

Abstract

The 2008 ABRF Proteomics Research Group Study offers participants the chance to participate in an anonymous study to identify qualitative differences between two protein preparations. We used a series of proteomics methods to identify the major protein components of the two samples. Analysis by SELDI MS and by SDS-PAGE both showed the presence of two major protein components in Sample A and four in Sample B. Peptide Mass Fingerprinting (PMF), using the SELDI reader, gave putative protein identification as human Advanced Glycation (or “Glycosylation” Endproduct-specific Receptor (RAGE). MS/MS analysis corroborated this result. Sequence comparisons show that both Sample A and B contain two identical fragments of the RAGE protein, with Sample B containing two extra unique fragments. In-solution digests allowed for an increase of sequence coverage from 79.29% to 91.6%.

Methods & Results

SDS-PAGE ANALYSIS

The first set of Samples A and B were run on 1-D SDS-PAGE. From Sample A two protein bands can be seen and from Sample B four bands are present. Their approximate masses are 70kDa, 33kDa, 26kDa, 15kDa. The two upper bands of Sample B matched the masses of the Sample A protein bands. From Sample A two protein bands can be seen and from Sample B four bands are present. Their approximate masses are 70kDa, 33kDa, 26kDa, 15kDa. The two upper bands of Sample B matched the masses of the Sample A protein bands.

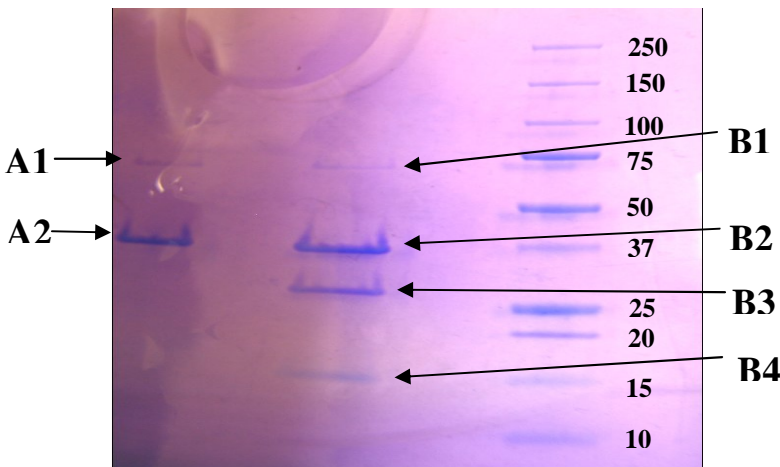


Figure 1. SDS-PAGE of Sample A and Sample B.

Samples were diluted straight into 20ul of XT Sample Buffer (Bio-Rad), heated for 10 minutes at 80°C and run alongside Precision Plus All Blue Protein Standards (Bio-Rad), on a 4-12% Bis-Tris XT Criterion gel with MES buffer (Bio-Rad). The gel was stained with Coomassie Safe Blue Stain (Bio-Rad) and destained with water.

SELDI MS ANALYSIS

The second set of samples were analyzed by SELDI MS, spectra are shown in Figure 2. Peaks corresponding to the masses of the bands on the SDS-PAGE are shown in Figure 2. Although MALDI typically results in single-charged ions, one double-charged peak for the 33kD peak can be seen. The masses do not match those of the bands from the gel precisely, however this shift is commonly seen when comparing MS data to SDS-PAGE.

