

# Proteomics of Small Proteins from Plant Tissues

Gregory B. Hurst, Trish K. Lankford, Xiaohan Yang, Ting Li, David J. Weston, Sara M. Allen, Timothy J. Tschaplinski, Gerald A. Tuskan  
Chemical Sciences Division, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge Tennessee



## OVERVIEW

- Several methods for enriching small proteins were evaluated using *E. coli* as a model system.
- Small proteins were enriched from *Arabidopsis thaliana* and analyzed by LC-MS-MS.

## INTRODUCTION

- Small genes and the proteins that they encode can play important biological roles including signaling, development, and mediation of plant-microbe interactions in organisms ranging from bacteria to plants to mammals (Frith *et al.*; Basrai *et al.*; Galindo *et al.*; Hemm *et al.* 2008, 2010; Kastenmeyer *et al.*). However, genes that encode proteins containing <100 residues are difficult to identify reliably solely by DNA sequence analysis (Dinger *et al.*)
- We previously described an approach to identify small-protein-encoding genes in the woody model species *Populus trichocarpa* that relied in part on proteomics to identify small proteins from unfractionated protein extracts (Yang *et al.*)
- To increase the sensitivity of proteomics toward small proteins, we sought to evaluate methods for enriching the low-molecular-weight proteome prior to LC-MS-MS analysis.
- Using *E. coli* as a model system, we evaluated several methods for enriching small proteins from cell lysates. We applied the most promising to fractionation of plant root and shoot tissues from *Arabidopsis thaliana* in order to increase the sensitivity of LC-MS-MS analysis toward small proteins.

## METHODS

### Evaluation of fractionation protocols (Figure 1)

- E. coli* used as a model system
- ACN:** removal of large proteins by acetonitrile precipitation (Aristoteli *et al.*)
- MWCO:** ultrafiltration using molecular weight cutoff filters (Aristoteli *et al.*)
  - 10 kilodalton and 30 kilodalton cutoffs evaluated
- In-Gel:** in-gel digestion of low-molecular-weight regions excised from SDS-PAGE gels (Shevchenko *et al.*)
- GelFree:** fractionation using the GelFree 8100 fractionation system (Protein Discovery, Knoxville TN)
- Full:** no fractionation (full lysate)

### Application to plant tissues (Figure 2)

- Root and shoot tissues from small laboratory-grown *A. thaliana* seedlings flash frozen in liquid nitrogen
- Protein extraction (Damerval *et al.*)
- In-Gel method to enrich small proteins (see above)

Each unfractionated lysate, liquid fraction, or gel slice was:

- digested using trypsin
- analyzed using LC-MS-MS (2D nanoLC interfaced with ThermoFinnigan LTO)
- Peptides and proteins identified using Sequest and DTASelect

Figure 1

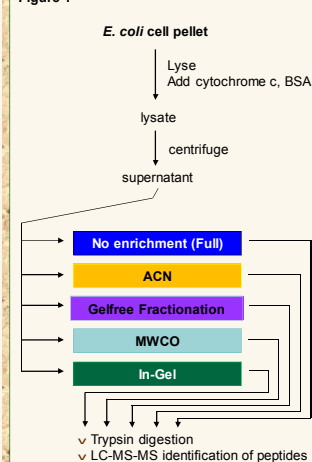
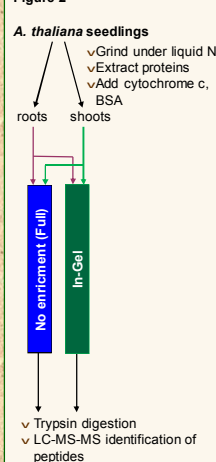


Figure 2



## RESULTS AND DISCUSSION

### Comparison of Small Protein Enrichment Methods

Figure 3 compares SDS PAGE analyses of *E. coli* proteins isolated using the various methods, and from the unfractionated proteome.

