

# Mining the Human Placenta Proteome > 2000 Proteins deep using CID/ETD on an Ion Trap

● IMSC 2009, PWA 113

Simone Lemeer<sup>1</sup>, Hannes Hahne<sup>1</sup>, Andrea Schneider<sup>2</sup>, Markus Lubeck<sup>2</sup>, Arnd Ingendoh<sup>2</sup>, Bernhard Kuster<sup>1,3</sup>

<sup>1</sup> Technische Universität München, Freising, GERMANY; <sup>2</sup> Bruker Daltonik GmbH, Bremen, Germany

## Introduction

A multi-tier analytical approach was chosen to identify as many proteins as possible from human placenta. The proteins were extracted under fully denaturing conditions and separated by 1D-SDS PAGE. The peptides were separated by high-resolution RP-HPLC connected to the amaZon ion trap which was operated in the Gas Phase Fractionation mode (GPF) to increase the depth at which ions can be picked for MS/MS. Both CID and ETD fragmentation were applied to sample alternative peptide pools. More than 2000 proteins were identified, the largest dataset from human placenta to date.

## Methods

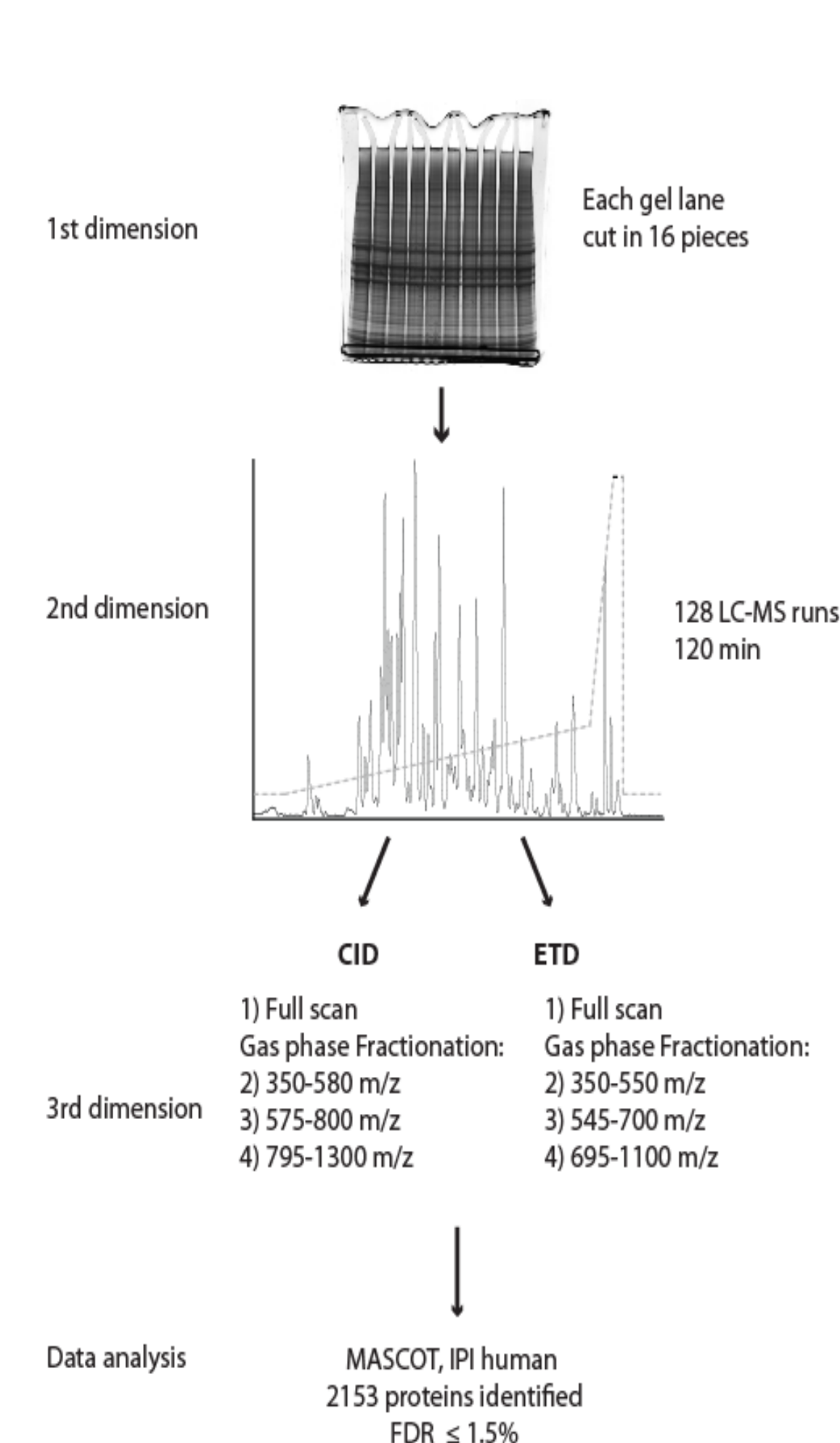


Fig. 1: Workflow

Post-delivery human placenta was obtained with written consent from a healthy female. Total protein extraction was performed using 8M urea, 100mM TEAB, 1% Triton X-100, 10mM NaF and 20 mM nitrophenylphosphate. Protein extracts were separated by 1D SDS gel electrophoresis and each lane was cut into 16 regions. All samples were digested with trypsin. Peptides were separated by a 120' gradient on a nanoscale RP-HPLC (Agilent nanoLC G2226 with G1376 loading pump) coupled online to an amaZon ion trap (Bruker). Each protein digest was measured eight times (4 gas phase fractions, each by CID and ETD). Proteins were identified from the human IPI database v.3.58 using Mascot 2.2 and further data analysis was performed with Scaffold v.2 and Excel 2007.

## Results

**Resolution:** ~2,500-5,000 (FWHM) at a scan speed of 8100 amu/sec, sufficient to resolve 3+ precursor ions within 100 ms scan time. Averaging ~5 scans increases resolution by ~20% (fig. 2).

**Mass accuracy:**  $0.12 \pm 0.08$  amu (fig. 3) for all identified proteins. This is improved to  $0.05 \pm 0.06$  amu by post DB search recalibration. >99% of peptide identifications are within 0.2 amu error enabling a 2nd pass search with higher identification confidence. The <sup>12</sup>C peak was correctly assigned in 94% of all cases.

**Sensitivity:** Robust identification of BSA is achieved at 250 fmol digest on column (fig. 4). In our hands, the limit of detection is in the range of ~50 amol.

