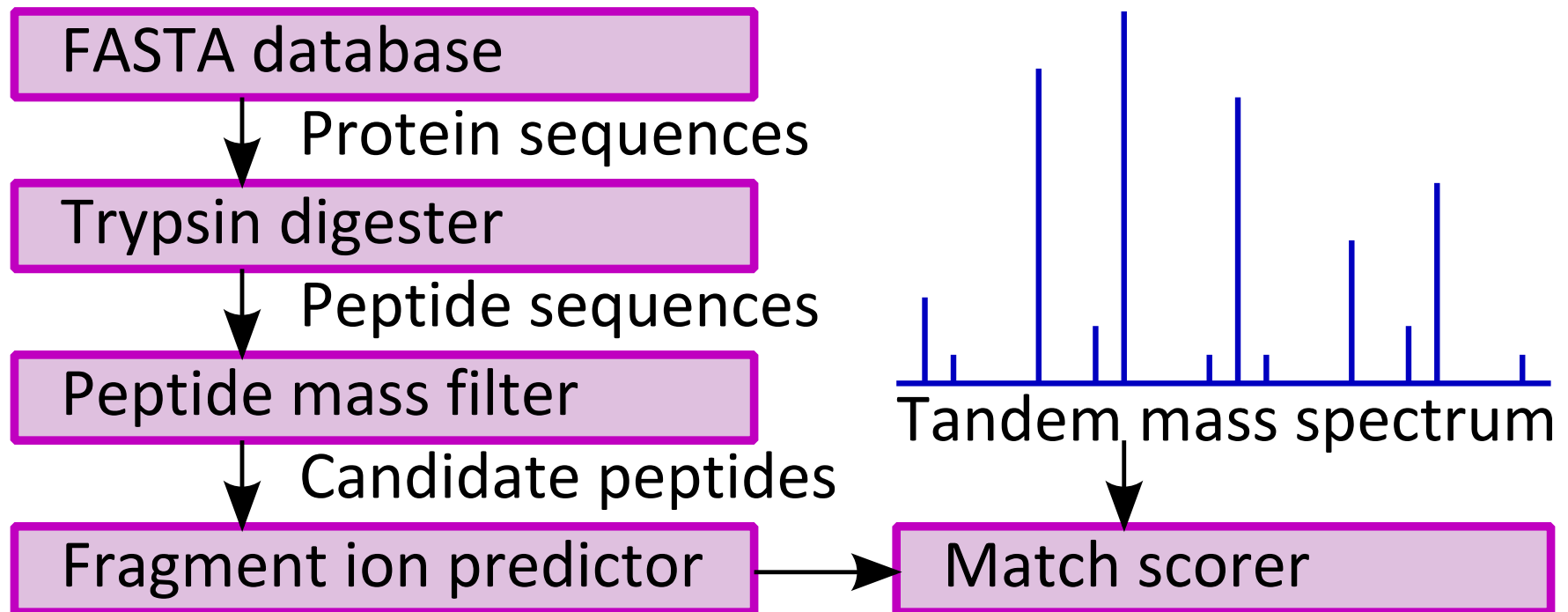
