# CMSC423: Bioinformatic Algorithms, Databases and Tools
# Lecture 24

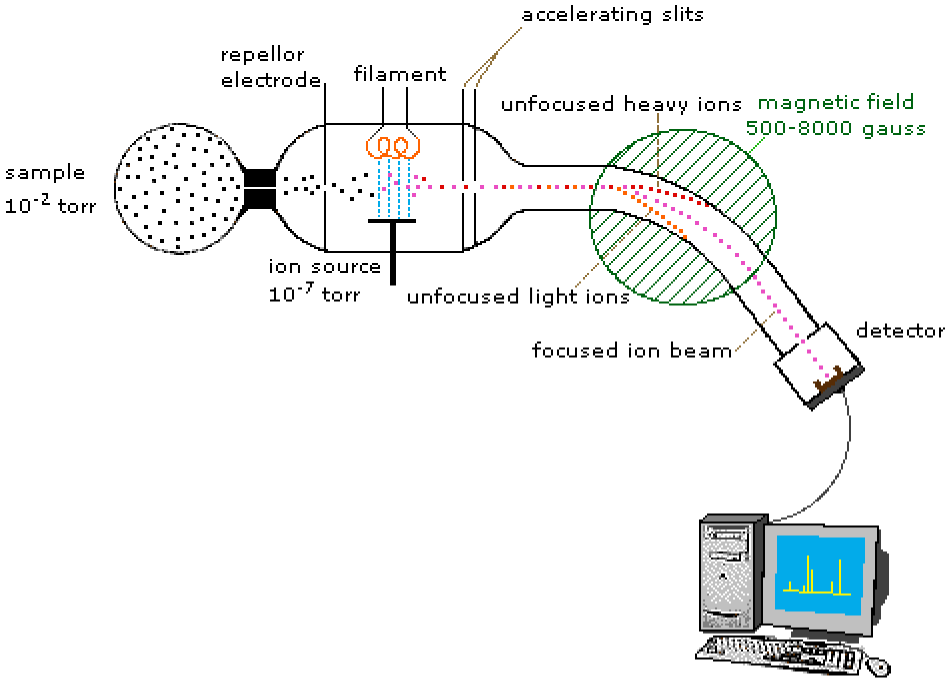
Mass spectrometry

Gene networks

# 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

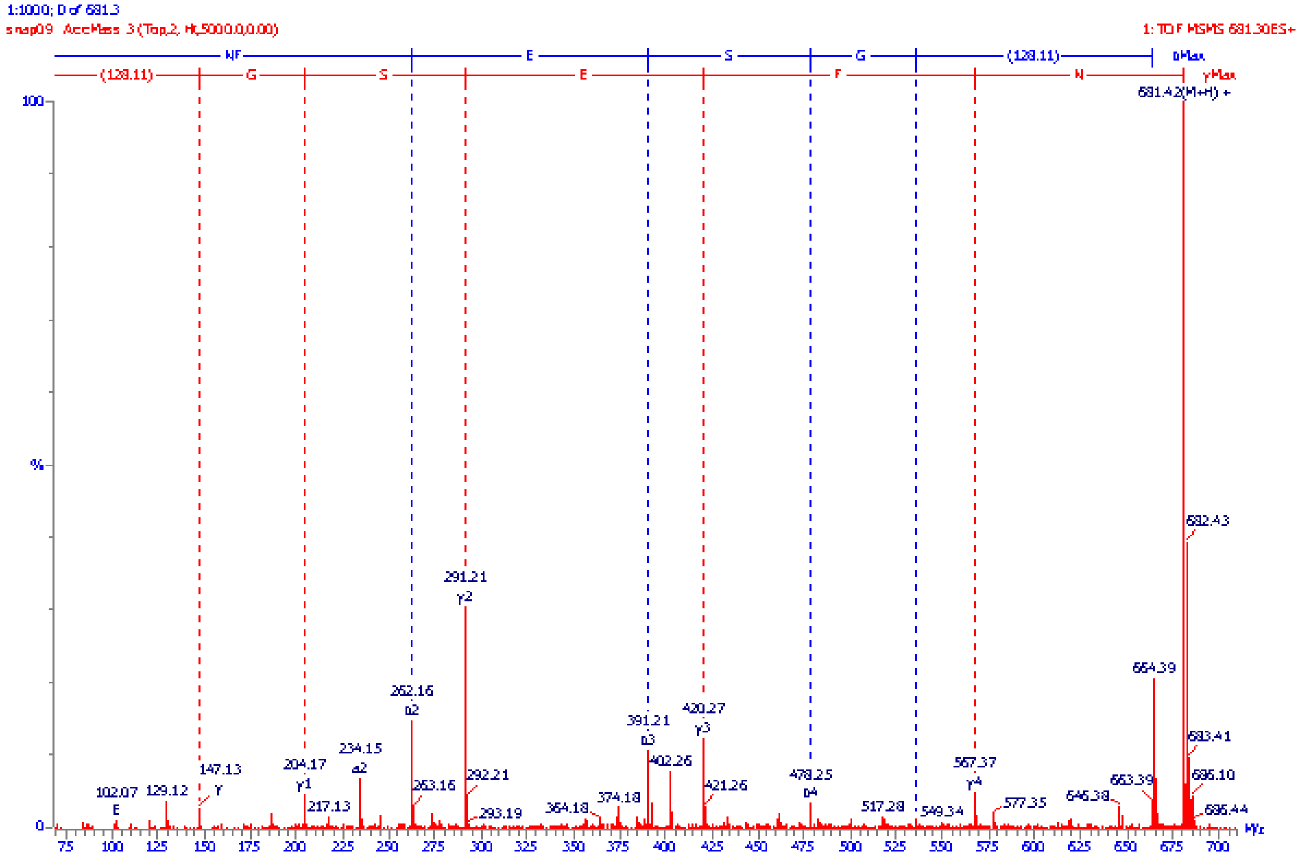
# Mass-spectrometry



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

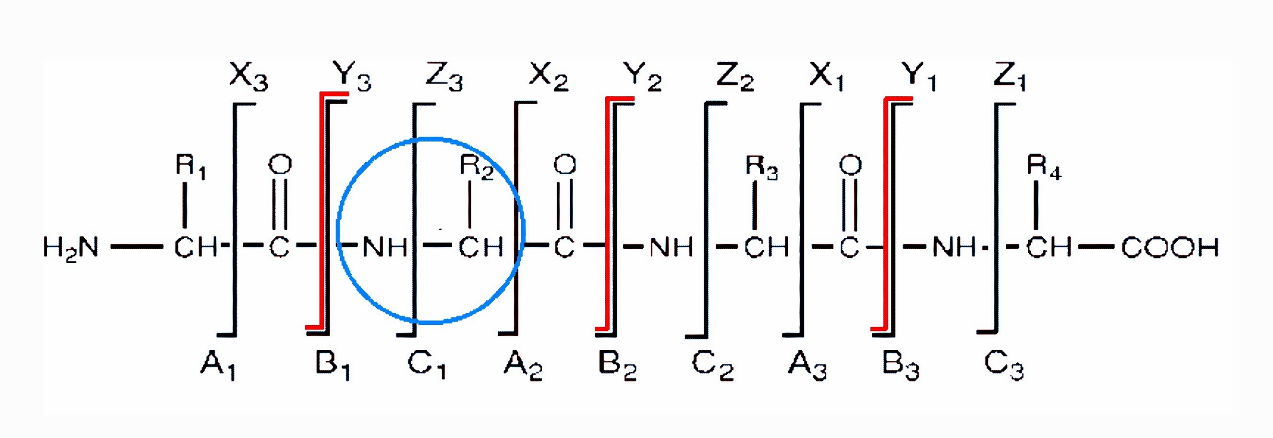
# Tandem Mass Spectrometry

- 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

# Example...

- peptide: GPFNA
- N-terminal fragments - G, GP, GPF, GPFN, GPFNA
- C-terminal fragments - A, NA, FNA, PFNA, GPFNA

- The spectrum will contain:
  - masses of both C- and N-terminal fragments
  - masses of “partial” fragments (missing H20, or NH3)
  - noise
  - missing peaks

- How can we re-construct the sequence of the peptide?

