Packings/Dionex (distributed by Kovalent AB, Hägersten, Sweden).

2.3. Sample preparation

2.3.1. Human proteome resource protein epitope signature tags (HPR PrESTs)

Protein epitope signature tags (PrESTs) were produced by the Swedish Human Proteome Resource [7] with an N-terminal His_6ABP (hexahistidine-albumin-binding protein) fusion partner using the pAff8c expression vector in *E. coli* BL 21 cells (DE3) as described elsewhere [8]. After an immobilized metal affinity chromatography (IMAC) purification step and elution of the PrEST with 2.5 ml (6 M urea, 50 mM NaH_2PO_4, 100 mM NaCl, 30 mM acetic acid, 70 mM Na-acetate, pH 5.0), reduction and alkylation was performed. To 10 μl of the PrEST (200 μM) containing IMAC eluate solution was added 90 μl 6 M urea/0.1 M NH_4HCO_3 and 1 μl 400 mM DTT followed by 1 h incubation at room temperature (ca. 20 °C). Thereafter, 400 mM IAA (2.5 μl) was added and the solution was incubated 30 min in the dark, followed by addition of 5 μl 400 mM DTT. A 20 μl sample of this solution was mixed with 480 μl 5% $ACN_{(aq)}$ containing 0.1% FA in a 96 well plate prior to LC–ESI-MS analysis.

2.3.2. Structural genomics consortium (SGC) proteins

Protein expression and purification was generally performed as described previously [9]. SGC proteins were analysed under reducing conditions using tris(2-carboxyethyl)phosphine (TCEP). To produce a protein concentration of 10 μg/ml in the MS sample solution, 0.5–5 μl protein solution from the final concentrated sample (20 mM HEPES, 300 mM NaCl, 10% glycerol, 0.5 mM TCEP) was mixed with 500 μl 5% $ACN_{(aq)}$ containing 0.1% FA and 1 mM TCEP in a 96 well plate prior to LC–ESI-MS analysis.

2.3.3. Carbohydrate-active enzymes (CAZymes)

Xyloglucan endotransglycosylase 16A from the hybrid aspen, *Populus tremula x tremuloides* (PttXET16A) was expressed in *Pichia pastoris* as previously described [10]. Expression and purification of the glycoside hydrolase family 36 α-galactosidase, GalA, from *Thermotoga maritima* was performed according to Miller et al. [11]. Carbohydrate active enzymes were analysed in their native, active forms, without reductive alkylation prior to LC–ESI-MS. These proteins, in 25–100 mM NaOAc buffer, pH 4.8, and ca. 300 mM NaCl after purificaton, were diluted to 1–5 μM in 5% $ACN_{(aq)}$ containing 0.1% FA in a 96 well plate for LC–ESI-MS analysis.

2.3.4. Horse heart myoglobin

Myoglobin from horse heart was obtained from Sigma–Aldrich Chemie Gmbh (Steinheim, Germany) as a powder and dissolved in ultrapure water (resistivity, ρ, ≥18.2 MΩ cm) prior to dilution to 6 μg/ml (0.36 μM) in 5% aqueous ACN containing 0.1% FA.

2.3.5. Column liquid chromatography

The CapLC system coupled on-line with the Q-Tof ^{TM} II mass spectrometer was used for LC–ESI-MS as schematically described in Fig. 1. The 96 well plate with one well loaded with 6 μg/ml (0.36 μM) HHM as a standard sample was placed in

