CMSC423: Bioinformatic Algorithms, Databases and Tools
Lecture 21

Protein folding

Proteomics

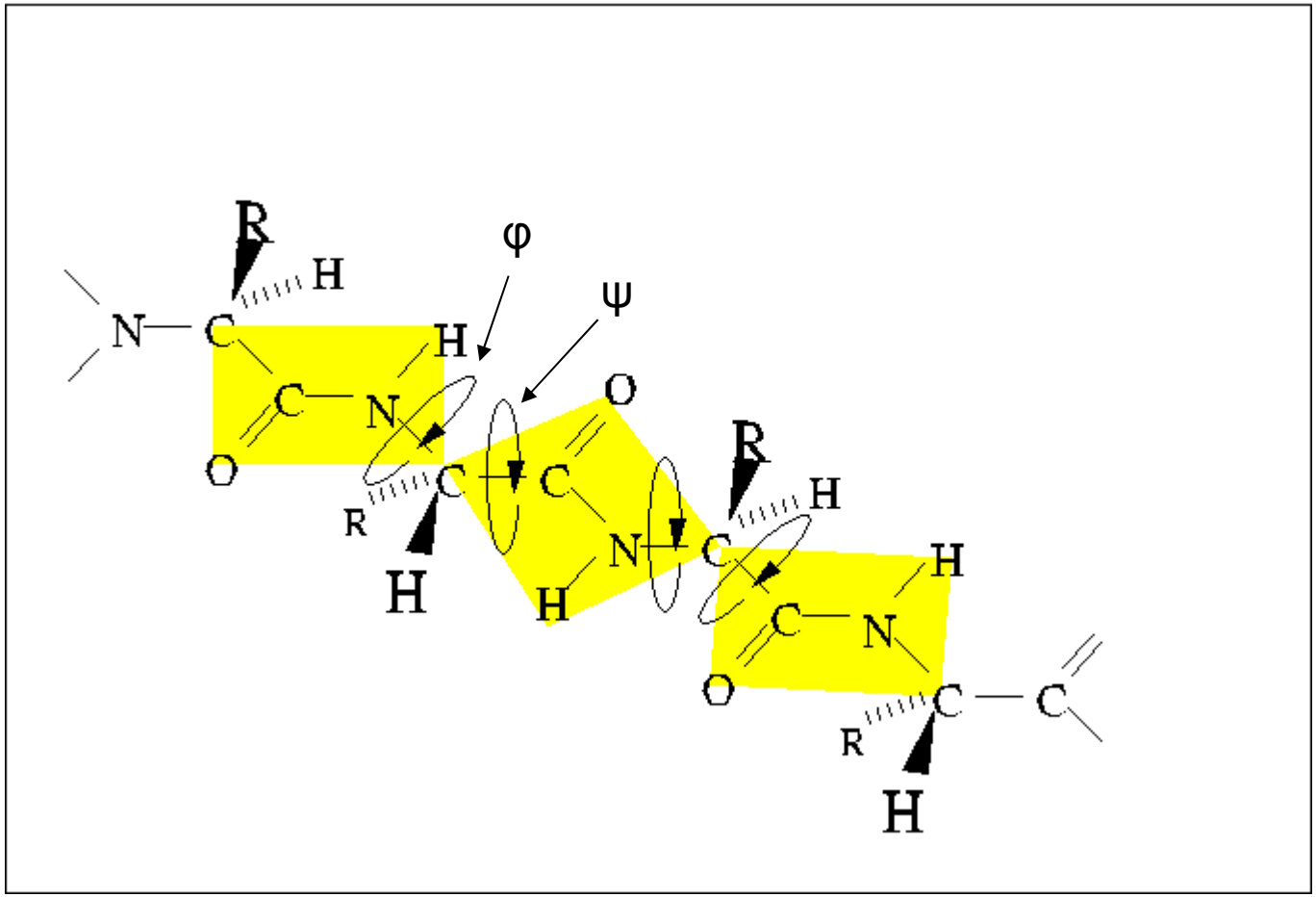
Mass spectrometry

Protein folding

- Note: mis-folded proteins may cause disease (e.g. Creutzfeld-Jakob a.k.a. mad cow)
- Drugs (e.g. antibiotics) often inhibit protein function – knowing structure can help design drugs

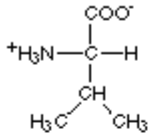
- Folding@home – lend your computer's unused cycles to help fold proteins (like SETI@home) (do you believe in evolution or aliens ?)

