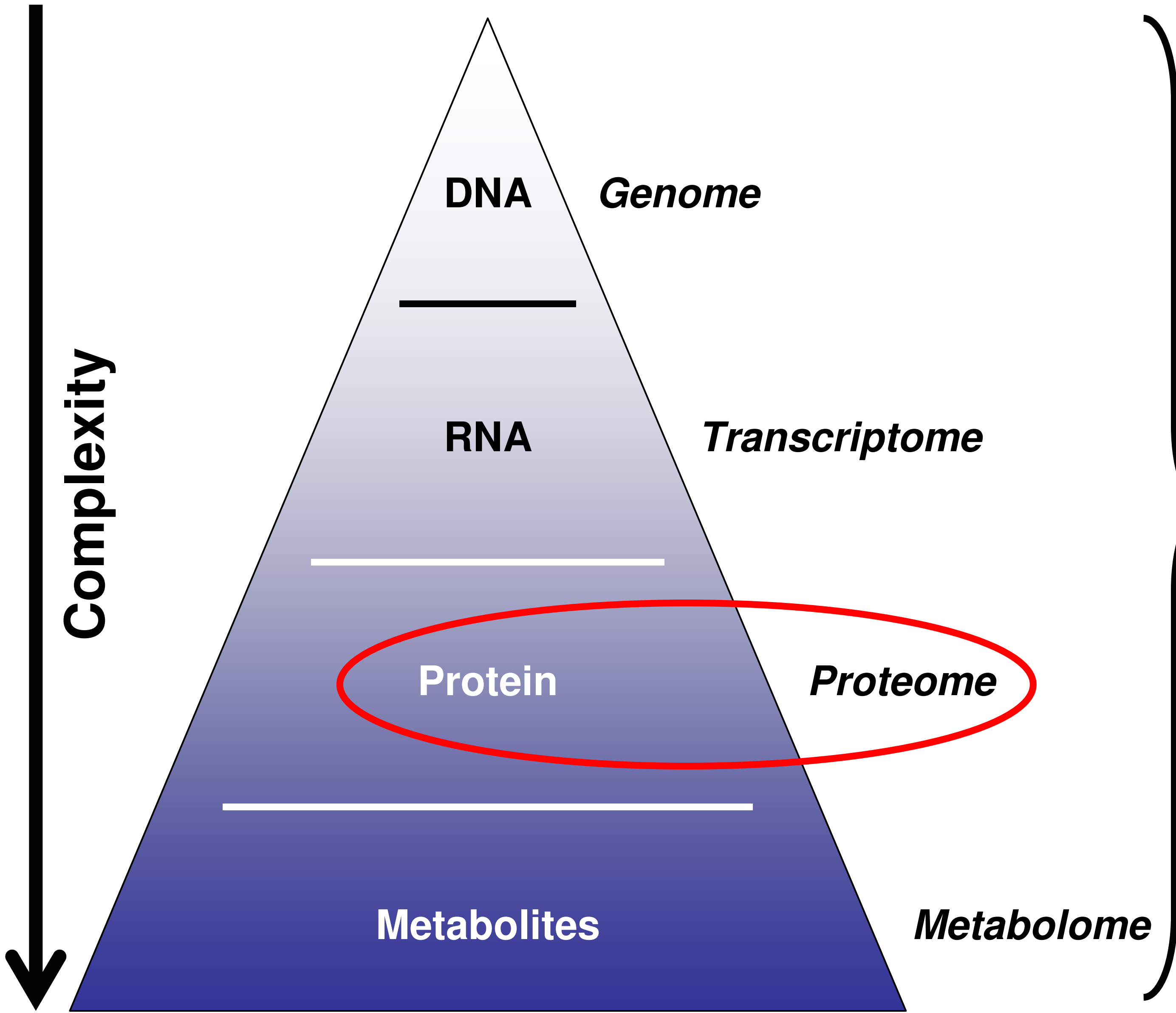


Systems Biology
Linking all the components of a cell in a quantitative and temporal manner

1 Genomic DNA

ccggtcgcccccacgtcgtccatgatgatgtgtcgggagtaccggtccagctgtgtcgcgtccaactccagcccgctcatgtggcggcttgggtggaccgcggcaaaag

2 Genomic DNA -> RNA [transcription]

ccggtcgcccccacgtcgtccatgatgatgtgtcgggagtaccggtccagctgtgtcgcgtccaactccagcccgctcatgtggcggcttgggtggaccgcggcaaaag

A->U G->C RNA
T->A C->G polymerase

3 RNA -> protein [translation]

TTT,TTC -> Phe (F); TTA,TTG-> Leu (L);
TCC,TCA, TCG,TCT -> Ser (S);
TGC,TGT -> Cys (C) ; etc...
Ribosome

KMVLFVYGYCDQRDCFYCPLGENRKNVQQVYANERPVEDDEDIISEAKLMDALGTSITGGEPQEVLDRTCRYLRLLKDEFGEDHHTHLYTGITGGRENMRR

4 protein -> folded protein [protein-folding]

Spontaneous and Chaperonins

Where are the functional elements?

Alternative splicing.

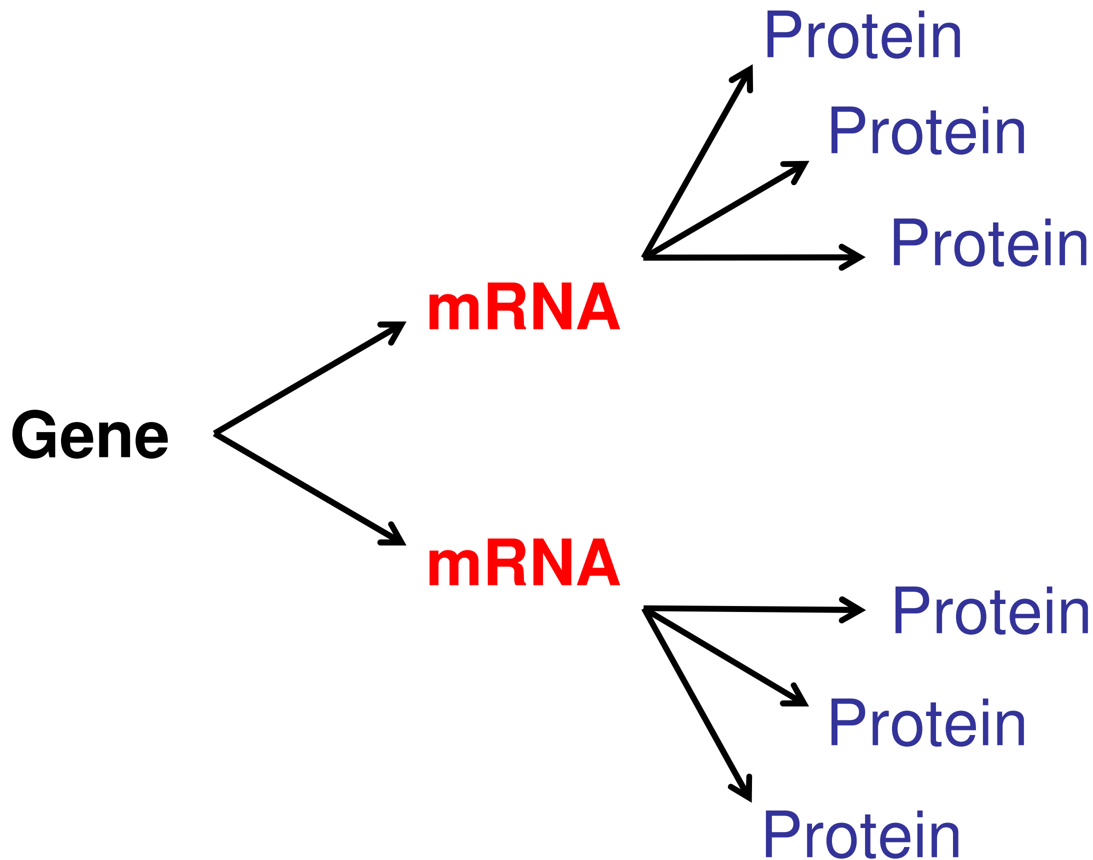
What are transcript levels under different conditions? Microarrays.