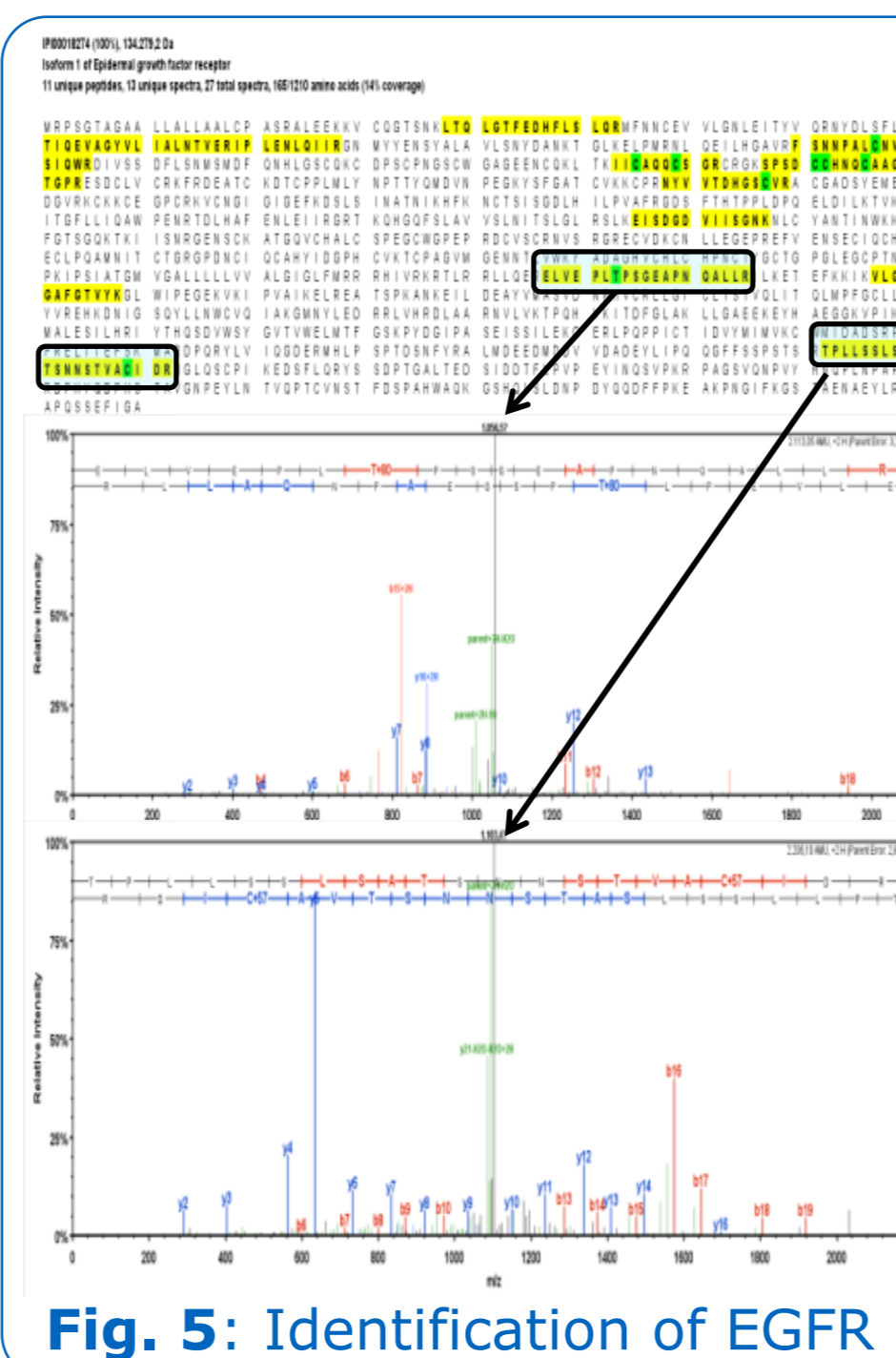
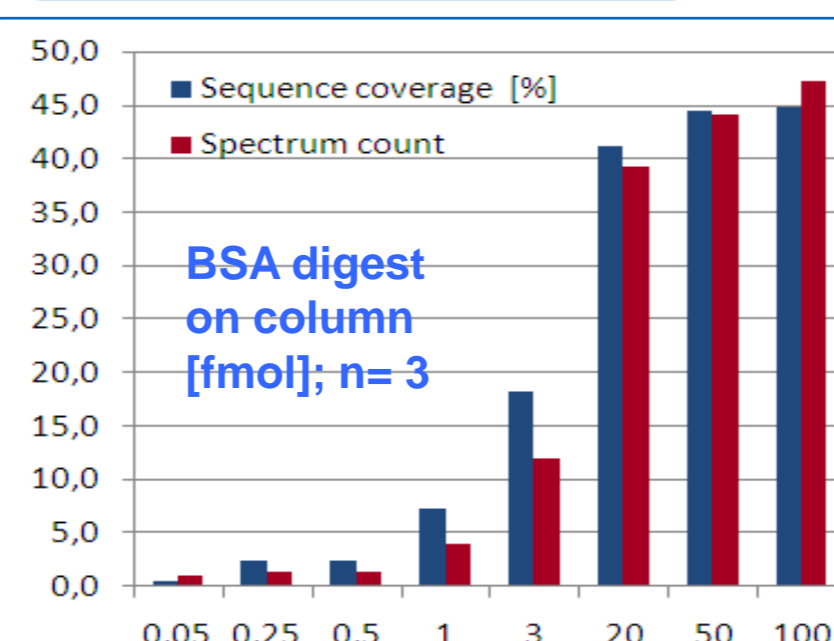
