

A Proposal for a Consortium Project

IBDC, Seattle • Oct 12, 2005

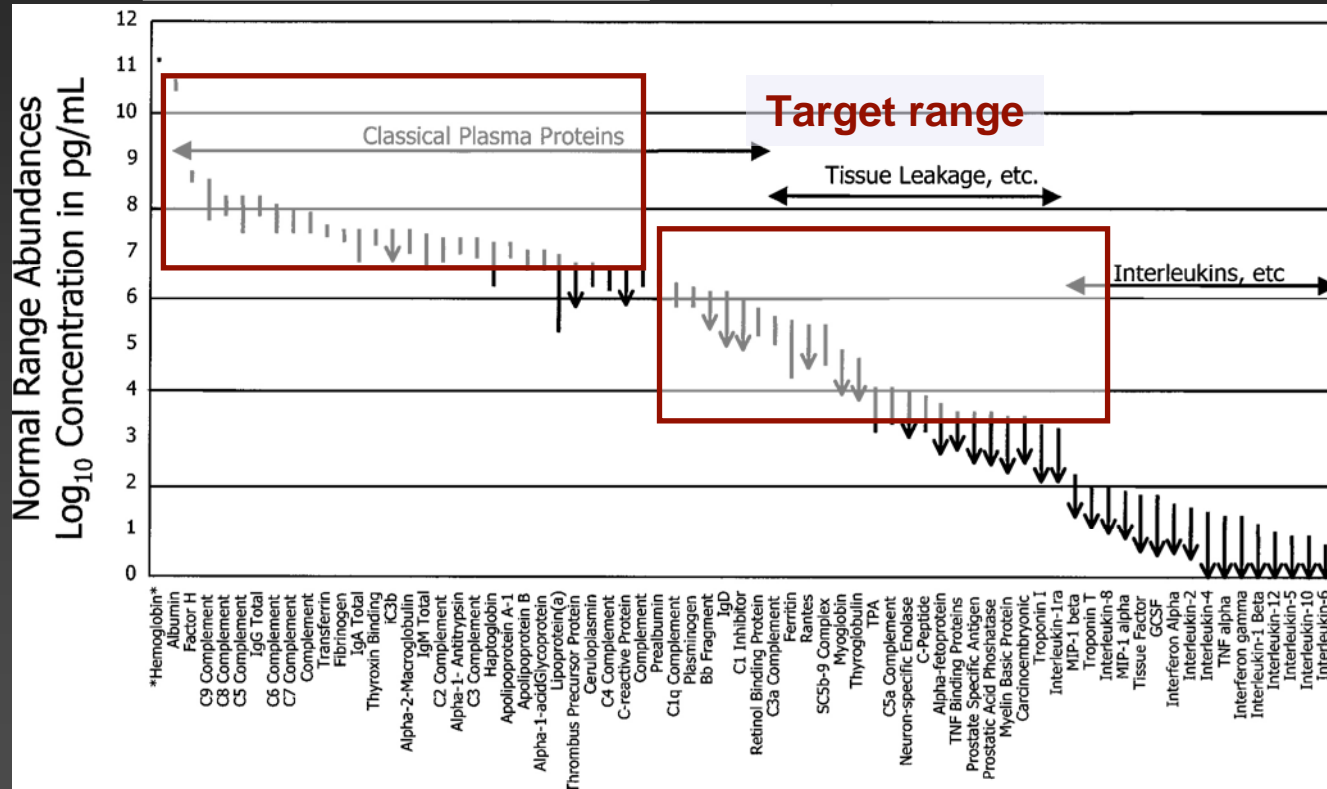
Ruedi Aebersold, Ph.D
Institute for Molecular Systems Biology
ETH-Zürich

and

Institute for Systems Biology
Seattle, Washington USA

The Serum Challenge : Significant Dynamic Range

Range of LC-MS and Most other methods



(From Anderson & Anderson (2002) MCP, 1, 845-67)

Some observations (also from HUPO PPP)

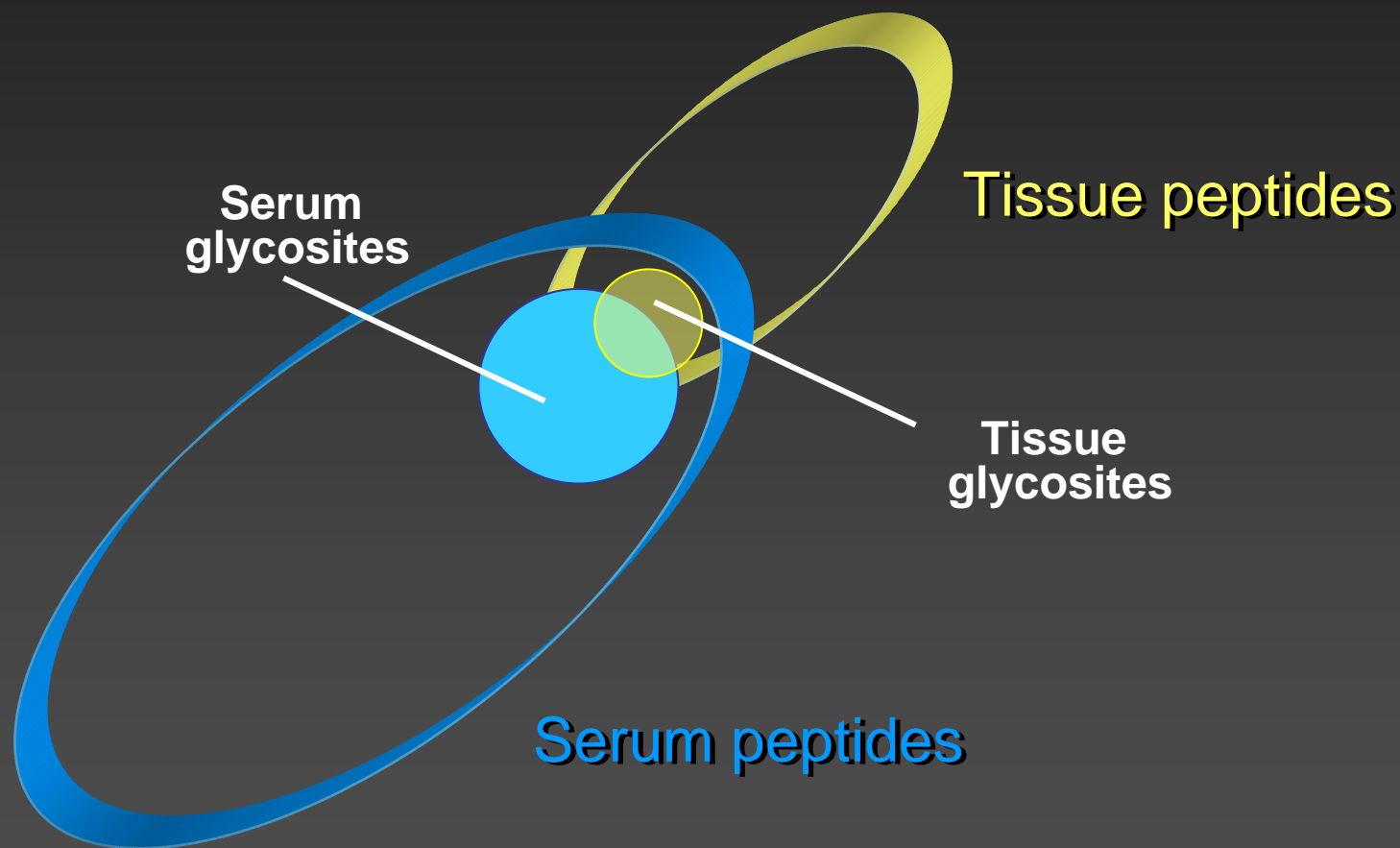
- No definitive method
- Difficult to compare methods and communicate across methods
- No firm quality standards
- Cannot impose standard method.