Protein structure.
PTMs
Protein localization.
Etc…

Origins of Cellular Complexity

Humans	*C. elegans*
• **~3,000 megabases**	• **~100 megabases**
• **~21,000-25,000 Genes**	• **~20,000 Genes**
• **~10,000,000,000,000 cells**	• **~1,000 cells**

- Alternative Splicing
- Improved and modular interaction between components
- Improved spatial relationship between components in the cell
- Improved interaction between intracellular signaling pathways
- Multiple protein forms and functions
- etc…



- Covalent Modification or Modifications
  - Phosphorylation
  - Glycosylation
  - Lipidation
  - Nitrosylation
  - Acetylation, etc…
- Cellular localization
  - Nucleous
  - Mitochondria
  - Endoplasmic Reticulum, etc…
- Presence or Absence of Ligands
  - Heme
  - Metal
  - ATP, etc…
- Proteolytic Cleavage
  - N-terminal and C-terminal truncations
- Oligomeric State
  - ~78% of yeast proteins act in part of a greater complex of proteins
- Protein Conformation

What needs to be measured?

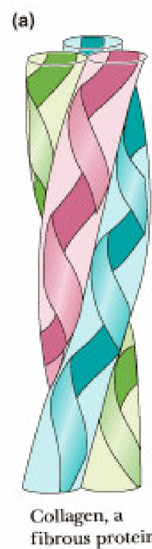
- Need methods that can detect and quantify all proteins (and modifications)
- Determine protein-protein interactions
- Determine protein localization
- Assess protein function
- Evaluate conformational changes

At X points in time and under Y different conditions

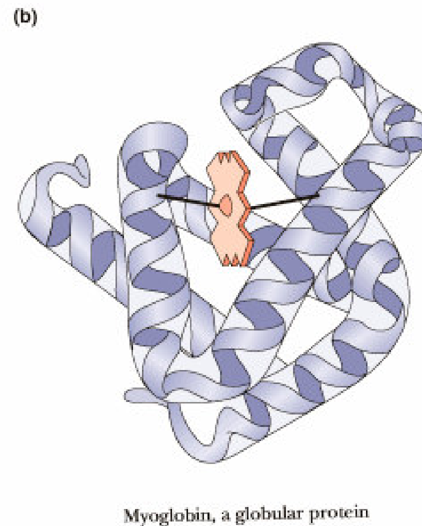
Must have methods that can measure more than one protein/modification/interaction/etc… at a time!!!!!

Dealing with protein is difficult!

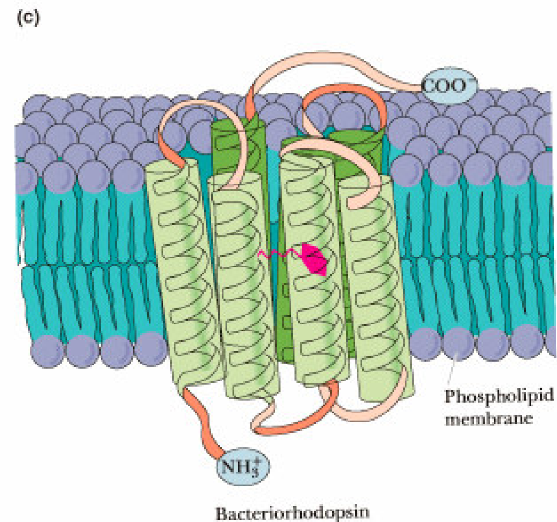
Proteins span an enormous range of physiochemical properties



Collagen, a fibrous protein



Myoglobin, a globular protein

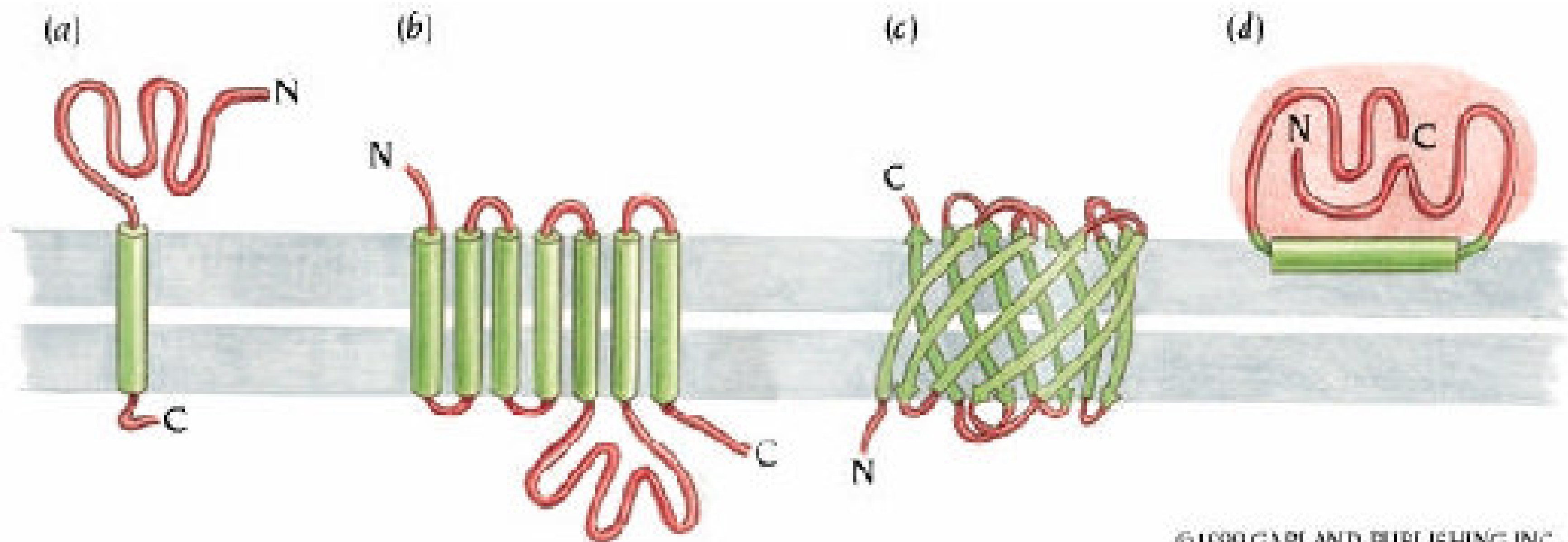


Bacteriorhodopsin

- 25-30% of all Eukaryote Proteins
- 50% of all drug targets
- Insoluble in aqueous solutions

~90% of all analyses are focused on soluble proteins!!

