Sample Information. The study samples were supplied in two vials (labeled "A" and "B") and consist of separate preparations of related human proteins that were expressed in *Escherichia 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 Histag. Vials A and B contain 5 µg and 10 µg of protein, respectively. The samples were prepared from aqueous solutions that also contained small amounts of salts. To the best of our knowledge, there are no appreciable quantities of interfering substances that contain primary amines and/or free thiols. The samples have been successfully dissolved in 25 – 50 mM ammonium bicarbonate with or without 20% acetonitrile; as well as 0.1% formic acid; we anticipate that other solvents can be used.

Sample Preparation. Both samples were dissolved in water containing 5% acetonitrile to a concentration of 1ug/uL.

In-solution Trypsin Digestion. 1ug of each sample was diluted to 10uL with 50mM ammonium bicarbonate pH=8.0. The samples were reduced with a final concentration of 10mM DTT for 30 minutes (RT) followed by alkylation with a final concentration of 50mM iodoacetamide for 30 minutes (RT). The samples were then digested with 0.1ug of Promega trypsin overnight at 37C in the dark. The resulting sample was acidified with 0.5uL of acetic acid.

LC-MS/MS Analysis of Tryptic Digest. The LC-MS/MS system consisted of a ThermoElectron LTQ-FT mass spectrometer system with a Protana nanospray ion source interfaced to a self-packed 8 cm x 75 um id Phenomenex Jupiter 10 um C18 reversed-phase capillary column. 10% of each sample was manually injected and the peptides eluted from the column by an acetonitrile/0.1 M acetic acid gradient at a flow rate of 0.4 µL/min (0-80% acetonitrile over 20 minutes). The nanospray ion source was operated at 2.8 kV. The digest was analyzed using the full data dependency (no repeat, 120s exclusion, 1Da window) of the instrument acquiring full scan mass spectra to determine peptide molecular weights (FT – 100,000 resolution) and product ion spectra to determine amino acid sequence in ten sequential scans (IT). This mode of analysis produced approximately 1000 CAD spectra of ions ranging in abundance over several orders of magnitude.

MALDI-TOF Analysis of Proteins. 1ug of each sample was diluted in sinapinic acid matrix (SA) 1:1. 0.5uL of this solution was spotted onto the plate and allowed to dry. The sample was analyzed on a Bruker Microflex by averaging 300 scans using external calibration and instrument defaults for the 5-50KD range. The matrix was saturated SA in a solution of 70% acetonitrile in 0.1% TFA.

Data Analysis. The LC-MS/MS data was analyzed using Bioworks 3.3.1 (Sequest) for the database search. The search was performed against the current version of IPI Human and NR. The parameters used were 10ppm parent mass, 1Da fragment mass, full tryptic, carboxyamidomethyl Cys, and variable oxidized Met. The search results were loaded

into Scaffold version 1_07_00. Peptide and protein identifications were accepted if they met 95% or greater probability.

Results

The database search (Human IPI) for both samples yielded one human protein with five or more peptides. Keratins and about dozen *E. coli* proteins were also present (NR search). Examination of the peptides in relation to ion current confirmed that only the human protein was at greater than 1% abundance. The major protein in each sample was established as Advanced Glycosylation End Product-specific Receptor Precursor (IPI00014810.3). To a first approximation, the samples contained the same peptides. A sub-database of this protein was made and all subsequent searches were against it. As PTMs in *E. coli* would be extremely rare and only one major protein was obtained, the next search focused on semi-tryptic peptides to look for a C-terminus. The third search focused on a no-enzyme search with a modification of the mass for GSHM for peptide N-termini. The following peptides were identified for the samples using these two searches:

Sample A – AVSISIIEPGEEG (C-terminus) and GSHMAQNITAR (N-terminus).
Theoretical mass – 33,169
Sample B – AVSISIIEPGEEG, VYQIPGKPEIVDSASE, VWEPVPLEEVQLVVE (C-termini) and GSHMAQNITAR (N-terminus).
Theoretical masses – 33,169; 24,585 and 12,525

The sequence coverage obtained for both samples was ~50-60% for a protein that appears to have been expressed minus some type of signal region and minus a C-terminal portion. As a quick confirmation, a standard SDS-PAGE gel was run and MWs were obtained by MALDI-TOF. Both analyses confirmed the LC-MS/MS results – sample A contains one major protein and sample B contains the same major protein and two C-terminal truncations of this protein. Further coverage could be obtained by either sub-digesting the remaining tryptic digest or using a different enzyme on another aliquot. Further digestion was deemed not useful as the first digest produced confirmation of the qualitative differences.