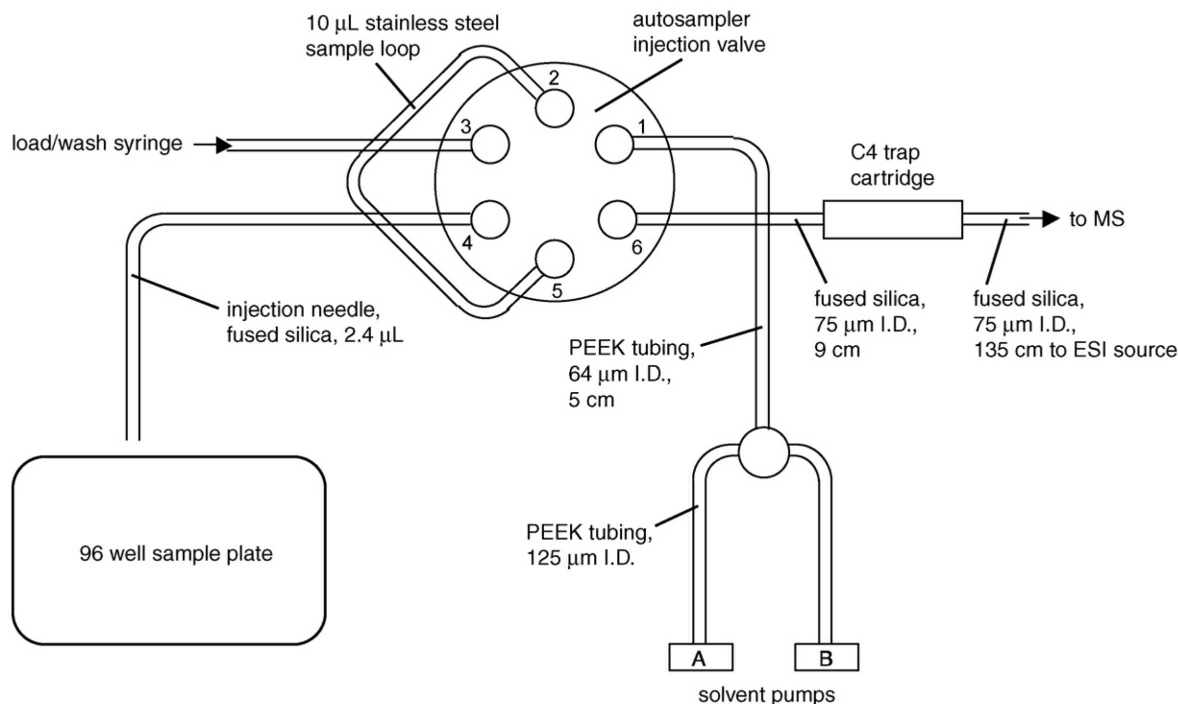

Fig. 1. Flow pattern schematic for preconcentration and desalting of proteins prior to ESI-MS analysis.

the sample holder of the CapLC system. Prior to each injection cycle, the system was equilibrated with solvent A (5% $ACN_{(aq)}$, 0.1% FA) for 1 min at a flow rate of 8 μl/min. Injections onto the C4 trap cartridge (5 μl) were performed in "partial loop mode" via a 2.4 μl fused silica capillary and 10 μl injection loop. Elution using solvents A and B (95% $ACN_{(aq)}$, 0.1% FA) was performed with the following gradient at a flow rate of 8 μl/min: 0% B (0–0.5 min), 0–90% B (0.5–1.5 min), 90% B (1.5–4.5 min) and 90–0% B (4.5–5 min).

2.3.6. Mass spectrometry

The sample flow from the C4 cartridge was coupled directly to the Q-Tof^{TM} II ESI interface consisting of the Z spray source fitted with an electrospray probe (source voltage 3.3 kV, source temp 80 °C, desolvation temp 140 °C, desolvation gas flow 175 l/h, cone voltage 35 V, cone gas flow 50 l/h). The quadrupole mass filter of the Q-Tof^{TM} II was operated in a wide band pass (RF only) mode when collecting TOF MS data. Collision energy was set to 10 V and argon was present in the collision cell to improve resolution by collisional cooling. TOF MS data were acquired over the *m/z* range 600–1300 at a resolution >10,000 FWHM. All data were collected using a scan time of 5 s. External TOF mass calibration was obtained over the *m/z* range 50–2000 using a solution of NaI (2 g/L) in 1:1 2-propanol/water prior to analysis of each 96 well plate.

2.3.7. Data analysis

Multiple-charged protein ion signals were deconvoluted to produce zero-charge spectra using the Maximum Entropy^{TM} 1 (MaxEnt1) algorithm in the Micromass MassLynx 4.0 software package. Observed protein mass values are reported as relative molecular mass (M_r) values, and are thus dimensionless [12]. For automated deconvolution of *m/z* spectra, sample lists created with MassLynx^{TM} were queued using AutoLynx^{TM} post analysis, and automatically MaxEnt1 processed using OpenLynx^{TM}. The OpenLynx^{TM} parameters included smoothing and combination of individual MS spectra in the chromatogram between 2.5 and 4.5 min. No background subtraction was used. MaxEnt1 was performed over the range 600–1300 *m/z* for all protein data with output ranges routinely set to M_r 15,000–40,000 for HPR-PrEST samples and M_r 15,000–100,000 for SGC samples. (Note that the MaxEnt1 algorithm makes use of a uniform gaussian damage model that requires the peak width at half height for each protein and that this value varies with protein M_r. It is therefore not strictly correct to apply such a wide output range, but for routine analysis this is the most practical configuration, using the HHM peak width at half height value of *m/z* 0.45.). For quality control assignments of the HPR-PrESTs, a Microsoft Excel-based navigator was designed in-house using Microsoft Visual Basic macros (details available on request to H.B.).