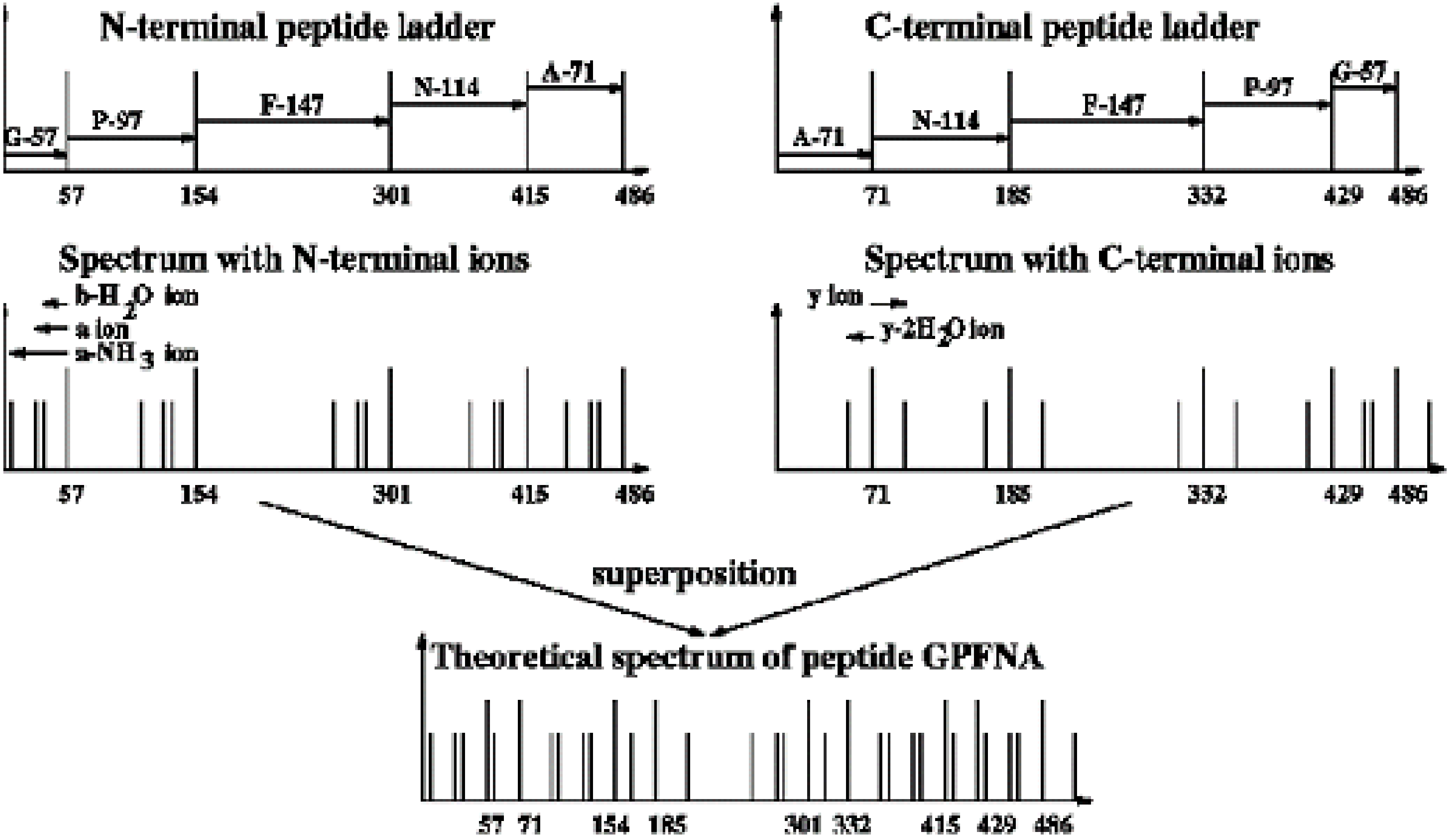
# Solutions?

- 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
- Problems:
  - how do we score alignments?
  - matching with dynamic programming can be slow
  - database can be very large (20^n n-length peptides, 10^18 - known peptides)

# De novo peptide sequencing

- Key idea: peptide ladder
  - adjacent fragments differ by exactly the mass of one amino-acid
- If we can identify pairs of masses in the spectrum that differ by an amino-acid's mass we can “read” the peptide
- Simple algorithm
  - Start with highest weight W
  - Find weight W' that differs from W by exactly the weight of an amino-acid AA
  - We know that peptide starts (or ends) with amino-acid AA
  - Repeat with W'...

