Prokaryotic proteins are more easily identified than eukaryotic ones by Peptide Mass Fingerprinting

Nancy Kendrick[1], Jon Johansen[1], Matt Hoelter[1] and Mary Ann Gawinowicz[2]
[1]Kendrick Laboratories Inc, [2]Columbia University Protein Core Facility

Abstract

A total of 560 proteins cut from silver-stained and 830 proteins cut from Coomassie blue-stained 2D gels were subjected to peptide mass fingerprinting by MALDI-TOF MS for identification over 3 years. Analysis of the collected results showed that the success of this approach is related to the source of the protein. If good mass spectra (GS) are defined as those that show 5 or more peptide peaks, then for silver-stained gels 95% of spectra from archaea proteins are GS, 66% from bacterial, 67% from plasmodia, and only 37% from mammalian proteins are GS. A similar trend was observed for proteins cut from Coomassie blue-stained 2D gels although the success rate was much higher. A proposed explanation stems from differing levels of post-translational modifications between organisms. Negatively charged modifications such as phosphate groups, known to be more abundant on mammalian proteins, would likely enhance silver deposition, giving a false indication of protein amount and interfering with mass spectrometry.

Introduction

MALDI-TOF MS (matrix-assisted laser desorption-ionisation/ time of flight mass spectrometry) may be used to identify proteins excised from acrylamide gels using the method of peptide mass matching. [1-3] In this method the excised polypeptide is digested with a protease and the molecular weights of the resultant peptides exactly determined. This "fingerprint" is matched to the genomic and proteomic databases using software such as MASCOT [4] and the protein identified from the best match. Since the databases are replete with sequenced genomes including human and mouse, this approach is highly successful.

Two-dimensional gel electrophoresis (2DE) is often used to find proteins of interest that change in either amount or mobility with changing experimental conditions or disease states. Excised spots can subsequently be enzymatically digested and analyzed by MALDI-TOF MS. Suitable staining techniques must be compatible with in-gel digestion and mass spectrometric analysis, and must also be sensitive enough to detect the less abundant proteins that are often of the greatest interest. Silver staining is 10-100 times more sensitive than Coomassie Brilliant Blue (CB) and is commonly used for 2-DE analyses. But results have been mixed with regard to protein identification. Silver staining has both been found to be compatible with [5,6] and to interfere with [2,7] mass spectrometry.

Numerous protocols for silver staining exist and it is likely that some methods are more conducive to peptide mass fingerprinting than others. We now report, however, that when silver staining and MS methods are held constant, the species of sample origin is an independent factor in MS success. Successful results are obtained in the order Archaea > Bacteria = Plasmodium >> Mammalian > Drosophila. The results for Coomassie blue staining shows the same pattern with a much higher success rate overall.

Material and methods

Two-dimensional electrophoresis: 2DE was performed using the carrier ampholine method of isoelectric focusing essentially according to the method of O'Farrell [8] with the following modifications. Samples were loaded onto 2.0 by 130 mm IEF tube gels containing 4% acrylamide, 2% IGEPAL, 9.5 urea and either 2.0% pH 3.5-10 ampholines (Amersham Biosciences, Piscataway, NJ) or 2% pH 4-8 ampholines (Gallard-Schlesinger Industries, Inc., Garden City, NY). Isoelectric focusing was carried out for 9600 Vh. After equilibration for 10 min in a buffer containing 10% glycerol, 50

mM fresh DTT, 2.3% SDS and 0.0625 M Tris, pH 6.8, the tube gels were sealed to the top stacking gels overlaying an 8%, 10% or 12% acrylamide slab gel (0.75 mm thick) containing 0.1 % SDS. Electrophoresis was carried out at 12.5 milliamps/gel until the bromophenol blue tracking dye had reached the bottom of the plate, about 4 hrs.

Gel Staining: Silver-stained gels were stained by a variation of the method of O'Connell and Stults [9]. Gels were fixed overnight in 50% methanol/10% acetic acid, rinsed 2 x 10 min with water, sensitized 1 min in sodium thiosulfate (0.2 g/L), again rinsed 2 x 10 min with water, stained for 30 min with silver nitrate (2 g/L), developed ~3 min to desired intensity in a solution of 37% formaldehyde (0.7 ml/L)/potassium carbonate (30 g/L)/sodium thiosulfate (0.01g/L), and the reaction stopped with 2.5% acetic acid for 5 min. Note that this method is similar to the method of Shevchenko et. al. [10].