3. Results and discussion

3.1. Method development—analysis of HPR PrESTs

The development of the present method was motivated by the need to perform accurate quality control analysis of human protein epitope signature tags produced using high-throughput techniques by the Swedish Human Proteome Resource (http://www.proteinatlas.org/) [7,13]. PrESTs are fragments of human proteins 25–200 amino acids in length derived from predicted open reading frames (ORFs) of the human genome, against which antibodies are raised for localization proteomics studies. As a consequence, quality assurance

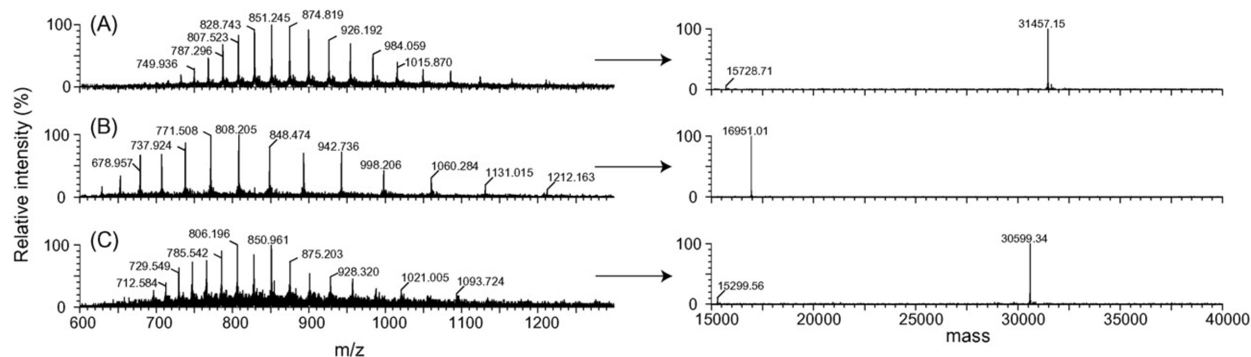

Fig. 2. Raw (left) and deconvoluted (right) ESI-MS spectra from consecutive injections of (A) PrEST1 (B) myoglobin and (C) PrEST2. Carry over of PrEST1 is less than 1% in the myoglobin standard and is not detectable in the PrEST2 spectra.

of recombinantly expressed PrESTs prior to immunization is essential to ensure that the antibodies generated have the expected specificity. PrESTs are produced as fusion proteins with sequential N-terminal hexahistidine and albumin binding protein (His_6-ABP) tags to facilitate purification and enhance immunogenicity, respectively [14]. Consequently, these PrEST constructs contain an invariant 18 kDa N-terminal portion coupled to a unique PrEST sequence. The constructs are expressed in *E. coli* dissolved under denaturing and reducing conditions (7 M guanidinium HCl, 20 mM β-mercaptoethanol), captured on immobilized metal affinity columns and eluted with a complex buffer solution (6 M urea, 50 mM NaH_2PO_4, 100 mM NaCl, 30 mM HOAc, 70 mM NaOAc, pH 5.0).

To achieve the desired accuracy and precision required to distinguish PrEST constructs which are often very similar in mass, electrospray ionization coupled with orthogonal acceleration time-of-flight mass spectrometry (oaTOF MS) was chosen. The use of non-volatile salts, including phosphate buffers and chaotropic agents, in the purification necessitated a desalting step prior to MS, as these are known to have deliterious effects on ESI [15]. Early attempts to develop an on-line desalting method using reverse phase trap columns in conjunction with standard ternary CapLC/Stream Select Module^{TM} programs met with limited success due to sample carry over. This observation was especially troublesome, as one of the key quality control goals was to ensure that sample cross contamination had not occurred in the PrEST production chain. The fluidic system was therefore simplified to a binary pump system in which the chromatographic gradient flushes the complete sample flow path, from injection loop to MS source (Fig. 1). A simple C4 reversed phase silica cartridge (300 μm × 5 mm) was used for robust and cost-effective sample desalting. The particular cartridges used in this study are available for less than 100 USD (ca. 10-fold cheaper than commerical C4 microbore analytical columns) and, in our experience, withstand 1000–5000 injections. Due to the low column cost, replacement upon failure, e.g., due to sample fouling, is trivial. The total cost of consumables for the method is thus low.