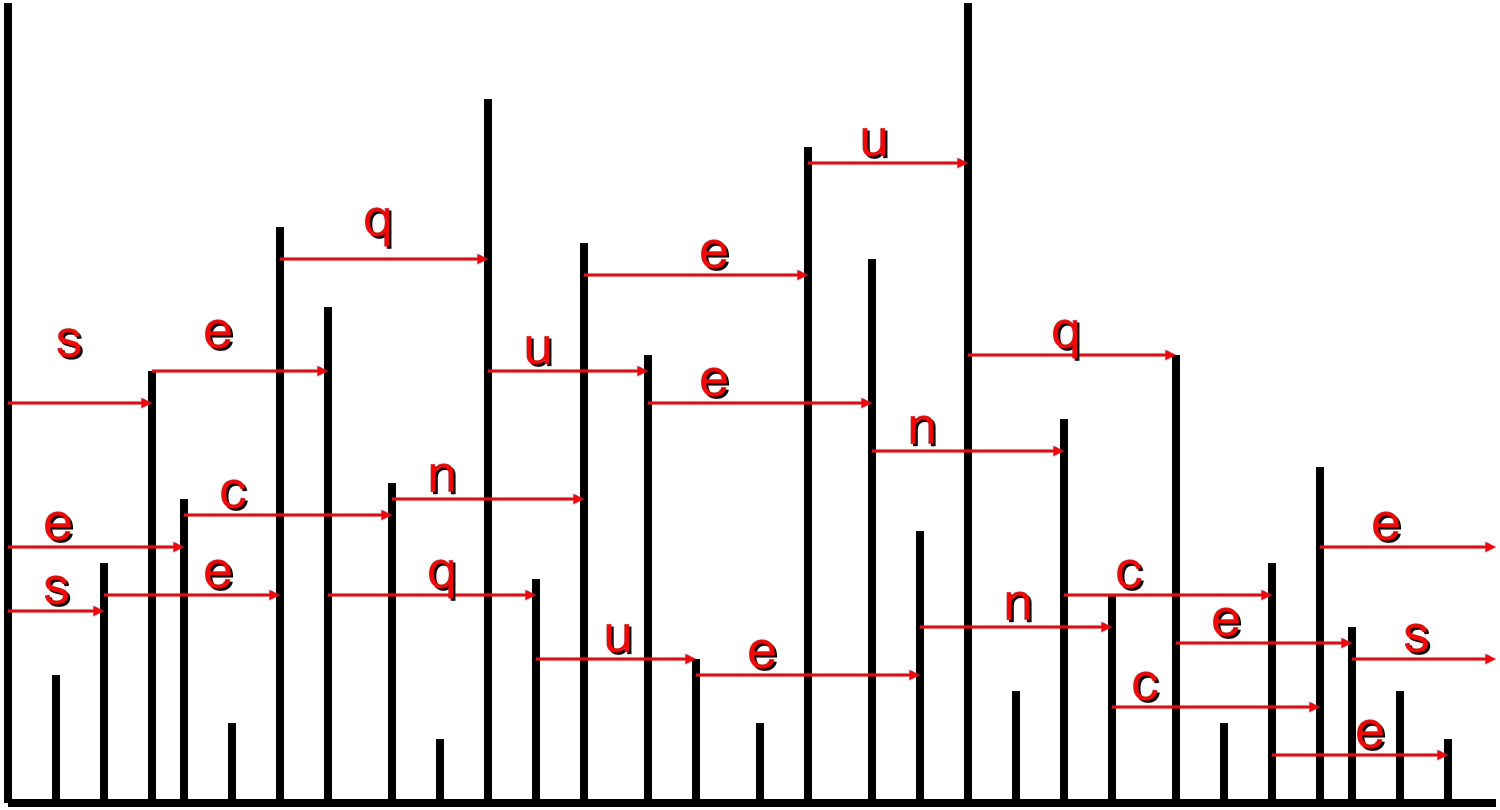
# Theoretical Spectrum



# Better algorithm

- Problems with simple algorithm
  - spectrum is mixture of C- and N-terminal fragments
  - spectrum contains masses for different ion types
  - spectrum contains errors
- Better solution
  - 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



# More problems...

- Spectral graph may contain many paths - many possible reconstructions
- Which path is the correct one?