Proposal

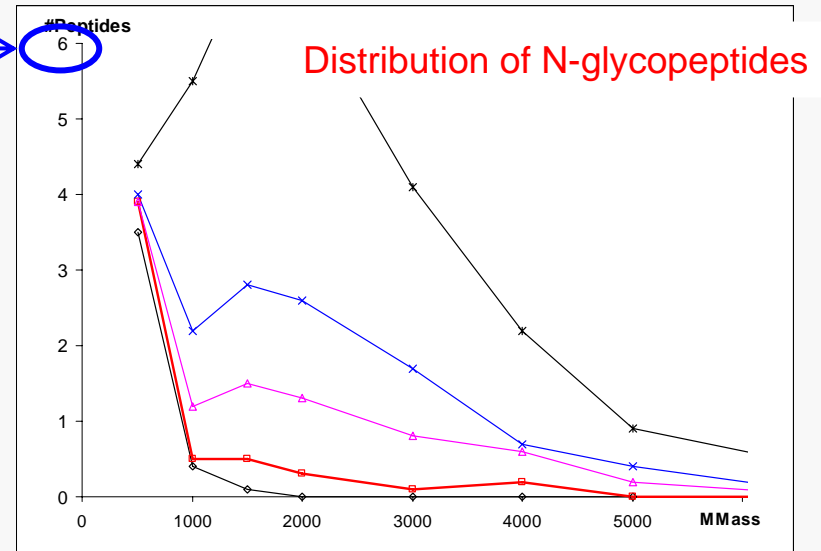
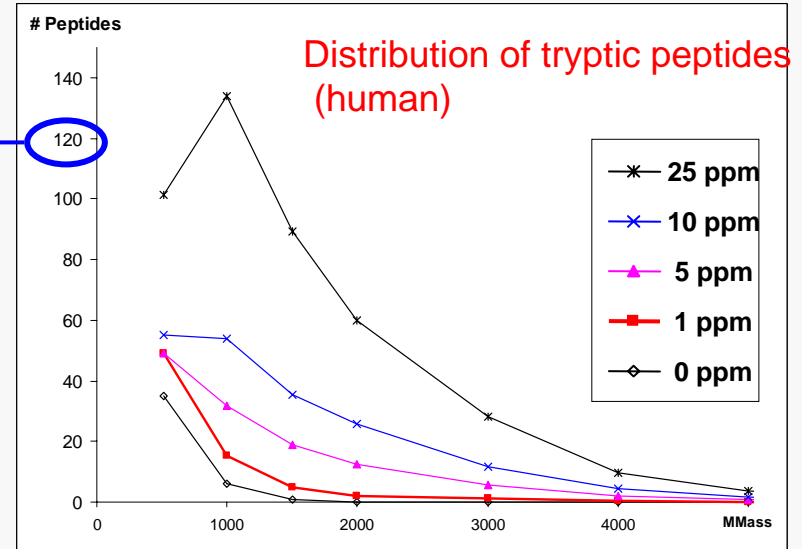
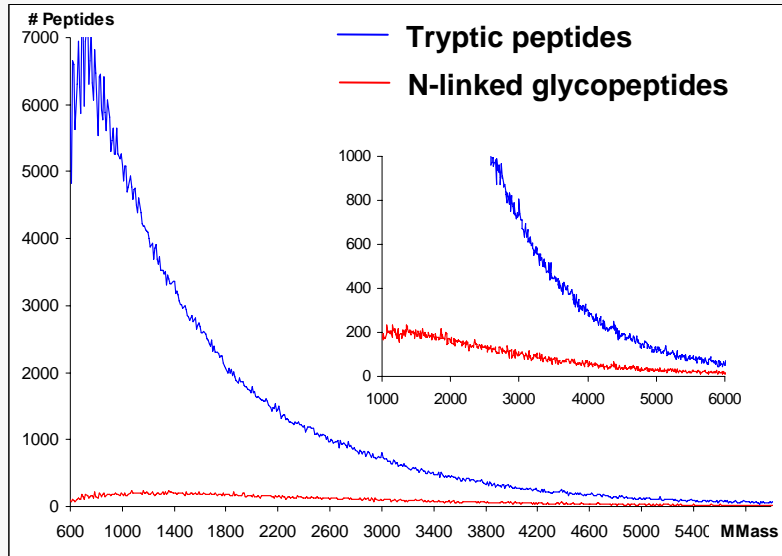
Establish a collaborative consortium biomarker project based on the analysis of *N*-glycosites isolated from serum and cell/tissue sources.