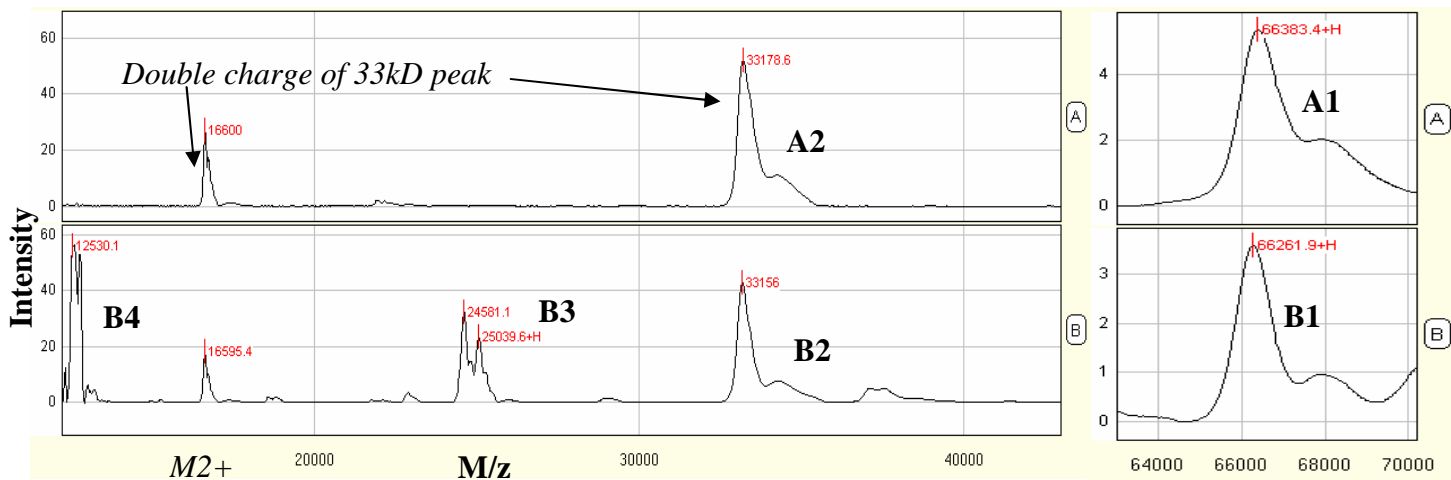


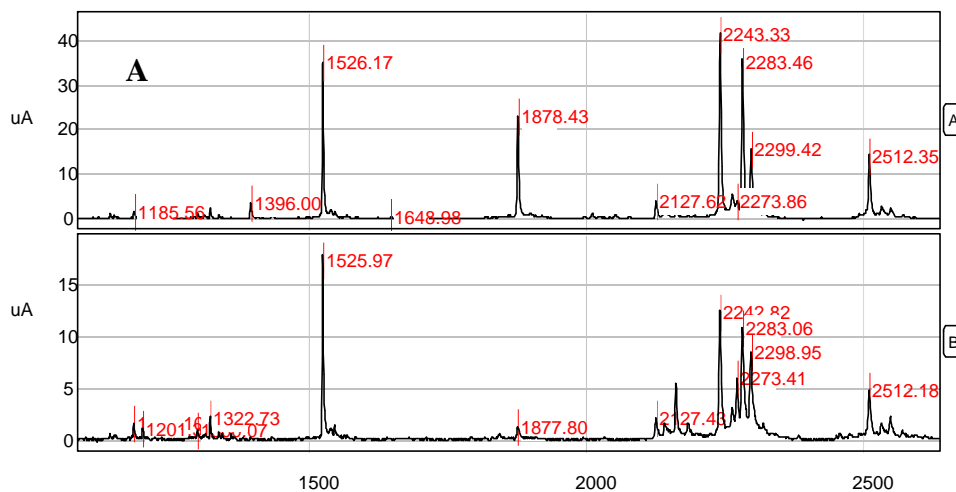
Figure 2. SELDI MS of Sample A and Sample B

Figure 2. SELDI MS of Sample A and Sample B

Samples A and B were dissolved in 20ul of water, then 1ul was spotted onto an NP20 ProteinChip array. The spot was allowed to dry, then matrix was added. The ProteinChip was read in the PBSIIc SELDI reader.

Peptide Mass Fingerprinting

Following in-gel digestion, peptide mass fingerprint (PMF) data was collected from the two major SDS-PAGE bands A2 and B2 using SELDI MS, see Figure 3. From the Mascot database, peptides were identified as belonging to the human RAGE protein, with 41% sequence coverage. PMF is a relatively low confidence technique requiring further validation, therefore we confirmed this identification by MS/MS analysis.



B

Q5JP24_HUMAN **Mass: 25586** **Score: 68** **Expect: 0.021** **Queries matched: 5** Advanced Glycation Endproduct-specific Receptor (Fragment).- Homo sapiens (Human).