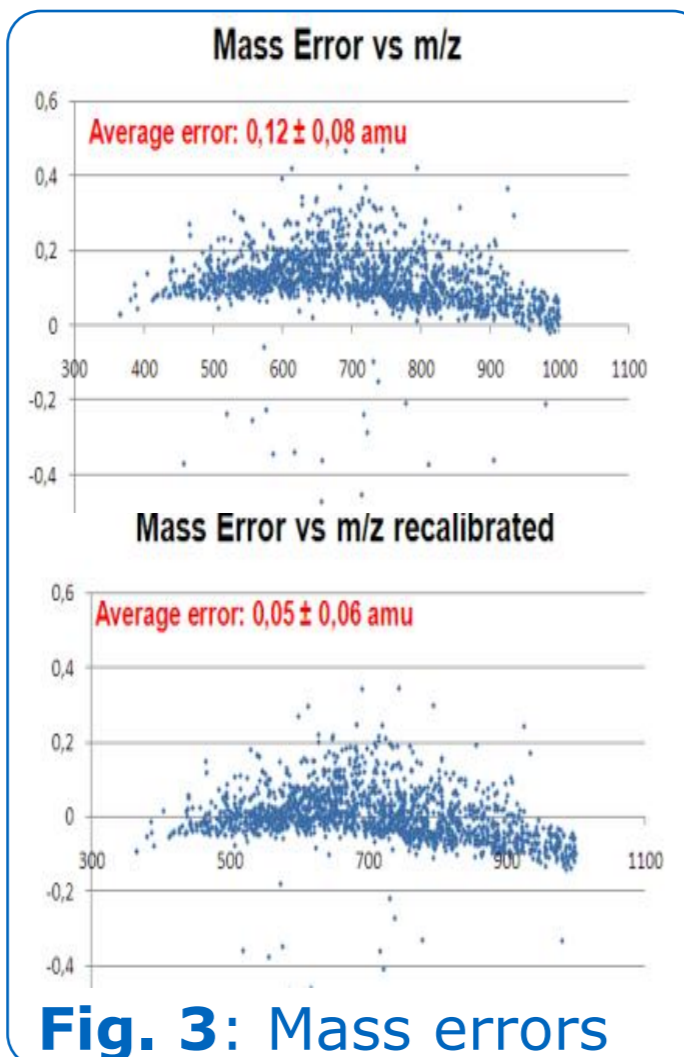
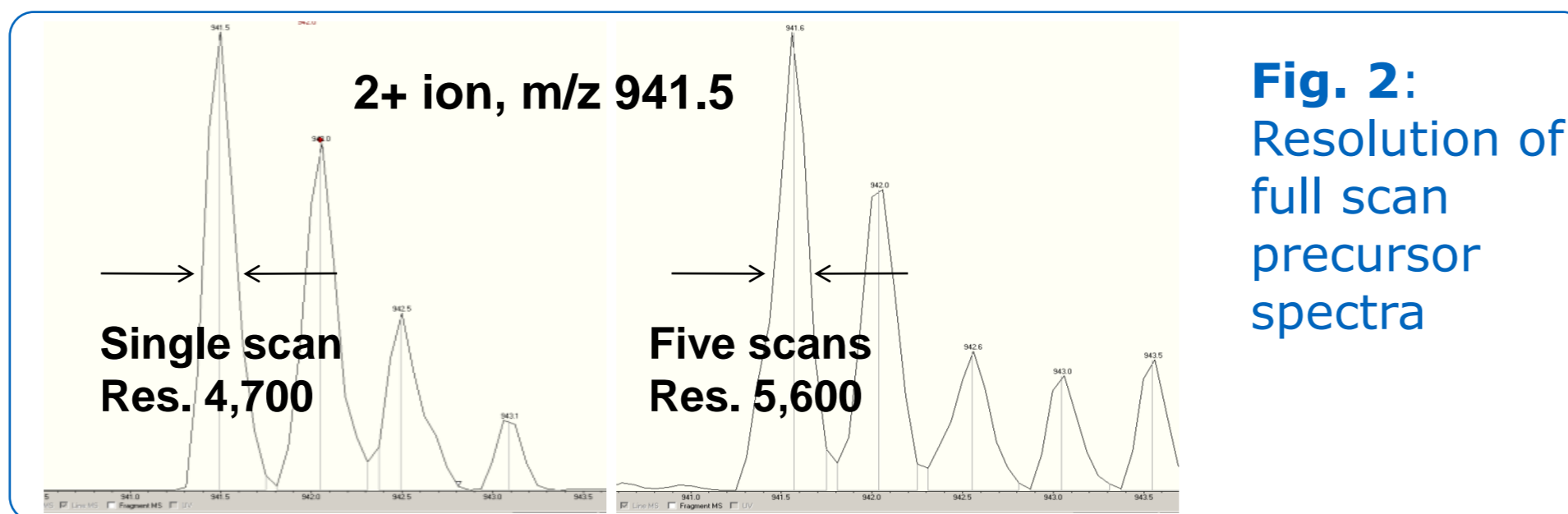