Protein structure
(primary structure = sequence)

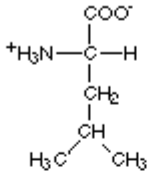


http://www.tulane.edu/~biochem/med/second.htm

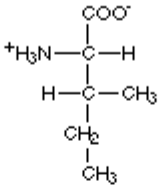
Amino acids with hydrophobic side groups



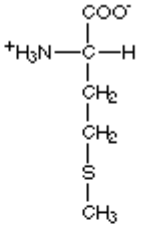
Valine
(val)



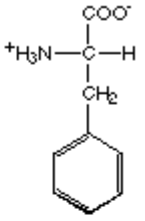
Leucine
(leu)



Isoleucine
(ile)



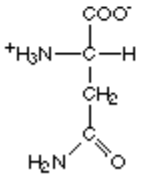
Methionine
(met)



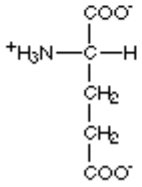
Phenylalanine
(phe)

hate water

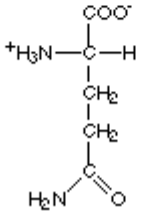
Amino acids with hydrophilic side groups



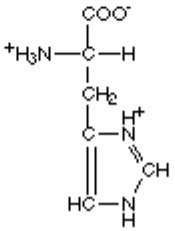
Asparagine
(asn)



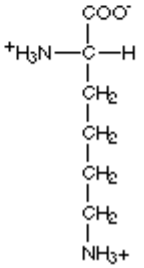
Glutamic acid
(glu)



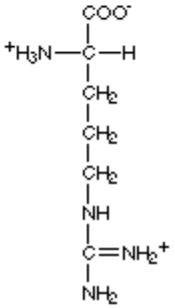
Glutamine
(gln)



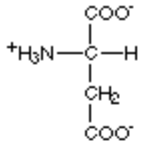
Histidine
(his)



Lysine
(lys)



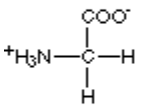
Arginine
(arg)



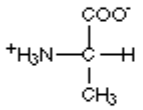
Aspartic acid
(asp)

like water

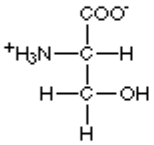
Amino acids that are in between



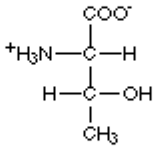
Glycine
(gly)



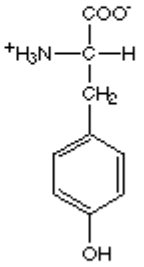
Alanine
(ala)



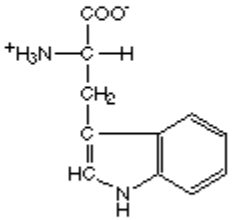
Serine
(ser)



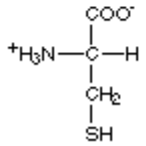
Threonine
(thr)



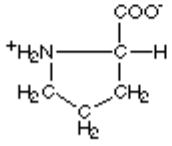
Tyrosine
(tyr)



Tryptophan
(trp)



Cysteine
(cys)

