Procedure outline:

The first approach was to identify by PMF and LC-MS/MS all the proteins present in Sample A and B, so we decided to resuspend each sample in 60 ul H_2O, 2M Urea, and make three aliquots of 20 ul each. The first aliquot was digested in trypsin, and analyzed by PMF (Peptide Mass Fingerprint) and ESI-QTOF. Combining these two techniques, and by using the MASCOT search engine against SwissProt database, we obtained only one entry in both samples (A and B): Q15109 RAGE_HUMAN "Advanced glycosylation end product-specific receptor", with the sequence coverage shown in **Fig.1**.

>Q15109|RAGE_HUMAN Advanced glycosylation end product-specific receptor - Homo sapiens (Human).
MAAGTAVGAWVLVLSLWGAVVGAQNITARIGEPLVLKC
KGAPKKPPQRLEWKLNTGRTEAWKVLSPQGGGPWDSVA
RVLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRV
RVYQIPGKPEIVDSASELTAGVPNKVGTCVSEGSYPAGTL
SWHLDGKPLVPNEKGVSVKEQTRRHPETGLFTLQSELMV
TPARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVWEPVPL
EEVQLVVEPEGGAVAPGGTVTLTCEVPAQPSPQIHWMKD
GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRA
VSISIIEPGEEGPTAGSVGGSGLGTLALALGILGGLGTAAL
LIGVILWQRRQRRGEERKAPENQEEEEERAELNQSEEPEA
GESSTGGP

Fig.1. PMF results shown the tryptic peptides obtained combining the two techniques (PMF and ESI-QTOF). The sequence coverage was around 51%, and none of the samples shown any characteristic peptide, so at this point, we were unable to differentiate sample A and B.

The second approach was to increase the sequence coverage and search possible differential peptides, by digesting with other proteases (for that we used the second aliquot). We used GluC (V8) and Chymotrypsin. Again, there was only one protein matched in both samples: the same Q15109, with no differential or characteristic peptides in any of the samples. The sequence coverage was only slightly increased, as shown in **Fig. 2**. Although we knew the recommendation of using SwissProt database, the protein database used at this stage was NCBInr, in order to find possible spliceoforms or any other alterations of the RAGE protein. In this database, two main isoforms of human RAGE and 21 possible spliceforms are founded. The *in silico* digestions (with the three proteases) of these variants revealed some unique peptides, but none of them were founded by MS.

>Q15109|RAGE_HUMAN Advanced glycosylation end product-specific receptor - Homo sapiens (Human).
MAAGTAVGAWVLVLSLWGAVVGAQNITARIGEPLVLKC
KGAPKKPPQRLEWKLNTGRTEAWKVLSPQGGGPWDSVA
RVLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRV
RVYQIPGKPEIVDSASELTAGVPNKVGTCVSEGSYPAGTL
SWHLDGKPLVPNEKGVSVKEQTRRHPETGLFTLQSELMV
TPARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVWEPVPL
EEVQLVVEPEGGAVAPGGTVTLTCEVPAQPSPQIHWMKD
GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRA
VSISIIEPGEEGPTAGSVGGSGLGTLALALGILGGLGTAAL
LIGVILWQRRQRRGEERKAPENQEEEEERAELNQSEEPEA
GESSTGGP

Fig.2. The sequence coverage was increased up to 58% with the use of GluC (V8) and Chymotrypsin proteases (in both A and B samples). No differential or characteristic peptides were founded in any of the samples, suggesting that only one protein was present in both samples.

The characteristic N-terminal of the protein was not founded at any previous stage, so due to the presence in the sequence of a signal peptide in position 1 to 22 (a fact revealed by the UniProt/SwissProt entry) we decided to remove it, and make a new database with the modified sequence, including the tag mentioned at the ABRF brochure letter (GSHM). At this point, the sequence, with 386 aminoacids, was left as shown in **Fig. 3**.

GSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGR
TEAWKVLSPQGGGPWDSVARVLPNGSLFLPAVGIQDEGI
FRCQAMNRNGKETKSNYRVRVYQIPGKPEIVDSASELTA
GVPNKVGTCVSEGSYPAGTLSWHLDGKPLVPNEKGVSV
KEQTRRHPETGLFTLQSELMVTPARGGDPRPTFSCSFSPG
LPRHRALRTAPIQPRVWEPVPLEEVQLVVEPEGGAVAPG
GTVTLTCEVPAQPSPQIHWMKDGVPLPLPPSPVLILPEIGP
QDQGTYSCVATHSSHGPQESRAVSISIIEPGEEGPTAGSVG
GSGLGTLALALGILGGLGTAALLIGVILWQRRQRRGEERK
APENQEEEEERAELNQSEEPEAGESSTGGP

Fig. 3. With this new sequence, we were able to find the N-terminal extreme peptide of the protein by MALDI-TOF and LC-MS/MS analysis (Score: 67 with statistical significance starting from 19), and in two of the three proteases used (in both A and B). With this new peptide, we achieved a sequence coverage of 64%.

Further analysis was needed in order to elucidate differences between the two samples. Thus we performed two different approaches: SDS-PAGE and intact protein MS determination by linear MALDI-TOF. SDS-PAGE obtained is shown in **Fig 4**.