Fig. 4: Sensitivity of protein ID

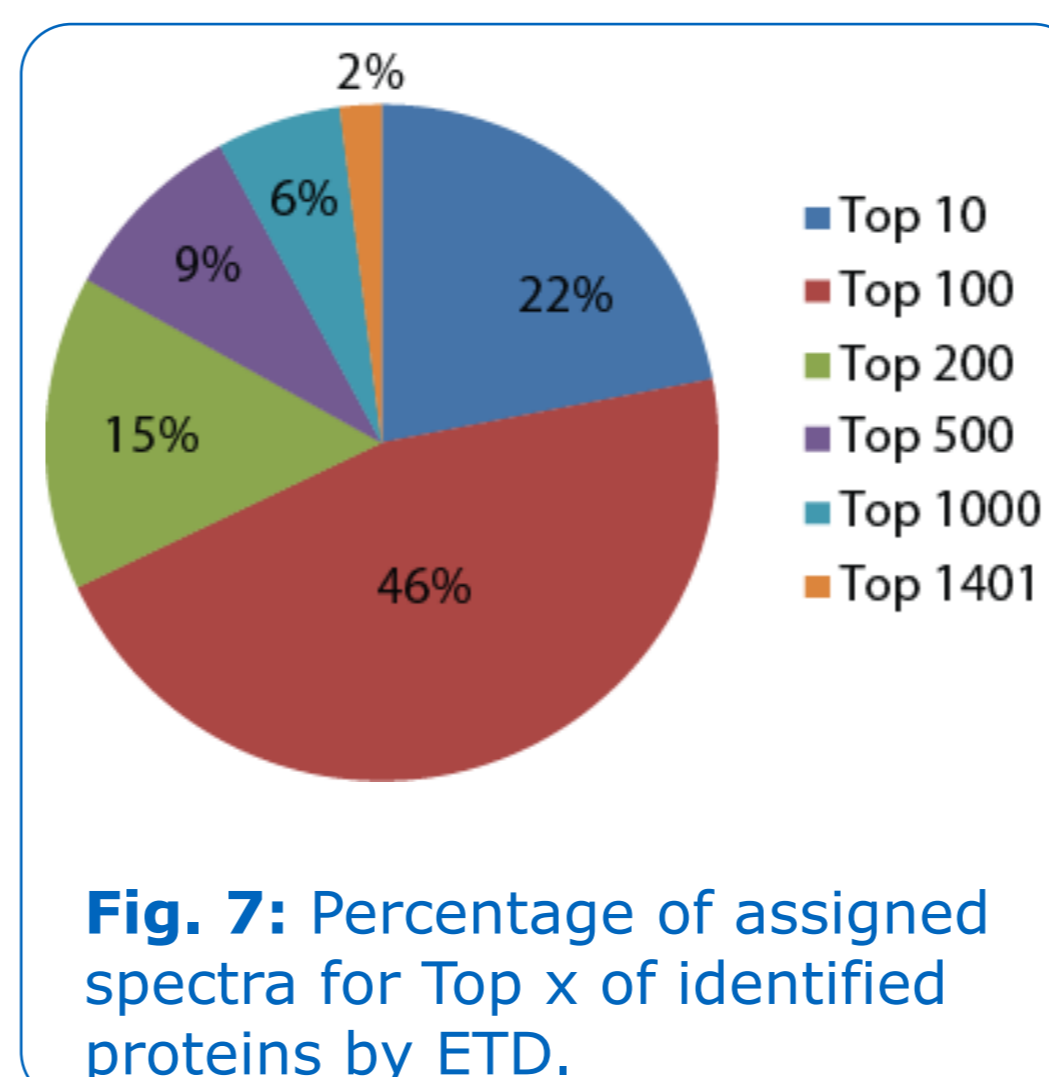