As such, this LC configuration typically resulted in undetectable or very little carry over (<1%) between samples. It was therefore possible to use a very short chromatography cycle time (6 min) without the need for an extensive column regeneration step [16]. Very rarely, however, carryover of 1–5% had been observed for certain protein samples, which was nonetheless unacceptable during the quality control of HPR PrESTs. Consequently, a standard practice of injecting horse heart myoglobin between each sample was initiated, the function of which was two-fold. Firstly, HHM facilitated the removal of particularly tenacious proteins synergistically with increasing concentration of organic modifier in the solvent. This phenomenon was attributed to soluble HHM functioning as an alternate binding surface to help prevent readsorption of "sticky" proteins while also helping to displace these proteins from the matrix surface by competitive binding. Fig. 2 shows three consecutive LC–MS runs in which a PrEST, HHM, and a second PrEST were analyzed consecutively. Carryover of PrEST1 into the HHM sample was less than 1%, while carryover of PrEST1 into the PrEST2 sample was undetectable. The second beneficial effect of alternating sample and standard injections was that instrument performance was readily monitored. Failed PrEST samples between positive HHM injections were unambiguously identified, while mass calibration drift due to electronic fluctuations and temperature-dependent expansion of the TOF tube [17] was easily observed. Typically, the relative standard deviation of M_r values for HHM during an overnight run of a 96 well sample plate (ca. 90 myoglobin injections) was 10–20 ppm. However, unexpected loss of TOF calibration on rare occasions during extended runs gave rise to much larger errors, which would have otherwise caused the PrEST samples to fail quality specifications (data not shown); recalibration versus alternating HHM injections avoided re-analysis of several large sample sets. Furthermore, external standard analysis was preferred to inclusion of HHM as an internal standard in PrEST samples due to a 4- to 5-fold reduction in signal due to ion supression effects (data not shown). Potential overlap of numerous protein ions in the range 600–1300 *m/z* which may confound analysis [18] is also avoided.

During the development of the method, it was found that permanent disruption of non-specific disulfide bonds by reduction and alkylation was essential to produce reproducible ESI-MS data. Due to the use of β-mercaptoethanol in the solubilization and purification buffers [8], PrESTs were observed to carry numerous β-mercaptoethanol adducts, often in proportion to the number of cysteine residues in the protein. However, more

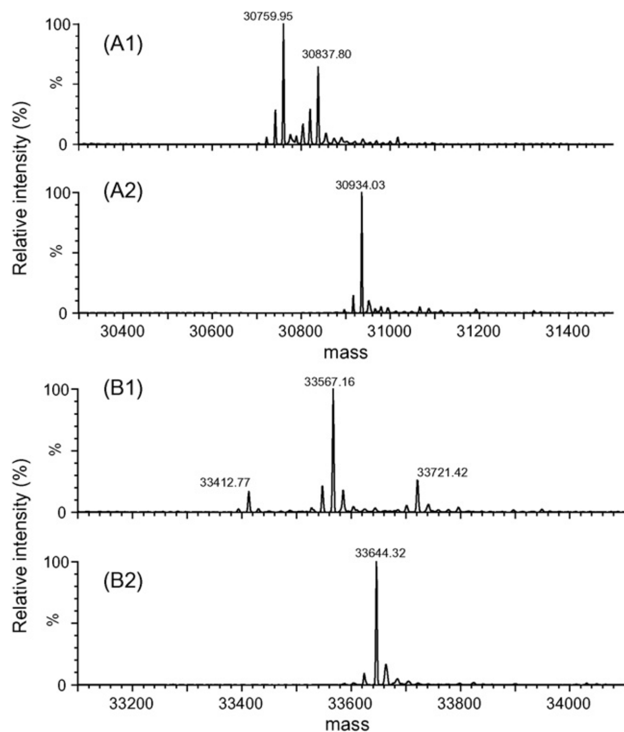

Fig. 3. Effect of dithiothreitol (DTT) reduction and iodoacetamide (IAA) alkylation of PrESTs on LC–ESI-MS analysis. (A1 and B1) deconvoluted spectra of PrESTs A and B after solubilization with β-mercaptoethanol and purification by IMAC, performed essentially as described in [8]. (A2 and B2) deconvoluted spectra of the same purified PrEST samples following reduction with DTT and alkylation with IAA.

