Protein1: Last week's take home lessons

- Protein interaction codes(s)?
- Real world programming
- Pharmacogenomics : SNPs
- Chemical diversity : Nature/Chem/Design
- Target proteins : structural genomics
- Folding, molecular mechanics & docking
- Toxicity animal/clinical : cross-talk

Protein2: Today's story & goals

- Separation of proteins & peptides
- Protein localization & complexes
- Peptide identification (MS/MS)
  - Database searching & sequencing.
- Protein quantitation
  - Absolute & relative
- Protein modifications & crosslinking
- Protein - metabolite quantitation

Why purify?

- Reduce one source of noise
  (in identification/quantitation)
- Prepare materials for in vitro experiments
  (sufficient causes)
- Discover biochemical properties

(Protein) Purification Methods

- Charge: ion-exchange chromatography, isoelectric focusing
- Size: dialysis, gel-filtration chromatography, gel-electrophoresis, sedimentation velocity
- Solubility: salting out
- Hydrophobicity: Reverse phase chromatography
- Specific binding: affinity chromatography
- Complexes: Immune precipitation (± crosslinking)
- Density: sedimentation equilibrium

Protein Separation by Gel Electrophoresis

- Separated by *mass*: Sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis.
  - Sensitivity: 0.02ug protein with a silver stain.
  - Resolution: 2% mass difference.

- Separated by *isoelectric point (pI)*: polyampholytes pH gradient gel.
  - Resolution: 0.01 pI.

Comparison of predicted with observed protein properties (localization, postsynthetic modifications)
E.coli

See Link et al. 1997 Electrophoresis 18:1259-313 (Pub)
(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9298646&dopt=Abstract)

Computationally checking proteomic data

Property	**Basis of calculation**
Protein charge	RKHYCDE (N,C), pKa, pH (Pub)
Protein mass	Calibrate with knowns (complexes)
Peptide mass	Isotope sum (incl.modifications)
Peptide LC	aa composition linear regression
Subcellular	Hydrophobicity, motifs (Pub)
Expression	Codon Adaptation Index (CAI)

(http://gcg.nhri.org.tw/isoelectric.html#algorithm)

Protein2: Today's story & goals

- Separation of proteins & peptides
- Protein localization & complexes
- Peptide identification (MS/MS)
  - Database searching & sequencing.
- Protein quantitation
  - Absolute & relative
- Protein modifications & crosslinking
- Protein - metabolite quantitation

Cell fraction: Periplasm
2D gel: SDS mobility
isoelectic pH

See Link et al. 1997 Electrophoresis 18:1259-313 (Pub)
(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9298646&dopt=Abstract)

Cell localization predictions

TargetP: using N-terminal sequence discriminates mitochondrion, chloroplast, secretion, & "other" localizations with a success rate of 85%. (pub) **(http://www.cbs.dtu.dk/services/TargetP/)**,
(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10891285&dopt=Abstract)

Gromiha 1999, Protein Eng 12:557-61. A simple method for predicting transmembrane alpha helices with better accuracy. (pub)
(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10436081&dopt=Abstract)

Using the information from the topology of 70 membrane proteins... correctly identifies 295 transmembrane helical segments in 70 membrane proteins with only two overpredictions.