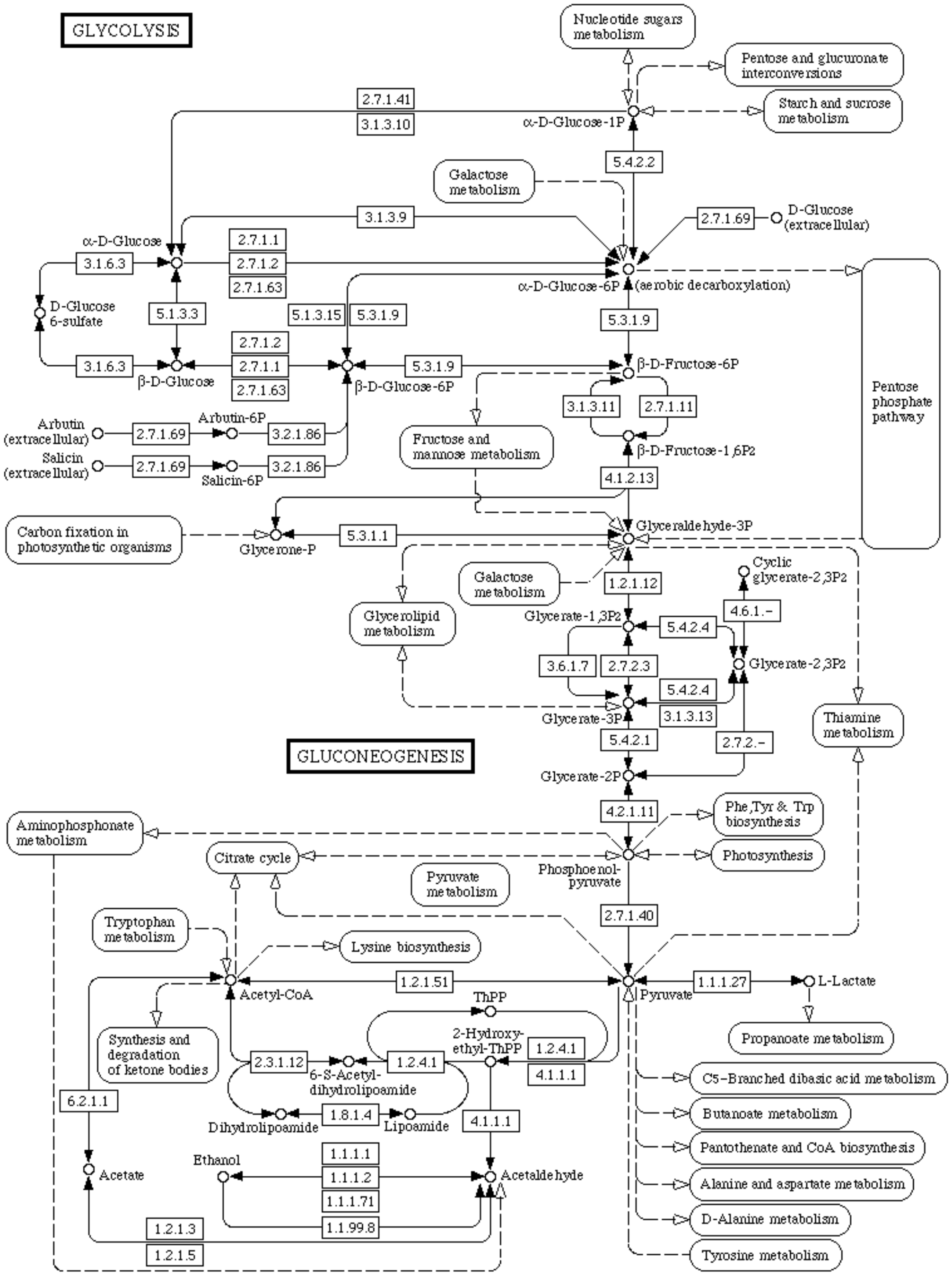
- Solution: estimate probability that peptide implied by path is consistent with spectrum
- Basic idea:
  - each peptide fragment contributes k different ions (masses)
  - for each ion/fragment combination we can estimate probability that ion is produced by the fragment

$$score(fragment_i) = (\prod_{observed\ peaks} p(ion|fragment))(\prod_{missing\ peaks} (1 - p(ion|fragment)))$$