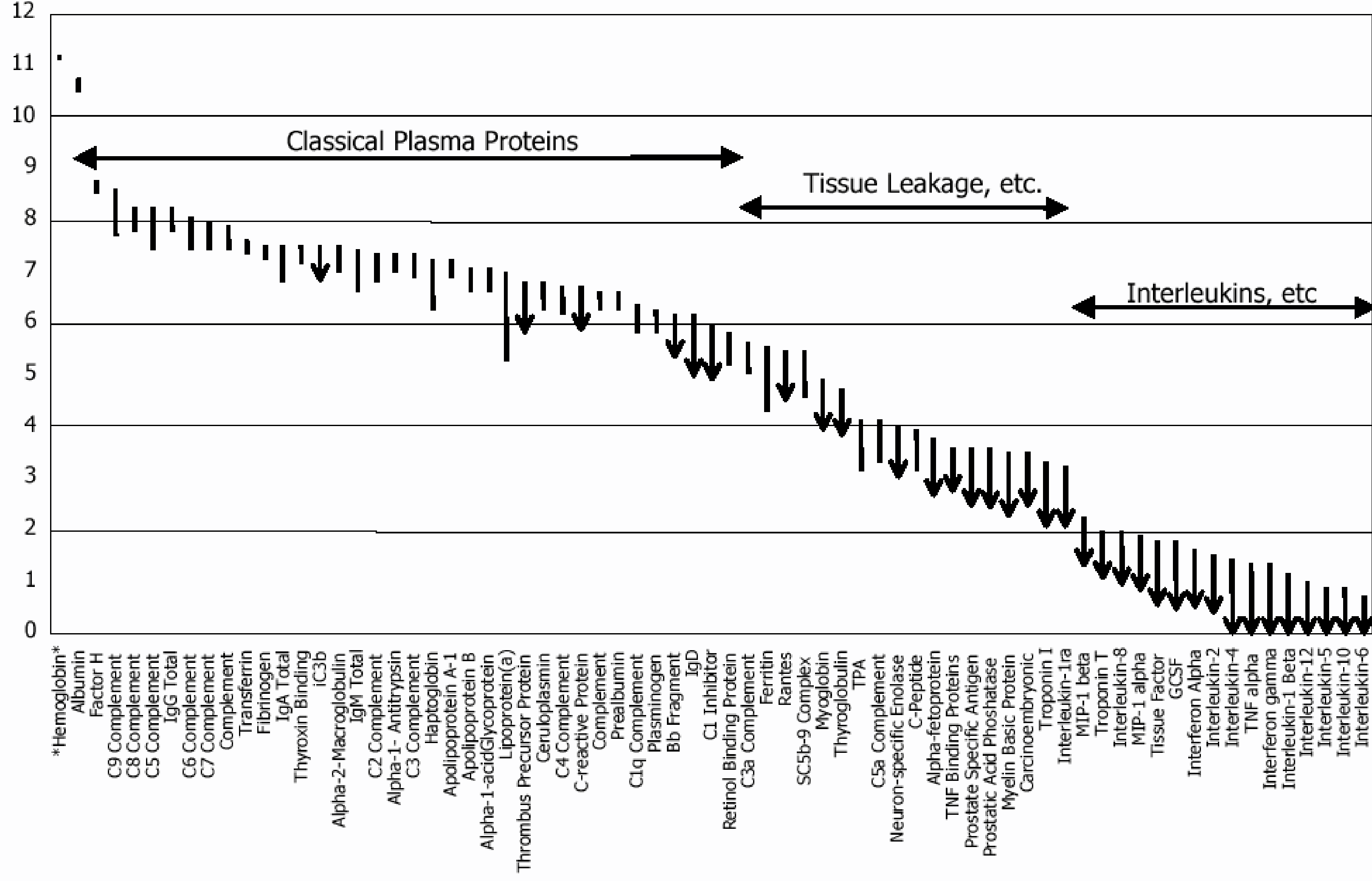
Arrangement of Proteins in a Lipid Bilayer



The Dynamic Range and Sensitivity is a Problem

Table 1. Detection limits for proteins assuming 100% recovery

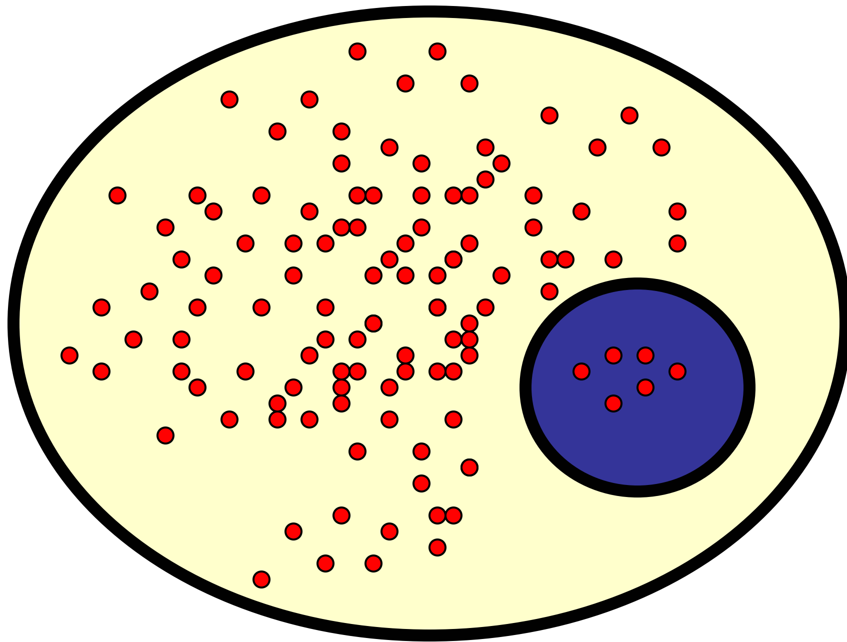
Number of cells	Protein copies/cell	Total number of **molecules**	Moles of protein
1.OE+09	1000000	1.OE+15	1600 pmole
1.OE+09	100000	1.OE+14	160 pmole
1.OE+09	10000	1.OE+13	16 pmole
1.OE+09	1000	1.OE+12	1.6 pmole
1.OE+09	100	1.OE+11	160 fmole
1.OE+09	10	1.OE+10	16 fmole
1.OOE+08	1000000	1.OE+14	160 pmole
1.OOE+08	100000	1.OE+13	16 pmole
1.OOE+08	10000	1.OE+12	1.6 pmole
1.OOE+08	1000	1.OE+11	160 fmole
1.OOE+08	100	1.OE+10	16 fmole
1.OOE+08	10	1.OE+09	1.6 fmole
1.OOE+07	1000000	1.OE+13	16 pmole
1.OOE+07	100000	1.OE+12	1.6 pmole
1.OOE+07	10000	1.OE+11	160 fmole
1.OOE+07	1000	1.OE+10	16 fmole
1.OOE+07	100	1.OE+09	1.6 fmole
1.OOE+07	10	1.OE+08	0.2 fmole
1.OOE+06	1000000	1.OE+12	1.6 pmole
1.OOE+06	100000	1.OE+11	160 fmole
1.OOE+06	**10000**	1.OE+10	16 fmole
1.OOE+06	1000	1.OE+09	1.6 fmole
1.OOE+06	100	1.OE+08	0.2 fmole



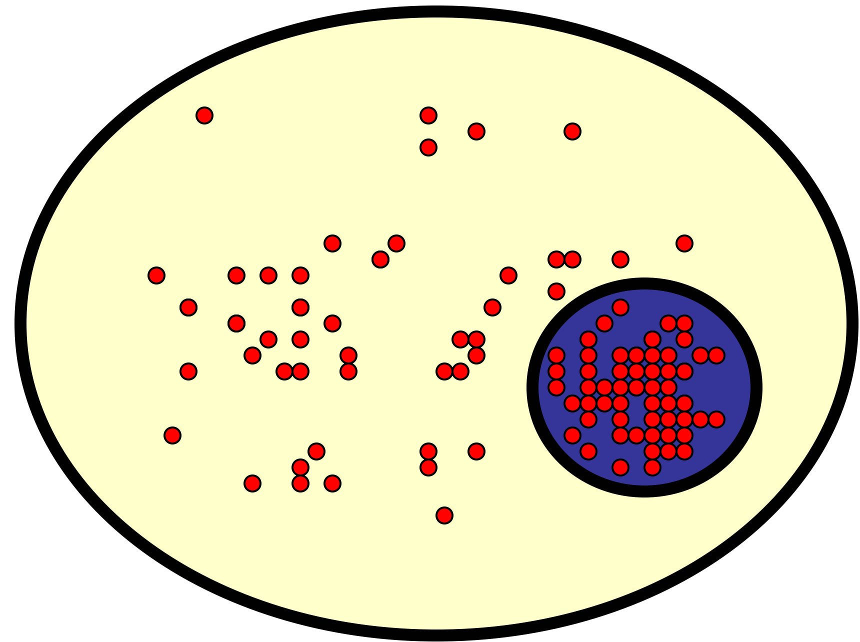
Normal Range Abundances
Log_{10} Concentration in pg/mL
Classical Plasma Proteins
Tissue Leakage, etc.
Interleukins, etc
12
11
10
9
8
7
6
5
4
3
2
1
0
Hemoglobin
Albumin
Factor H
C9 Complement
C8 Complement
C5 Complement
IgG Total
C6 Complement
C7 Complement
Complement
Transferrin
Fibrinogen
IgA Total
Thyroxin Binding
iC3b
Alpha-2-Macroglobulin
IgM Total
C2 Complement
Alpha-1- Antitrypsin
C3 Complement
Haptoglobin
Apolipoprotein A-1
Apolipoprotein B
Alpha-1-acidGlycoprotein
Lipoprotein(a)
Thrombus Precursor Protein
Ceruloplasmin
C4 Complement
C-reactive Protein
Complement
Prealbumin
C1q Complement
Plasminogen
Bb Fragment
IgD
C1 Inhibitor
Retinol Binding Protein
C3a Complement
Ferritin
Rantes
SC5b-9 Complex
Myoglobin
Thyroglobulin
TPA
C5a Complement
Neuron-specific Enolase
C-Peptide
Alpha-fetoprotein
TNF Binding Proteins
Prostate Specific Antigen
Prostatic Acid Phoshatase
Myelin Basic Protein
Carcinoembryonic
Troponin I
Interleukin-1ra
MIP-1 beta
Troponin T
Interleukin-8
MIP-1 alpha
Tissue Factor
GCSF
Interferon Alpha
Interleukin-2
Interleukin-4
TNF alpha
Interferon gamma
Interleukin-1 Beta
Interleukin-12
Interleukin-5
Interleukin-10
Interleukin-6