Isotope calculations

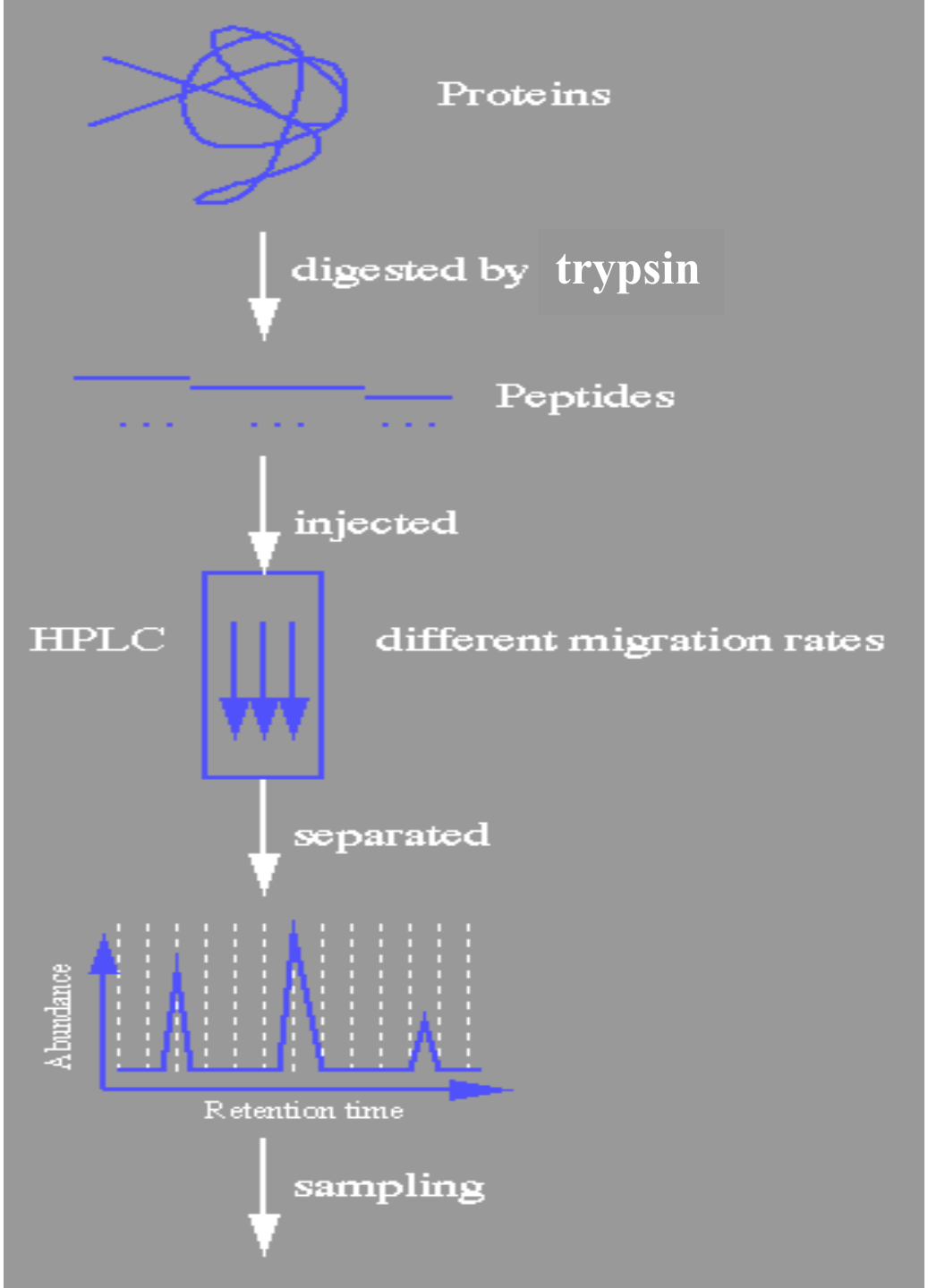
Mass resolution 0.1% vs. 1 ppm

Symbol	Mass	Abund.	Symbol	Mass	Abund.
H(1)	1.007825	99.99	H(2)	2.014102	0.015
C(12)	12.000000	98.90	C(13)	13.003355	**1.10**
N(14)	14.003074	99.63	N(15)	15.000109	0.37
O(16)	15.994915	99.76	O(17)	16.999131	0.038
S(32)	31.972072	95.02	S(33)	32.971459	0.75

Computationally checking proteomic data

Property	**Basis of calculation**
Protein charge	RKHYCDE (N,C), pKa, pH (Pub)
Protein mass	Calibrate with knowns (complexes)
Peptide mass	Isotope sum (incl.modifications)
Peptide LC	aa composition linear regression
Subcellular	Hydrophobicity, motifs (Pub)
Expression	Codon Adaptation Index (CAI)

(http://gcg.nhri.org.tw/isoelectric.html#algorithm)



High
Performance
Liquid
Chromatography

Mobile Phase of HPLC

- The interaction between the mobile phase and sample determine the migration speed.
  - Isocratic elution: constant migration speed in the column.
  - Gradient elution: gradient migration speed in the column.