  - optimal path can now be computed with dynamic programming

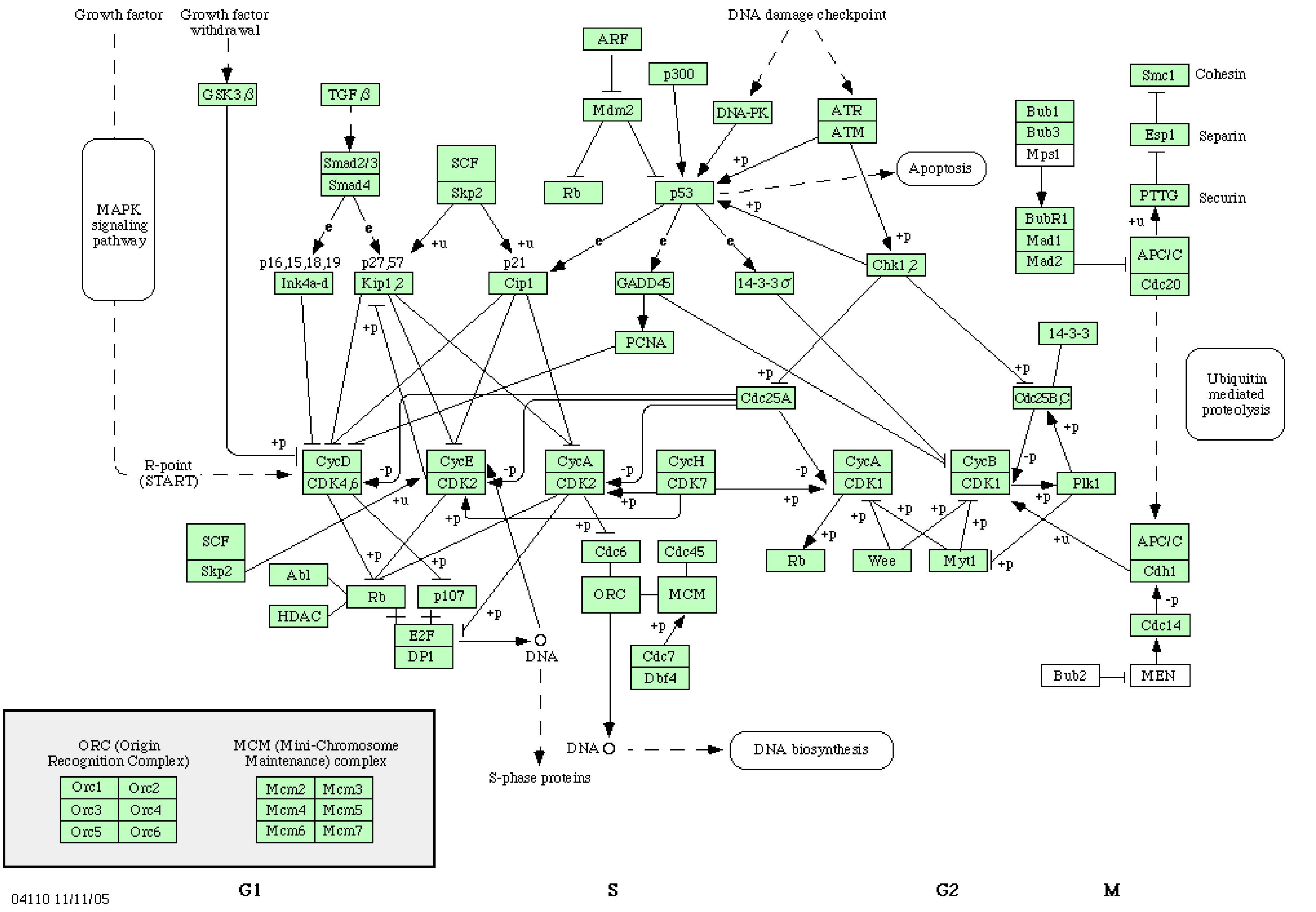
# Biological networks

- Genes/proteins do not exist in isolation
- Interactions between genes or proteins can be represented as graphs
- Examples:
  - metabolic pathways
  - regulatory networks
  - protein-protein interactions (e.g. yeast 2-hybrid)
  - genetic interactions (synthetic lethality)