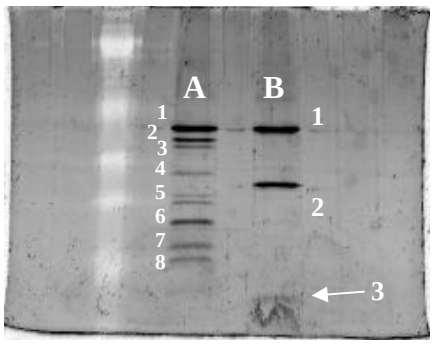

Fig. 4. The SDS-PAGE (Sypro staining) shows as much as 8 bands for Sample A, and only 2 for Sample B, running from around 20 kDa up to around 42 kDa (apparent molecular weight). Each band was isolated and digested with trypsin (*in gel* digestion).

The intact protein MS determination by linear mode MALDI-TOF for Sample B shows three main peaks at 33182.28, 24600.54 and 12541.36 Da (+/- 30 Da). These results can be compared with the corresponding SDS-PAGE result, assuming that there is a little shift in the mobility of each band in the SDS-PAGE. Also, the peak at 12541.36 Da was lately assumed as the small band near the edge at lane B (**Fig. 4**, marked with an arrow). Unfortunately, this band could not be isolated. For Sample A, the MALDI-TOF analysis revealed more masses, as shown also in SDS-PAGE: The first of them at 33180.28 Da, then several peaks at 31812.49, 30146.19, 25547.04, 18893.08, 15251.72, and a bunch of peaks below 12000 Da (not shown at SDS-PAGE probably because of their small size).

As mentioned in **Fig. 4**, each band of the SDS-PAGE was isolated, digested with trypsin, and analyzed by PMF and LC-MS/MS. Several databases were elaborated with the information obtained by the linear MALDI-TOF results, narrowing the sequence of RAGE (without the signal peptide, and with GSHM tag) up to the masses observed by intact mass determination, in order to find the characteristic C-terminal extreme of each fragment. The sequences are summarized in **Table 1**.

Sample	Sequence	Position	TM	OM	Δ mass
A1/B1	GSHMAQNITARIGEPL….AVSISIIEPGEEG	1-309	33168.1	33182.8	14.7
B2	GSHMAQNITARIGEPL….PLEEVQLVVE	1-255	24585.2	24600.5	15.3
B3	GSHMAQNITARIGEPL….GKPEIVDSASE	1-114	12524.4	12541.4	17.0
A2	GSHMAQNITARIGEPL….SSHGPQESR	1-296	31886.5	31812.5	74.0
A3	GSHMAQNITARIGEPL….PEIGPQDQGT	1-280	30158.6	30146.2	12.4
A4	GSHMAQNITARIGEPL….HRALRTAPIQ	1-208	22584.9	22547.0	37.9
A5	GSHMAQNITARIGEPL….GLFTLQSEL	1-174	18924.6	18893.1	31.5
A6	GSHMAQNITARIGEPL….PAGTLSWHL	1-141	15251.5	15202.7	48.8

Table.1 Estimated sequences of each fragment predicted by the linear MALDI-TOF results. TM: teorical mass. OM: observed mass. Δmass: mass difference between TM and OM

For A1 and B1, the C-terminal peptide was detectable by LC-MS/MS, validating the previous assumption, and confirming the sequence of the main protein as presented in **Fig. 5**.

GSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWK
VLSPQGGGPWDSVARVLPNGSLFLPAVGIQDEGIFRCQAMNRNGK
ETKSNYRVRVYQIPGKPEIVDSASELTAGVPNKVGTCVSEGSYPAG
TLSWHLDGKPLVPNEKGVSVKEQTRRHPETGLFTLQSELMVTPAR
GGDPRPTFSCSFSPGLPRHRALRTAPIQPRVWEPVPLEEVQLVVEPE
GGAVAPGGTVTLTCEVPAQPSPQIHWMKDGVPLPLPPSPVLILPEI
GPQDQGTYSCVATHSSHGPQESRAVSISIIEPGEEG

Fig. 5. Final sequence of the main protein present at both samples A and B. The final sequence coverage was 80%

For B2, we could not find the characteristic C-terminal peptide predicted by the results of intact mass determination (including a +/- 1 aminioacid C-terminal sequence variation, way above our mass precision of 15 Da). The last peptide founded was slightly behind the alleged C-terminal sequence (**Fig. 6**), so it was impossible to fully characterize the sequence of this truncate.

GSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWK
VLSPQGGGPWDSVARVLPNGSLFLPAVGIQDEGIFRCQAMNRNGK
ETKSNYRVRVYQIPGKPEIVDSASELTAGVPNKVGTCVSEGSYPAG
TLSWHLDGKPLVPNEKGVSVKEQTRRHPETGLFTLQSELMVTPAR
GGDPRPTFSCSFSPGLPRHRALRTAPIQPRVWEPVPLEEVQLVVE

Fig. 6. Alleged sequence of protein B2, based on intact mass determination (by linear MALDI-TOF). From the last peptide obtained (highlighted), there are 15 aminoacids not characterized by neither PMF nor LC-MS/MS.

We could not isolate band B3 from SDS-PAGE, so it was impossible to obtain PMF or LC-MS/MS to characterize the band.

Bands A2 to A6 could be isolated, but each corresponding C-terminal peptides was not founded, so the unambiguous characterization of these truncates was impossible.

The conclusions we could offer from the approach we have taken for this experiment are:

- There is only one gene product present in both samples, corresponding to Q15109 RAGE_HUMAN “Advanced glycosylation end product-specific receptor”
- The sequence of this protein does not correspond to the whole sequence in the Q15109 entry, but to a portion of 309 aminoacids with a molecular weight of 33168.1 Da. The deletion includes the whole N-terminal signal peptide sequence and the last 75 aminoacids.
- The main difference between the two samples (A and B) is the complexity, with more truncates present in A (main protein and no less than 7 truncates) than in B (main protein plus two truncates)