- Number of mass values searched: **7** Number of mass values matched: **5**
- Sequence Coverage: **41%**
- Matched peptides shown in **Bold**

1 MAAGTAVGAW VLVLSLWGAV VQAQNTARI GEPLVLKCKG APKKPPQRLE **51** WKLNTGRTEA
WKVLSPOGGG PWDSVARVLP NGSFLPAVG IQDEGIFRCQ 101 AMNRNGKETK SNYRVRVYQI
PGKPEIVDSA SELTAGVPNK VVEESRRSRK 151 RPCEQEVGTC VSEGSYPAGT LSWHLDGKPL
 VPNEKGVSVK EQTRRHPETG **201** LFTLQSELMV TPARGGDRP TFSCSFSPGL PRHRAL

C

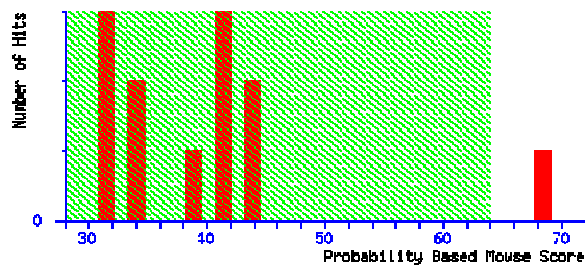


Figure 3. SELDI PMF results from in-gel digest of bands A2 and B2.

2ul of digests from excised bands A2 and B2 were applied to a Normal Phase (NP20) ProteinChip. After allowing the digest to dry, 2x1ul of 20% CHCA matrix was applied to each spot. The samples were analyzed by the SELDI PCS 4000 instrument with the following instrument settings: Mass range 0-10kDa, focus mass 1.5kDa, laser setting 300nJ. Following data acquisition, each spectrum was internally calibrated and then the unique peptide masses were used to search the Matrix Science Mascot database. Screen shot of the Mascot search is shown above: A) SELDI spectra of PMF analysis of bands A2 and B2, B) Sequence information data, C). Probability scores for the Mascot search.

TOF MS

Analysis of samples A and B, dissolved in water and spotted onto NP20 ProteinChips, as described above, were then run in MS mode on the QSTAR XL fitted with the ProteinChip Interface revealing identical spectra, suggesting that these are the same protein (protein fragment).

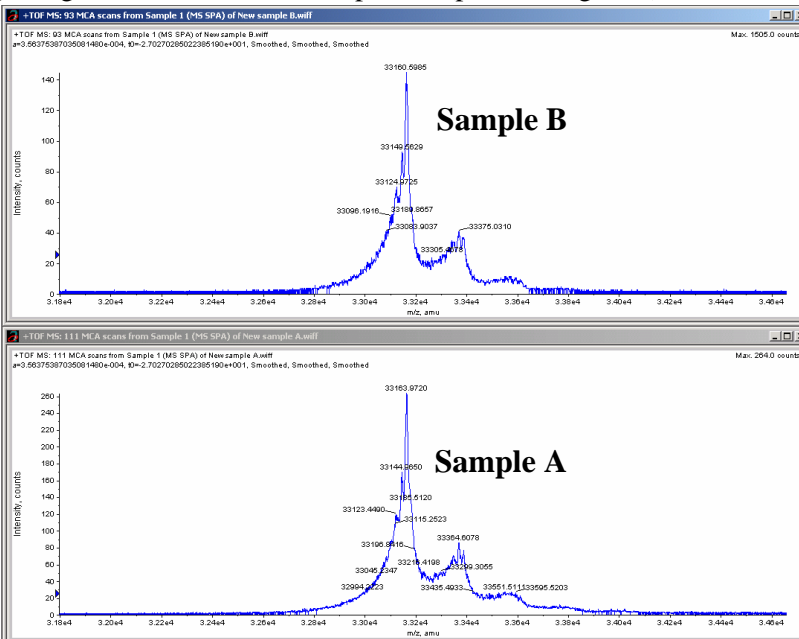


Figure 4. TOF-MS of Sample A and Sample B

Samples A and B were dissolved in 20ul of water. 1ul was then spotted onto an NP20 ProteinChip array, the spot was allowed to dry, and matrix was then added. The ProteinChip was read in the QSTAR XL instrument fitted with a ProteinChip interface.

MS/MS

The in-gel trypsin digests from each band on the SDS-PAGE, Figure 1, were spotted onto NP20 ProteinChips. MS/MS analysis was performed on a QSTAR XL Quadrupole MS instrument, fitted with a ProteinChip Interface. Sample A exhibited two bands on non-reducing SDS-PAGE. Band A2 was identified as the 33,168.85 Da extracellular domain of Advanced Glycation Endproduct-specific Receptor (309 amino acids, 79.29% coverage by MS/MS), see figures 5(a) and 6. The first four amino acids for all the bands were from a plasmid for expression in E.coli (see Discussion).