Stationary Phase of HPLC

- The degree of interaction with samples determines the migration speed.
  - Liquid-Solid: polarity.
  - Liquid-Liquid: polarity.
  - Size-Exclusion: porous beads.
  - Normal Phase: hydrophilicity and lipophilicity.
  - Reverse Phase: hydrophilicity and lipophilicity.
  - Ion Exchange.
  - Affinity: specific affinity.

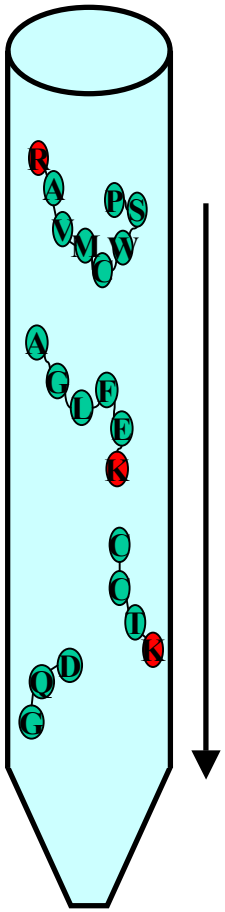
See

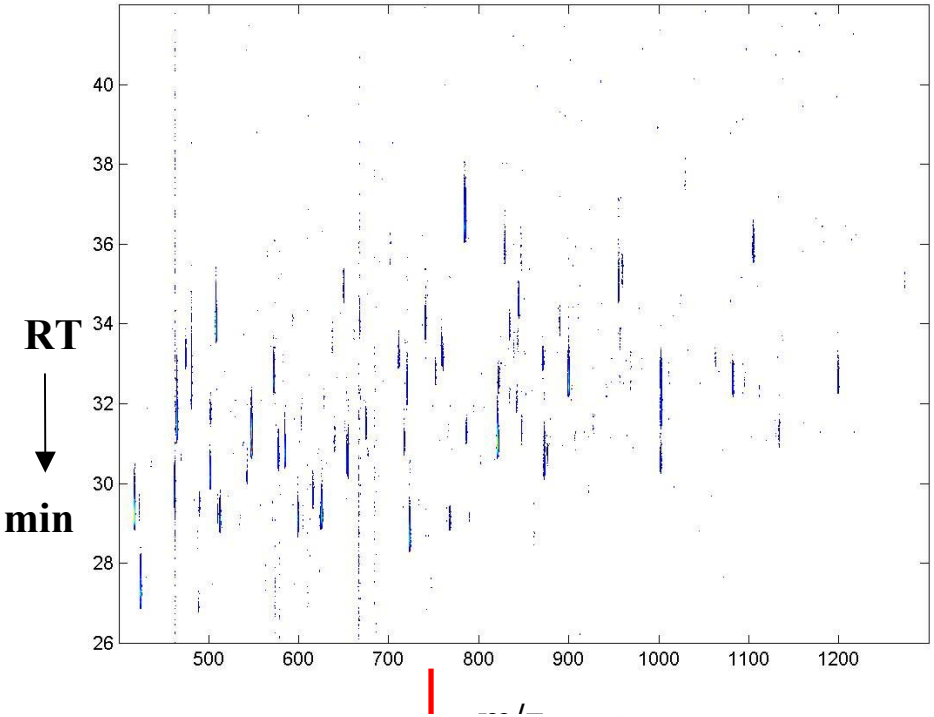
Sereda, T. et al. "Effect of the α-amino group on peptide retention behaviour in reversed-phase chromatography.

Wilce, et al. "High-performance liquid chromatography of amino acids, peptides and proteins." *Journal of Chromatography*, 632 (1993) 11-18.