than one protein species was typically observed as a result of a combination of non-specific β-mercaptoethanol adduct and intramolecular disulfide bond formation. In addition, severe problems with protein aggregation leading to reduced or eliminated signal was observed for some PrESTs, especially those with a high cysteine content. Fig. 3 shows the mass spectra for two typical PrESTs: PrEST A, which contains three cysteine residues, and PrEST B, which contains four cysteine residues. For purified PrEST A (Fig. 3, A1), the observed peak at 30759.95 corresponds to the calculated protein mass minus two protons lost by intramolecular disulfide bond formation; the peak at 30837.8 corresponds to the β-mercaptoethanol adduct of this species. Such protein-mercaptoethanol adduct formation has been previously observed for proteins with $M_r < 10^4$ using MALDI-TOF MS [19]. Lower intensity peaks surrounding each main peak are attributed to dehydrated/deamidated and oxidized protein species and MaxEnt1 artifacts due to deconvolution of spectra with noisy baselines. Upon reduction with dithiothreitol and alkylation with iodoacetamide, the reconstructed spectrum of PrEST A (Fig. 3, A2) is noticably improved. Reduction of the complexity of the protein sample gave rise to raw MS spectra with higher signal-to-noise values, which has been commonly observed for the vast majority of samples analyzed (data not shown). The base peak in the reconstructed spectrum results from the reduced and alkylated form (calc. M_r 30933.9, obs. 30934.0), with minor peaks resulting from dehydration/deamidation and oxidiation of the parent protein. An unalkylated sample of PrEST B exhibited greater complexity, including peaks due to the parent protein (calc. M_r 33415.7, obs. 33412.8) and species with two (calc. M_r 33568.0, obs. 33567.2) and four (calc. M_r 33720.3, obs. 33721.4) β-mercaptoethanol adducts (Fig. 3, B1). Likewise, this complexity was reduced by reduction and alkylation to yield a single protein species (Fig. 3, B2) with the expected M_r (calc. M_r 33643.9, obs. 33644.3). An additional benefit of the reduction and alkylation is that the storage stability of PrEST (and other proteins) dissolved in the injection solution in multiwell plates is increased from ca. 1 day to up to 1 month. It was likewise noted that samples exhibit higher stablility in 5% acetonitrile than in aqueous buffer solutions.

The number of HPR PrESTs analyzed during the most recent 6 month period (non-continuous operation) was ca. 3500 sample injections (ca. 7000 including HHM standard injections). The simple LC–ESI-MS system has proven to be robust, and may be classified as medium-throughput, with a single chromatography column having a cycle time of 6 min. As the baseline chromatographic peak width of the method is 30–60 s, transferring the method to a multi-column format with simultaneous column regeneration could be used to achieve much higher sample throughput. Alternatively, the requirements for alternating HHM standard/sample injections could be relaxed in applications where sample carryover is not a concern. Using stringent quality control standards (e.g., good spectral quality and undesired protein contaminants present at less than 30% of the MaxEnt 1 base peak intensity) ca. 80% of HPR PrESTs give acceptable spectra; the success rate of HHM standard injections is 100%.

3.2. Extension of the method to correctly folded, soluble proteins

3.2.1. SGC proteins

One benefit of the present method is its applicability to various protein types from different sources. One such example is a diverse group of human proteins produced for structural determination by the Stockholm node of the Structural Genomics Consortium (http://sgc.ki.se/). These samples are native proteins of M_r up to 10^5 produced in *E. coli* at a rate of 20–30 soluble constructs per week.

As an alternative to β-mercaptoethanol, the SGC uses the reducing agent tris(2-carboxyethyl)phosphine in protein purification buffers to maintain a reducing environment. Water soluble phosphines have the advantage over thiol reducting agents, such as β-mercaptoethanol, that adducts do not form between the protein and the reducing agent. The TCEP protocol employed produces reproducible ESI-MS data and is a viable alternative to the longer two-step reduction and alkylation used for denatured proteins such as the HPR PrESTs. A potential limitation, however, is the possibility of re-oxidation of cysteine residues, although this has not been observed due to the TCEP/protein stoichiometry used, together with typically short storage times between sample preparation and analysis.