- All fractionation methods show depletion of large proteins.
- ACN and MWCO methods appear also to entail significant losses of the small protein complement.
- Low molecular weight fractions from In-Gel and GelFree methods appear to contain highest abundance of small proteins.

Figure 4 shows molecular weight distributions for proteins detected by each method

- Proteins identified by LC-MS-MS from each of the fractionation methods exhibited molecular mass distributions with medians significantly lower than those of unfractionated lysates
- The largest numbers of small protein identifications from LC-MS-MS analysis were obtained from
  - In-gel digestion of the low molecular weight range of SDS-PAGE separated proteins, and
  - the GelFree system

Consistent with SDS-PAGE analysis, the ACN and MWCO methods provided the lowest numbers of identified small proteins.

Figure 5 compares Spectrum Count values for the In-gel digestion (0-20 kDa region) to the unfractionated proteome. While 26 "small" (i.e., <= 100 amino acids in length) proteins were more abundant in the In-gel digestion, 58 small proteins were less abundant in the In-gel digestion. Table 1 summarizes results of this analysis for the various fractionation methods. Additionally, Table 1 shows that the In-Gel and GelFree methods yielded the highest fraction of Spectrum Count from small proteins.

Table 2 lists selected small *E. coli* proteins that were more abundant in one or more enrichment samples than in the unfractionated proteome.

Enrichment of small proteins appears to depend both on the particular protein, and on the fractionation method used. One factor that may affect the results is the participation of small proteins in large complexes such as the ribosome, and how such complexes survive the various fractionation protocols.

Additional improvements may be gained in the future by considering consequences of the smaller numbers of distinct tryptic peptides obtained from digestion of small proteins.

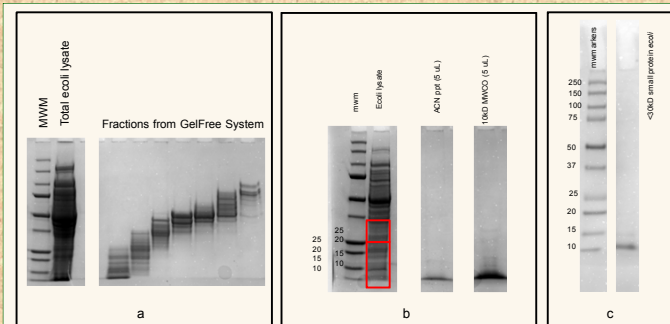


Figure 3. SDS PAGE analyses of (a) fractions from GelFree fractionation system, (b) supernatant from acetonitrile precipitation (ACN ppt) and flow-through from 10 kDa MWCO separation, and (c) flow-through from 30 kDa MWCO separation. Gel slice locations for the In-Gel digestion are shown approximately by the red boxes in panel b; lower box is the 0-20 kDa range, upper box is the 20-35 kDa range. MWM: molecular weight markers.

Table 1. Comparison of Enrichment Methods for Small Proteins from *E. coli*

Isolation Method	Number of proteins that contain:				Proteins Identified in Isolate (average)	n**	Fraction of Total Spectrum Count in Isolate From Proteins with ≤100 AA
	Enriched in Isolate	Depleted or Not Detected in Isolate	Enriched in Isolate	Depleted or Not Detected in Isolate			
ACN	25	64	43	1125	221	3	21%
MWCO 10 kDa	9	74	9	1149	91	3	15%
MWCO 30 kDa	0	81	9	1148	24	1	4%
GelFree fraction 2	26	64	122	1077	449	3	16%
In-Gel digestion, <20 kDa	26	58	80	1089	176	1	25%
unfractionated proteome	-	-	-	-	1236	4	4%

\* Spectrum Count is higher (Enriched) or lower (Depleted) for isolation method compared to unfractionated proteome  
\*\* n = number of replicate LC-MS-MS measurements

Table 2. Selected *E. coli* proteins detected more abundantly following enrichment