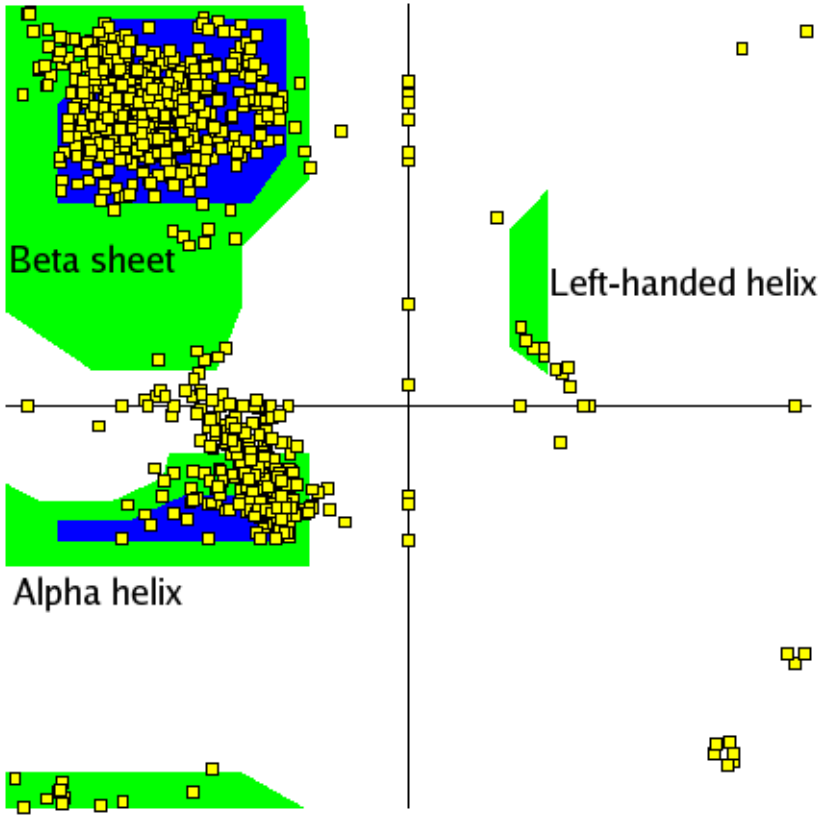


Proline
(pro)

can't decide

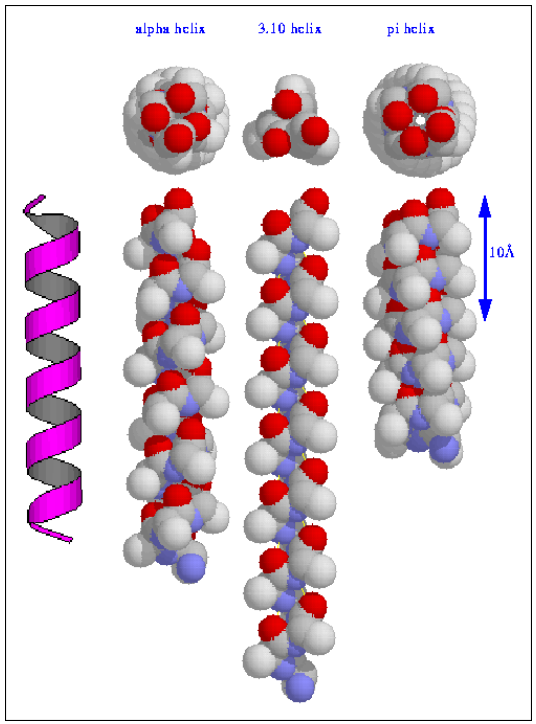
http://web.mit.edu/esgbio/www/lm/proteins/aa/aminoacids.html

Not all bends equally likely
Ramachandran plot

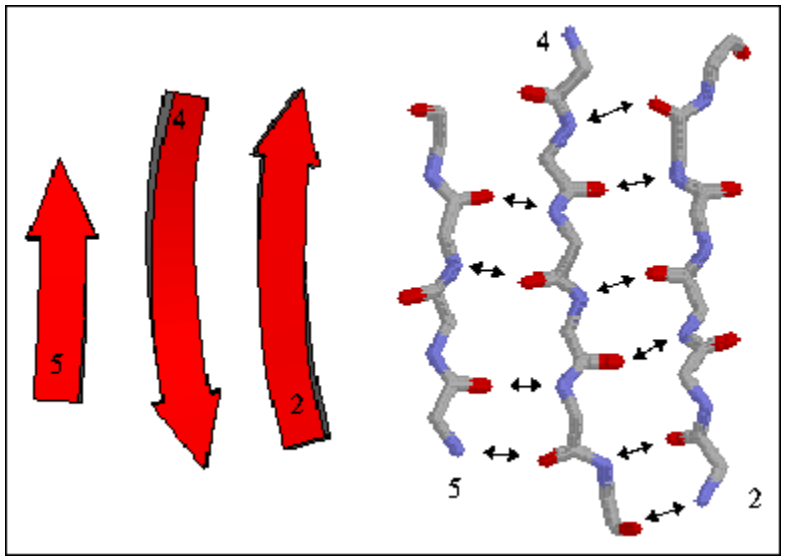


Secondary structure (motifs)