The general applicability of the LC–ESI-MS method to high M_r proteins has been demonstrated in the analysis of the SGC

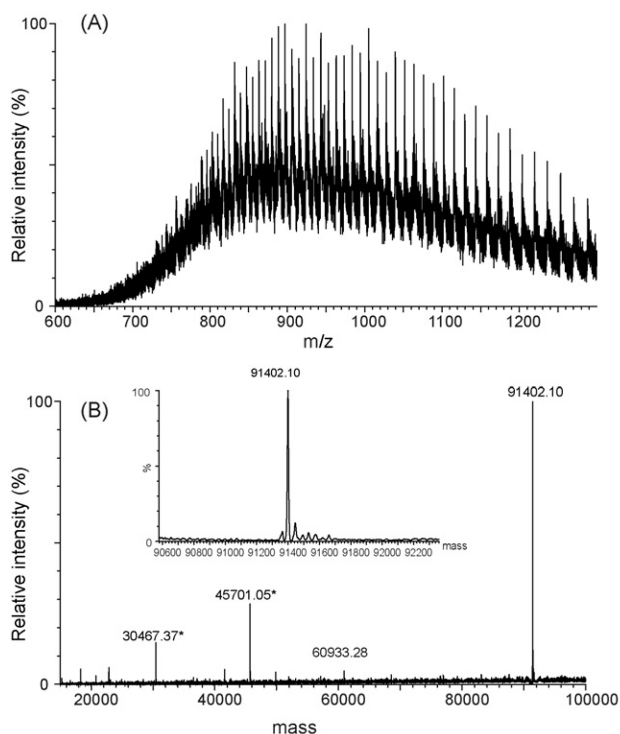

Fig. 4. MS spectra corresponding to high molecular weight SGC protein showing (A) raw spectrum and (B) deconvoluted spectrum. Peaks marked with asterisks at one-half and one-third the base peak M_r value are harmonic artifacts due to the broad MaxEnt1 output range used. The inset shows the MaxEnt1 output over a more appropriate, limited mass range.

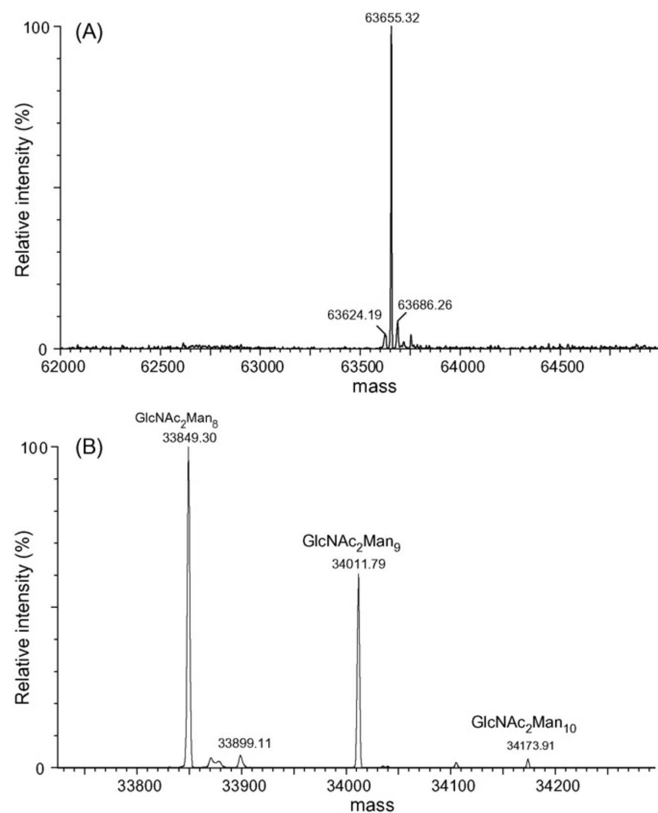

Fig. 5. Deconvoluted ESI-MS spectra of *T. maritima* α-galactosidase A (A) and *Ptt*XET16A glycoforms (B).

proteins. Whereas the HPR PrEST sample set is comprised of constructs with ca. M_r 35000, SGC proteins with broadly distributed M_r values in the range 20000–90000 are routinely analyzed. Fig. 4 shows a representative MS analysis of a high mass SGC protein (calc. M_r 91404, obs. M_r 91402.1). The success rate of SGC proteins is similar to that of the HPR PrESTs (ca. 80%). In both cases, failed samples correspond to those which yielded no interpretable mass spectrum, multiple protein peaks, or a single peak which deviated from the calculated mass by a threshold value (typically > 5 units). Success or failure in MS analysis was often correlated with SDS-PAGE results, thus the failure HPR and SGC samples is attributed to individual protein production problems. The method is extremely robust for standard samples (100% success rate for HHM).