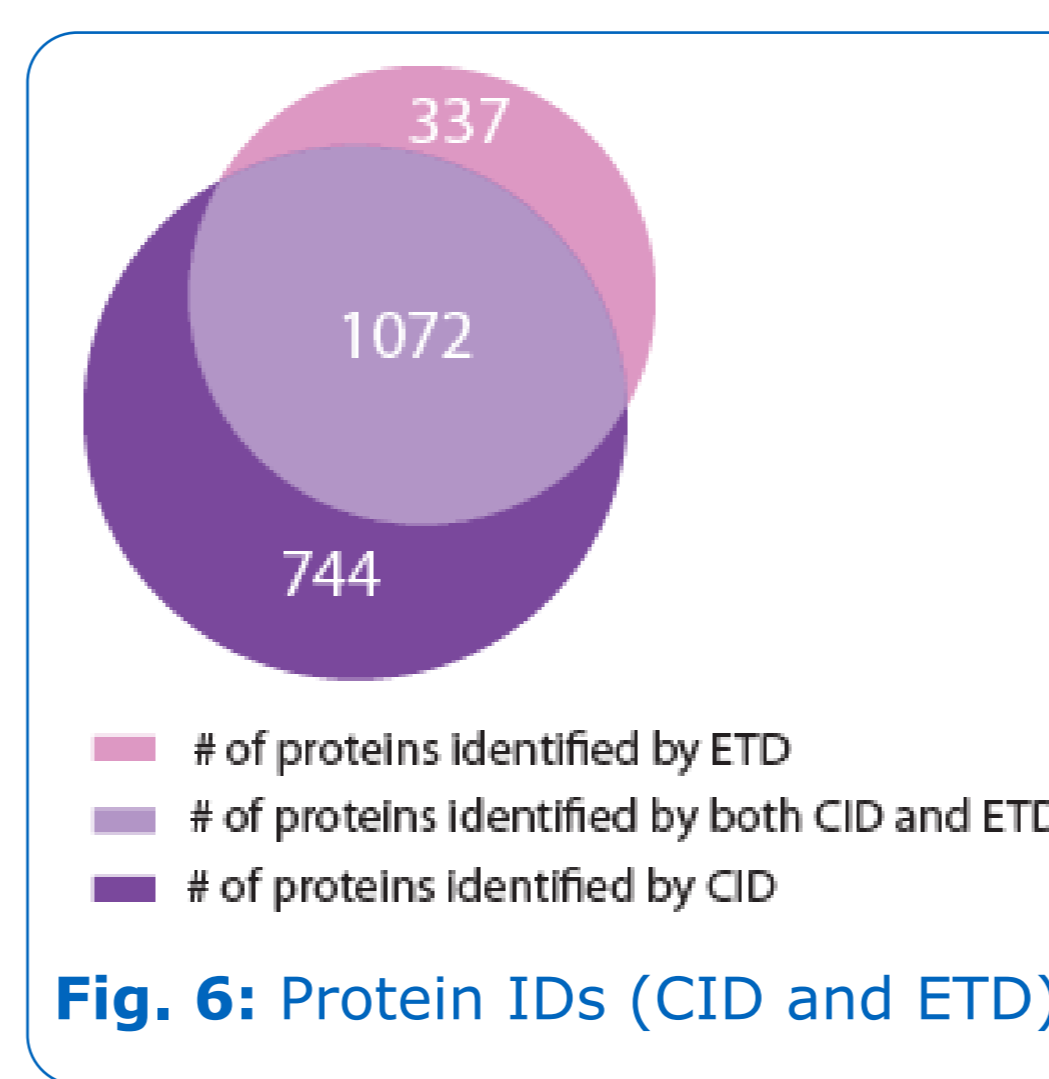