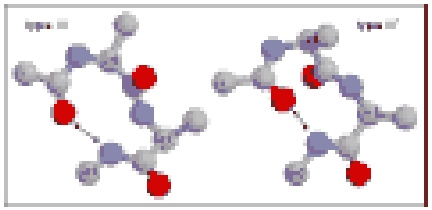
helix



sheet

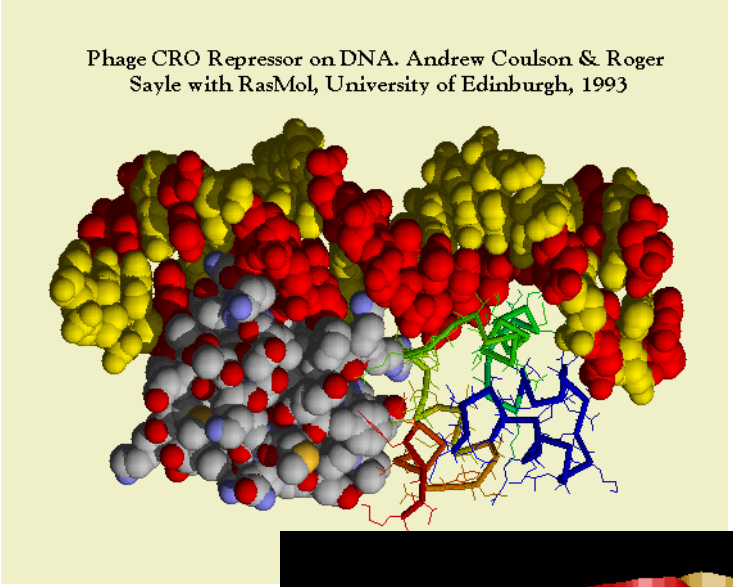


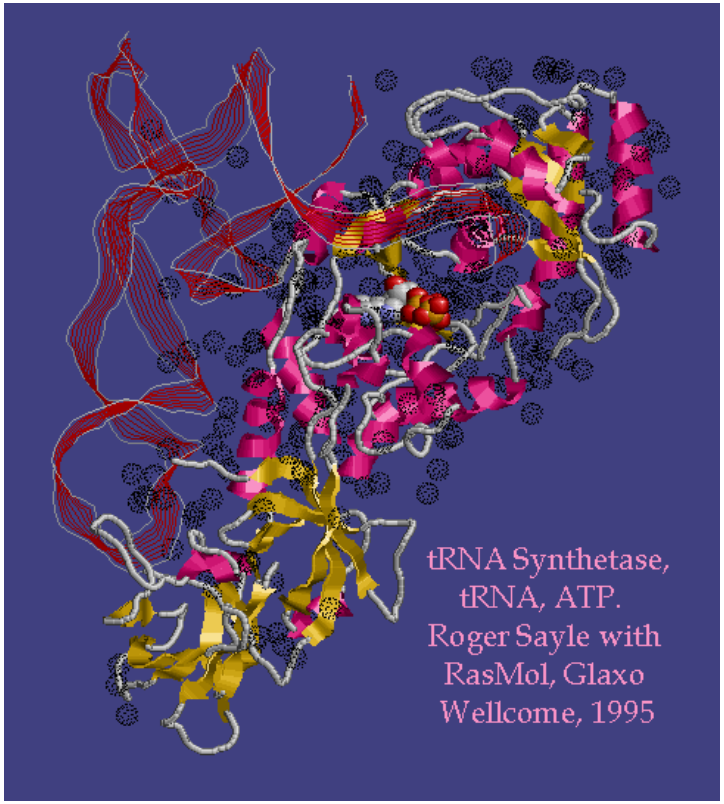
turn

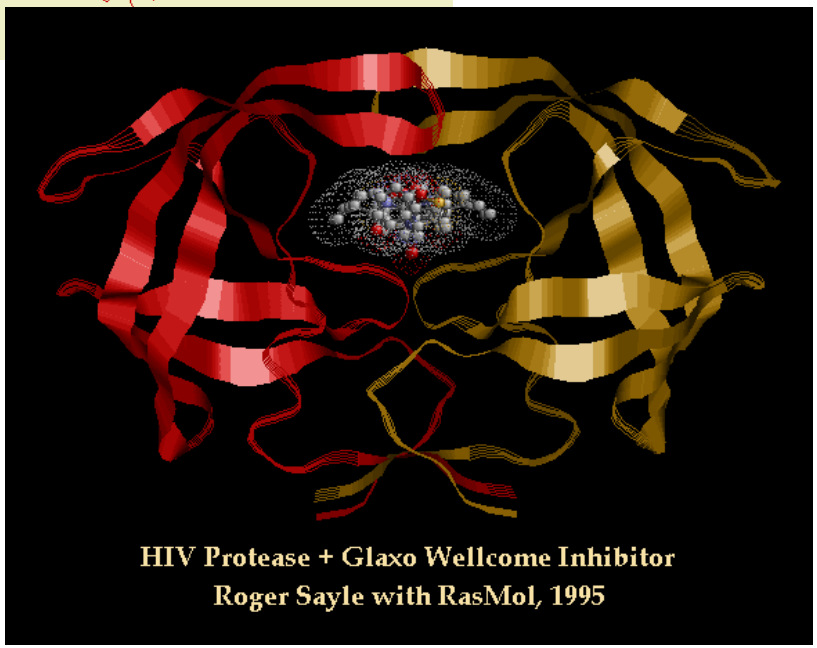


http://alpha2.bmc.uu.se/~kenth/bioinfo/structure/secondary/01.html

Tertiary structure (3D shape)







http://www.umass.edu/microbio/rasmol/sayle1.htm

Folded shape: lowest free energy

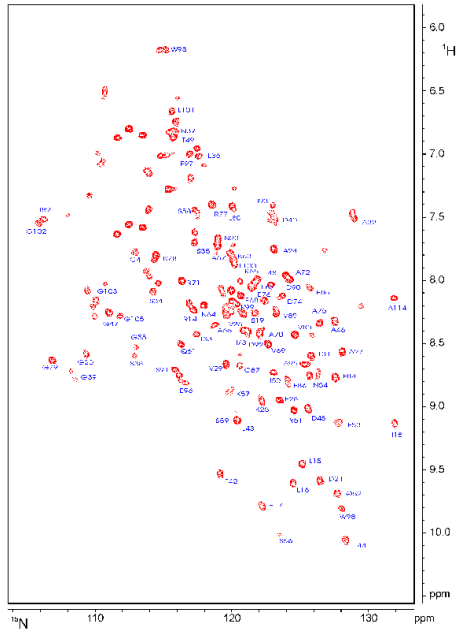
- Energy components
  - electrostatic (~$1/D^2$) (n^2 terms)
  - van der Waals (n^2 terms)
  - hydrogen bonding (n terms)
  - “bending” (n terms)
  - solvent (water/salt) (?? terms)
  - exclusion principle (no two atoms share same volume)
- Energy minimzation
  - small perturbations & computation: hill climbing, simulated annealing, etc.
- Molecular dynamics

How do we know the truth?

- X-ray crystallography
  - crystallize protein
  - shine X-rays
  - examine diffraction patterns

http://www.cryst.bbk.ac.uk/BBS/whatis/cryst_an.html

- Nuclear Magnetic Resonance (NMR)
  - no crystallization necessary
  - magnetic field “vibrates” hydrogen atoms
  - Nobel prize: Kurt Wuethrich



