The Proteomics Research Group Qualitative Proteomics Challenge 2007

A common challenge for proteomics laboratories today is determining qualitative differences among samples. Not infrequently, the samples are putatively the same but produce different results in functional or biological assays. Often, there are questions about the identities of the proteins present and any differences in these proteins among the samples submitted for analysis. Resolving these issues can require the optimization of methods to maximize sequence coverage in addition to manual examination o f the data.

The primary goal of this study is to document the breadth of approaches used by the ABRF community and highlight the type of information obtained. Participants will be asked to:
. Identify the proteins present in two similar samples
. Determine any qualitative differences between the proteins in the samples
. Provide information about methods used to maximize protein sequence coverage

When submitting their results, participants will be asked to report both the identities of the major protein components in the samples and any alterations that were detected.

The study samples consist of separate preparations of related human proteins that were expressed in *E. coli* and affinity purified. The cloning vectors that were used produced an N-terminal His_6-tagged fusion protein that included a thrombin cleavage site in the His-tag. In both preparations, the sequence GSHM was retained with the protein after thrombin cleavage to remove the His-tag. 5 and 10 µg of protein were supplied, respectively.

Approach

In order to allow a wide community access to mass spectrometric service, this laboratory needs to distinguish first pass experiments which are typically sufficient in proteomic service, and advanced experiments which are part of extensive collaborative projects. The methods of gel electrophoresis and peptide mapping applied to the challenge are standard in our service. Extensive further analysis was intentionally not performed for this study for economical reasons.

Results

The samples were separated on mini-2D-PAGE using pH strips 3-10. Representative spots were excised as documented in the figure below. All spectra share a number of significant peaks as is exemplified on a few spots. The protein was digested in the gel using trypsin and the peptides were measured with MALDImicroMX (Waters; *m/z* 800-3500). All spectra could be assigned to RAGE protein (Q15109; 42803 Da). Its large peptides were not detected in this mass window. For verification of the assignment several masses on the A3-digest were subjected to manual nanoMS/MS using the $Esquire_{3000}$ ion trap. For *m/z* 1525.96 the expected sequence was

confirmed. Other data obtained in the time frame of the experiment were not as instructive. Further attempts were not made, although plenty of material is still available for advanced analysis.

Reasons for the appearance of several spots on the gel are likely the presence of RAGE at different lengths as well as the formation of stable multimers in the original sample. For the chain-like spots at 40 kDa, carbamidomethylation is responsible which is commonly associated with the PAGE process.

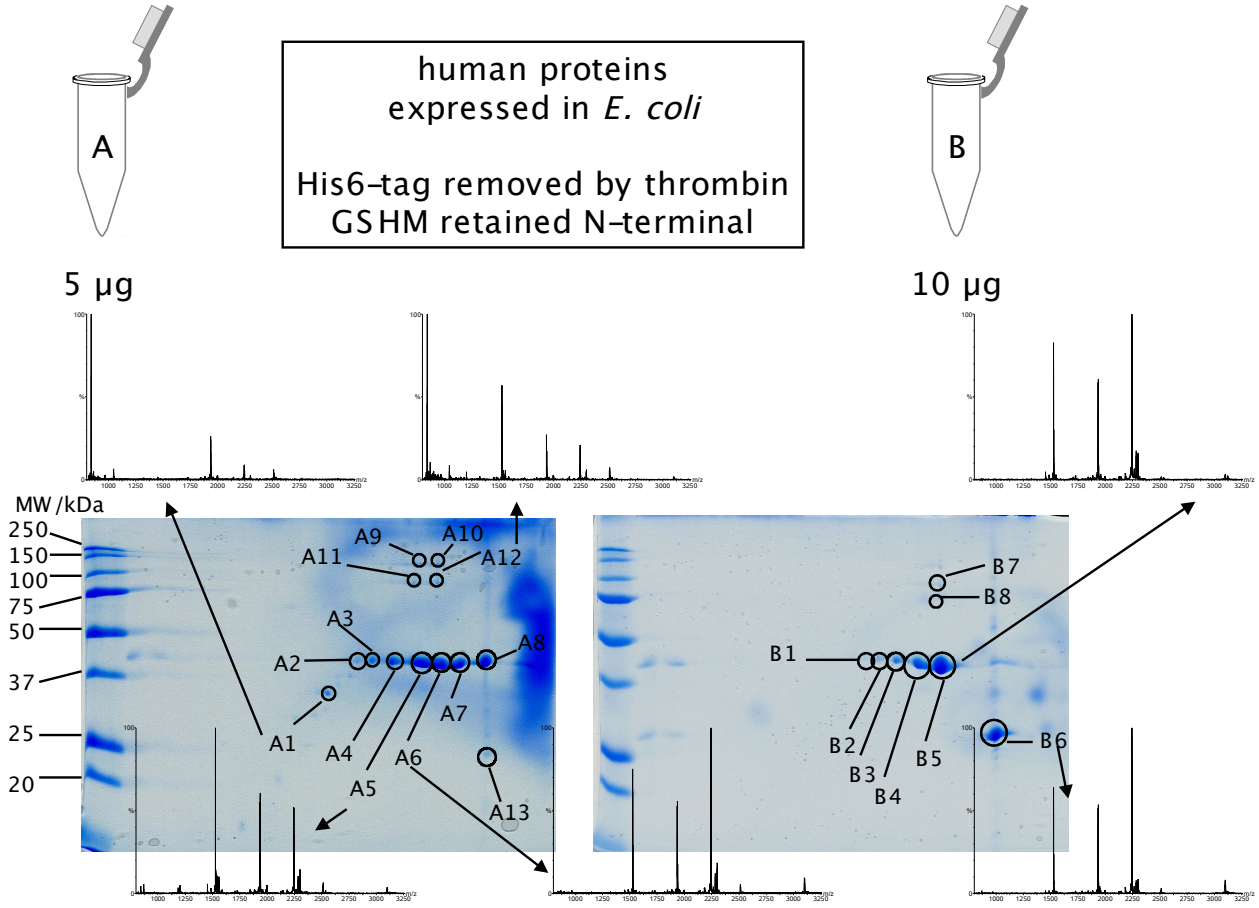

Fig.: Overview of the experimental results.