Coomassie Brilliant Blue R-250 (CB) [11] stained gels were stained overnight with gentle orbital shaking in a solution of 0.013% W/V dye, 50% methanol, and 10% glacial acetic acid. The next morning they were rehydrated for 2 hr in 0.013% W/V CB, 10% glacial acetic acid, and then destained in 10% glacial acid for 4-6 hr. All CB and silver stained gels were dried between sheets of cellophane paper with the acid edge to the left prior to excision of the protein spots for mass spectrometry.

Spot excision, destaining and in-gel digestion:

The silver- or Coomassie blue-stained spots were excised from the dried gels with a clean scalpel blade, transferred to an Eppendorf tube and express shipped to the Columbia University Protein Core Facility. At Columbia the spots were rehydrated in water and the gel slices transferred to new microcentrifuge tubes that had been cleaned by pre-digestion with trypsin. A reduction / alkylation step was eliminated because cysteine residues were consistently alkylated with acrylamide under the electrophoresis conditions used. CB stained acrylamide pieces were washed with 200 µl of 0.05M Tris, pH 8.5/50% acetonitrile for 20 min with shaking. The supernatant was discarded, the wash repeated, and the washed gel pieces dried for 30 minutes in a Speed-Vac concentrator. Then 15 µl of digestion buffer containing 0.06 µg trypsin in 0.025 M tris, pH 8.5 was added and the tube incubated for 20 hr at 32^{o} C.

Silver stained gel spots were destained according to Gharahdaghi et al. [12] and washed with several changes of water until no yellow color was visible, then dehydrated with 200 µl acetonitrile. The acetonitrile was removed and the gel piece completely dried for 30 minutes in a Speed-Vac concentrator. The gel pieces were rehydrated with 0.02 µg modified trypsin and 0.1 µg endoproteinase Lys-C in 12 µl of 0.025 M Tris, pH 8.5, and the tubes placed in a 32^{o} C heating block for 20 hours. For comparison of stains in Table 3, silver stained spots were digested with trypsin only as for CB stained spots. After digestion, the peptides were extracted by adding 50 µl of 50% acetonitrile/2% trifluoroacetic acid (TFA) followed by shaking for 30 min. The supernatant was removed to a clean tube. The extraction was repeated and the combined supernatants dried in a Speed-Vac concentrator.

MALDI-TOF MS: Matrix solution for MALDI analysis was prepared by making 10 mg/ml solution of 4 hydroxy-α-cyanocinnamic acid in 50% acetonitrile/0.1% TFA and adding two internal standards (angiotensin and bovine insulin) to the matrix. The dried digest was dissolved in 4 µl matrix/standard solution and 0.8 µl spotted on the sample plate. When the spot was completely dried, it was washed twice with water and dried again. MALDI TOF MS analysis was performed on the digest using a PerSeptive Voyager DE-RP mass spectrometer (Applied Biosystems, Foster City, CA) in the linear mode [10,13-15]

Results

All mass spectra obtained at the Columbia University Protein Core Facility for Kendrick Labs, Inc (a Contract Research Organization) over a 3-year period (2000-2003) were collected for review. The spectra were grouped based on the kingdom, phylum and species of the original sample. Spectra from unusual species with low numbers were excluded, including 13 silver and 33 CB stained results. Forty-eight spectra from plants and 58 from unnamed species were also excluded.

Some proteins giving strong mass spectra with numerous peptide masses could not be identified. Conversely, some proteins with poor spectra showing only 3-4 peptide masses could be tentatively identified because these peaks matched a protein in the database. Thus, successful identification of a protein was not the best criterion for mass spectrum quality. We defined "Good Spectra" (GS) operationally as those showing 5 or more unambiguous peptide masses and "Poor Spectra" (PS) as those showing 4 or fewer peptide masses after standardized digestion with trypsin followed by MALDI-TOF MS. This designation is based on the presence or absence of enough clear, unambiguous peptide masses, excluding known artifacts and standards, that when interpreted manually by an experienced operator would give a reasonable chance of positively identifying a known protein.