http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/2dnmr.htm

Simpler problems

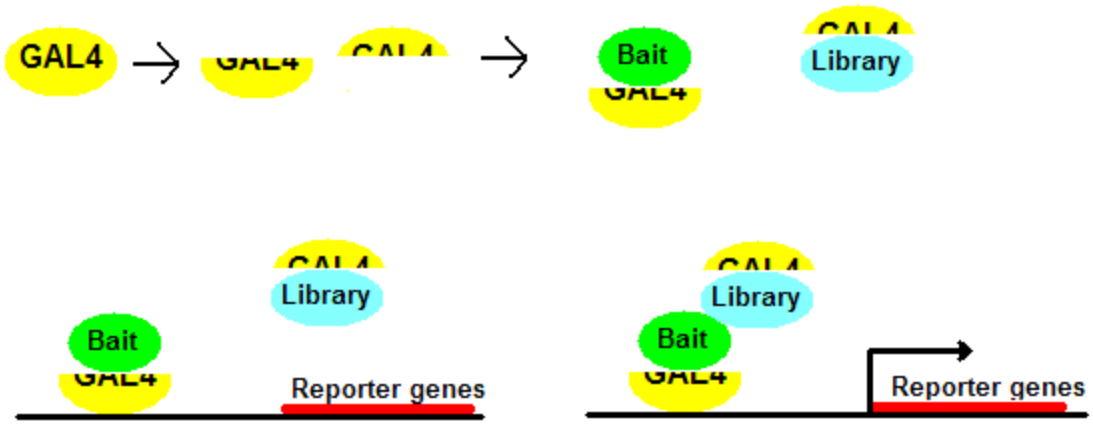
- Secondary structure prediction
- Side-chain conformation (assuming fixed backbone)
- Protein docking (how do proteins interact)
- Database searches (protein threading)

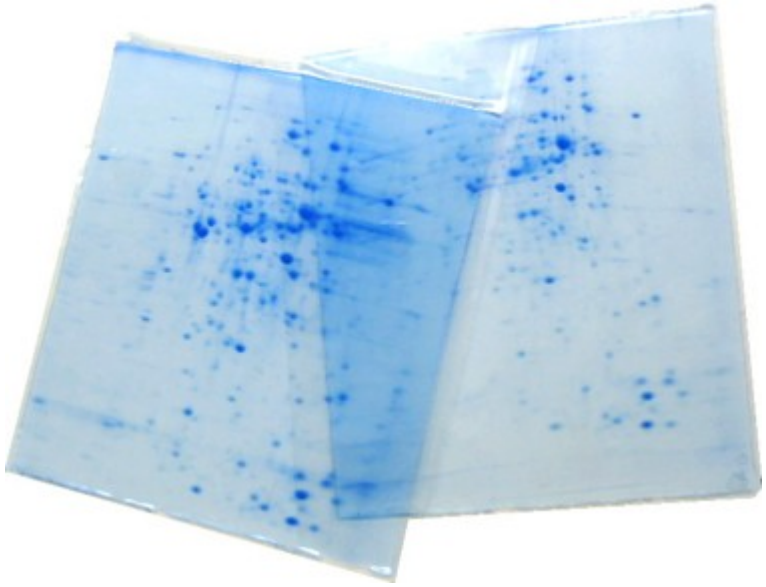
- Simpler energy functions
- Folding on a lattice (theoretical approximation)

- Critical Assessment of Fully Automated Structure Prediction – competition on proteins with unpublished 3D structure

Proteomics

- Large-scale analysis of proteins
  - protein-protein interactions (e.g. yeast 2-hybrid)
  - 2D gels (mass vs. isoelectric point)
  - Mass-spectrometry
  - Protein microarrays
  - etc.