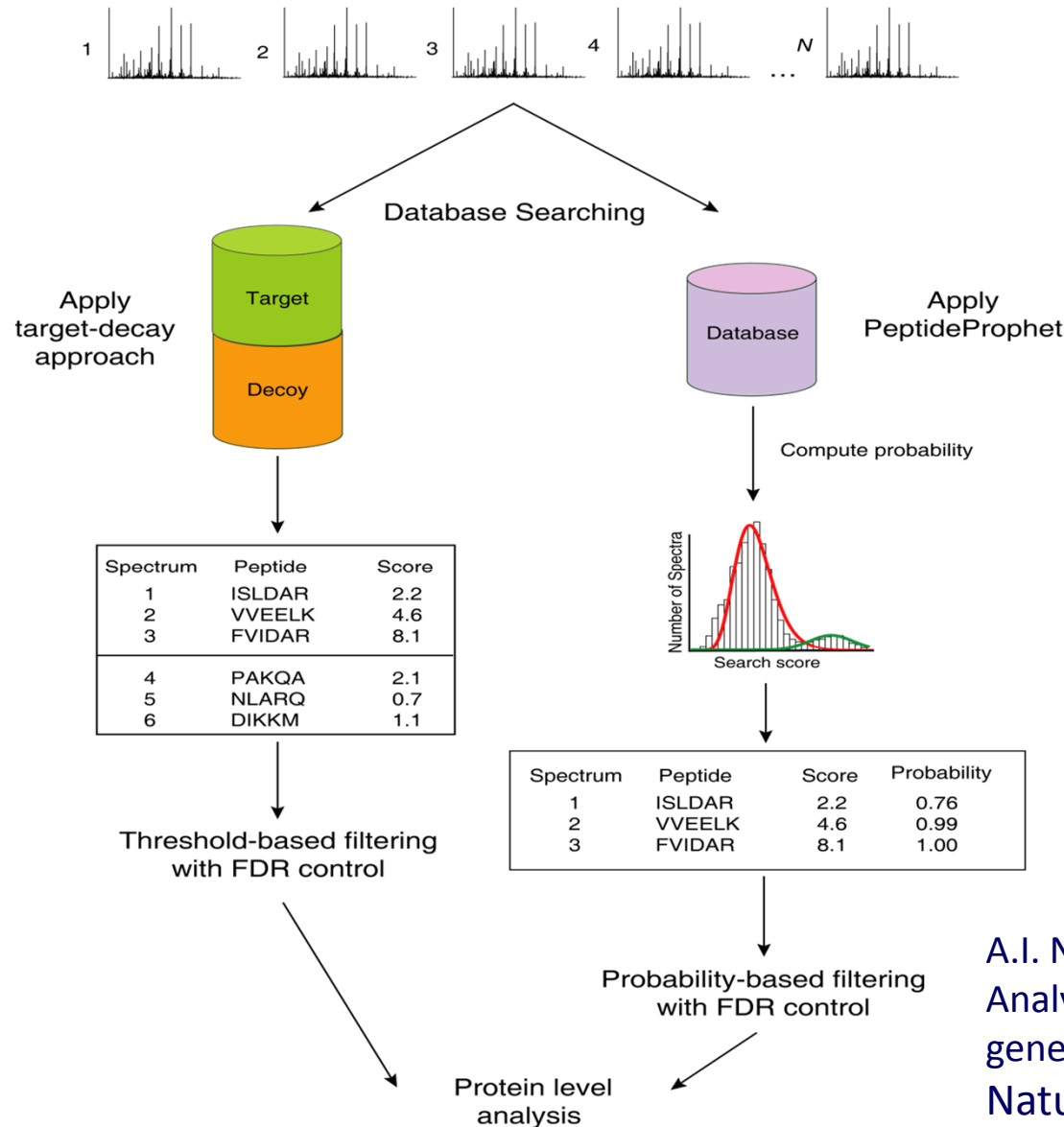