Initially results were broken down into two major groups, prokaryotic and eukaryotic proteins. Proteins from prokaryotes showed an average of 72% GS while those from eukaryotes showed 44% GS, a substantial difference. Further breakdown showed that % GS varied by organism with archea > bacteria = plasmodia > mammals > drosophila as shown in Figure 1.

All plasmodia results are from a single species, the malaria-causing *Plasmodium falciparum*, a single celled eukaryote. The conclusion that *Drosophila* results are the poorest rests on few samples and is regarded as provisional. However, the results for archaea versus bacteria ($p < 0.0003$), and for bacteria versus mammals ($p < 0.0001$) are highly significant (Anova).

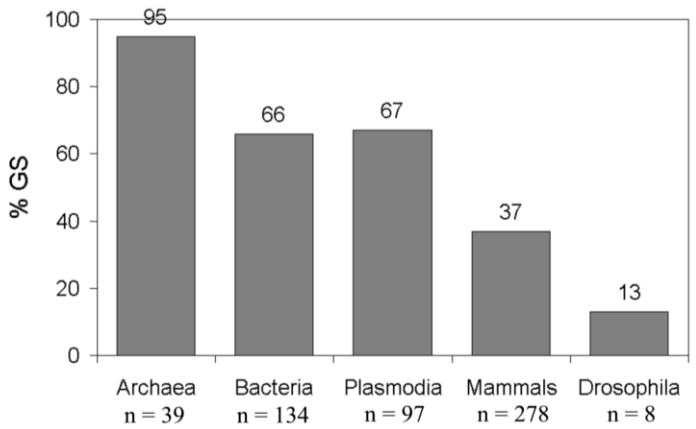

Figure 1. Percent good spectra for silver-stained proteins from 2D Gels. "Good Spectra" are defined as those with 5 or more peptide peaks observable by MALDI TOF mass spectrometry after trypsin diges-tion of a protein spot. The differences between bacteria and mammals and also between archaea and bacteria are highly significant.

To view the results in a different way, percent GS values were calculated for individual species within the groups where the number of proteins analyzed was eight or greater as shown in Table 1 on the next page. This created two subgroups for archaea, six for bacteria and three subgroups for mammals. When the sub-groups values were used to perform Student's t test, then p was < 0.028 for % GS from 3 mammalian species versus 6 bacterial species and p was < 0.01 for 3 mammalian versus 8 prokaryote species (archaea plus bacteria combined). Thus the observed differences between the major groups are statistically significant by a second test.

Archea	# GS	#PS	Total	%GS
Methanococcus	18	1	19	95
Methansarcina acetivorans C2A	18	0	18	100
Ferroplasma acidarmanus	1	1	2	
All archaea combined	**37**	**2**	**39**	**95**
Bacteria	# GS	#PS	Total	%GS
Bacillus subtilis	5	3	8	63
Chlamydia pneumoniae	2	3	5	
Clostridium acetobutylicum	6	1	7	
Escherichia coli	1	0	1	
Enterococcus faecalis	5	0	5	
Geobacter	5	2	7	
Methylomonas	8	0	8	100
Mycobacterium smegmatis	3	0	3	
Neisseria gonorrhoeae	10	7	17	59
Pasteruella multocida	3	1	4	
Pseudomonas	10	7	17	59
Salmonella typhimurium	8	0	8	100
Staphylococcus aureus	22	20	42	52
Streptococcus pneumoniae	0	2	2	
All bacteria combined	**88**	**46**	**134**	**66**
Plasmodium	# GS	#PS	Total	%GS
All *Plasmodium falciparum*	**65**	**32**	**97**	**67**
Mammals	# GS	#PS	Total	%GS
Dog	2	1	3	
Hamster	0	5	5	
Human	54	78	132	41
Mouse	35	55	90	39
Rat	12	36	48	25
All mammals combined	**103**	**175**	**278**	**37**

Table 1. Results for individual species for silver stained gels. The number of GS, PS and % GS for sets of proteins cut from silver stained 2D gels are shown for species where n, the number of samples, was 8 or greater. The % GS values were used to calculate statistical significance between the groups.