Sample A

Peptide	MH+	(ppm)	z	XC
R.IGEPLVLK.C	868.5502	-4.0	2	2.72
K.LNTGRTEAWK.V	1175.6167	0.0	2	2.52
K.KPPQRLEWK.L	1181.6790	-0.7	2	2.58
-.GSHMAQNITAR.I	1185.5793	-1.7	2	3.45
-.GSHM*AQNITAR.I	1201.5742	-2.0	2	3.18
R.AVSISIIEPGEEG.-	1300.6631	4.1	2	3.21
R.PTFSCSFSPGLPR.H	1452.6940	2.8	2	4.51
K.VLSPQGGGPWDSVAR.V	1525.7758	1.2	3	5.41
R.GGDPRPTFSCSFSPGLPR.H	1934.9178	-1.5	3	4.93
R.HPETGLFTLQSELMVTPAR.G	2127.0903	1.7	3	3.41
R.VYQIPGKPEIVDSASELTAGVPNK.V	2512.3293	7.9	2	6.24
K.VGTCVSEGSYPAGTLSWHLDGKPLVPNEK.G	3098.5252	5.1	4	5.06

Sample B

Peptide	MH+	(ppm)	z	XC
K.PLVPNEK.G	796.4563	0.7	2	2.61
R.IGEPLVLK.C	868.5502	0.8	2	2.94
K.LNTGRTEAWK.V	1175.6167	-0.2	2	2.49
K.KPPQRLEWK.L	1181.6790	-0.7	2	2.51
-.GSHMAQNITAR.I	1185.5793	-2.4	2	3.91
-.GSHM*AQNITAR.I	1201.5742	-0.7	2	3.43
R.AVSISIIEPGEEG.-	1300.6631	-2.2	2	3.49
R.PTFSCSFSPGLPR.H	1452.6940	7.4	2	4.39
K.VLSPQGGGPWDSVAR.V	1525.7758	4.0	3	5.25
R.VYQIPGKPEIVDSASE.L	1731.8800	4.0	2	4.31
R.VWEPVPLEEVQLVVE.P	1764.9418	6.4	2	4.31
R.GGDPRPTFSCSFSPGLPR.H	1934.9178	-3.6	3	4.48
R.HPETGLFTLQSELMVTPAR.G	2127.0903	2.6	2	5.00
R.RHPETGLFTLQSELMVTPAR.G	2283.1914	0.3	2	5.04
R.RHPETGLFTLQSELM*VTPAR.G	2299.1863	1.4	3	4.27
R.VYQIPGKPEIVDSASELTAGVPNK.V	2512.3293	8.0	2	6.25
K.VGTCVSEGSYPAGTLSWHLDGKPLVPNEK.G	3098.5252	0.4	4	4.49

Sample A

IPI:IPI00014810.3|SWISS-PROT:Q15109-1| AGER Isoform 1 of Advanced glycosylation end product-specific receptor precursor