Why *N*-glycosites

Focus on sub-proteome increases likelihood to identify cell/tissue derived peptides in serum



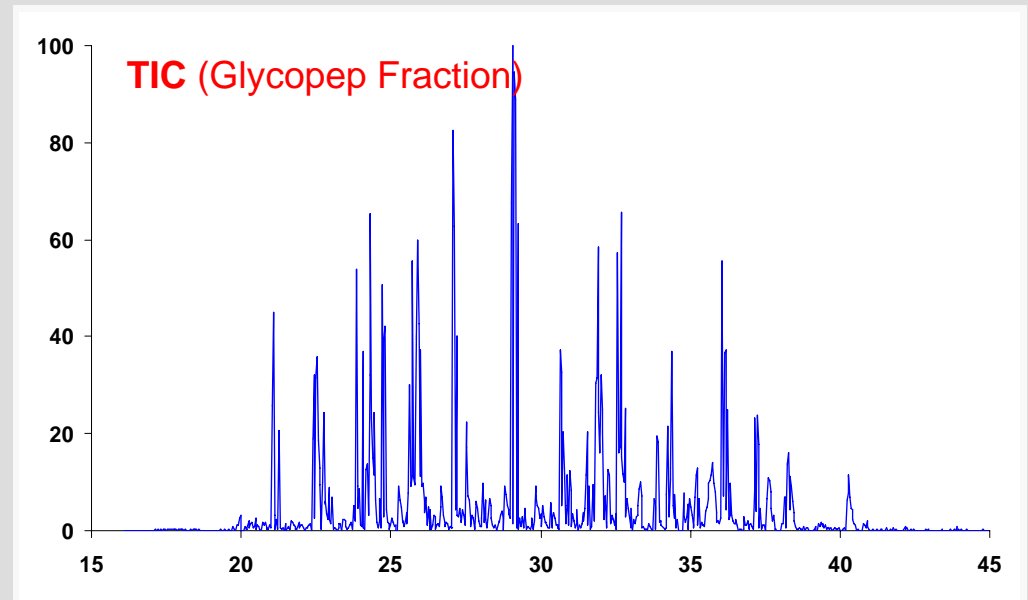
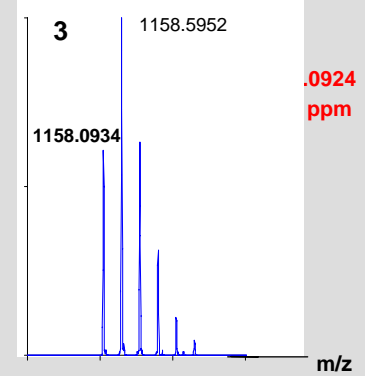
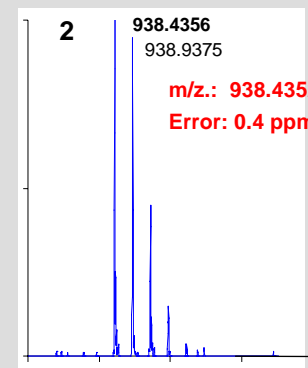
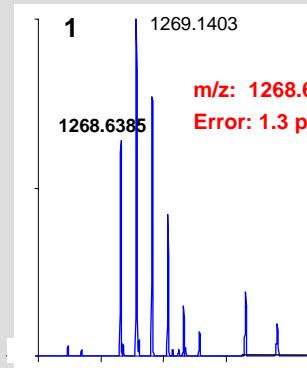
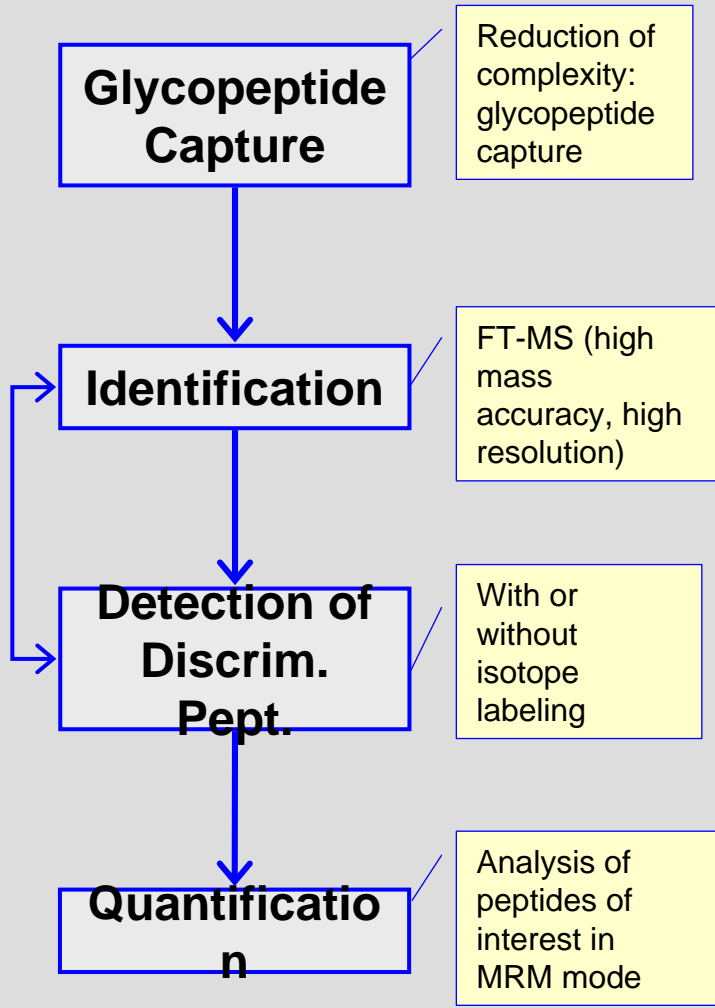
Mass Accuracy and Subproteome