Coomassie blue Results: A substantial number of proteins from Coomassie blue-stained 2D gels were subjected to peptide mass finger-printing over the same 4-year period. The % GS was 100% for proteins from archaea, 99% for proteins from bacteria and 89% for proteins from mammalian samples. To view the results for individual species and to generate groups for statistical analysis using Student's t test, the % GS was calculated for those species when n was 8 or greater. This created one subgroup for archaea, five for bacteria and three subgroups for mammals. Table 2 shows the collected data from these results. The same trend is observed, although overall the % GS values are much higher for proteins from Coomassie blue than silver stained gels. Student's t test indicates that the differences are significant with $p < 0.008$ for 5 bacterial species averaging 99% GS versus 3 mammalian species averaging 89% and $p < 0.003$ for % good spectra from 6 prokaryotes versus 3 mammalian species.

Archea	# GS	#PS	Total	%GS
Methanococcus	3	0	3	
Ferroplasma acidarmanus	11	0	11	100
All archaea combined	**14**	**0**	**14**	**100**
Bacteria	# GS	#PS	Total	%GS
Bacillus subtilis	213	2	215	99
Borrelia burgdorferi	1	0	1	
Deinococcus radiodurans	8	0	8	100
Escherichia coli	15	1	16	94
Mycobacterium paratuberculosis	17	1	18	94
Pseudomonas aeruginosa	5	0	5	
Staphylococcus aureus	9	0	9	100
Streptococcus pneumoniae	1	0	1	
Vibrio cholerae	1	0	1	
All bacteria combined	**270**	**4**	**274**	**99**
Mammals	# GS	#PS	Total	%GS
Cow	2	1	3	
Hamster	5	0	5	
Human	238	31	269	88
Mouse	148	17	165	90
Opossum	7	0	7	
Pig	1	1	2	
Rat	83	8	91	91
Sheep	1	0	1	
All mammals combined	**468**	**58**	**527**	**89**
Fruit Fly	**16**	**8**	**24**	**67**

Table 2. Results from individual species for Coomassie blue stained 2D gels. Number of GS, PS and % GS were determined for polypeptide spots cut from CB stained gels. Values for % good spectra were calculated for species where n, the number of spots, was 8 or more and used in Student's t test. The difference between bacteria and mammals was statistically significant.

Examination of Table 1 (silver staining) and Table 2 (CB staining) reveals a striking difference in outcome for the two stains. The silver stain differentially interacts with bacterial versus mammalian proteins given constant load and staining conditions, but in addition, the silver is dramatically interfering with MS relative to CB. However, different amounts of protein are loaded for the two stains. Less protein is applied for silver stain (50 μg) than for CB (200 μg) to maximize resolution and minimize spot spreading. The latter can be a formidable problem because of silver's great sensitivity. Also, as a practical matter, silver stain is preferentially used for samples with too little material for CB.

Because of the different protein loads, it is unclear if insufficient protein is causing a poor result for silver, or if the silver metal is interfering with mass spectrometry. To address this question the same fruit fly lysate for which 7 out of 8 proteins of interest had given no peptide peaks with silver stain, was run with 200 μg applied to duplicate 2D gels. One gel was CB stained and the other gel silver stained. Six

of the seven unidentified spots could be found on both gels. They were excised and subjected to MALDI-TOF MS using standard methods and the spectra examined for number of peptide masses and percent coverage.

The six protein spots were found to contain 8 polypeptides of which 7 yielded GS with CB and 5 yielded GS with silver stain (Table 3). This suggests that protein abundance is an important factor since none showed GS from a silver stained gel with low load. Two of the three polypeptides giving poor spectra with silver staining, #1a (ATP synthase B-subunit) and #5a (cyclophilin) gave relatively poor results with CB while the third #5 (nucleoside diphosphate kinase) was disparate. However four silver stained proteins, #1 (CG9466 gene product C-terminal fragment). #6 (phosphoglycerate mutase) and #7 & #8 (isoforms of alcohol dehydrogenase) showed only modest reductions in peptide mass number and coverage while one protein, #2 (CGp468 gene product C-terminal fragment) showed no change.