L	MW	Gene Symbol	Description
55	6507	rmf	b0953 ribosome modulation factor
63	7281	yaiA	b0389 predicted protein
63	7273	rpmC	b3312 50S ribosomal subunit protein L29
66	7892	glgS	b3049 predicted glycogen synthesis protein
69	7463	cspE	b0623 DNA-binding transcriptional repressor
69	7402	cspC	b1823 stress protein, member of the CspA-family
70	7781	cspG	b0990 cold shock protein homolog, cold-inducible
70	7403	cspA	b3556 RNA chaperone and anti-terminator, cold-inducible
71	8500	rpsU	b3065 30S ribosomal subunit protein S21
72	8250	infA	b0884 translation initiation factor IF-1
77	8639	yedF	b1930 conserved protein, UPF0033 family
84	9704	rpsQ	b3311 30S ribosomal subunit protein S17
85	9119	ptsH	b2415 phosphohistidinophosphatase component of PTS system (Hpr)
90	9226	hupB	b0440 HU, DNA-binding transcriptional regulator, beta subunit
90	9535	hupA	b4000 HU, DNA-binding transcriptional regulator, alpha subunit
97	10387	groS	b4142 Cpn10 chaperonin GroES, small subunit of GroESL
99	10776	yjiS	b3922 conserved protein, UPF0381 family

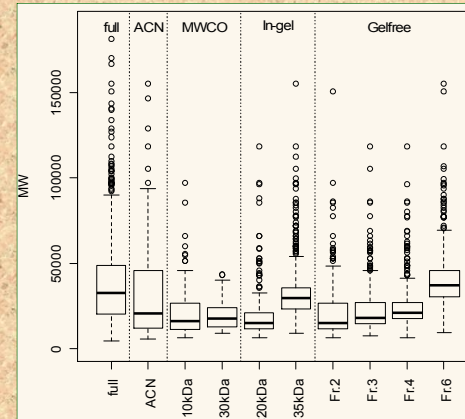


Figure 4. Molecular weight distributions of LC-MS-MS protein identifications resulting from various fractionation protocols applied to the *E. coli* proteome. Boxplots show molecular mass distributions of proteins identified in proteomics measurements. Dark horizontal bars: median molecular mass for identified proteins. Box: 25<sup>th</sup> and 75<sup>th</sup> percentile molecular masses. Whiskers: 1.5 x interquartile range. Circles mark molecular masses of any outliers more extreme than the whiskers.

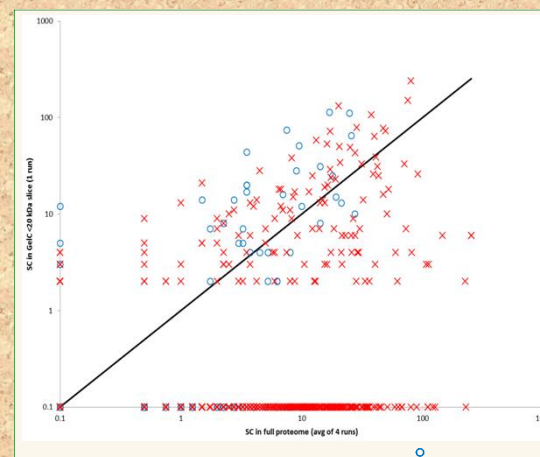
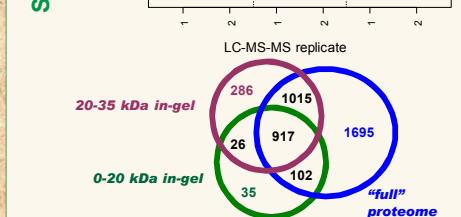
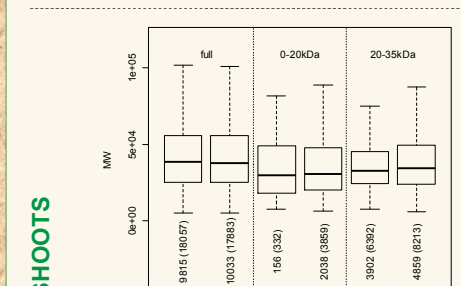
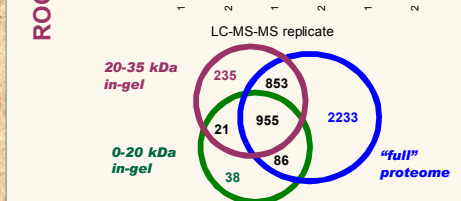
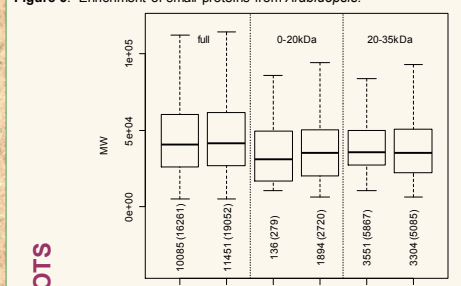