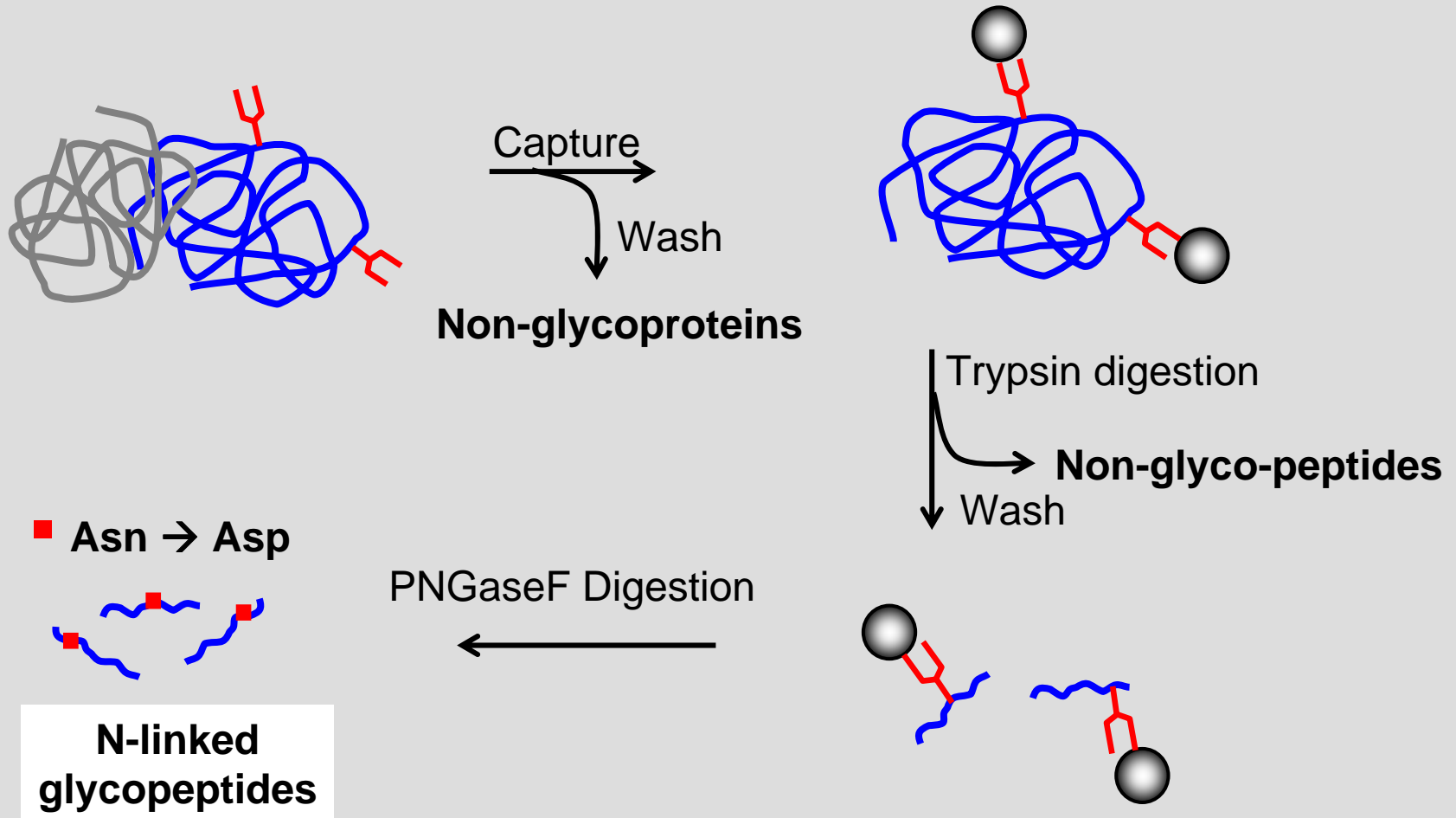
Occurrences of peptides and N-linked glycopeptides (human proteins; fully tryptic)

- Glycopeptide capture (N-linked) greatly reduces sample complexity (~20 fold)
- Accurate mass measurements allows identification with higher confidence
- Faster database searches