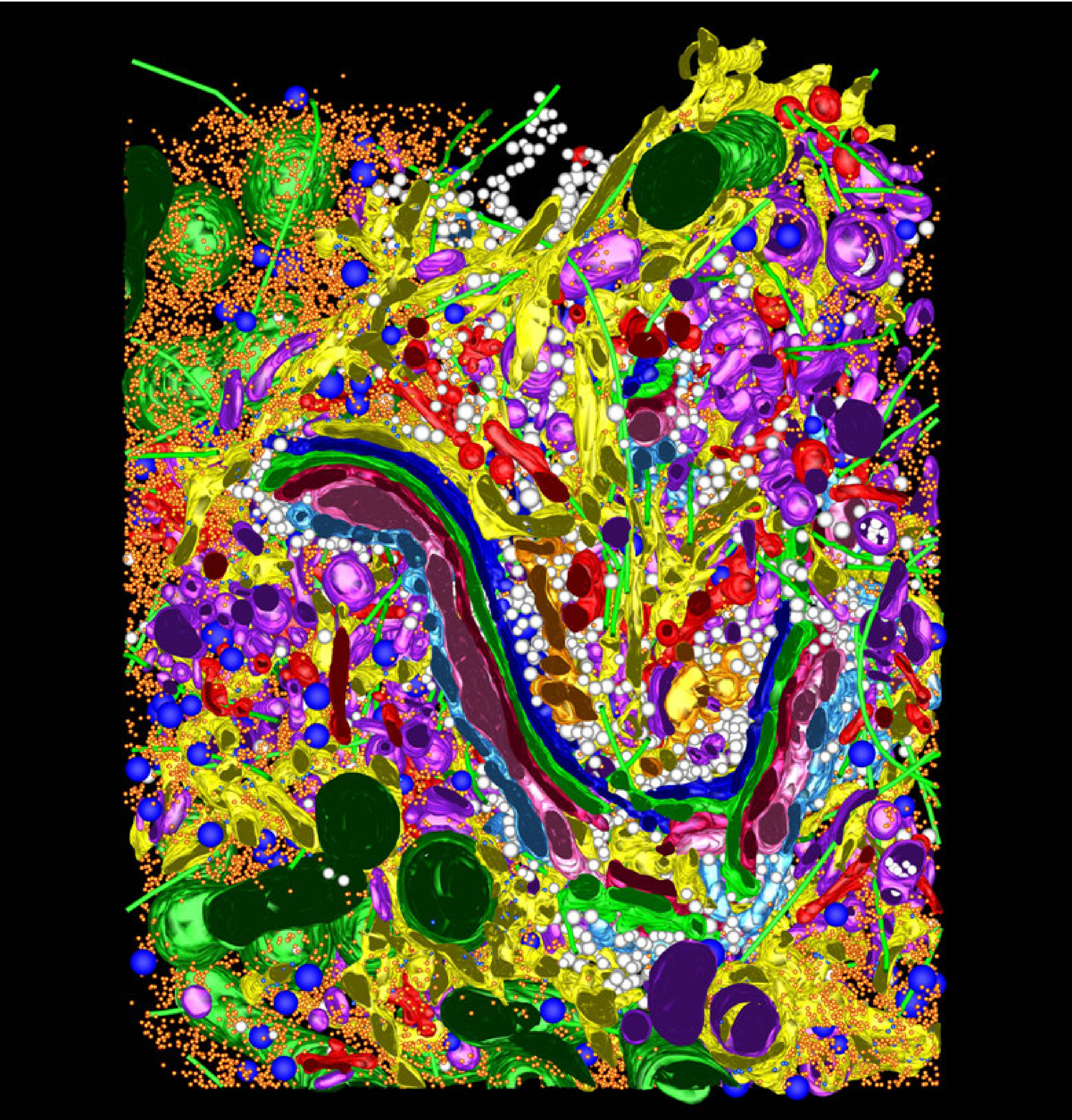
We have a protein in multiple compartments and the localized level changes but not the overall cellular level. How do we measure this?

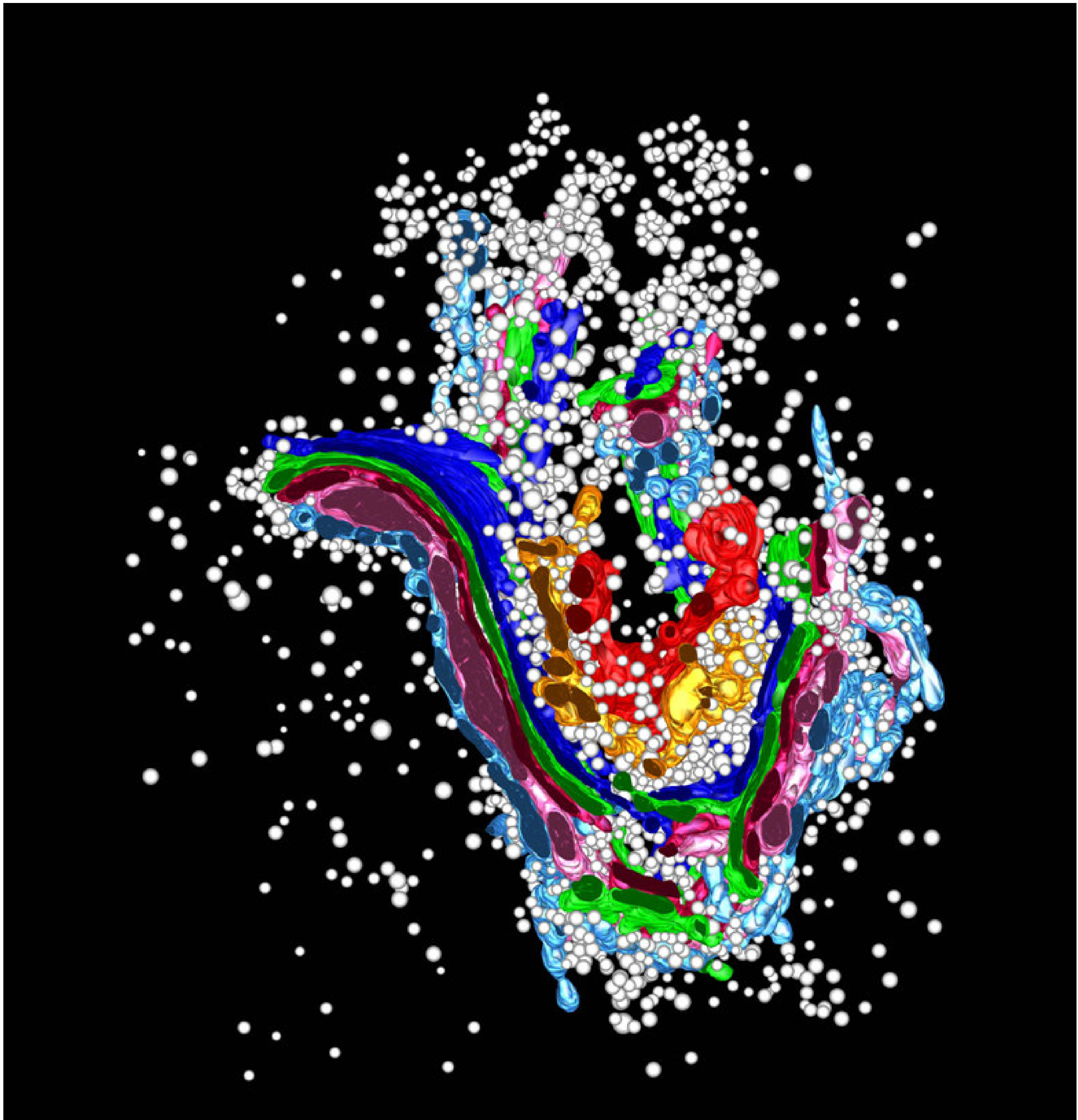
1.