GLYCOLYSIS
Nucleotide sugars metabolism
Pentose and glucuronate interconversions
Starch and sucrose metabolism
2.7.1.41
3.1.3.10
α-D-Glucose-1P
5.4.2.2
Galactose metabolism
3.1.3.9
2.7.1.69
D-Glucose (extracellular)
α-D-Glucose
2.7.1.1
2.7.1.2
2.7.1.63
3.1.6.3
α-D-Glucose-6P
(aerobic decarboxylation)
D-Glucose 6-sulfate
5.1.3.3
5.1.3.15
5.3.1.9
5.3.1.9
2.7.1.2
2.7.1.1
2.7.1.63
3.1.6.3
β-D-Glucose
5.3.1.9
β-D-Glucose-6P
β-D-Fructose-6P
Pentose phosphate pathway
3.1.3.11
2.7.1.11
Arbutin (extracellular)
2.7.1.69
Arbutin-6P
3.2.1.86
Salicin (extracellular)
2.7.1.69
Salicin-6P
3.2.1.86
Fructose and mannose metabolism
β-D-Fructose-1,6P2
4.1.2.13
Glyceraldehyde-3P
Carbon fixation in photosynthetic organisms
Glycerone-P
5.3.1.1
Galactose metabolism
1.2.1.12
Cyclic glycerate-2,3P2
4.6.1.-
Glycerolipid metabolism
Glycerate-1,3P2
5.4.2.4
3.6.1.7
2.7.2.3
Glycerate-2,3P2
5.4.2.4
Glycerate-3P
3.1.3.13
Thiamine metabolism
GLUCONEOGENESIS
5.4.2.1
2.7.2.-
Glycerate-2P
4.2.1.11
Phe,Tyr & Trp biosynthesis
Aminophosphonate metabolism
Citrate cycle
Phosphoenol-pyruvate
Photosynthesis
Pyruvate metabolism
2.7.1.40
Tryptophan metabolism
Lysine biosynthesis
1.2.1.51
Acetyl-CoA
Pyruvate
1.1.1.27
L-Lactate
ThPP
Propanoate metabolism
Synthesis and degradation of ketone bodies
2-Hydroxy-ethyl-ThPP
1.2.4.1
2.3.1.12
1.2.4.1
6-S-Acetyl-dihydrolipoamide
4.1.1.1
C5-Branched dibasic acid metabolism
6.2.1.1
Dihydrolipoamide
1.8.1.4
Lipoamide
4.1.1.1
Butanoate metabolism
1.1.1.1
Pantothenate and CoA biosynthesis
Ethanol
1.1.1.2
Acetate
Acetaldehyde
Alanine and aspartate metabolism
1.1.1.71
1.2.1.3
1.1.99.8
D-Alanine metabolism
1.2.1.5
Tyrosine metabolism

00010 3/23/06

# CELL CYCLE


Growth factor
Growth factor withdrawal
GSK3β
TGFβ
Smad2/3
Smad4
MAPK signaling pathway
SCF
Skp2
ARF
Mdm2
Rb
p300
p53
DNA damage checkpoint
DNA-PK
ATR
ATM
Apoptosis
Smc1
Cohesin
Esp1
Separin
PTTG
Securin
Bub1
Bub3
Mps1
BubR1
Mad1
Mad2
APC/C
Cdc20
+u
e
p16,15,18,19
Ink4a-d
p27,57
Kip1,2
+u
p21
Cip1
GADD45
14-3-3σ
Chk1,2
+p
PCNA
14-3-3
Cdc25A
Cdc25B,C
Ubiquitin mediated proteolysis
R-point (START)
CycD
CDK4,6
-p
CycE
CDK2
CycA
CDK2
CycH
CDK7
CycA
CDK1
CycB
CDK1
Plk1
SCF
Skp2
Abl
HDAC
Rb
p107
E2F
DP1
DNA
Cdc6
ORC
Cdc45
MCM
Cdc7
Dbf4
Rb
Wee
Myt1
APC/C
Cdh1
Cdc14
Bub2
MEN
S-phase proteins
DNA
DNA biosynthesis
ORC (Origin Recognition Complex)
Orc1
Orc2
Orc3
Orc4
Orc5
Orc6
MCM (Mini-Chromosome Maintenance) complex
Mcm2
Mcm3
Mcm4
Mcm5
Mcm6
Mcm7
04110 11/11/05
G1
S
G2
M