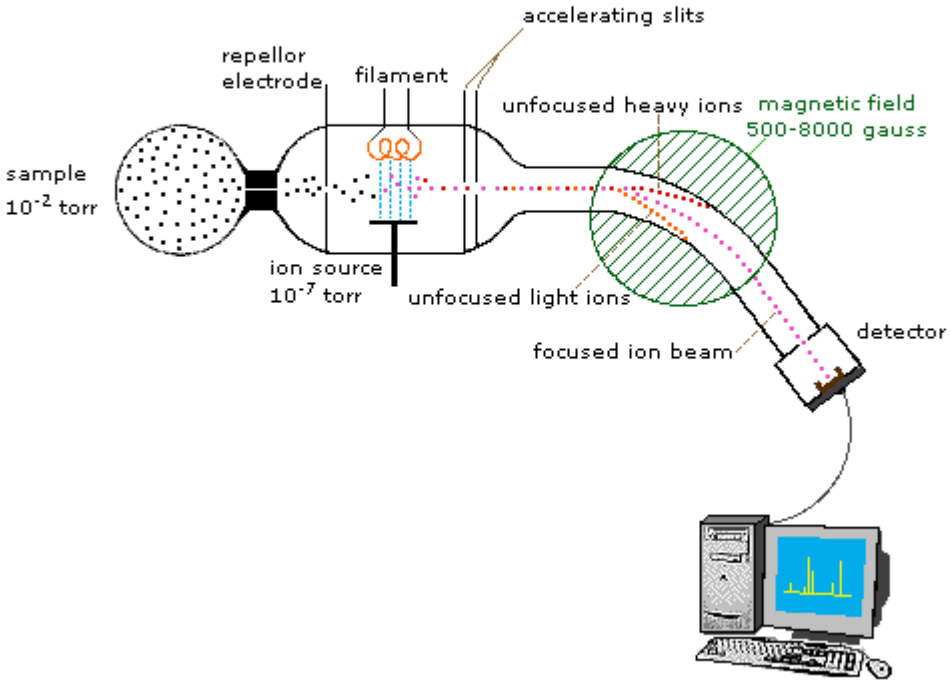
Proteomics

- Why proteomics? Are DNA/RNA microarrays not sufficient?
- RNA abundance is not necessarily related to protein abundance
- Many proteins are modified post-translation
  - addition of additional molecules (phosphate, sugars, etc.)
  - creation of complexes (hemoglobin is actually 4 molecules)

Mass spectrometry

- Technique for measuring the mass-to-charge ratio of ions
- Basic idea
  - shoot ions into a magnetic field
  - deflection depends on mass
- Output of a mass-spectrometer
  - ions “sorted” by mass
  - for each mass bucket - number of ions with that specific mass