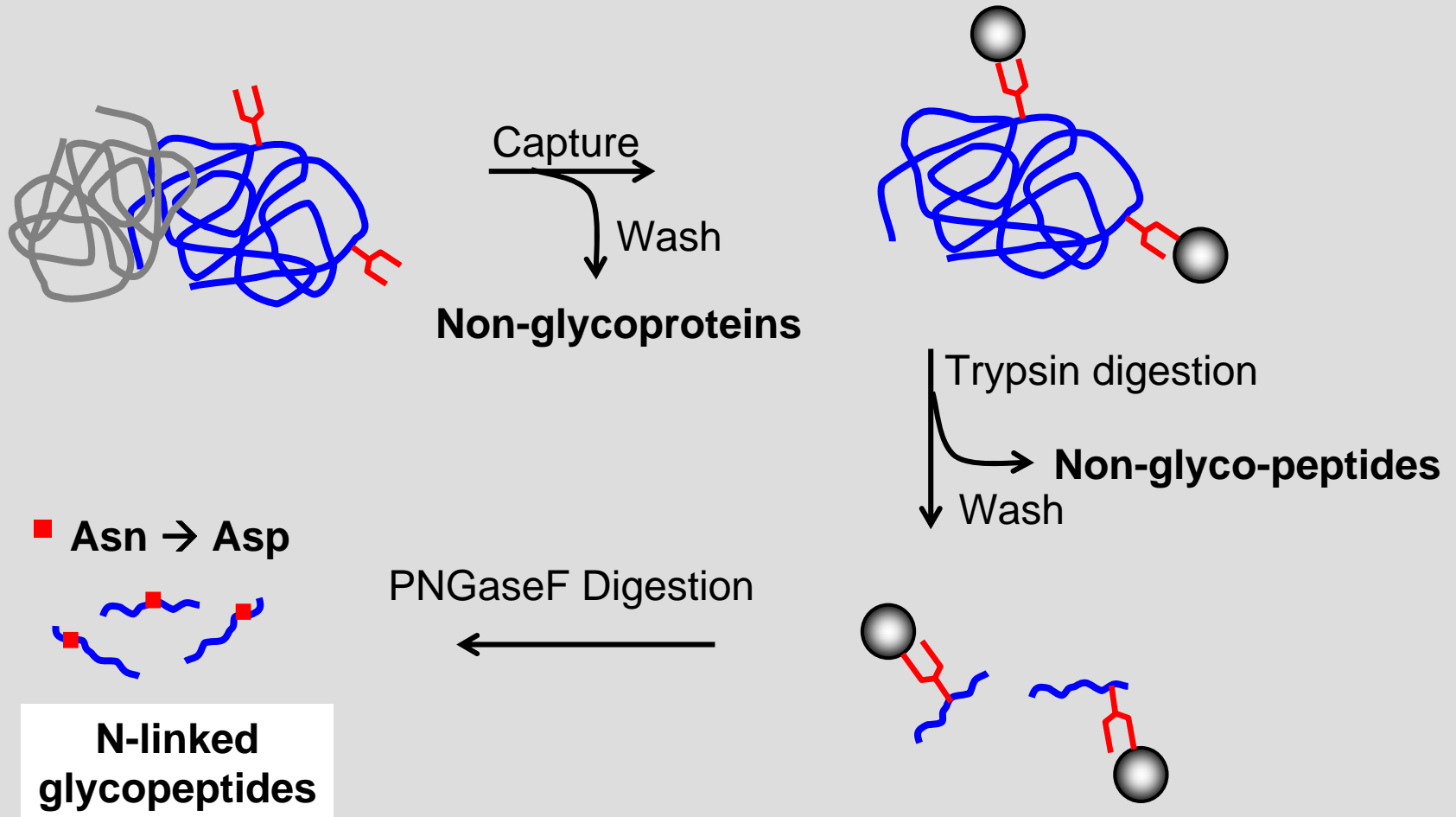
Principle of the approach



Glycopeptide Isolation



Glycopeptide Isolation



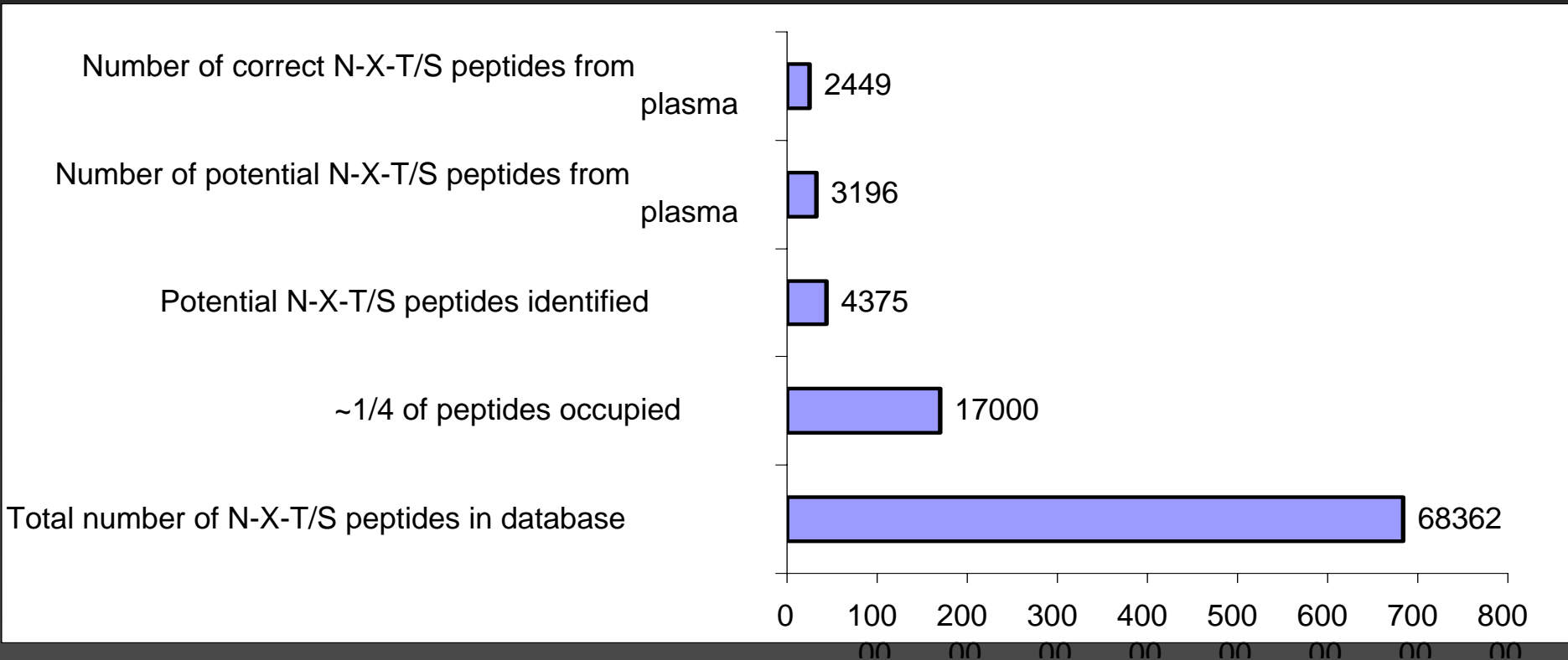
Isolation method is well established for serum, cell and tissue samples

Aims:

- Discovery phase: Establish a database of ALL *N*-glycosites in the human proteome
- Analysis phase:
 - Identify discriminate *N*-glycosites from respective tissue.
 - Detect and quantify discriminate *N*-glycosites in serum
 - Develop various platforms for detecting patterns (MRM-MS, affinity based...)

Discovery phase

Current status: Discovery



The total number of peptides containing N-X-T/S motif from database, the estimated number of peptides occupied, and the number of peptides identified in all tissues and in plasma.

N-glycosite sequencing

- Ruedi Aebbersold group
- Matthias Mann group
- Dick Smith?
- Merck?

Protein Info

IPI ID	IPI00007221
Protein Name	Plasma serine protease inhibitor precursor
Protein Symbol	SERPINA5
Subcellular Location	Secreted
Swiss Prot ID	P05154
Synonyms	Plasma serine protease inhibitor precursor (PCI) (Protein C inhibitor) (Plasminogen activator inhibitor-3) (PAI3) (Acrosomal serine protease inhibitor).
Protein Summary	

Predicted N-linked Proteotypic Glycopeptides

NXS/T	Predicted Sequence	Predicted Mass	Number Proteins with Peptide
249	R. NLS CR.V	591.27	2 Hits
262	R.WGVVPYQG NAT ALFILPSEK.M	2159.15	2 Hits
338	K.VLPSLGISNVFTSHADLSGIS NHS NIQVSEMVHKA	3616.82	1

Identified N-linked Proteotypic Glycopeptides

NXS/T	Identified Sequence	PeptideProphet Score	Tryptic Ends	Peptide Mass	Tissues
249	R.EDQYHYLLDR NLS CR.V	1.00	2	1924.87	serum
262	R.WGVVPYQG NAT ALFILPSEK.M	0.99	2	2160.14	serum, other
338	K.VLPSLGISNVFTSHADLSGIS NHS NIQVSEMVHKA	1.00	2	3617.81	serum