Database search overview



Error Rate Assessment Methods

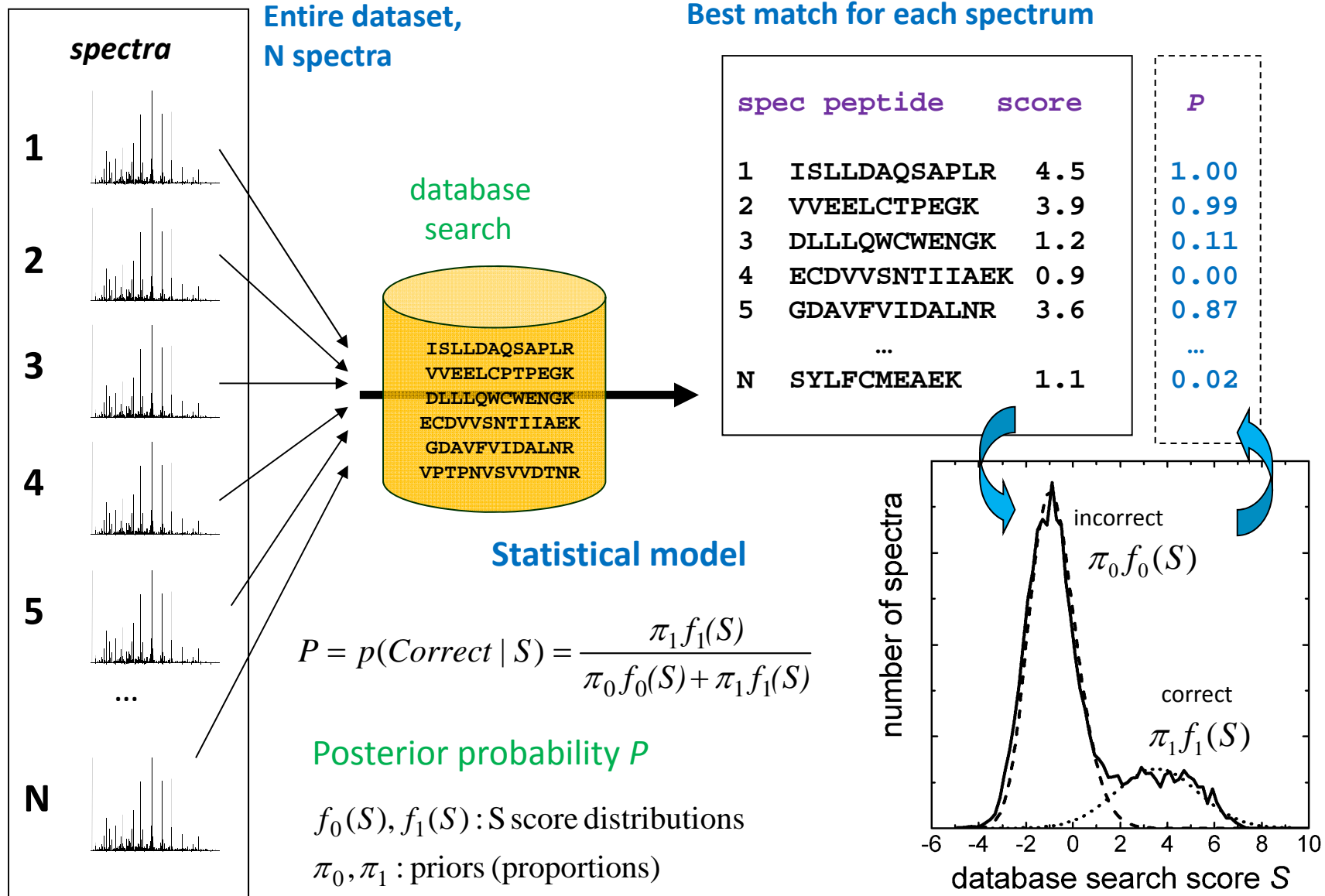


Commonly used methods:

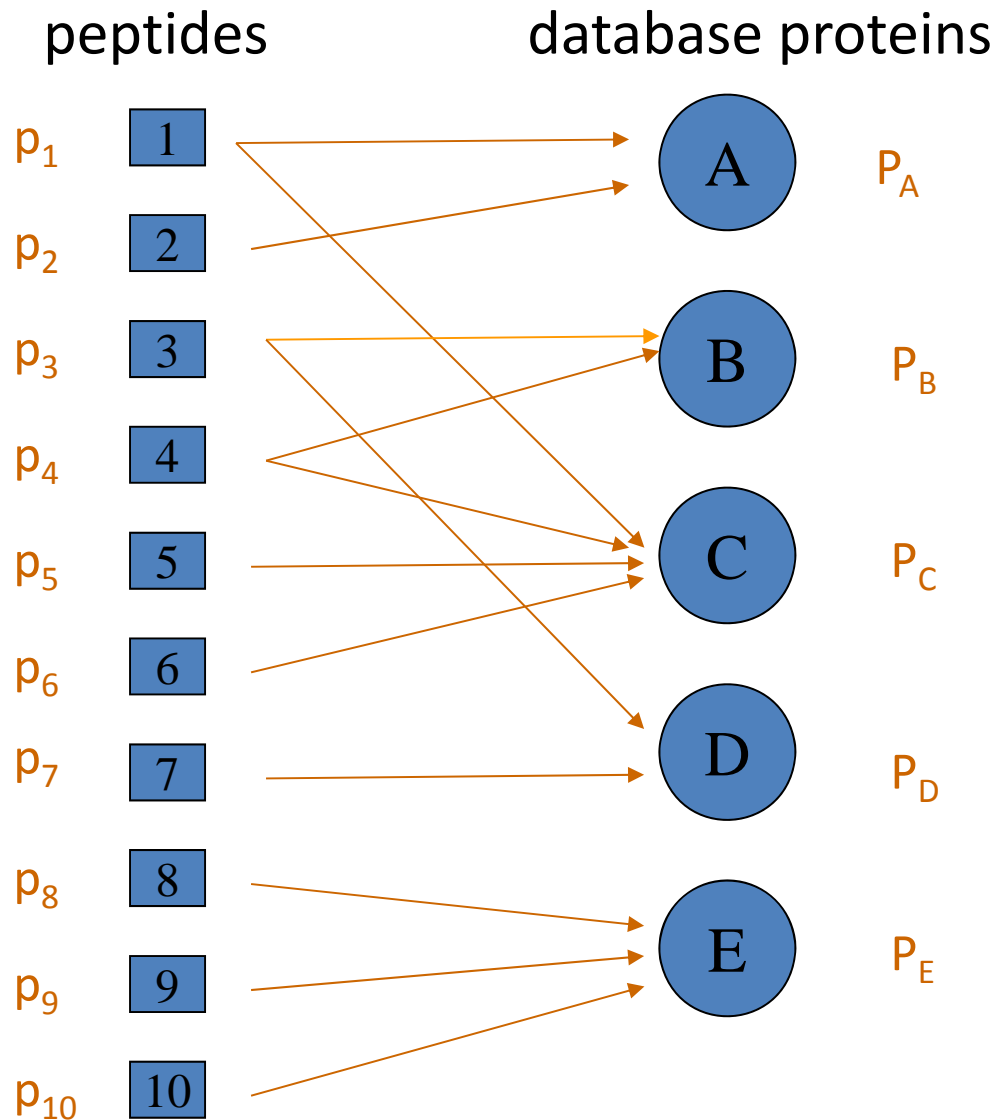
1. Target-decoy database searching (global FDR)
2. Statistical modeling, e.g. PeptideProphet (probability – localFDR)

A.I. Nesvizhskii, O. Vitek, R. Aebersold
 Analysis and validation of proteomic data generated by tandem mass spectrometry
 Nature Methods 4, 787-97 (2007)

Mixture Model-based Methods



Protein Inference



Task for computational tools:

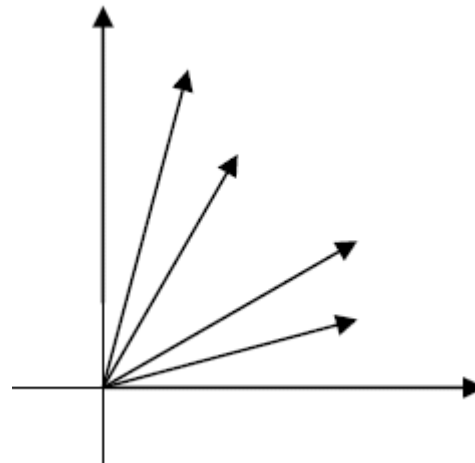
Probability-based model that aims at deriving minimum list of proteins that can explain all observed peptides