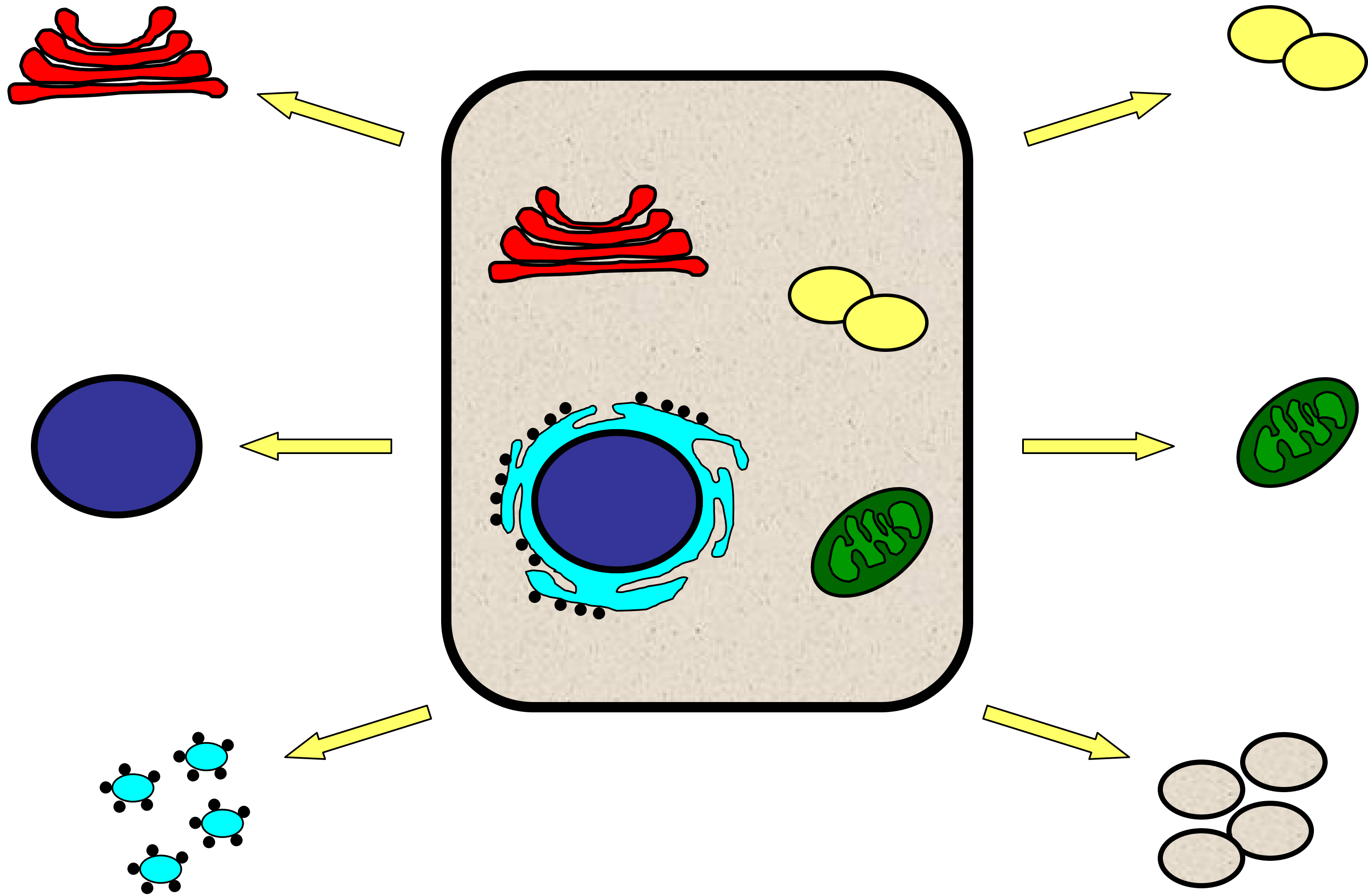
2.