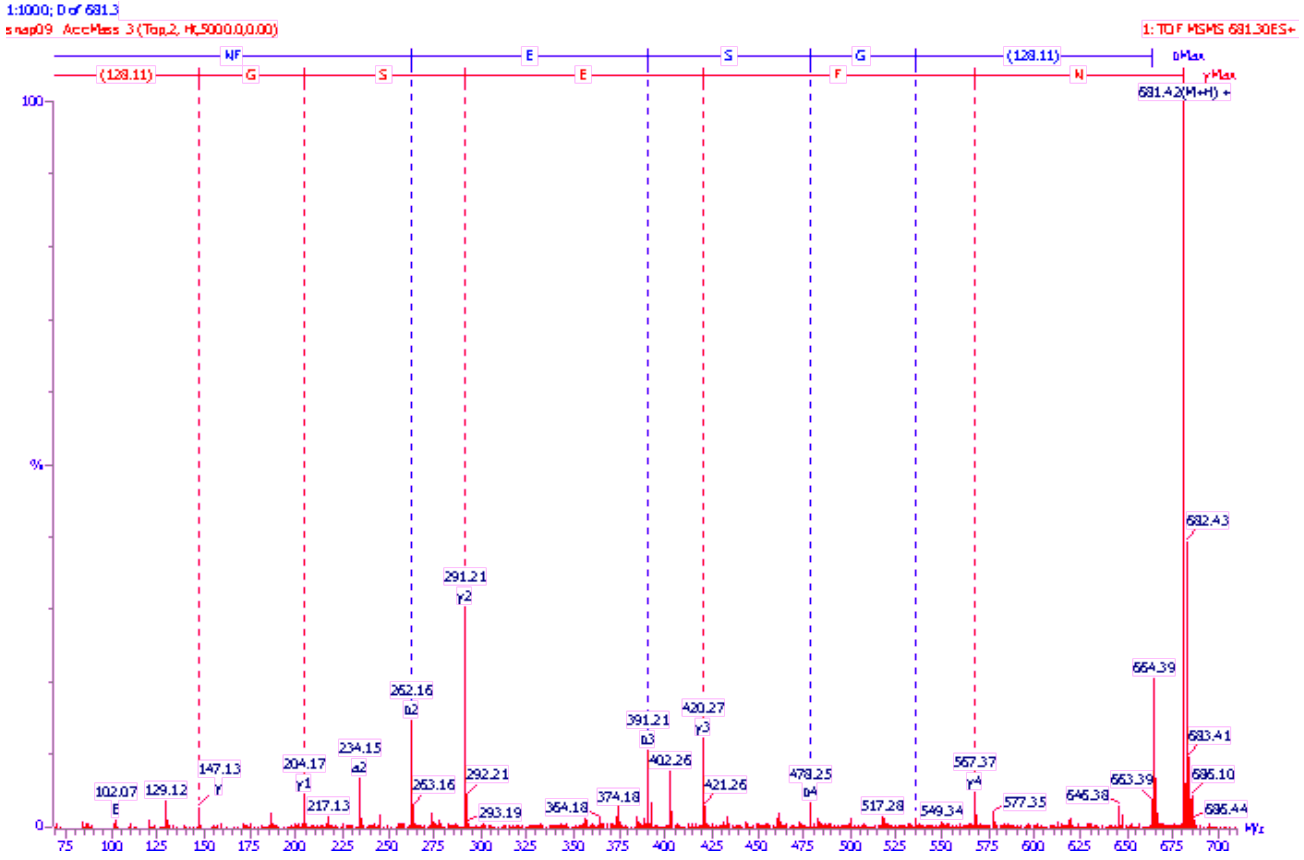
Mass-spectrometry



http://www.cem.msu.edu/~reusch/VirtualText/Spectrpy/MassSpec/masspec1.htm

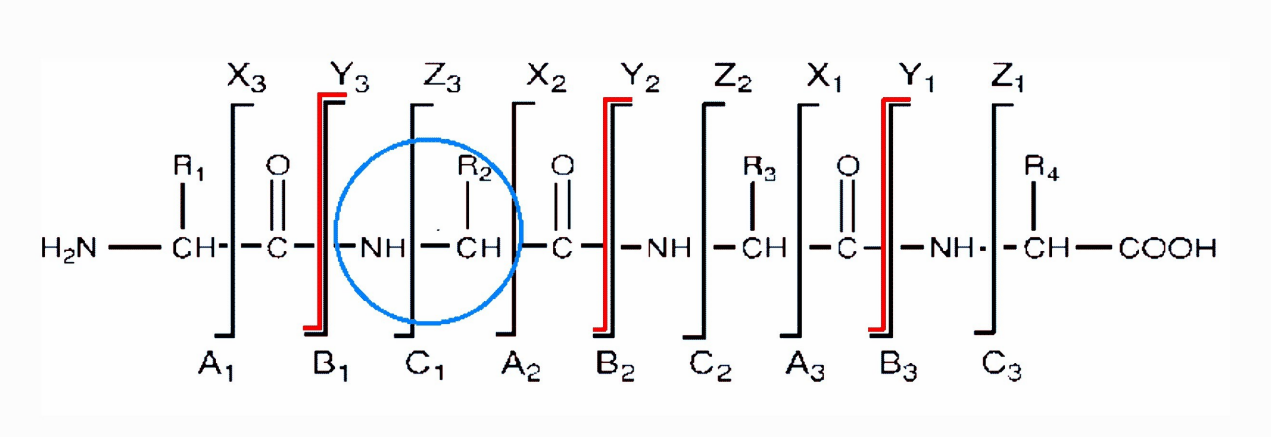
Tandem Mass Spectrometry

- First mass-spectrometer “focuses” on a specific protein
- Second mass-spectrometer breaks the protein into smaller chunks
- Problem: given the chunks, what was the original protein?



Peptide sequencing

- Peptide - a chunk of a protein, usually obtained by enzymatic cleavage of the protein (using trypsin)



- Problem: Given an MS spectrum (weights of fragments), what was the sequence of the peptide?
- Or: find the peptide (of mass m) that best matches the experimental data

MS Algorithms

- Database search
  - build database of “all possible” peptides (better - all peptides observed in known proteins)
  - match experimental spectrum to the database
  - closest hit is our peptide
- “Assembly” (de novo sequencing)
  - start with a spectrum
  - identify masses that likely represent the same fragment
  - build graph (spectral graph) that represents adjacency of fragments
    - nodes = fragments (or ion types)
    - edges = edge v->w indicates fragments v and w differ by exactly 1 amino-acid
  - path through this graph represents a peptide sequence

Some Mass Differences between Peaks Correspond to Amino Acids