Resolve ambiguous cases (when possible) and presents results in a convenient “biologist-friendly” format

Spectral Library Matching

- SQRT normalization of intensities boosts small peak value in matching.
- Dot products of spectra are related to angle between two vectors.
- A smaller angle is a better match.
- Recognizing *chimeric* scans becomes feasible.

$$\cos(A, B) = \frac{A \bullet B}{\|A\| \|B\|}$$

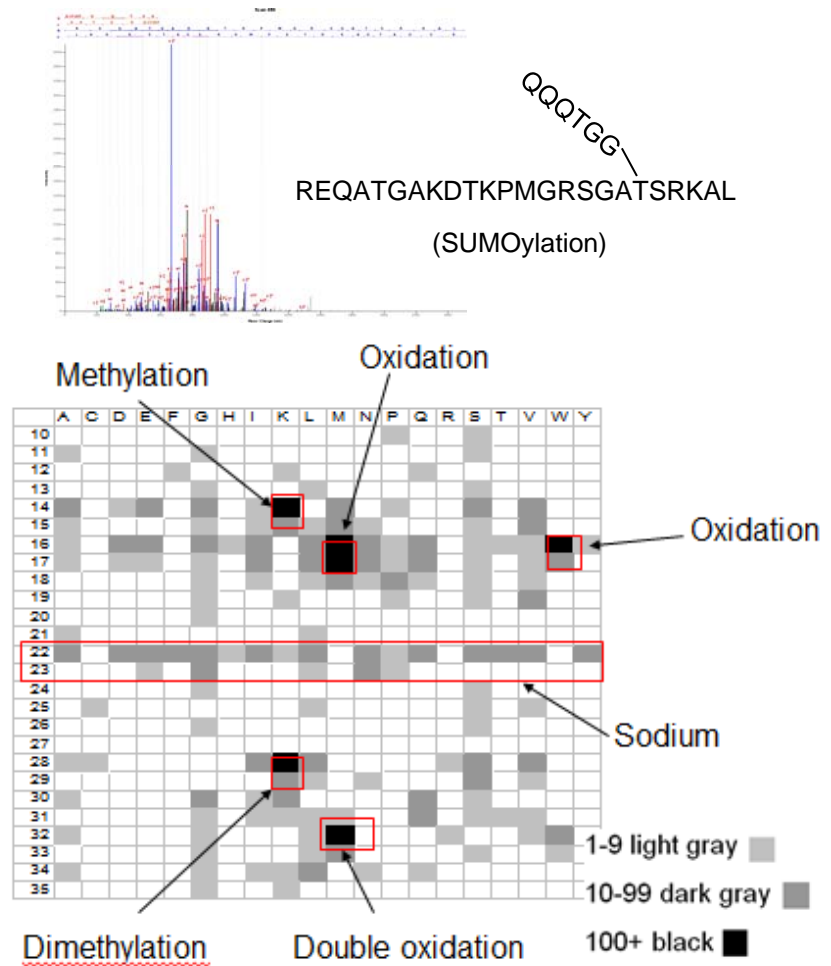


Stein and Scott, *JASMS* (1994) 5: 859-866.

Wang et al, *Mol. Cell. Proteomics* (2010) 9: 1476-1485.

Exotic PTM Recognition

- Profiling PTM spectra recognizes distinctive fragmentation patterns.
- Sequence tags discover PTMs of unknown mass and specificity.
- Patterns of mass shifts reveal dominant PTMs in samples.



Bandeira et al, *PNAS* (2007) 104: 6140-6145.

Dasari et al, *Chem. Res. Tox.* (2011) 24: 204-216.