3.2.2. Carbohydrate-active enzymes

A third class of proteins that have successfully been analyzed using the present method are native CAZymes involved in the re-organization and degradation of polysaccharides. CAZymes are widely distributed in Nature, from *Archaea* to eukaryotes, including mammals and plants (http://afmb.cnrs-mrs.fr/CAZY/). Prior to further biochemical studies, LC–ESI-MS analysis was successfully performed on the glycoside hydrolase family 36 α-galactosidase from the thermophilic bacterium *Thermotoga maritima* (*Tm*GalA), that was recombinantly expressed in *E. coli* in a soluble, catalytically active form [11] (Fig. 5A). The observed mass (M_r 63655.3) is two units lower than the calculated mass (M_r 63657.3), in agreement with the formation of a disulfide linkage between C188 and C428 revealed by X-ray crystallography ([20], PDB entry 1zy9).

The method is similarly applicable to glycoproteins produced in other expression systems. The glycoside hydrolase family 16 xyloglucan *endo*-transglycosylase from the hybrid aspen *Populus tremula x tremuloides* (*Ptt*XET16A) requires *N*-glycosylation of a conserved site for proper protein folding and expression [10]. Consequently, this and related proteins must be expressed in the methylotrophic yeast *P. pastoris* to produce functional enzymes. During expression, *P. pastoris* installs an *N*-glycan on *Ptt*XET16A (calc. polypeptide M_r 32,151) that contains a conserved di-*N*-acetylglucosamine (GlcNAc) core bearing variable oligo-mannose (Man) branches. Fig. 5B shows the different recombinant *Ptt*XET16A glycoforms including $GlcNAc_2Man_{8-10}$ (calc. M_r 33853.8, 34015.9 and 34178.0). The observed masses are ca. 4 units lower than the calculated masses due to the presence of two disulfide linkages between residues C266–C253 and C207–C216 ([21], PDB code 1un1) under the nonreducing sample conditions used.

4. Conclusions

A straightforward LC–ESI-MS method is presented which is capable of analyzing intact proteins of different origin and broad mass distribution. The inherent robustness and simple configuration of the LC system, including the use of cost-effective, disposable columns and short gradient cycle times, make it suitable for use as a general method for rapid LC–ESI-MS analysis

of recombinant proteins, including large and glycosylated proteins.

Acknowledgements

We thank the Knut and Alice Wallenberg Foundation (supporting the Swedish Human Proteome Resource, HPR) for funding. Purchase of the LC–MS equipment was funded by the Wallenberg Consortium North for Functional Genomics. Mathias Uhlén and Sofia Hober of the HPR and Johan Weigelt of the Structural Genomics Consortium are thanked for helpful discussions. Prof. Robert Kelly (North Carolina State University) is gratefully acknowledged for providing the *T. maritima* α-galactosidase expression construct; Alex Selvanayagam (KTH Wood Biotechnology) is thanked for performing the expression and purification of this enzyme. Dr. Kathleen Piens (KTH Wood Biotechnology) is thanked for providing the *Ptt*XET16A sample. Wald Lodewyckx (KTH Wood Biotechnology) is thanked for skilled technical and programming assistance. H.B. is a fellow of the Swedish Research Council (*VR Rådsforskare*).