Fig. 7: Percentage of assigned spectra for Top x of identified proteins by ETD.

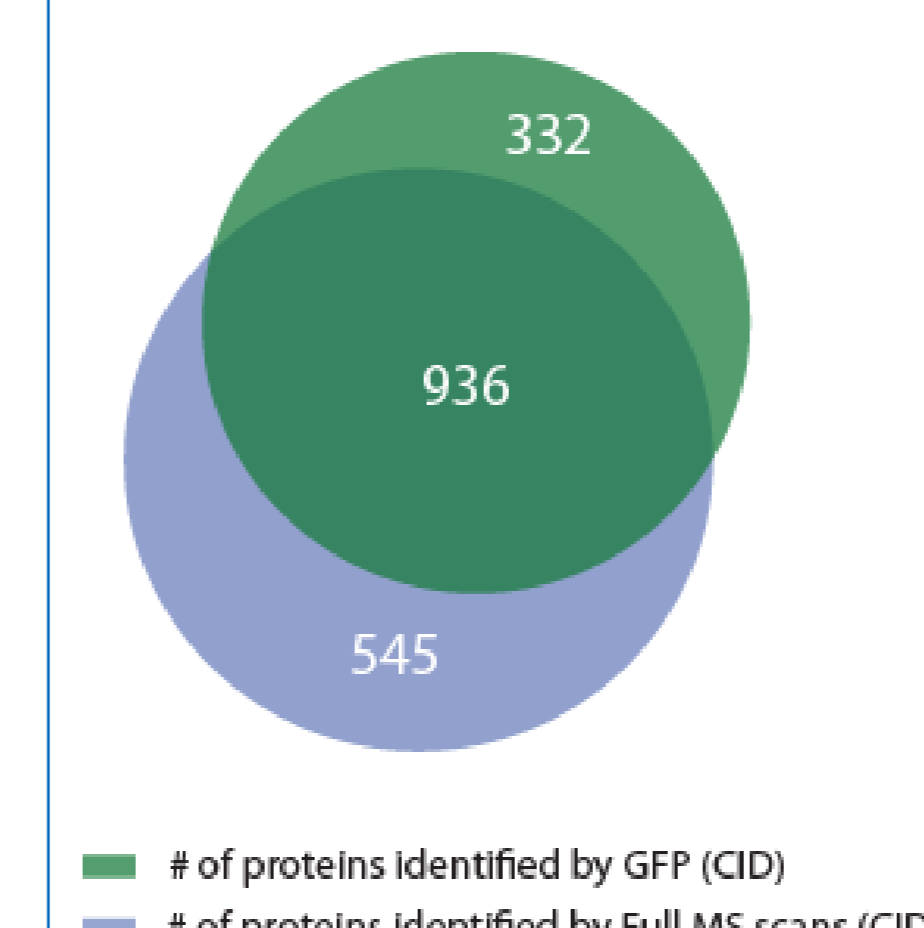
**In-depth analysis:** The EGF receptor was identified from a complex lysate of human placenta with 14% sequence coverage (fig. 5). In addition a Threonine phosphorylation site on the receptor was identified in one of the gas phase fractions.