Details

Data analysis was performed using Mascot in house. No extra attempt was made to consider terminal sequence variations. The information given below corresponds to the output of the algorithm.

Table 1: Mascot output for the analyses of the different spots according to Q15109. Modifications: carbamidomethylation C, methionine oxidation M. Sequence coverage is given at the bottom of the column in %.

In A1 m/z 1525.8 is missing. It is suspected that the minor peaks labelled with m are not derived from the sequence and that the protein on this spot is N-terminally shorter by possibly up to 77 residues. MS/MS on *

Mr	Peptide	Md.	A1 A2	A3	A4	A5 A6	A7 A8 A9	A10	A11	A12	A13
867.54	IGEPLVLK 30-37		m	X	X	X	X	X	X	X	X
1175.70	LNTGRTEAWK 53-62							X	X		
1451.85 1451.66	IGEPLVLKCKGAPK 30-43 CQAMNRNGKETK 99-110	 C,M	m	X	 X	 X	X	X		 X	
1524.77	VLSPQGGGPWDSVAR* 63-77			X	X	X	X	X	X	X	X
1876.89	GGDPRPTFSCSFSPGLPR 199-216			X							
1933.91	GGDPRPTFSCSFSPGLPR 199-216	C	X	X	X	X	X	X	X	X	X
2126.08	HPETGLFTLQSELMVTPAR 180-198					X					
2142.08	HPETGLFTLQSELMVTPAR 180-198	M	X	X	X	X				X	X
2241.22	VLPNGSLFLPAVGIQDEGIFR 78-98		X	X	X	X	X	X	X		
2282.18	RHPETGLFTLQSELMVTPAR 179-198			X	X	X	X				
2298.18	RHPETGLFTLQSELMVTPAR 179-198	M	X	X	X	X	X	X	X	X	
2511.32	VYQIPGKPEIVDSASELTAGVPNK 117-140		X	X	X	X	X	X	X	X	X
2766.49	VRVYQIPGKPEIVDSASELTAGVPNK 115-140		X		X						
3097.52	VGTCVSEGSYPAGTLSWHLDGKPLVPNEK 141-169	C	X	X	X	X	X			X	
			28	34	36	36	36	28	28	29	21

Mr	Peptide	Md.	B1	B2	B3	B4 B5	B6	B7 B8
867.54	IGEPLVLK 30-37		X	X	X	X	X	X
1175.70	LNTGRTEAWK 53-62							
1180.67	KPPQRLEWK 44-52				X			
1451.85 1451.66	IGEPLVLKCKGAPK 30-43 CQAMNRNGKETK 99-110	 C,M		 X	 X	 X	 X	
1524.77	VLSPQGGGPWDSVAR* 63-77		X	X	X	X	X	X
1876.89	GGDPRPTFSCSFSPGLPR 199-216							
1933.91	GGDPRPTFSCSFSPGLPR 199-216	C	X	X	X	X	X	X
2126.08	HPETGLFTLQSELMVTPAR 180-198			X	X	X	X	
2142.08	HPETGLFTLQSELMVTPAR 180-198	M			X	X	X	
2241.22	VLPNGSLFLPAVGIQDEGIFR 78-98		X	X	X		X	
2282.18	RHPETGLFTLQSELMVTPAR 179-198			X	X	X	X	
2298.18	RHPETGLFTLQSELMVTPAR 179-198	M		X	X	X	X	
2511.32	VYQIPGKPEIVDSASELTAGVPNK 117-140		X	X	X	X	X	X
2766.49	VRVYQIPGKPEIVDSASELTAGVPNK 115-140				X		X	
3097.52	VGTCVSEGSYPAGTLSWHLDGKPLVPNEK 141-169	C		X	X	X	X	
			21	36	39	31	36	28

Table 2: RAGE sequence as obtained from SwissProt. Parts in red correspond to peptides found with MALDI-TOF.

```
        10          20          30          40          50          60
MAAGTAVGAW VLVLSLWGAV VGAQNITARI GEPLVLKCKG APKKPPQRLE WKLNTGRTEA

        70          80          90         100         110         120
WKVLSPQGGG PWDSVARVLP NGSLFLPAVG IQDEGIFRCQ AMNRNGKETK SNYRVRVYQI

       130         140         150         160         170         180
PGKPEIVDSA SELTAGVPNK VGTCVSEGSY PAGTLSWHLD GKPLVPNEKG VSVKEQTRRH

       190         200         210         220         230         240
PETGLFTLQS ELMVTPARGG DPRPTFSCSF SPGLPRHRAL RTAPIQPRVW EPVPLEEVQL

       250         260         270         280         290         300
VVEPEGGAVA PGGTVTLTCE VPAQPSPQIH WMKDGVPLPL PPSPVLILPE IGPQDQGTYS

       310         320         330         340         350         360
CVATHSSHGP QESRAVSISI IEPGEEGPTA GSVGGSGLGT LALALGILGG LGTAALLIGV

       370         380         390         400
ILWQRRQRRG EERKAPENQE EEEERAELNQ SEEPEAGESS TGGP
```