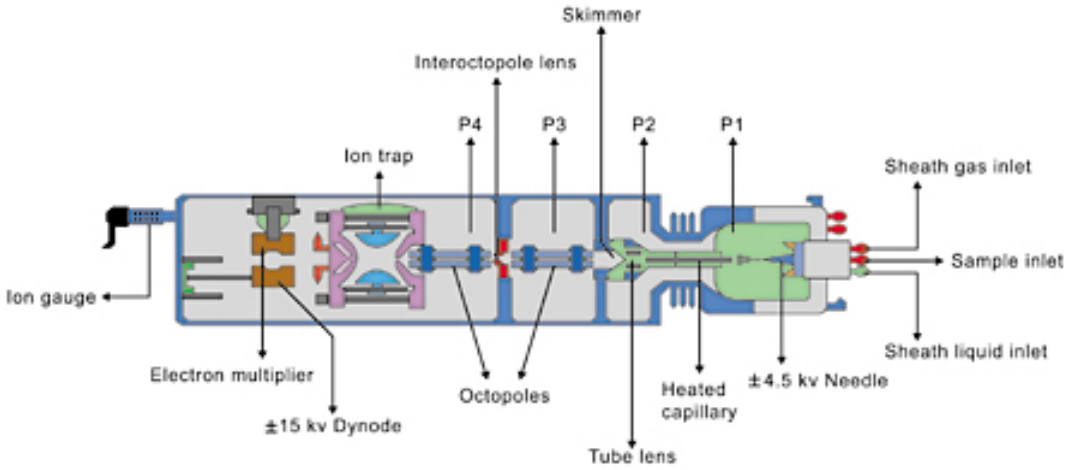
A Map is Like a 2D Peptide Gel

First Dimension:
Reverse Phase Chromatography
Separation By Hydrophobicity

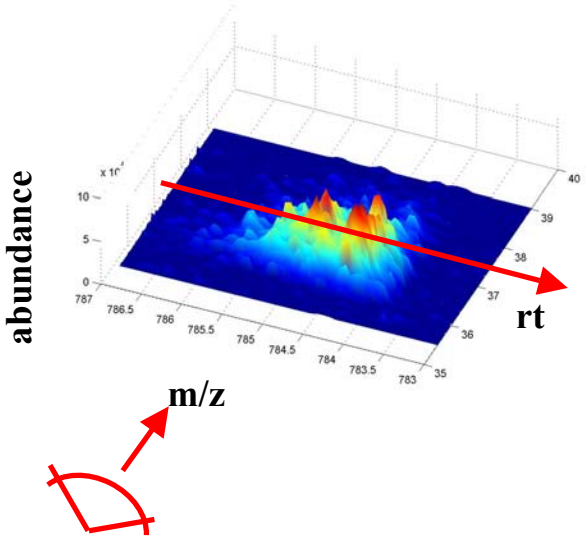


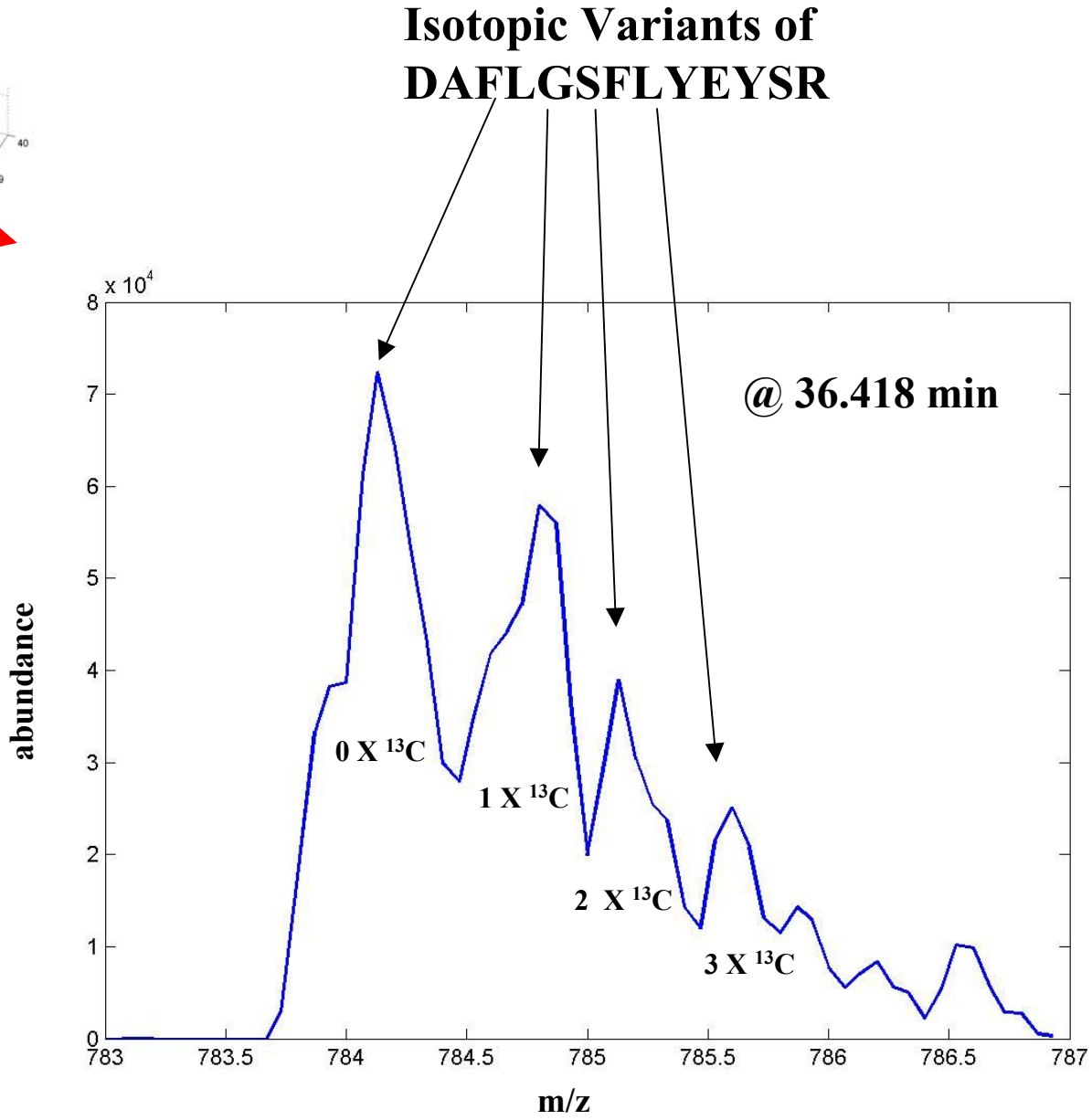


Second Dimension:
Mass Spectrometry
Separation by Mass

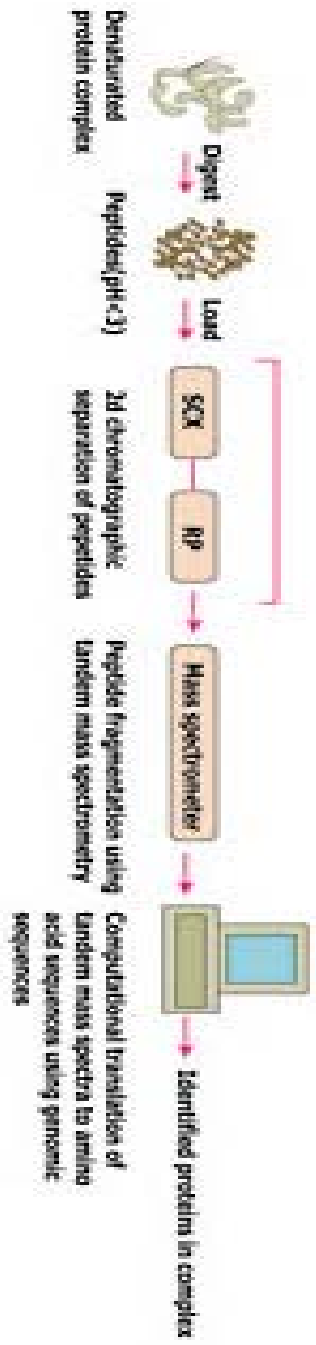


What Information Can Be Extracted From A Single Peptide Peak





K.Leptos 2001



Directed Analysis of Large Protein Complexes by 2D separation: strong cation exchange and reversed-phased liquid chromatography.

Link, et al. 1999, **Nature Biotech.** 17:676-82. (Pub)

(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10404161&dopt=Abstract)

A new 40S subunit protein

See Link AJ, et al., Nat Biotechnol. 1999 Jul;17(7):676-82. Direct analysis of protein complexes using mass spectrometry.