Band A1 showed the same tryptic fragments as the Band A2. MW of the band A1 is approximately 2-fold higher than the MW of the band A2. Most likely, the band A1 is a storage artifact of the bacterially-expressed 33.2 kDa protein (spontaneous cross-linking of free cysteines under aerobic conditions). This is also seen in sample B, bands B2 and B1.

Sample B exhibited four bands on non-reducing SDS-PAGE. The tryptic map of band B2 was found to be identical to the tryptic map of the band A2. Most likely, this is the same protein, i.e., the 33,168.85 Da extracellular domain of Advanced Glycation Endproduct-specific Receptor (309 amino acids, 79.29% coverage by MS/MS), as previously shown by TOF-MS, see figure 5(b). Proteins in the bands A2 and B2 have identical N-termini and identical C-termini, and show no difference in their peptide maps.

When in-gel and in-solution digest data is combined, there is an increase of overall amino acid sequence coverage from 79.29% to 91.26% and thus an increase in confidence for protein identification. Band B3 was unique for the sample B. It was identified as a 24,585.18 Da truncated extracellular domain of Advanced Glycation Endproduct-specific Receptor (225 amino acids, 71.56% coverage by MS/MS), see figure 7.

Band B4 was unique for the sample B. It was identified as a 12,524.39 Da truncated extracellular domain of Advanced Glycation Endproduct-specific Receptor (114 amino acids, 62.28% coverage by MS/MS), see figure 8.