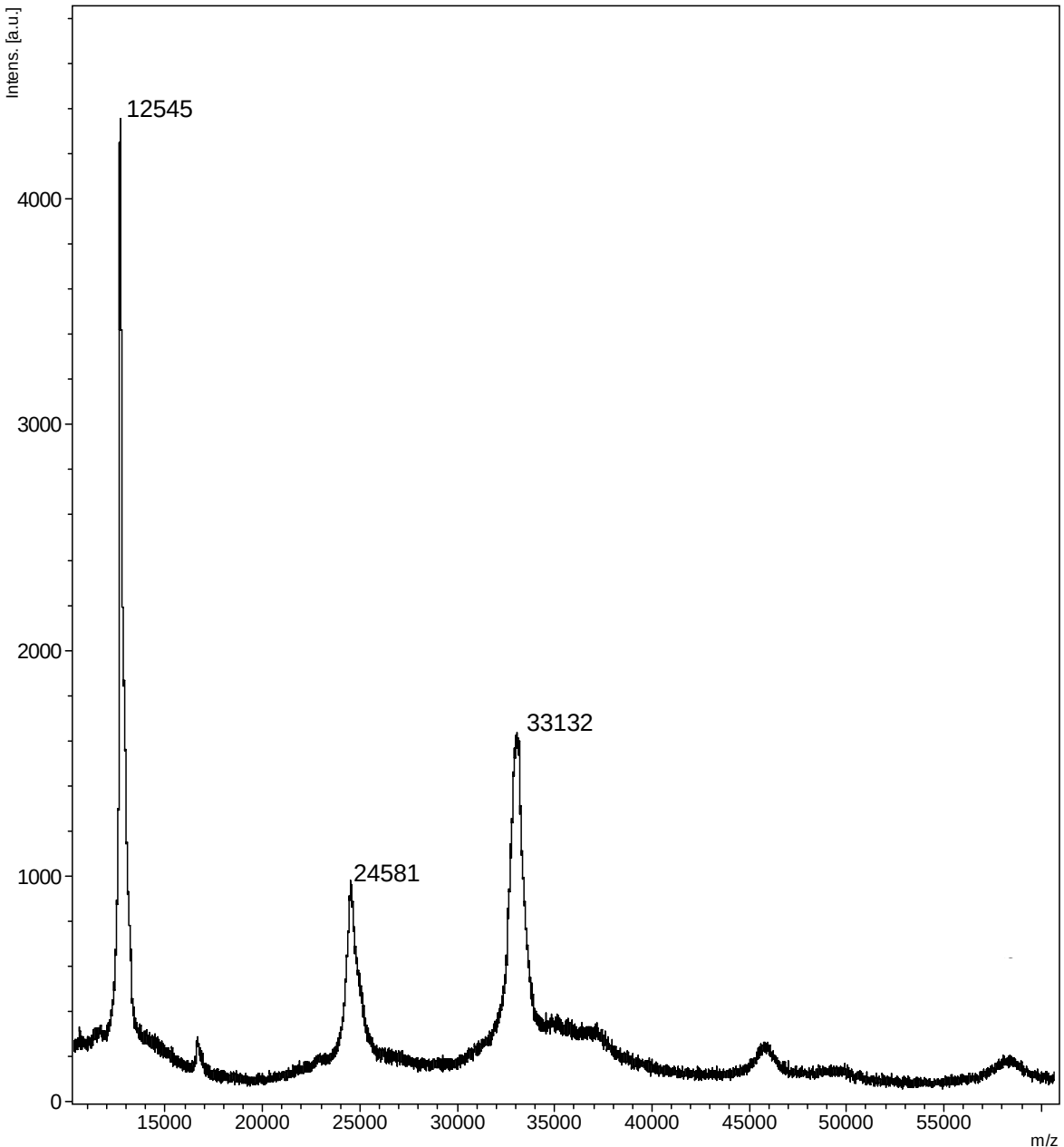
GSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWKVLSPQGGGPWDSVAR
VLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRVRVYQIPGKPEIVDSASELTAG
VPNKVGTCVSEGSYPAGTLSWHLDGKPLVPNEKGVSVKEQTRRHPETGLFTLQSELMVT
PARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVWEPVPLEEVQLVVEPEGGAVAPGGT
VTLTCEVPAQPSPQIHWMKDGVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQES
RAVSISIIEPGEEG

Sample B

IPI:IPI00014810.3|SWISS-PROT:Q15109-1| AGER Isoform 1 of Advanced glycosylation end product-specific receptor precursor

GSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWKVLSPQGGGPWDSVAR
VLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRVRVYQIPGKPEIVDSASELTAG
VPNKVGTCVSEGSYPAGTLSWHLDGKPLVPNEKGVSVKEQTRRHPETGLFTLQSELMVT
PARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVWEPVPLEEVQLVVEPEGGAVAPGGT
VTLTCEVPAQPSPQIHWMKDGVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQES
RAVSISIIEPGEEG

MALDI-TOF


Intens. [a.u.]
12545
24581
33132
4000
3000
2000
1000
0
15000
20000
25000
30000
35000
40000
45000
50000
55000
m/z

SDS-PAGE