Protein quality assurance at the Center for Eukaryotic Structural Genomics (CESG)

David J. Aceti[1], Grzegorz Sabat[2], Amy C. Harms[2], Brian G. Fox[1], George N. Phillips[1], Jr., John L. Markley[1], and Michael R. Sussman[1]

The Center for Eukaryotic Structural Genomics[1] and the Mass Spectrometry/Proteomics Facility[2], University of Wisconsin-Madison, 433 Babcock Drive, Madison Wisconsin, USA 53706-1549

Abstract

We describe CESG's procedures for quality assurance. All purified proteins from the *E. coli* cell-based pipeline and all protein candidates for structure determination produced from the wheat germ cell-free pipeline are analyzed by MALDI-TOF and ESI mass spectrometry (MS) to confirm target protein identity and integrity and, as relevant, to determine incorporation of selenomethione (Se-Met) or isotopes. Also analyzed are proteins chemically modified by reductive methylation (to enhance crystallization). Cases in which the initial MS results fail to agree with the predicted mass are investigated further through peptide sequencing by tryptic digest/LC-MS/MS. The efficiency of reductive methylation ranged from 85% to greater than 95%. Incorporation of Se-Met was greater than 90% in the great majority of cases while incorporation of stable isotopes (^{15}N and/or ^{13}C) was generally greater than 95%. Incorporation of ^{15}N appears to be as efficient wheat germ cell-free protein synthesis as in *E. coli* - based synthesis. Use of ICP-MS to detect metal ligands resulted in the identification of proteins with bound Ni^{2+}, Fe^{2+} or Ca^{2+}.

Introduction

The Center for Eukaryotic Structural Genomics (CESG), a Protein Structure Initiative center, has focused on the plant *Arabidopsis thaliana* as its initial structural genomics target but has expanded to mouse, human, and other organisms. This and other structural genomics projects aim to describe new protein folds so that, ultimately, examples of all folds found in nature are available to researchers.

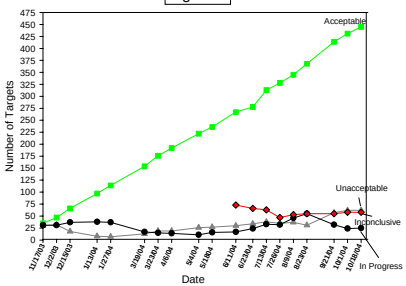
Mass spectrometry has become an extremely useful technique in protein research of all kinds. Electrospray Ionization (ESI) MS can determine the mass of whole proteins to within several Daltons. Proteolytic digestion combined with Matrix-Assisted Laser Desorption/Ionization Time-Of- Flight (MALDI-TOF) MS and Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) can determine amino acid sequence and protein identity.

It was expected that routine use of MS in this project would safeguard against misidentified proteins, determine incorporation of selenomethione, ^{15}N and ^{13}C labels, and provide evidence for post-translational modifications and covalently bound cofactors. As well, MS was expected to be useful in solving less routine, more research-oriented challenges.

Overview of quality assurance by mass spectrometry

Between July, 2003 to October, 2004, 590 protein samples were examined. Of these, 75% were deemed acceptable for structural analysis, 11% were unsuitable, 10% gave inconclusive results, and 4% are currently in progress (Figure 1). Those that were unsuitable were degraded (28%), poorly incorporated with Se-Met (2%) mixtures of two target proteins (3%), insufficiently purified from *E. coli* expression host proteins (2%), or of incorrect identity due to mishandling errors in the production pipeline (65%); the latter were were renamed and re-entered into the pipeline whenever possible.

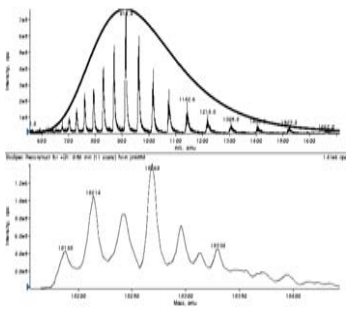
Figure 1



Analysis of efficiency in reductive methylation

Reductive methylation of 14 proteins was carried out with the goal of improving crystallizability. Each lysine and the N-terminal residue are capable of two methylation events, each adding 14 Da to the mass of the protein. Shown in Figure 2 is an ESI-MS spectrum of the protein encoded by the At1g70850.1 locus which has 13 lysines; thus, with the N-terminus,14 residues x 2 sites/residue may be methylated. Complete methylation would result in the addition of 490 Da to the 17812 Da protein for a total of 18302 Da.Peaks representing partially methylated species are seen at 18188 Da, 18214 Da, 18245 Da, 18269 Da, 18297 Da, 18315 Da. The difference between sequential peaks is 26, 31, 24, 28, 17, and 15 Da, indicating successive single or double methylation events. The 18330 Da Peak represents fully methylated protein.