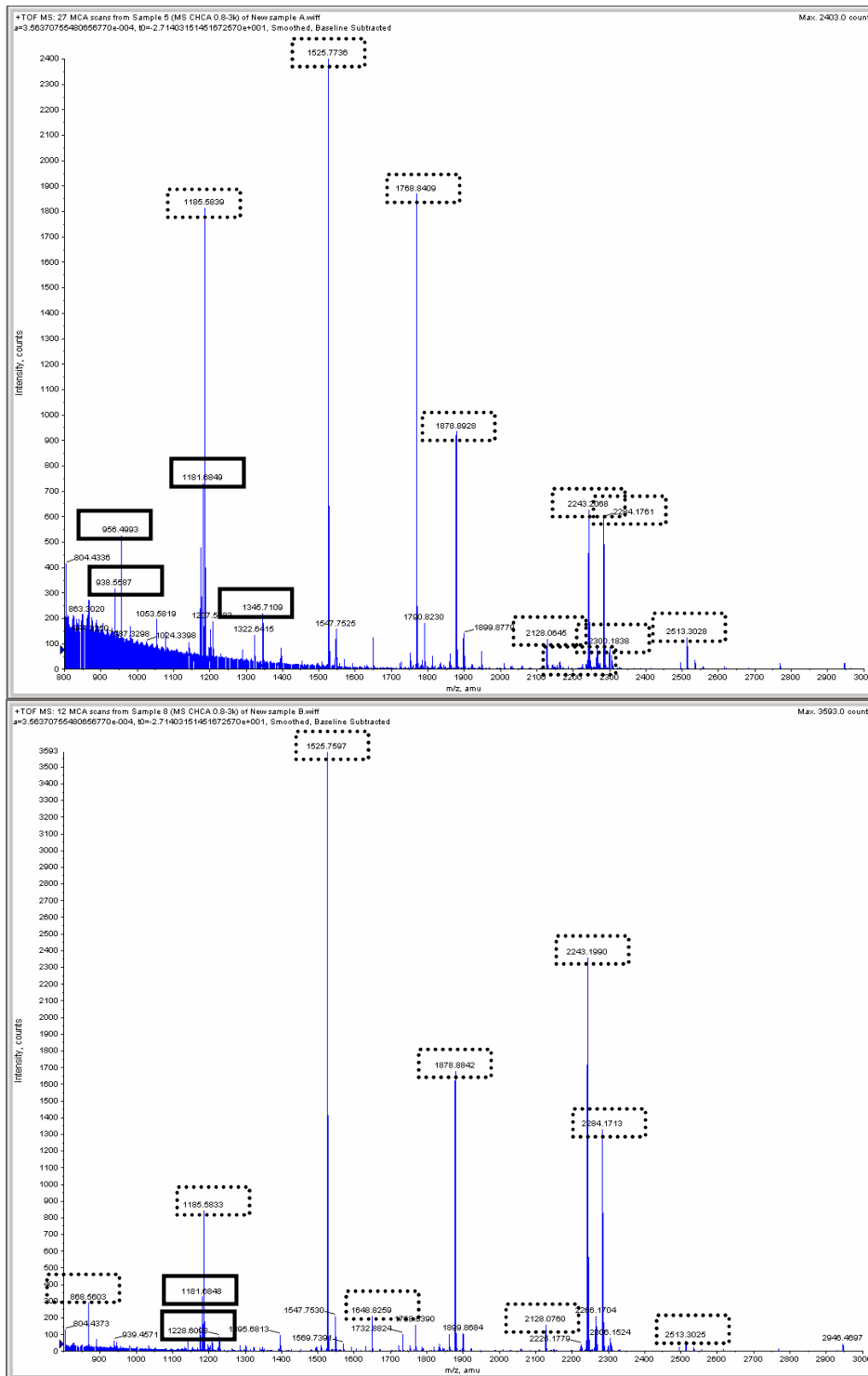


Figure 5. Screen shots of in-solution digest peptide analysis, 0.8-3kDa.

Samples A and B were dissolved in 20ul of water and 1ul spotted onto an NP20 ProteinChip array, then the spot was allowed to dry and matrix added. The ProteinChip was read in the QSTAR XL instrument fitted with a ProteinChip interface. A) Sample A and B) Sample B. A box with a dashed line indicates that this peak is exactly the same as one sequenced in the in-gel digest and matches the theoretical digest of human RAGE. Only new ions were sequenced, i.e., those that were absent in the in-gel digest but could be detected in the in-solution digest. We believe they may be fragments with missed trypsin cleavages. They are marked with solid rectangles.

10	20	30	40	50	60	70
<u>GSHMAQ</u> NITA	RIGEPLVLK C	kgapk KPPQR	LEWKLNTGRT	EAWKVLSPQG	GGPWDSVARV	LPNGSLFLPA
80	90	100	110	120	130	140
VGIQDEGIFR	CQAMNRNGKE	TKSNYRVRVY	QIPGKPEIVD	SASELTAGVP	NKVGTCVSEG	SYAGTLSWH
150	160	170	180	190	200	210
LDGKPLVPNE	KGVSVKEQTR	RHPETGLFTL	QSELMVTPAR	GGDPRPTFSC	SFSPGLPR hr	alrtapiqpr
220	230	240	250	260	270	280
VWEPVPLEEV	QLVVEPEGGA	VAPGGTVTLT	CEVPAQSPQ	IHWMKDGVPL	PLPPSPVLIL	PEIGPDQGT
290	300	310	320	330	340	350
YSCVATHSSH	GPQESRAVSI	SIIEPGEEGP	tagsvgs g l	gtl al algi l	ggl ta alli	gvilw qrr qr
360	370	380				
rgeerkapen	qeeeeerael	nqseepeage	sstggp			

Figure 6. Peptide sequence coverage of 79.29% from band A2.

Sample A was dissolved in 20ul of water and 1ul spotted onto an NP20 ProteinChip array, then the spot was allowed to dry and matrix added. The ProteinChip was read in the QSTAR XL instrument fitted with a ProteinChip interface. Sequence of bold underlined peptide is predicted from *de novo* sequencing. Biologically relevant amino acid sequence (AQNITA) starts from Ala5. The protein was identified in the Swiss-Prot database as Q15109, Advanced Glycosylation Endproduct-specific Receptor. Tryptic peptides identified by MS/MS with high confidence are bold red (in-gel digestion) or bold green (in-solution digestion). AspN peptide identified by MS/MS with high confidence is bold pink (in-solution digestion)

10	20	30	40	50	60	70
<u>GSHMAQ</u> NITA	RIGEPLVLK C	KGAPK KPPQR	LEWKLNTGRT	EAWKVLSPQG	GGPWDSVARV	LPNGSLFLPA
80	90	100	110	120	130	140
VGIQDEGIFR	CQAMNRNGKE	TKSNYRVRVY	QIPGKPEIVD	SASELTAGVP	NKVGTCVSEG	SYAGTLSWH
150	160	170	180	190	200	210
LDGKPLVPNE	KGVSVKEQTR	RHPETGLFTL	QSELMVTPAR	GGDPRPTFSC	SFSPGLPR HR	ALRTAPIQPR
220	230	240	250	260	270	280
VWEPVPLEEV	QLVVE ←	VAPGGTVTLT	CEVPAQSPQ	IHWMKDGVPL	PLPPSPVLIL	PEIGPDQGT
290	300	310				
YSCVATHSSH	GPQESRAVSI	SIIEPGEEGP	C terminus			

Figure 7. Peptide sequence coverage of 71.56% from band B3.