Protein/Peptide Sequence

>[IPI00007221](#) | Plasma serine protease inhibitor precursor
MQFLLLLCIVLLSPQGASLHRHHPREMKRVEDLHVGATVAPSSRRDFTFDLYRALASAA
PSQNIFFSPVSISSMLAMLSLGAGSSTKMQILEGLGLNLQKSSEKELHRGFQQLLQELNQ
PRDGFQLSLGNALFTDLVVDLQDTFVSAMKTLYLADTFPTNFRDSAGANKQINDYVAKQT
KGKIVDLLKNLDSNAVYIMVNYIEFFKAKVETSENHKGTQEQDFYVTSETVVRVPMMSRED
QYHYLLDR**NLS**CR.V**VGVVPYQGNAT**ALFILPSEK**MOOVENGLSEKTLRKVLKMFKKRQLE**
LYLPKFSIEGSYQLEK**VLPSLGISNVFTSHADLSGISNHSNIQVSEMVHKA**AVVEVDESQT
RAAAATGTIFTFRSARLNSQRLVFNRPFLHFIVDNNILFLGKVNRP

Click a button to highlight different sequence features

Glyco Site

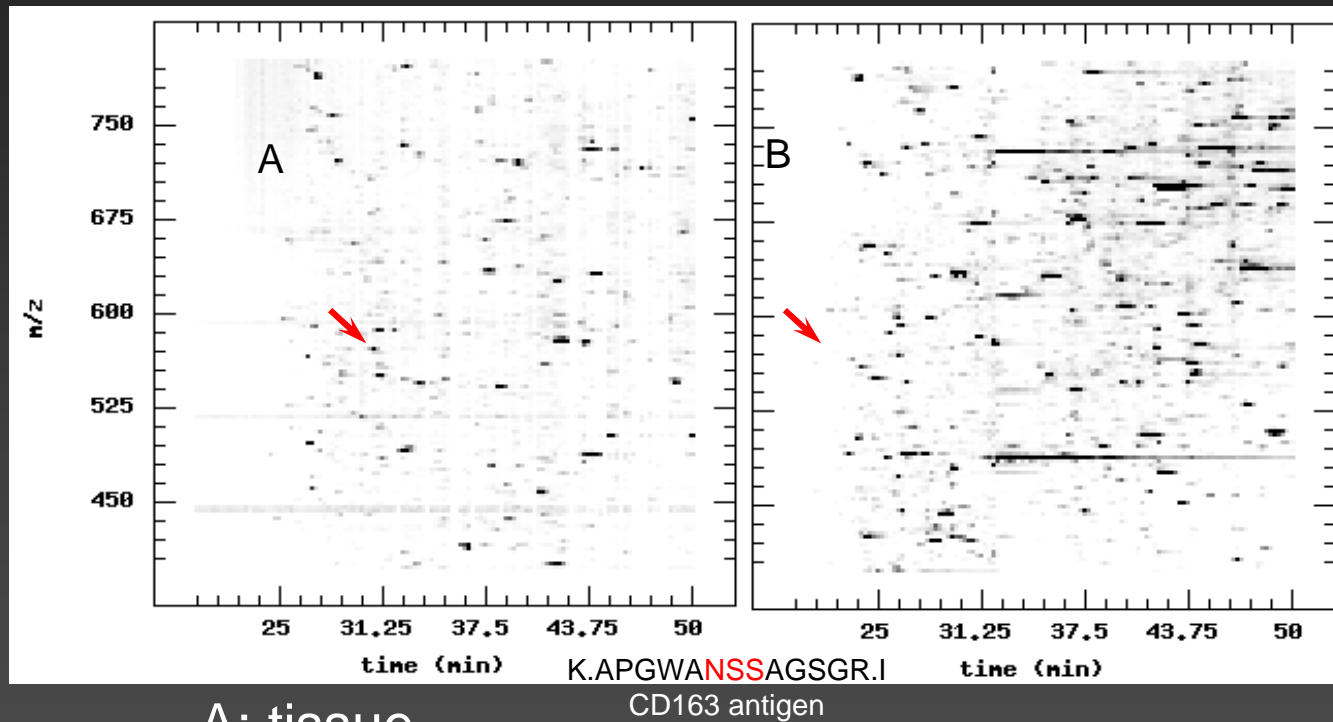
Predicted Peptide

Identified Peptide

Signal Sequence