**Overlap CID & ETD:** 2153 proteins were identified using CID and ETD full scans and GPF (Fig.6). The majority of proteins were identified by both CID and ETD. However, ETD still has added value for the identification of PTMs (Table 1).

**Dynamic range:** Top 10 proteins already comprise 22% of the assigned spectra (Fig. 7). However, low abundant proteins (abundance estimated to be 1000x lower) are identified with high confidence. A similar graph can be obtained for CID data.

Method	# of unique phosphopeptides identified
CID	58
GPF CID	52 (35 GPF only)
ETD	35 (31 ETD only)

Table 1: Unique number of id'ed phosphorylation sites by the different methods.



Method	# of assigned spectra	# of identified proteins
GPF 350-580	1051	411
GPF 575-800	3686	808
GPF 795-1300	3335	721
Full scan	17802	1846

Table 2: Number of assigned spectra and IDs for the different methods.

**Identification of phosphorylation sites:** In total 93 unique phosphorylation sites were identified by CID. In addition, 31 unique phosphorylation sites could be identified by ETD (Table 1).

**Overlap Full scan MS & GPF:** 1813 proteins were identified using MS and GPF CID (fig. 8 & 9, table 2). GPF adds little to the amount of proteins identified. However, it improves sequence coverage (Fig. 10).

**Improved sequence coverage:** Using GPF in combination with full scan CID, sequence coverage of medium abundant proteins is improved significantly. In addition the GPF adds 35 unique phosphorylation sites to the total number of phosphorylation sites identified (Table 1).