Figure 5. Comparison of Spectrum Count for proteins ≤100 aa (•) and >100 aa (x) for 0-20 kDa gel slice versus unfractionated proteome. Diagonal black line shows equal average Spectrum Count for the two methods. Data points above the diagonal are "enriched" in the gel slice relative to the unfractionated proteome, while data points below the diagonal are "depleted" in the gel slice. Spectrum Count for proteins that were not detected was replaced with a value of 0.1.

### Small Proteins in *Arabidopsis thaliana*

Shoot and root tissues from *Arabidopsis thaliana* were separated by SDS-PAGE, and in-gel digestion performed on slices corresponding to <20 kDa and 20-35 kDa ranges. LC-MS-MS analyses of these fractions and unfractionated protein extracts were performed in duplicate. Figure 6 summarizes the results for molecular weight distributions (boxplots) and numbers of proteins identified in the various fractions (Venn diagrams.)

Figure 6. Enrichment of small proteins from *Arabidopsis*.



## CONCLUSIONS

Results from enrichment of *E. coli* proteins suggests that improved sensitivity toward small proteins is feasible with an MS-based approach, especially with further optimization of the In-Gel and GelFree isolation protocols.

Hemm *et al.* (2008, 2010) have employed other approaches to the detection of small proteins in *E. coli*, and showed that expression of a number of these proteins required subjecting the cells to various stresses.

Several *E. coli* proteins were identified more abundantly in small protein enrichment fractions than in unfractionated proteomes. Among these proteins are several with annotations indicating that their functions are not yet characterized. Evidence for expression of these proteins from LC-MS-MS identification supports improved annotation of the corresponding small genes, and also provides candidates for further studies of the biological functions of these small proteins.

Further investigation of small proteins identified from *Arabidopsis thaliana* will complement our ongoing research, which integrates informatics and experimental approaches for identifying genes that encode small proteins in plants (Yang *et al.*).

## REFERENCES

Aristoteli *et al.*, *J. Proteome Research* 2006, 6, 571.  
Basrai, M.A. *et al.*, *Genome Research* 1997, 7, 768.  
Damerval, C. *et al.*, *Electrophoresis* 1986, 7, 52.  
Dinger, M.E. *et al.*, *PLoS Comput. Biol.* 2008, 4, e1000177.  
Frith, M.C. *et al.*, *PLoS Genetics* 2006, 2, 515.  
Galindo, M.I. *et al.*, *PLoS Biology* 2007, 5, 1052.  
Hemm, M.R. *et al.*, *Mol. Microbiol.* 2008, 70, 1487.  
Hemm, M.R. *et al.*, *J. Bacteriology* 2010, 192, 46.  
Kastenmeyer, J.P. *et al.*, *Genome Research* 2006, 16, 365.  
Shevchenko, A. *et al.*, *Nature Protocols* 2006, 1, 2856.  
Yang, X. *et al.*, *Genome Research* 2011, 21, 634.

## ACKNOWLEDGMENTS

We thank Dr. Karuna Chourey (ORNL) for advice on protein extraction from plant tissues. Research sponsored by the Genomic Science Program, Office of Biological and Environmental Research, U.S. Department of Energy, under contract No DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.