Visual examination of the original MS spectra revealed that reduction in peak height was a related and important factor. Figure 2 shows corresponding spectra from silver and CB stained protein spots for three of the proteins in Table 3. The top pair of spectra shows protein #2, which had 13 peptide masses each for spots cut from silver and CB gels. The peak heights of some but not all of the masses are dramatically reduced by silver staining suggesting moderate silver interference. For protein #6 (Fig. 2, middle), for which the number of peptide masses was reduced by just one, the silver quenching of peak height is severe and fairly uniform. Most of the peaks are reduced in size by over 75%. For protein #7 (Fig. 2, bottom), the number of peptide masses is moderately reduced from 16 to 13 by silver staining, but the height of the peptide mass peaks in the spectrum is decreased by over 50%. Silver staining is causing a reduction in peak height for many but not all peptide masses that causes some peaks to disappear from the spectra at low protein l

Spot	Protein Identified	Coomassie blue		Silver	
		# peptides[a]	% coverage	# peptides[a]	% coverage
1	CG9466 g.p.C-terminal fragment[b]	15	20.8	9	14.4
1a	ATP synthase β-subunit	7	21.0	2	5.8
2	CG9468 g.p. C-terminal fragment[b]	13	20.0	13	20.0
5	Nucleoside diphosphate kinase	7	39.9	0	0
5a	Cyclophilin	4	22.0	2	6.2
6	Phosphoglycerate mutase	7	41.3	6	35.0
7	Alcohol dehydrogenase	16	88.7	13	76.6
8	Alcohol dehydrogenase	13	84.0	9	63.3

Table 3. Comparison of MALDI-TOF MS results from six Drosophila protein spots cut from CBB and silver stained 2-DE gels loaded equally with 200 µg protein. The six spots contained 8 proteins total. G.p. = gene product. [a]Peptides assigned to the identified protein based on masses from the MALDI-TOF spectrum with signal to noise greater than 2. [b]All peptides identified originate from the C-terminal region of the gene product. Coverage is calculated from the total gene product. Proteins 7 and 8, both identified as alcohol dehydrogenase, were the same molecular weight of 27 kDa, but separated by about 3 cm in the pI dimension, suggesting they were charge isoforms.