Protein2: Today's story & goals

- Separation of proteins & peptides
- Protein localization & complexes
- Peptide identification (MS/MS)
  - Database searching & sequencing.
- Protein quantitation
  - Absolute & relative
- Protein modifications & crosslinking
- Protein - metabolite quantitation

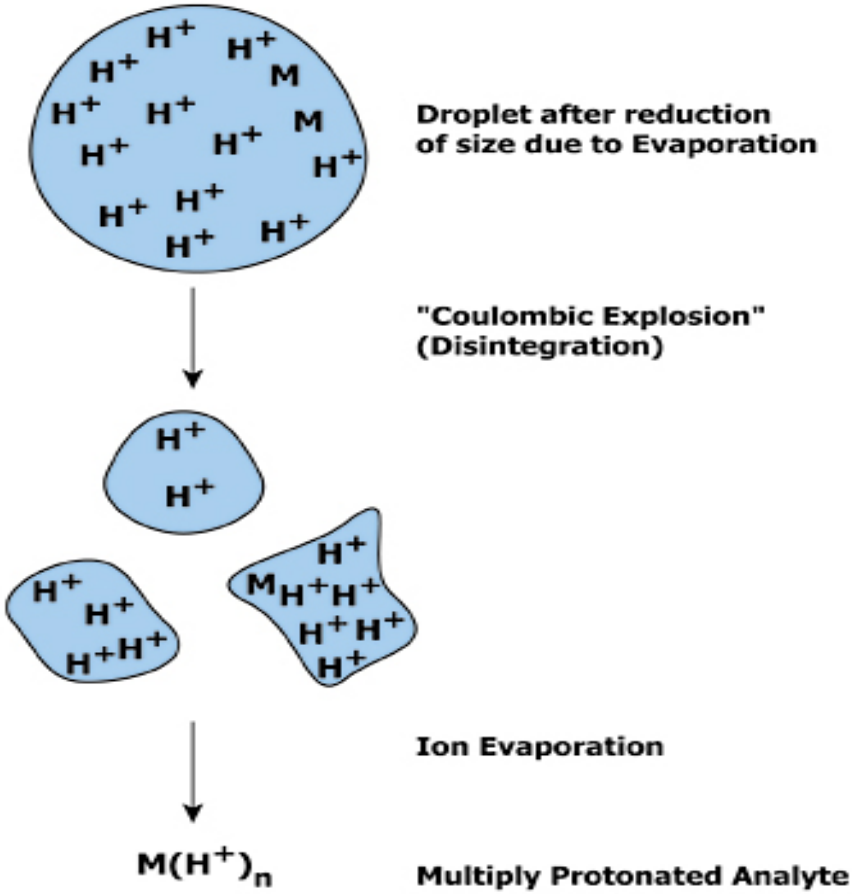
Tandem Mass Spectrometry

See Siuzdak, Gary. "The emergence of mass spectrometry in biochemical research." *Proc. Natl. Acad. Sci.* 1994, *91*, 11290-11297.