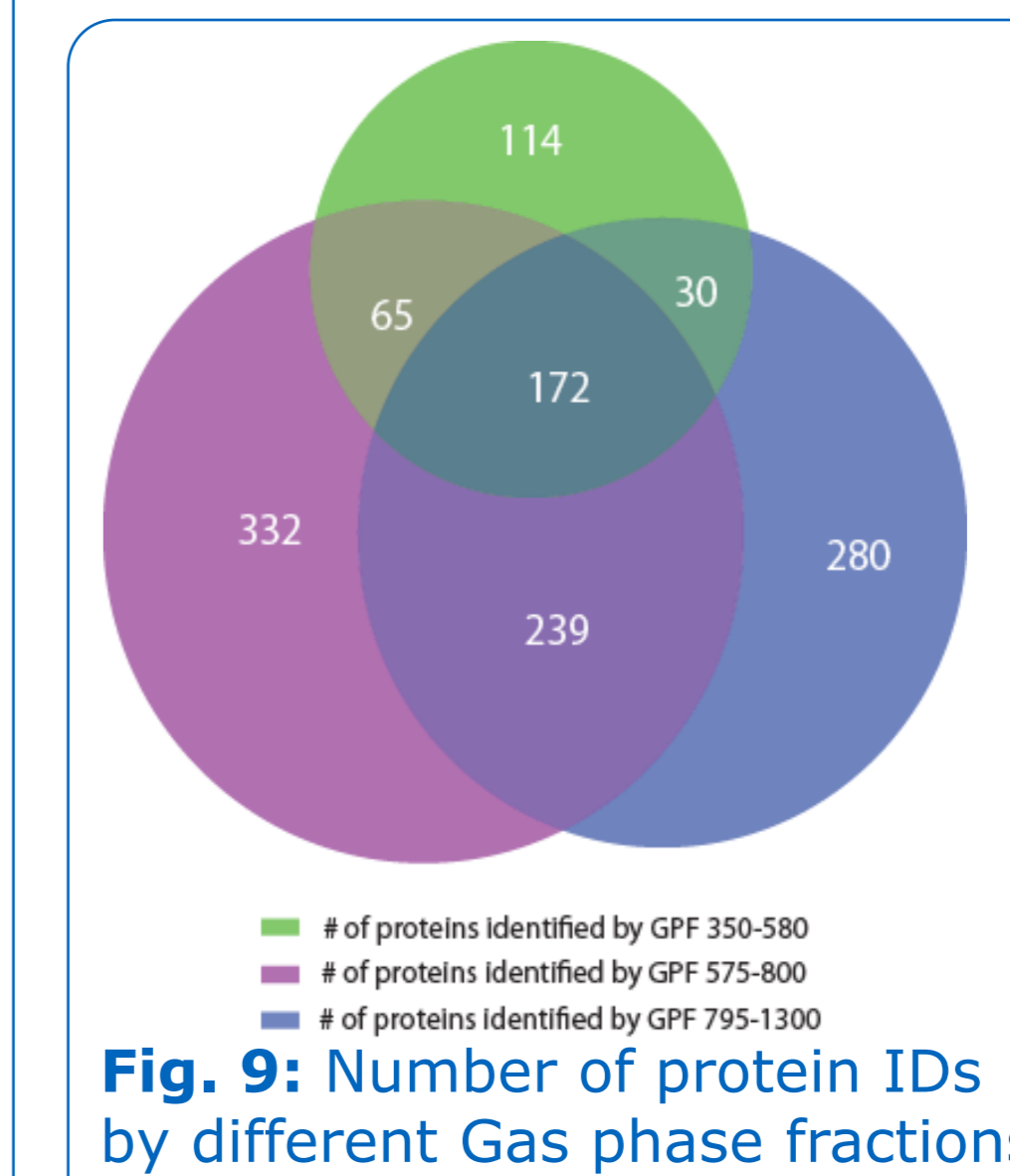
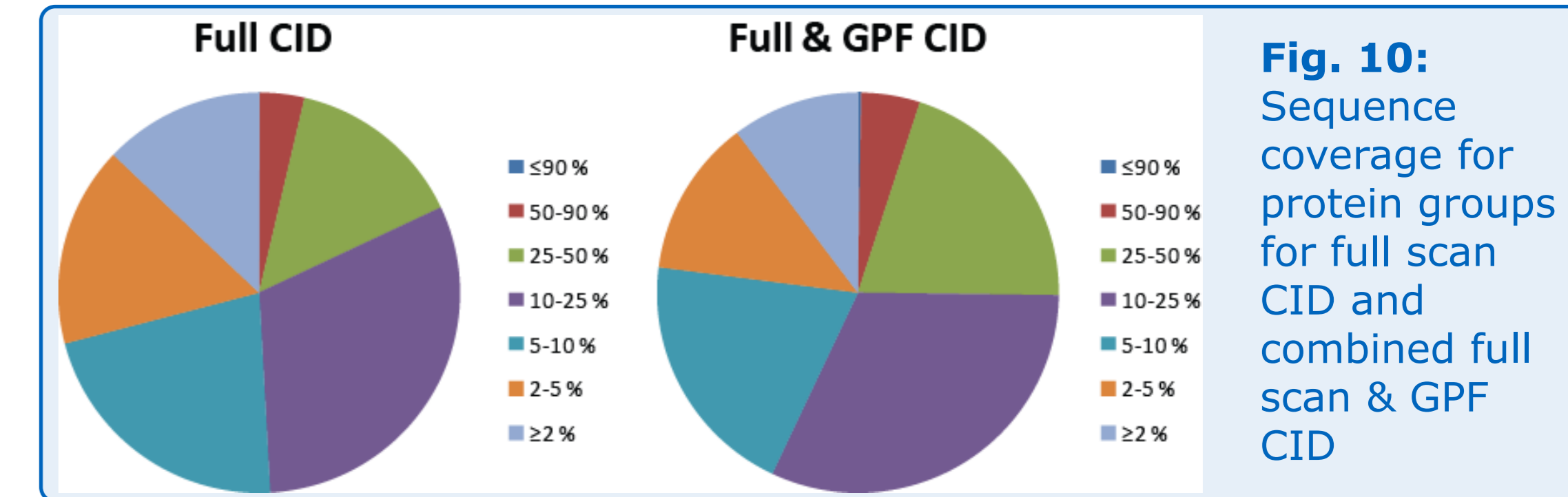


Fig. 9: Number of protein IDs by different Gas phase fractions

## Conclusions

- The amazon ion trap delivers up to 5,000 resolution at 8,100 amu/sec scan speed, 0.2 amu mass accuracy and attomol sensitivity for protein ID from complex proteomes.
- ~2,200 proteins were identified using a 5-tier approach, the largest set of protein ID from human placenta to date.
- CID identifies more proteins than ETD which adds 20% extra proteins compared to CID alone.
- Gas phase fractionation increases protein ID by a very modest ~10% compared to full scan analysis, but it mostly improves sequence coverage.
- The combination of CID, ETD and gas phase fractionation doubles the number of identified phosphorylation sites.