References

[1] M. Karas, F. Hillenkamp, Anal. Chem. 60 (1988) 2299.
[2] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Science 246 (1989) 64.
[3] M. Prochazka, I. Campuzano, A. Sage, T. McKenna, J. Langridge, M. Kennedy, H. Brumer, K. Piens, J. Peptide Sci. 10 (2004) 121.
[4] Y.-Q. Xia, D.B. Whigan, M.L. Powell, M. Jemal, Rapid Commun. Mass Spectrom. 14 (2000) 105.
[5] B. Feng, A.H. Patel, P.M. Keller, J.R. Slemmon, Rapid Commun. Mass Spectrom. 15 (2001) 821.
[6] J.P. Whitelegge, C.B. Gundersen, K.F. Faull, Protein Sci. 7 (1998) 1423.
[7] M. Uhlén, E. Björling, C. Agaton, C.A.-K. Szigyarto, B. Amini, E. Andersen, A.-C. Andersson, P. Angelidou, A. Asplund, C. Asplund, L. Berglund, K. Bergström, H. Brumer, D. Cerjan, M. Ekström, A. Elobeid, C. Eriksson, L. Fagerberg, R. Falk, J. Fall, M. Forsberg, M.G. Björklund, K. Gumbel, A. Halimi, I. Hallin, C. Hamsten, M. Hansson, M. Hedhammar, G. Hercules, C. Kampf, K. Larsson, M. Lindskog, W. Lodewyckx, J. Lund, J. Lundeberg, K. Magnusson, E. Malm, P. Nilsson, J. Ödling, P. Oksvold, I. Olsson, E. Öster, J. Ottosson, L. Paavilainen, A. Persson, R. Rimini, J. Rockberg, M. Runeson, Å. Sivertsson, A. Sköllermo, J. Steen, M. Stenvall, F. Sterky, S. Strömberg, M. Sundberg, H. Tegel, S. Tourle, E. Wahlund, A. Walden, J. Wan, H. Wernérus, J. Westberg, K. Wester, U. Wrethagen, L.L. Xu, S. Hober, F. Pontén, Mol. Cell Proteomics 4 (2005) 1920.
[8] M. Larsson, S. Gräslund, Y.B. Li, E. Brundell, M. Uhlén, C. Höög, S. Ståhl, J. Biotechnol. 80 (2000) 143.
[9] A. Magnusdottir, P. Stenmark, S. Flodin, T. Nyman, M. Hammarström, M. Ehn, M.A. Bakali H, H. Berglund, P. Nordlund, J. Biol. Chem. 281 (2006) 22434.
[10] Å.M. Kallas, K. Piens, S.E. Denman, H. Henriksson, J. Fäldt, P. Johansson, H. Brumer, T.T. Teeri, Biochem. J. 390 (2005) 105.
[11] E.S. Miller Jr., K.N. Parker, W. Liebl, D. Lam, W. Callen, M.A. Snead, E.J. Mathur, J.M. Short, R.M. Kelly, Methods Enzymol. 330 (2001) 246.
[12] I. Mills, T. Cvitaš, K. Homann, N. Kallay, K. Kuchitsu, IUPAC Physical Chemistry Division-Quantities, Units and Symbols in Physical Chemistry, Blackwell Science, Oxford, 1993.
[13] M. Uhlén, F. Pontén, Mol. Cell Proteomics 4 (2005) 384.
[14] C. Agaton, J. Galli, I.H. Guthenberg, L. Janzon, M. Hansson, A. Asplund, E. Brundell, S. Lindberg, I. Ruthberg, K. Wester, D. Wurtz, C. Höög, J. Lundeberg, S. Ståhl, F. Pontén, M. Uhlén, Mol. Cell Proteomics 2 (2003) 405.
[15] T.L. Constantopoulos, G.S. Jackson, C.G. Enke, J. Am. Soc. Mass Spectrom. 10 (1999) 625.
[16] R.E. Majors, LC–GC Europe 16 (2003) 404.
[17] J.-C. Wolff, C. Eckers, A.B. Sage, K. Giles, R. Bateman, Anal. Chem. 73 (2001) 2605.
[18] K.F. Blom, J. Am. Soc. Mass Spectrom. 9 (1998) 789.
[19] G.E. Begg, J. Biomol. Technol. 10 (1999) 17.
[20] S.A. Lesley, P. Kuhn, A. Godzik, A.M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H.E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L.S. Brinen, M.D. Miller, T.M. McPhillips, M.A. Miller, D. Scheibe, J.M. Canaves, C. Guda, L. Jaroszewski, T.L. Selby, M.A. Elsliger, J. Wooley, S.S. Taylor, K.O. Hodgson, I.A. Wilson, P.G. Schultz, R.C. Stevens, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 11664.
[21] P. Johansson, H. Brumer III, M.J. Baumann, Å.M. Kallas, H. Henriksson, S.E. Denman, T.T. Teeri, T.A. Jones, Plant Cell 16 (2004) 874.