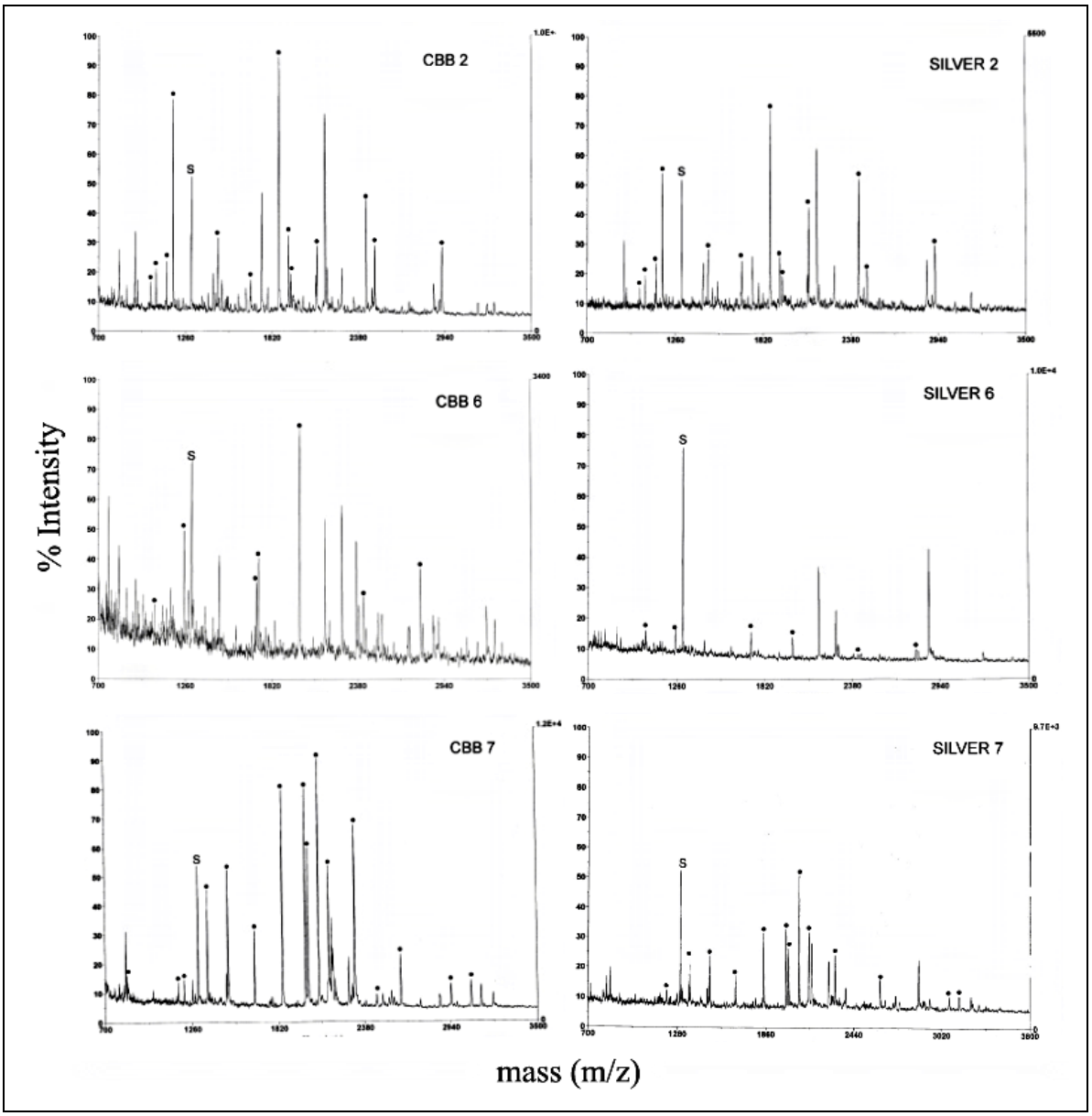

Figure 2. Comparison of MALDI-TOF MS spectra obtained from corresponding silver and CB stained spots #2, 6 and #7 in Table 3. The peak marked with an "S" is the angiotensin internal standard added after the trypsin digestion. The spectra were normalized by adjusting the angio-tensin standard to be of similar intensity. Peptide masses counted in Table 3 are marked by dots.

Discussion

Key observations reported here are that bacterial and archaeal proteins show a significantly higher % GS than mammalian proteins when cut from identically loaded silver-stained 2D gels and also from identically loaded CB stained gels. If silver deposited more readily on mammalian proteins than on archaea or bacterial protein then the former might appear darker and in the range for successful identification by MALDI-TOF MS, but in fact would be too scant.

A mechanism whereby this might happen can be envisioned based on post-translational

modifications. Rabilloud, in a comprehensive review on mechanisms of protein silver staining [16], pointed out that increased silver ion concentration near the protein versus the background area of the gel is a key factor. Silver is known to bind strongly to carboxylic acid groups of proteins through salt formation; other negatively charged post-translational modifications would be expected to enhance silver deposition as well. The sialic acids are a family of 9-carbon negatively charged sugars commonly found on proteins of higher organisms, but rarely found in archaea and bacteria [17,18] that might disproportionately bring down silver.

In addition, phosphate groups might serve this function. The *Homo sapiens* genome, for example, contains 518 known protein kinases that regulate protein phosphorylation, or about 2% of the total number of known genes [19]. *E. coli* contains only four protein kinases [20]. Furthermore, Saiardi et al. discovered a protein phosphorylation pathway through the high energy intermediate of inositol pyrophosphates [21]. They observed that phosphorylation by inositol pyrophosphate 7 appeared to be selective for eukaryotic organisms since no proteins were phosphorylated in bacterial extracts under conditions that worked well for mammalian tissue extracts. Olsen et. al using sophisticated mass spectrometry methods found that at least 70% of proteins are phosphorylated during the cell cycle of cultured HeLa cells [22], an amount far higher than that observed in bacteria and archaea [23].

A second possibility, supported by the data in Fig. 2 and Table 3, is that silver binding to mammalian proteins disproportionately decreases their ability to be good substrates for mass spectroscopy. The fact that % GS for CB stained bacterial proteins (99%) is significantly higher than for similarly stained mammalian proteins (89%) suggests that silver exacerbates an already existing problem. Presumably post-translational modifications such as glycosylation and phosphorylation would be the culprits in this case as well.