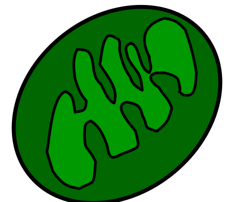


Cell Fractionation

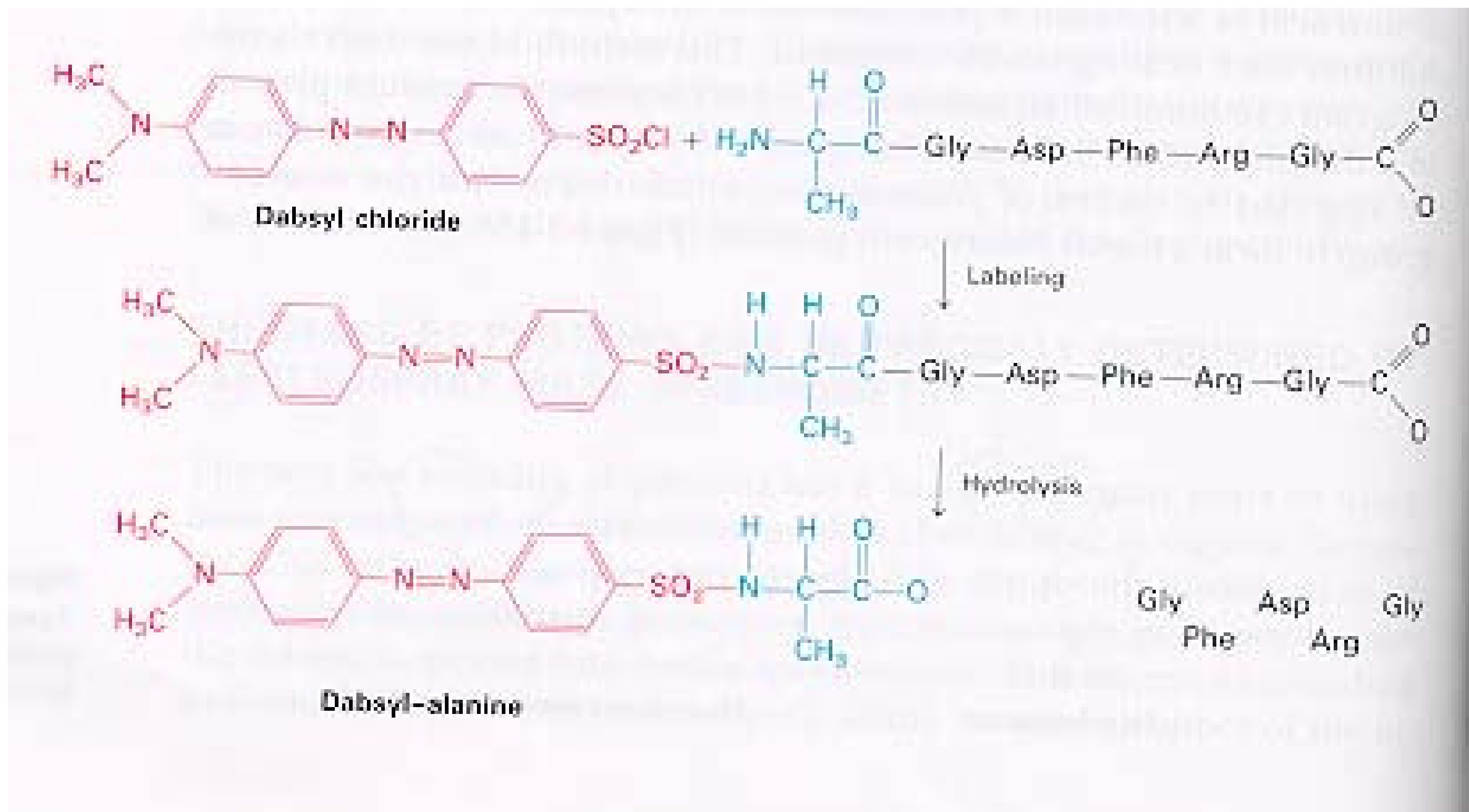


How do we distinguish between a high abundance contaminant or a low abundance specific protein?

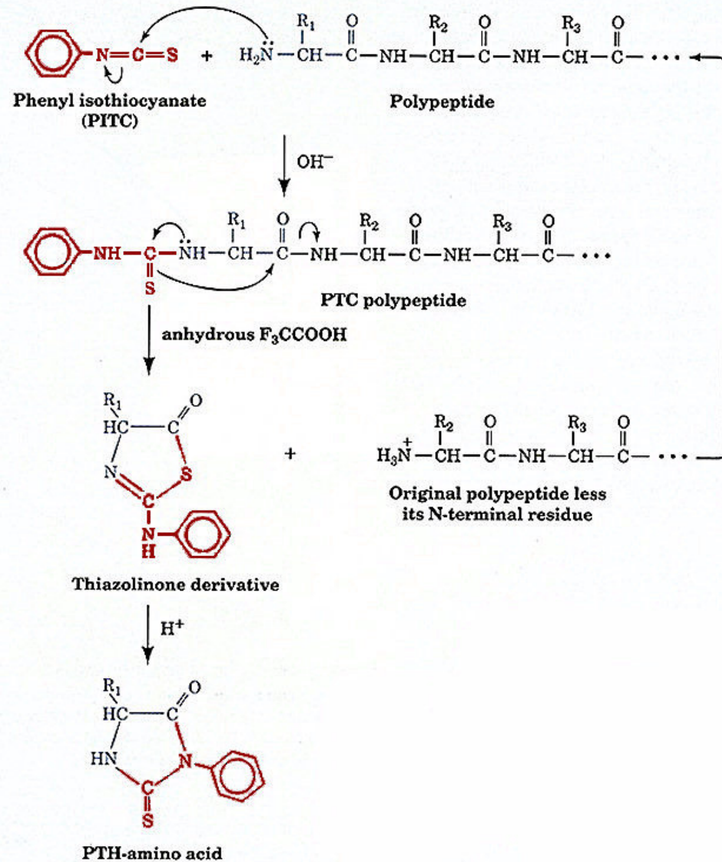
- Consider two proteins present in a cell.
  - Protein A is localized to the cytosol and is present at 1,000,000 copies/cell
  - Protein B is localized to the mitochondria and is present 100 copies/cell
- If the mitochondria is then enriched 100x
  - Protein A will be depleted 100x => 10,000
  - Protein B will be enriched 100x => 10,000



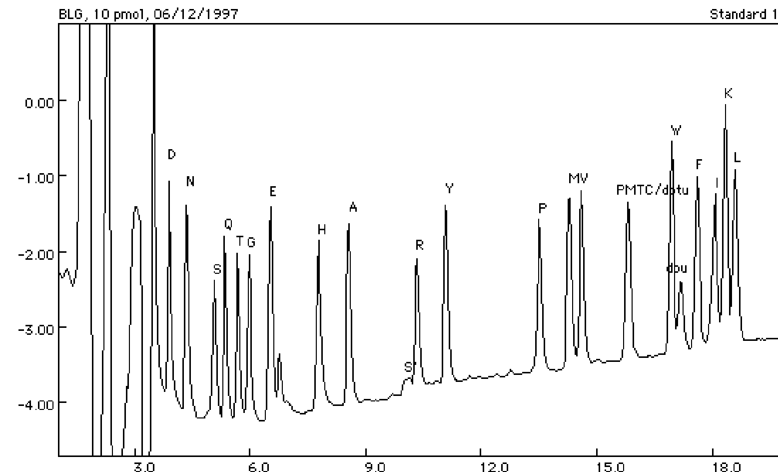
N-terminal amino acid can be identified by reaction with dabsyl chloride or dansyl chloride followed by acid hydrolysis