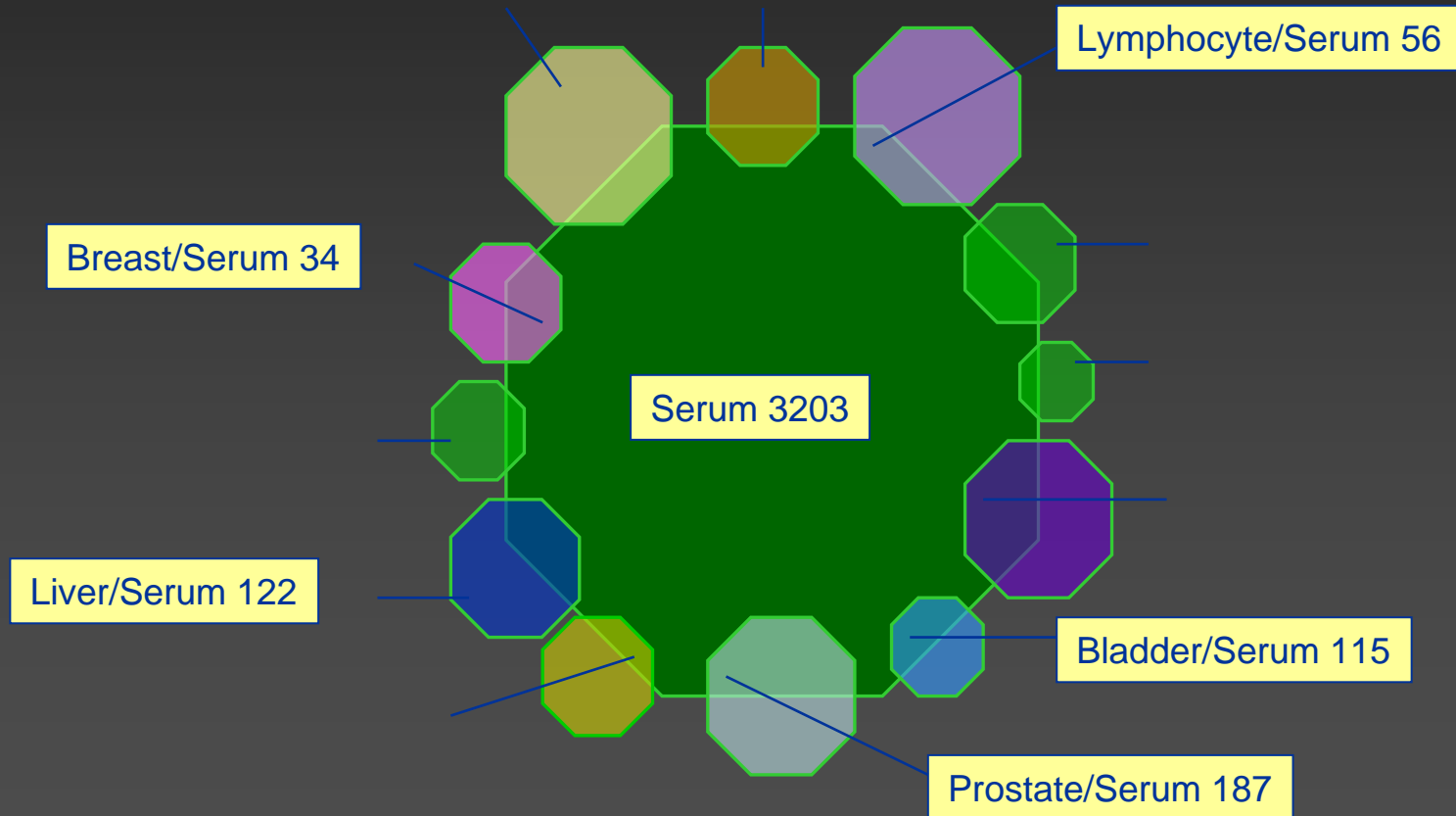
Analysis phase

Current status: Serum-tissue



Peptide patterns and identification of shared peptides from ovarian cancer tissue (left) and plasma (right).

Current status: Glycosite overlap between serum and tissue



Absolute Quantification

- Method

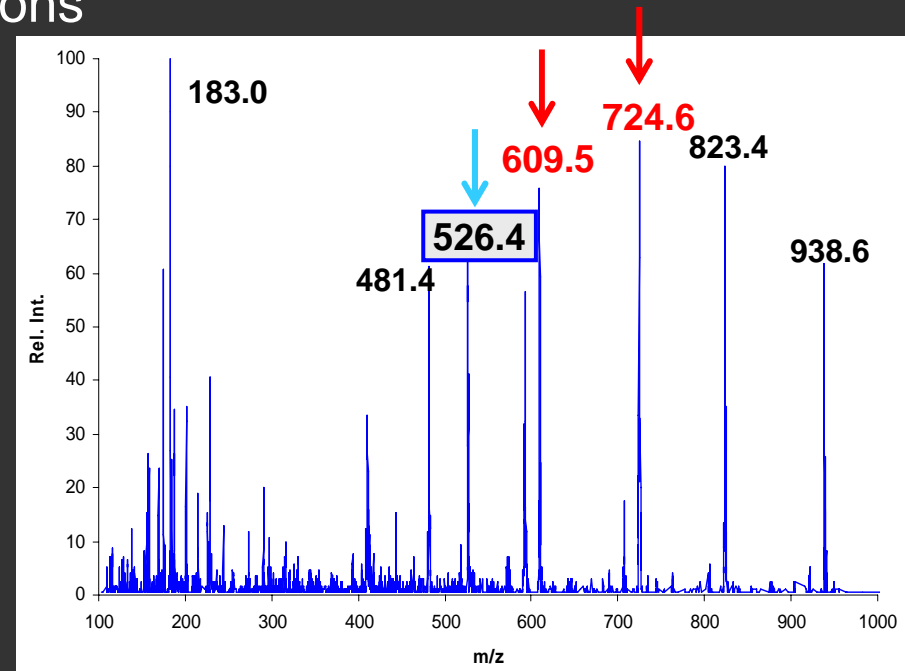
- Addition of defined amount internal standards into glycopeptide fraction
- Analysis performed on a QTrap instrument
- Monitor fragmentation pathways specific to each peptide (MRM)

- Multiple Reaction Monitoring

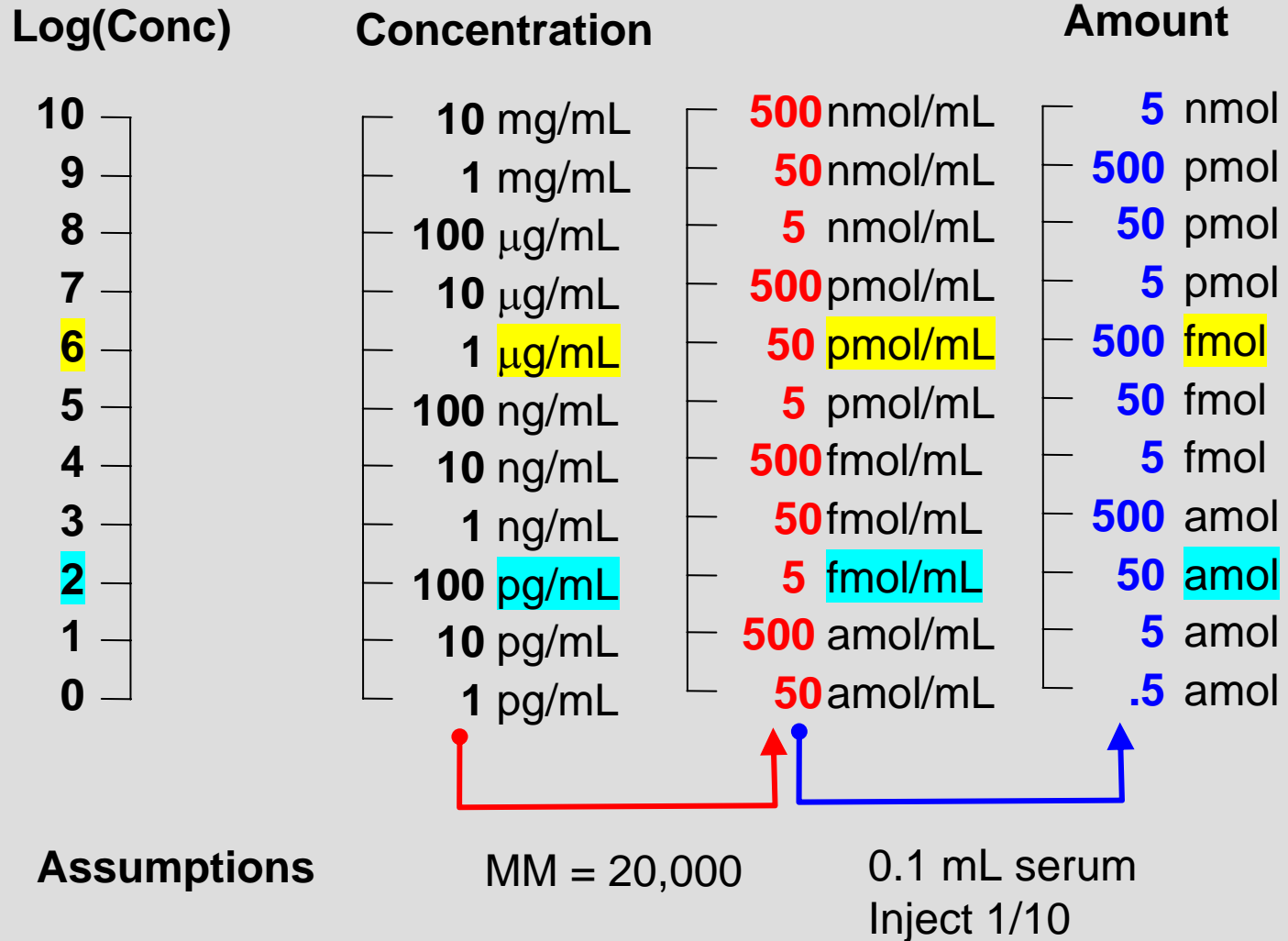
- Typically, doubly charged precursors
- High intensity, non-degenerated y-ions
- Two transitions per peptide