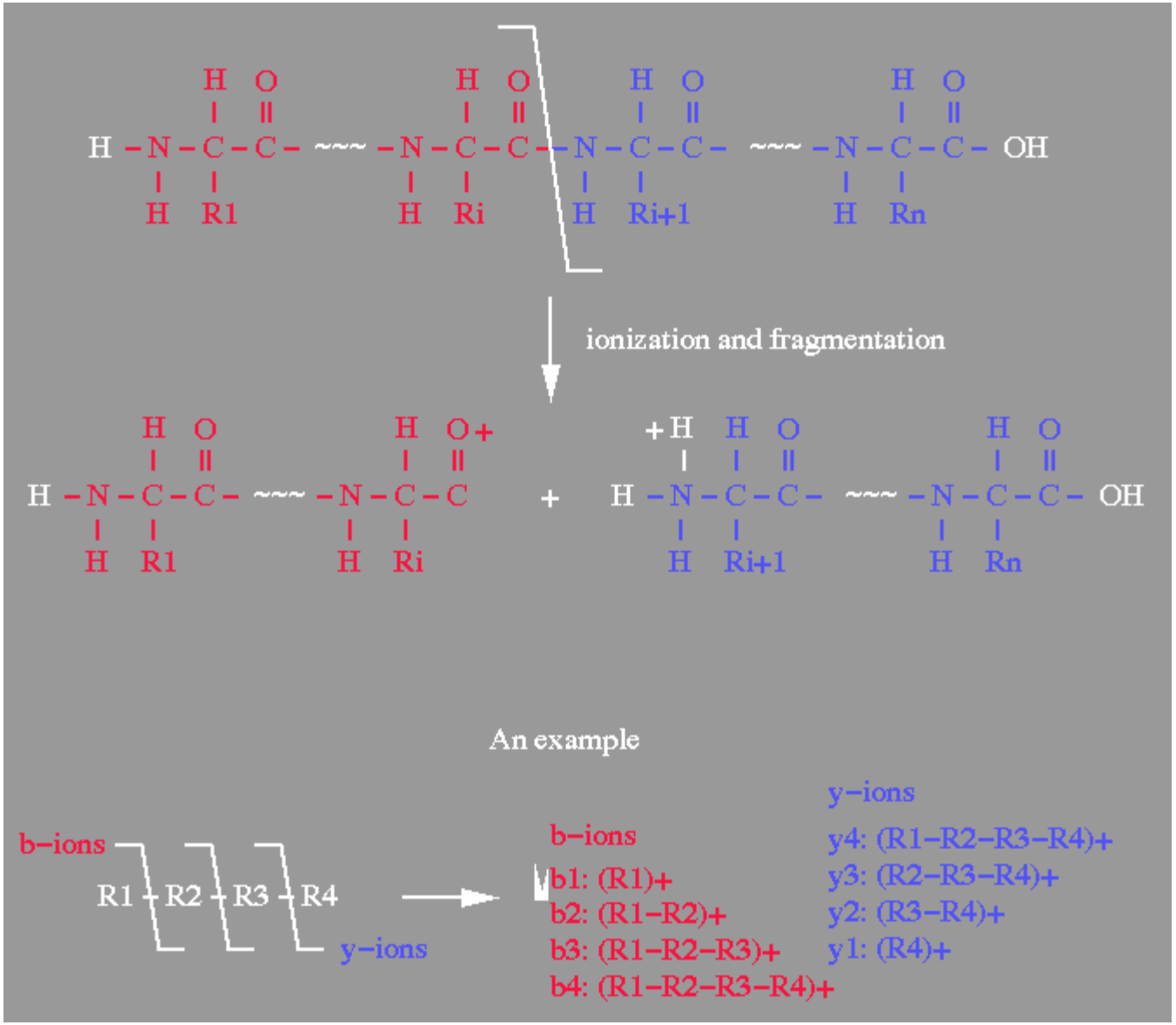
Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* 1994, *11*, 601.

Quadrople Q1 scans or selects m/z.
Q2 transmits those ions through collision gas (Ar).
Q3 Analyzes the resulting fragment ions.

Ions


Droplet after reduction of size due to Evaporation
"Coulombic Explosion" (Disintegration)
Ion Evaporation
$M(H^+)_n$
Multiply Protonated Analyte

Peptide Fragmentation and Ionization



Tandem Mass Spectra Analysis

See Gygi *et al*. Mol. Cell Bio. (1999)

Mass Spectrum Interpretation Challenge

- It is unknown whether an ion is a b-ion or an y-ion or else.
- Some ions are missing.
- Each ion has multiple of isotopic forms.
- Other ions (a or z) may appear.
- Some ions may lose a water or an ammonia.
- Noise.
- Amino acid modifications.

A dynamic programming approach to de novo peptide sequencing via tandem mass spectrometry

See Chen et al 2000. 11th Annual ACM-SIAM Symp. of Discrete Algorithms pp. 389-398.

SEQUEST: Sequence-Spectrum Correlation

Given a raw tandem mass spectrum and a protein sequence database.

- For every protein in the database,
- For every subsequence of this protein
  - Construct a hypothetical tandem mass spectrum
  - Overlap two spectra and compute the correlation coefficient (CC).
- Report the proteins in the order of CC score.