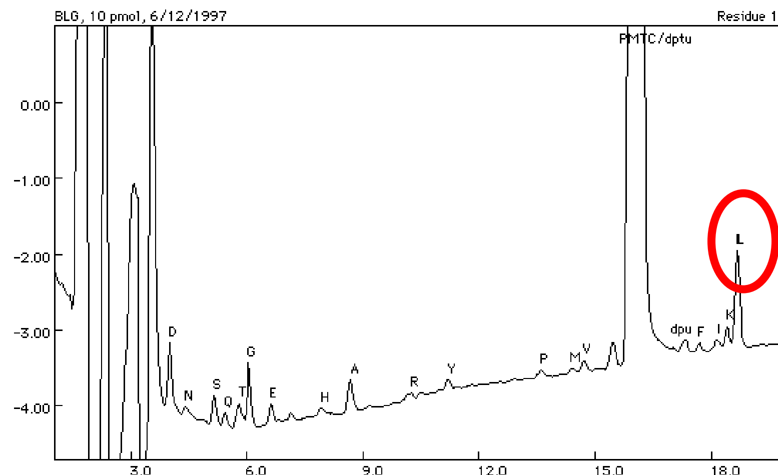
N-Terminal Protein Sequencing



Standard



Residue 1



30-50 min per cycle

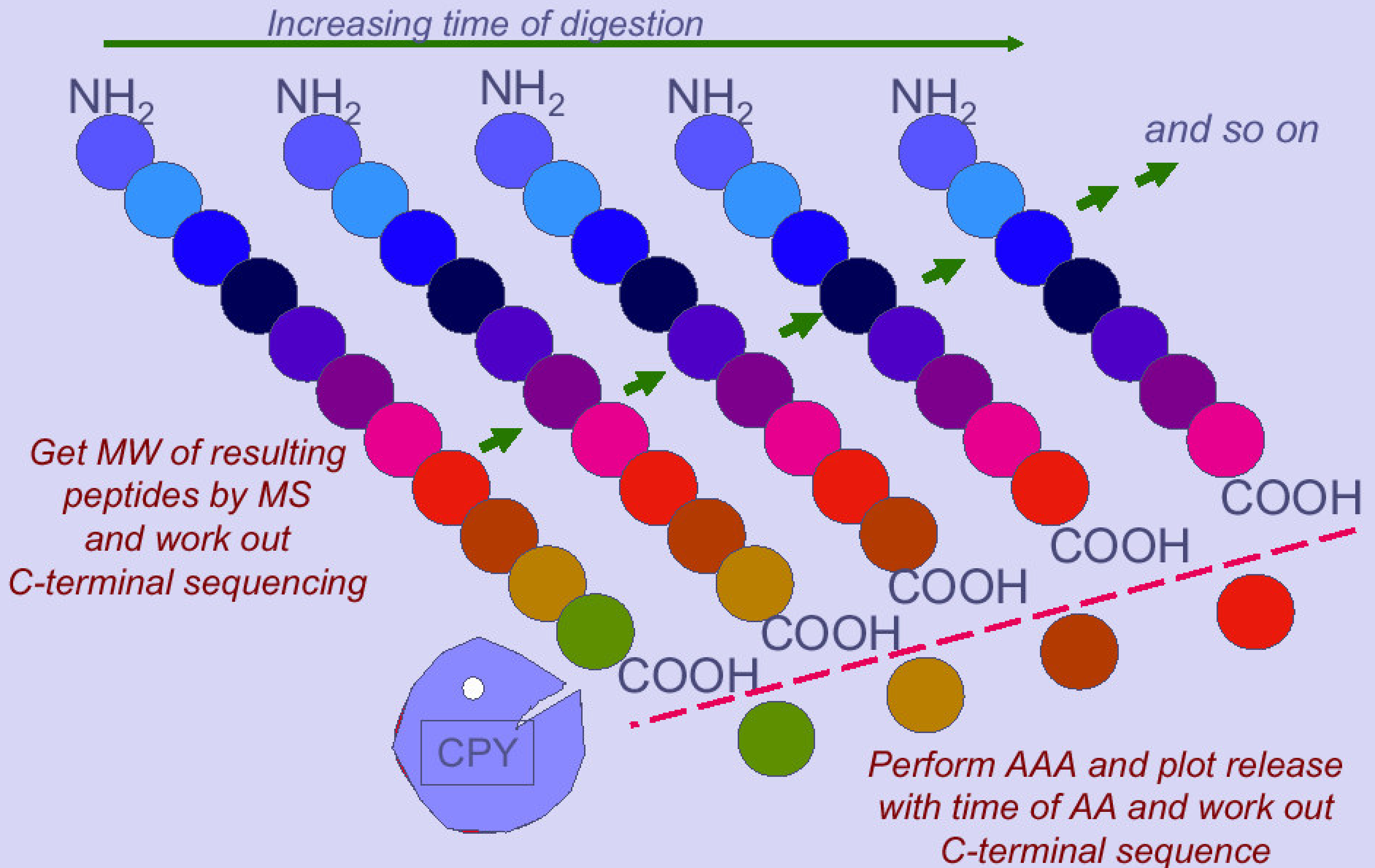
Each cycle is ~98% effective.
Capable of sequencing ~50 AA.

Edman P. A method for determination of the amino acid sequence in peptides, *Arch. Biochem.* **22**: 475-476 (1949)

C-terminal Methods are Almost Non Existent

- Most methods involve aminopeptidases (carboxypeptidases) to selectively cleave one amino acid at a time from the C-terminus.
  - This amino acid is then determined by chromatographic methods or
  - remaining protein sequence is measured by mass spectrometry.

C-terminal Sequencing by MS



Determining the Primary Structure of a Protein

Protein sequencing has become largely obsolete because of large scale DNA sequencing and protein coding gene predictions. Most strategies now focus on detection instead of sequencing.

E

E

Post-Translational Modifications

http://www.abrf.org/index.cfm/dm.home

351 Modifications in Delta Mass. Not all of these occur *in vivo*

ABRF Delta Mass Reference - Mozilla

File Edit View Go Bookmarks Tools Window Help

http://www.abrf.org/index.cfm/dm.home Search

Home Bookmarks mozilla.org mozillaZine mozdev.org

Search

GO

Advanced Search

- About ABRF
- Forms and Documents
- News and Announcements
- Communications
- Research Groups
- Open Research Studies
- Committees
- Reference
- Sponsorship

Flyout Menu Off

Delta Mass

A Database of Protein Post Translational Modifications

Contributors

This page is dedicated to the memories of Finn Wold and Ken Mitchelhill

Please send new data or corrections to Len Packman

The ABRF accepts no responsibility for the accuracy of these data which are freely donated by individuals

Search the Delta Mass Database View All

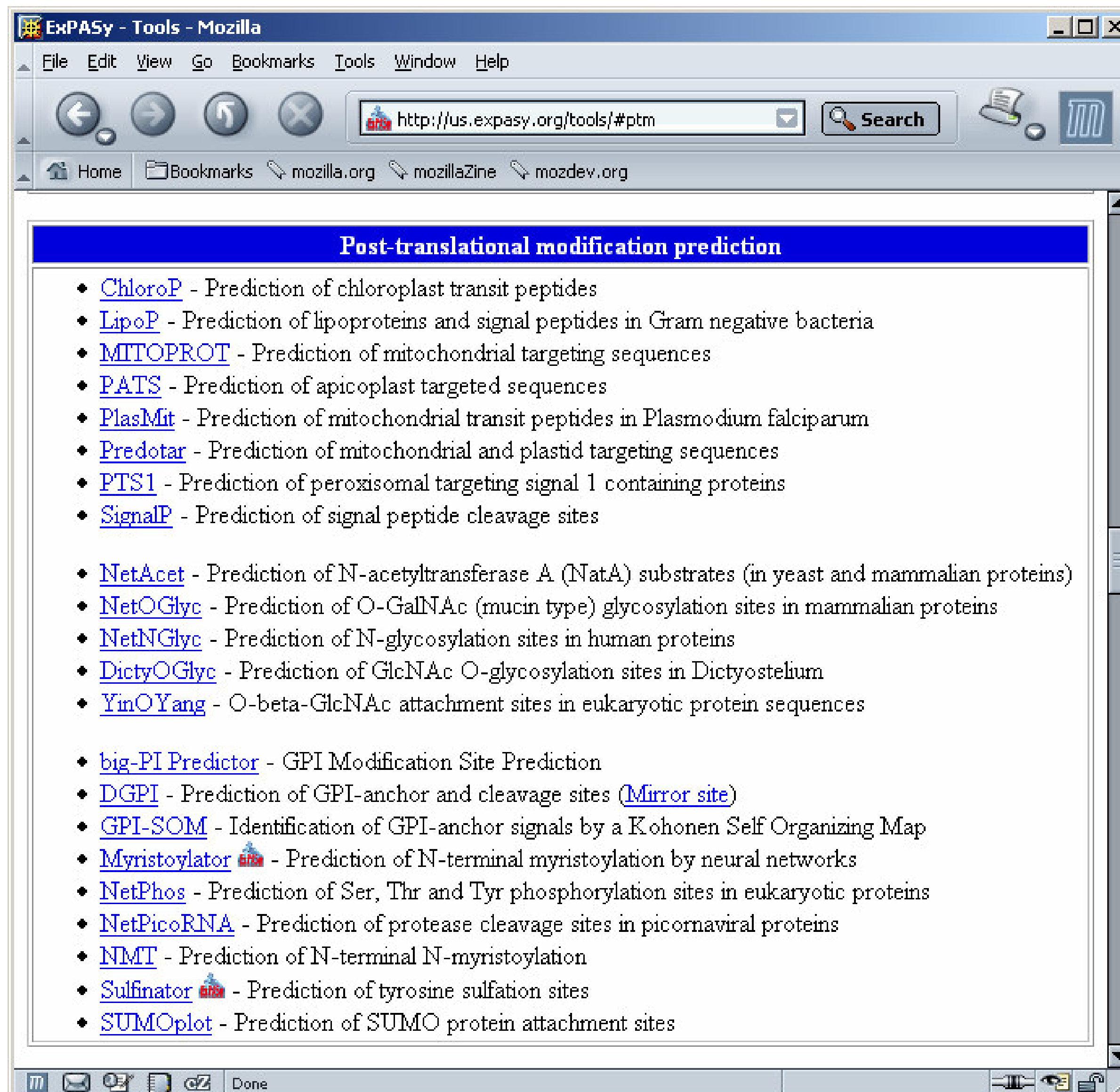
Enter a +/- integer to search on, and select a margin of error (range of search).

Avg. Mass Change: Error Margin: Exact Search!

Transferring data from www.abrf.org...

Post-Translational Modification Site Prediction

http://us.expasy.org/tools/#ptm



Mostly stores motifs.