Sample B was dissolved in 20ul of water and 1ul spotted onto an NP20 ProteinChip array, then the spot was allowed to dry and matrix added. The ProteinChip was read in the QSTAR XL instrument fitted with a ProteinChip interface. The first four amino acids come from an expression plasmid (see “RAGE Protein” note before Conclusions). Sequence of bold underlined peptide is predicted from *de novo* sequencing. Biologically relevant amino acid sequence (AQNITA) starts from Ala5. Protein was identified in the SwissProt database as Q15109, Advanced Glycation Endproduct-specific Receptor. Tryptic peptides identified by MS/MS with high confidence are bold red (in-gel digestion) or bold green (in-solution digestion).

10	20	30	40	50	60	70
<u>GSHMAQ</u> NITA	RIGEPLVLK C	KGAPK KPPQR	LEWKLNTGRT	EAWKVLSPQG	GGPWDSVARV	LPNGSLFLPA
80	90	100	110	120	250	260
VGIQDEGIFR	CQAMNRNGKE	TKSNYRVRVY	QIPGKPEIVD	SASELTAGVP	CEVPAQSPQ	IHWMKDGVPL
270	280	290	300	310		
PLPPSPVLIL	PEIGPDQGT	YSCVATHSSH	GPQESRAVSI	SIIEPGEEGP	C terminus	

Figure 8. Peptide sequence coverage of 62.28% from band B4.

Sample B was dissolved in 20ul of water and 1ul spotted onto an NP20 ProteinChip array, then the spot was allowed to dry and matrix added. The ProteinChip was read in the QSTAR XL instrument fitted with a ProteinChip interface. The first four amino acids come from an expression plasmid (see “RAGE Protein” note before Conclusions). Sequence of bold underlined peptide is predicted from *de novo* sequencing. Biologically relevant amino acid sequence (AQNITA) starts from Ala5. Protein was identified in the SwissProt database as Q15109, Advanced Glycation Endproduct-specific Receptor. Tryptic peptides identified by MS/MS with high confidence are bold red (in-gel digestion) or bold green (in-solution digestion).

DISCUSSION

We suspect that the source protein, RAGE, or Receptor for Advanced Glycation Endproducts, was expressed in E.Coli with a His-tag in pET28 or similar plasmid, purified using Ni-NTA or a similar resin, then the His tag was cleaved by thrombin, leaving the tetrapeptide GSHM upstream of the RAGE coding sequence.

NB: SwissProt identifies the protein Advanced *Glycosylation* Endproduct-specific Receptor, while the literature typically refers to this protein as Advanced *Glycation* Endproduct-specific Receptor. The distinction lies in whether the native protein receives sugar moieties via enzymatic reactions in the endoplasmic reticulum or Golgi apparatus (glycosylation), or from simple chemical reactions based on encountering reducing sugars (glycation). In any event, no post-translational modification would have occurred with prokaryotic expression of the protein.

RAGE is a cell surface protein with a single membrane-spanning region containing a small, ~40 residue, cytosolic domain and a large extracellular region composed of 3 Ig-like domains. Because of its link to the development of diseases such as rheumatoid arthritis and diabetes, this protein is of great interest to the medical field. It is proposed that the interaction of RAGE with its ligands results in pro-inflammatory gene activation. sRAGE, a soluble form of the receptor lacking the transmembrane and signaling domain, is thought to counteract the affects of the full length receptor and is being investigated as a potential therapeutic agent.

CONCLUSION

We performed multiple types of analysis serially, with the findings of each analysis confirming the next. Although multiple bands were seen on the gel, with corresponding multiple peaks on the SELDI spectra, we consider that the actual number of proteins identified was one, RAGE. Combinations of multiple methods in proteomics research not only allow for a wider coverage of the information to be determined (in this case combining the different methods increased the overall sequence coverage from 79.29% to 91.6%), but also lend validity to the findings of each individual method.