Success rates found by others are consistent with our results. In 1996, Shevchenko et. al [10] presented results for 10 yeast proteins analyzed from bands on an SDS slab gel, and found a GS rate of 60%, similar to the other unicellular organisms (except archaea) reported here. In mouse liver, O'Connell and Stults [9] found that 22 out of 35 protein spots analyzed gave little or no signal (37% GS), in agreement with 37% GS we observed for mammals.

Our results suggest that the specific protein(s) used to validate new proteomics methods should be chosen with care. Bovine serum albumin is the most common test protein because of ease of availability in a pure form, and its acceptance as a universal standard. However one protein will never represent the variety of proteins present in a cell, or the variety encountered in core facilities using the test. A set of several bacterial and mammalian cellular proteins would better test the robustness of any general proteomics method.

References

1. Bakhtiar R., Nelson R. W., *Mol Pharmacol*. 2001, 60, 405-415.
2. Gevaert K., Vandekerckhove J., *Electrophoresis*. 2000, 21, 1145-1154.
3. Michael Kinter N. E. S. Protein Sequencing and Identification Using Tandem Mass Spectrometry: WileyEurope; 2000.
4. Perkins D. N., Pappin D. J., Creasy D. M., Cottrell J. S., *Electrophoresis*. 1999, 20, 3551-3567.
5. Alge C. S., Suppmann S., Priglinger S. G., Neubauer A. S., May C. A., *et al.*, *Invest Ophthalmol Vis Sci*. 2003, 44, 3629-3641.
6. Yan J. X., Wait R., Berkelman T., Harry R. A., Westbrook J. A., *et al.*, *Electrophoresis*. 2000, 21, 3666-3672.
7. Richert S., Luche S., Chevallet M., Van Dorsselaer A., Leize-Wagner E., *et al.*, *Proteomics*. 2004, 4, 909-916.
8. O'Farrell P. H., *J Biol Chem*. 1975, 250, 4007-4021.
9. O'Connell K. L., Stults J. T., *Electrophoresis*. 1997, 18, 349-359.
10. Shevchenko A., Wilm M., Vorm O., Mann M., *Anal Chem*. 1996, 68, 850-858.

11. Burgess-Cassler A., Johansen J. J., Santek D. A., Ide J. R., Kendrick N. C., *Clin Chem*. 1989, 35, 2297-2304.

12. Gharahdaghi F., Weinberg C. R., Meagher D. A., Imai B. S., Mische S. M., *Electrophoresis*. 1999, 20, 601-605.

13. Henzel W. J., Billeci T. M., Stults J. T., Wong S. C., Grimley C., *et al.*, *Proc Natl Acad Sci U S A*. 1993, 90, 5011-5015.

14. Mann M., Hojrup P., Roepstorff P., *Biol Mass Spectrom*. 1993, 22, 338-345.

15. Cottrell J. S., *Pept Res*. 1994, 7, 115-124.

16. Rabilloud T., *Electrophoresis*. 1990, 11, 785-794.

17. Jung E., Veuthey A. L., Gasteiger E., Bairoch A., *Proteomics*. 2001, 1, 262-268.

18. Apweiler R., Hermjakob H., Sharon N., *Biochim Biophys Acta*. 1999, 1473, 4-8.

19. Manning G., Whyte D. B., Martinez R., Hunter T., Sudarsanam S., *Science*. 2002, 298, 1912-1934.

20. Krupa A., Abhinandan K. R., Srinivasan N., *Nucleic Acids Res*. 2004, 32, D153-155.

21. Saiardi A., Bhandari R., Resnick A. C., Snowman A. M., Snyder S. H., *Science*. 2004, 306, 2101-2105.

22. Olsen, J.V., Vermeulen M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg,E.A., Brunak, S. Mann, M. Science signaling, 2010. 3(104): p. ra3.

23. Kennelly P. J., *Biochem J*. 2003, 370, 373-389.

Acknowledgements

We thank Yelena Milgrom for technical assistance with MALDI-TOF MS, Gerardus Tromp (Wayne State University) for help with statistical analyses, and Larry Grossman (Wayne State University) for critical reading of the manuscript.