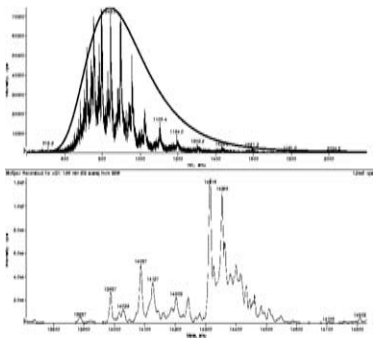
Figure 2



^{15}N incorporation in samples produced by wheat germ cell-free synthesis

^{15}N incorporation in proteins produced by cell-free synthesis is generally as efficient as in cell-based systems. The ESI-MS spectrum of ^{15}N-labeled *Homo sapiens* protein BC059385 shown in Figure 3 shows a peak at 14316 Da (+/- 2.8 Da). 100% incorporation is predicted to result in a 14148 Da protein. If this discrepancy is completely due to unlabeled nitrogens, the efficiency of incorporation is between 91% and 94% (11 to 17 unlabeled nitrogens out of 182 total nitrogens). The large peak at 14355 Da represents protein with a potassium adduct, predicted to add 38 Da. The small peaks at lower MW indicate minor degradation; this is often seen in such samples, which undergo a two week stability test before being submitted for mass spectrometry.

Figure 3



Multimeric proteins discovered by mass spectrometry

Three proteins examined by MALDI-TOF and ESI-MS appear to be multimers, two dimers and one trimer. All appeared at monomer size on SDS-PAGE. Since non-covalent bonds usually dissociate during preparation of samples for mass spectrometry, At5g17110.1 and At1g78780.1 subunits are most likely attached by disulfide bonds; this was supported by 2-mercaptoethanol-induced dissociation of the subunits when visualized by native gel electophoresis (data not shown). At2g28230.1 did not dissociate upon reduction, consistent with a lack of cysteines to form disulfide bonds.

Table 1

Arabidopsis Genome Initiative ID	Expected monomer mass (Da)	Actual Mass by ESI-MS	Ratio of Actual/Expected	# Cys	Comment
At5g17110.1	23765	47548 Da (+/- 10)	2.0	2	to monomer upon reduction
At2g28230.1	19484	57238 (+/- 11)	2.9	0	Does not reduce to monomer. ID confirmed by MS
At1g78780.1	28553	57101 (+/- 11)	2.0	1	to monomer upon reduction

Metal analysis by ICP-MS

Metals were identified by ICP-MS (Inductively-Coupled Plasma Mass Spectrometry in the *Arabidopsis* target proteins listed in Table 2. The presence of an FeS center in At3g17210.1 is supported by the crystal structure solved by CESG (PDB 1Q4R). The presence of iron in At5g51720.1 is suggested by the red color of the protein, absorbance peaks at ~333 nm and ~458 nm, and release of sulfur upon acidification with HCl. The presence of calcium in At3g03410.1 was confirmed by NMR.

Table 2

ID	Predicted Function	Metal Detected
At5g51720.1	unknown	Fe
At3g17210.1	unknown	Ni
At3g03410.1	calmodulin	Ca
At5g11950.1	lysine decarboxylase-like	Ni

Summary

Mass spectrometry analysis has eliminated 54 problem targets, the majority of which were degraded or of incorrectly identified. Three multimers were identified by ESI-MS and later confirmed by other methods. Mass spectrometry was also used to determine Se-Met, ^{15}N, and ^{13}C incorporation, and efficiency of reductive methylation. ICP-MS was useful in detecting metals in several proteins.

Acknowledgements

We would like to thank James Brown for mass spectrometry support, Martin Shafer and Joel Overdier of the UW Engineering Experiment Station for ICP-MS, and the entire CESG team for supplying proteins and helpful advice for these studies.

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (P50 GM 64598).