For Example:
Protein kinase C (PKC) recognizes [ST]-x-[RK]

Thus any protein with:

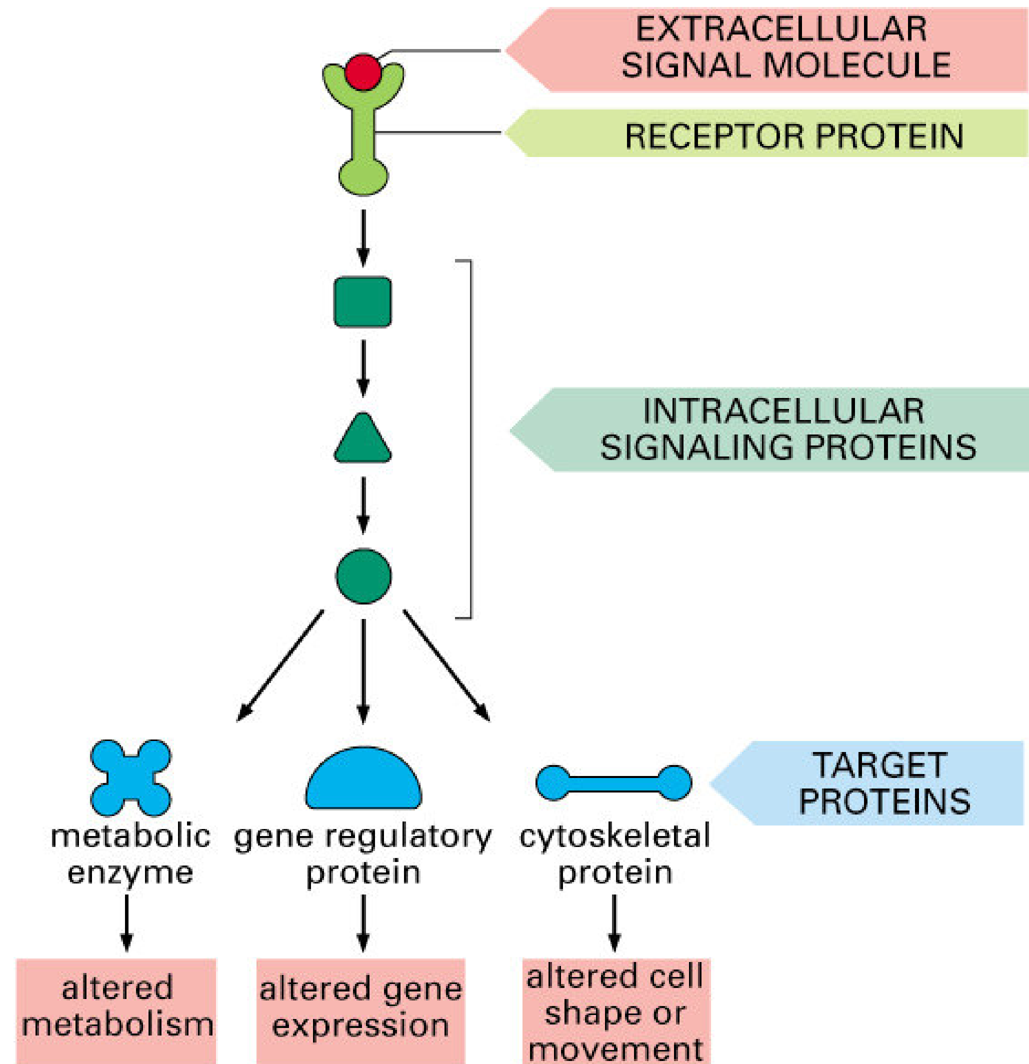
S-X-R
T-X-R
S-X-K
T-X-K

Would come back with a predicted phosphorylation site.

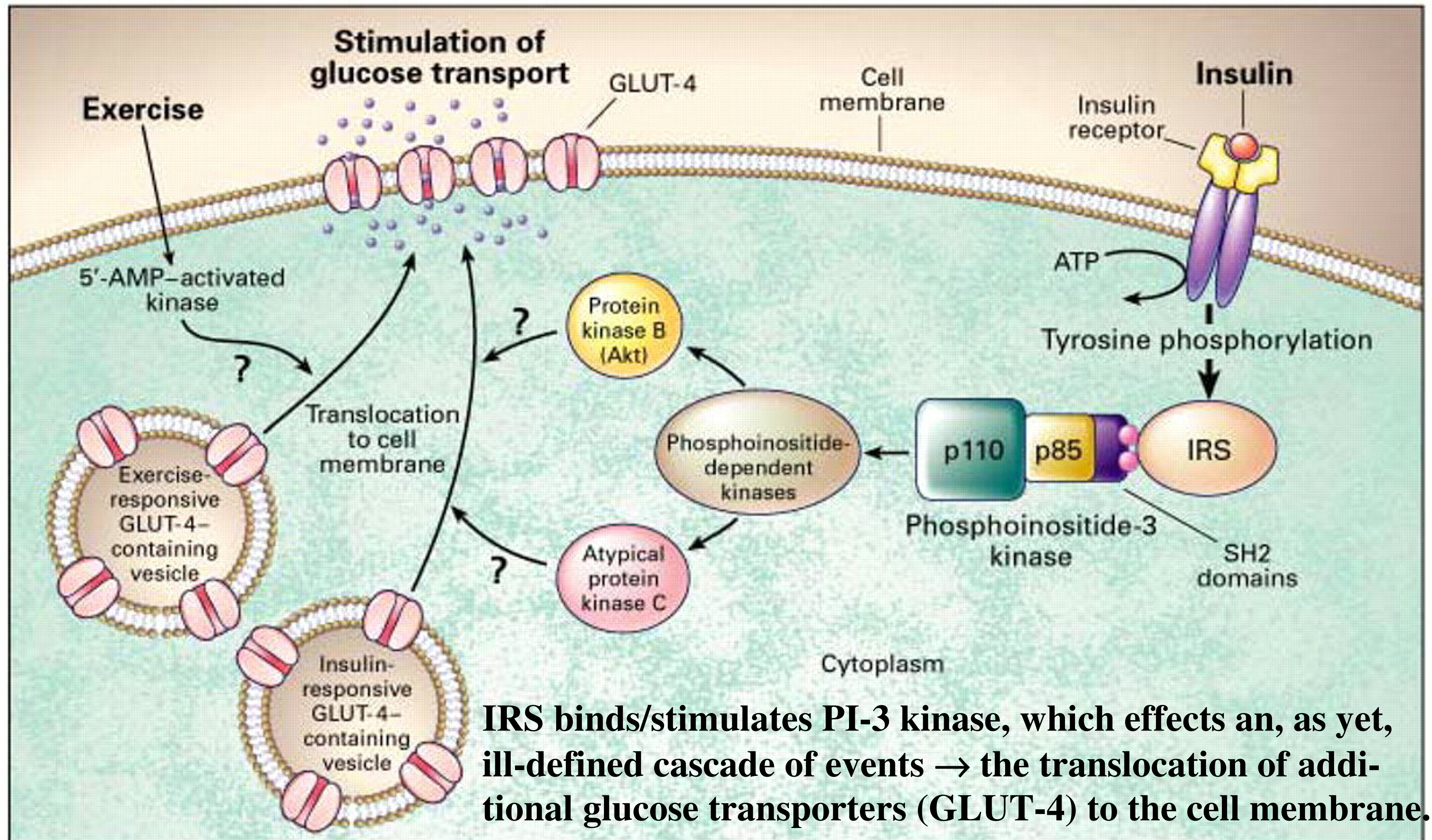
INTRACELLULAR SIGNAL TRANSDUCTION:
A Journey from the Plasma Membrane to the Nucleus

How do we determine the components?

How do we determine the mechanism of action of a signaling pathway?



Example: Insulin Stimulation of Glucose Transport



New Engl J Med **341**:248, 1999

GLUT-4 Translocation to the Plasma Membrane

How do we determine the substrates for a kinase, ubiquitin ligase, methyl transferase?

What if the stoichiometry of the modification is low?

What if the phosphorylation / dephosphorylation event is very rapid?

How would we know it even happened?