- Internal standards

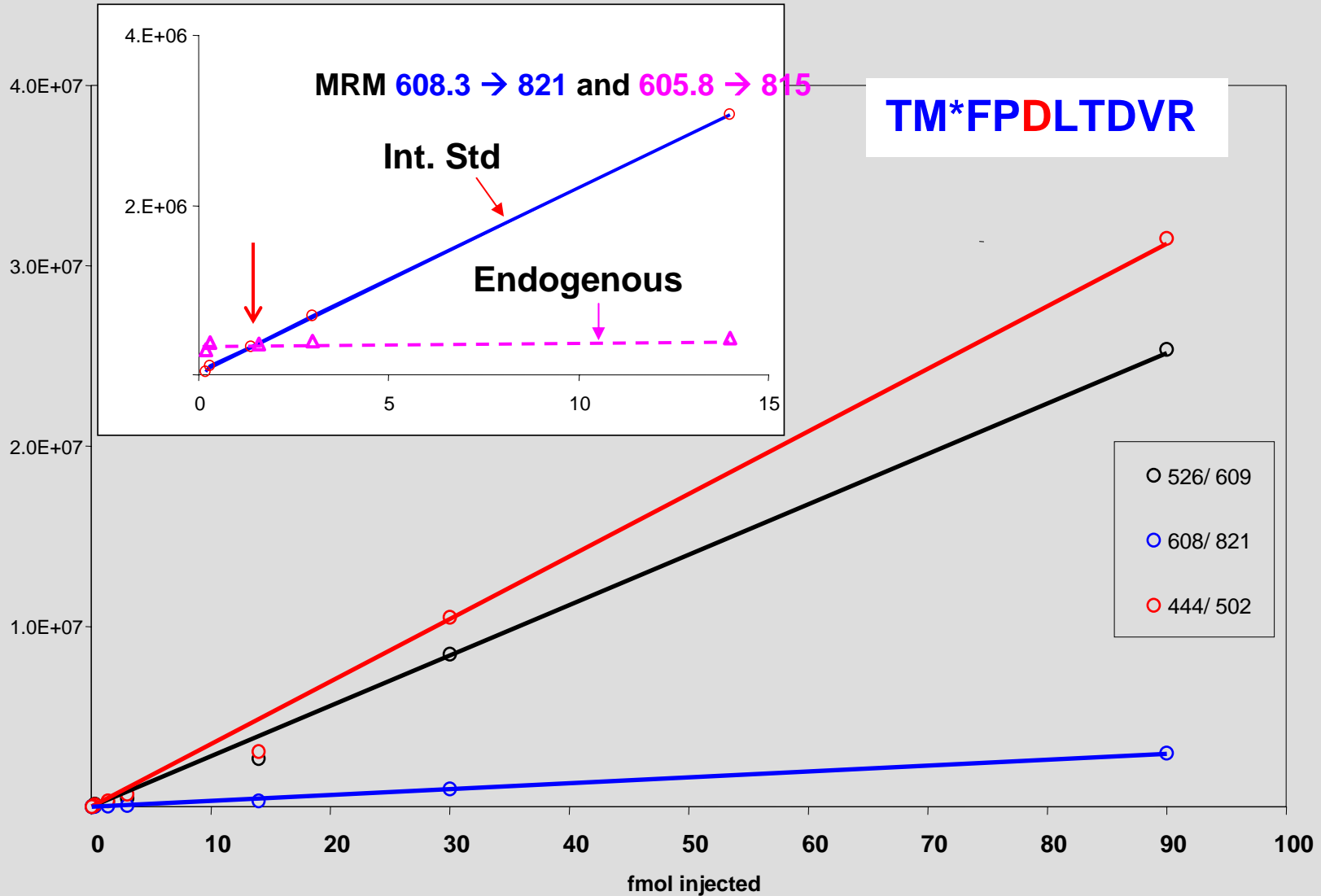
- Proteotypic peptides
- Synthetic $^{13}\text{C}/^{15}\text{N}$ labeled peptides
- Coeluting



Number Game



Absolute Quantification



Summary

- After the discovery phase all experiments will navigate in known territory (analogous to SNP scoring).
- Navigating in a fully explored space is critical for dealing with the undersampling problem
- The fact that all patterns (markers) are contained in the same mapped space means that the data between disease types are comparable between studies.
- Methods well worked out
- Peptide/protein identification problem reduced
- Compatible with different analysis methods because the unit of information, the glycosite is precisely defined (and can be chemically synthesized).
- Reagents , synthetic peptides, antibodies etc. against well defined entities can be synthesized.
- Data drive prioritization of marker development
- Very easy to share data.