Eng, et al. 1994, **Amer. Soc. for Mass Spect.** 5: 976-989 (Sequest)
(http://thompson.mbt.washington.edu/sequest/)

Protein2: Today's story & goals

- Separation of proteins & peptides
- Protein localization & complexes
- Peptide identification (MS/MS)
  - Database searching & sequencing.
- Protein quantitation
  - Absolute & relative
- Protein modifications & crosslinking
- Protein - metabolite quantitation

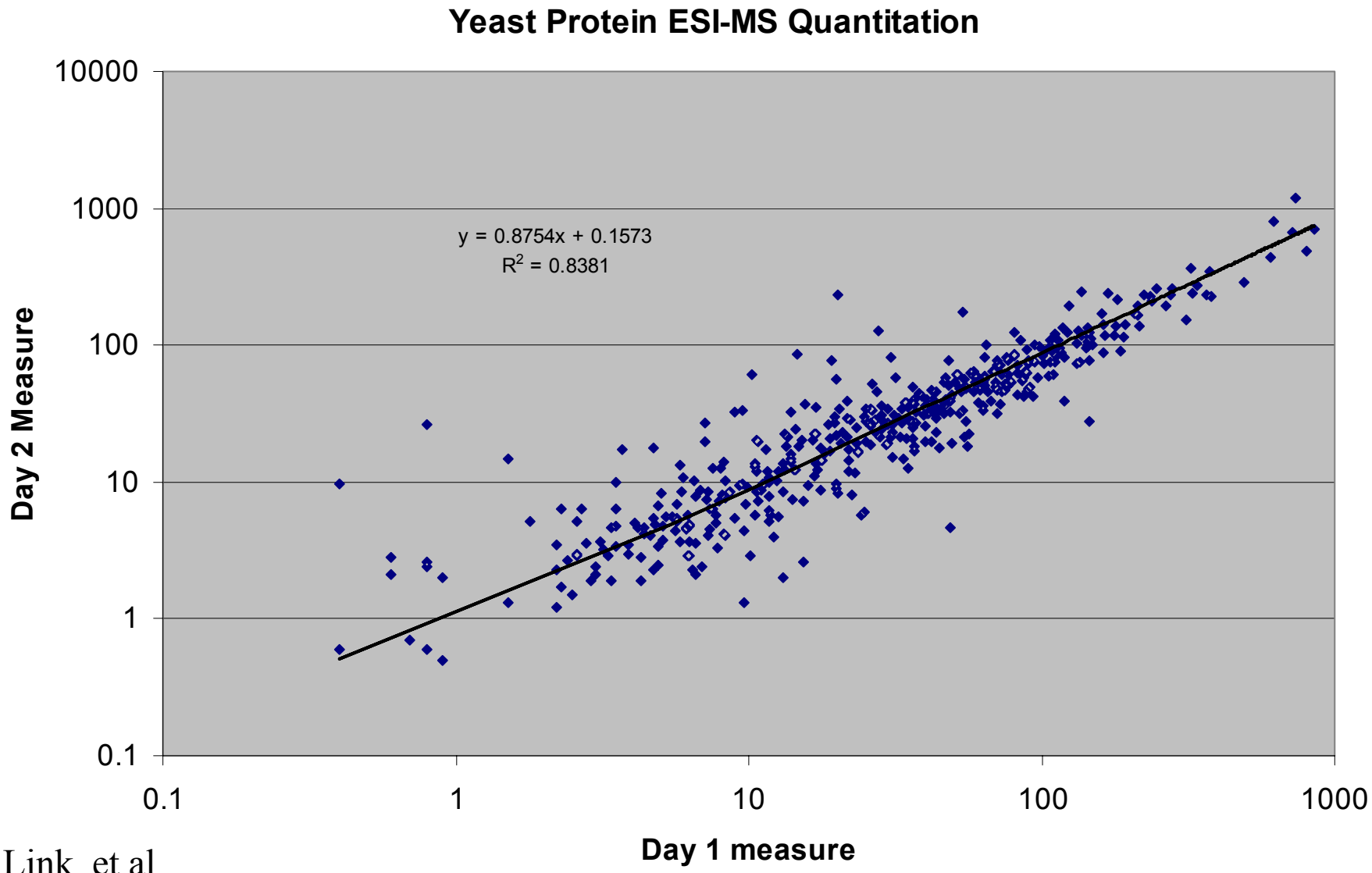
Expression quantitation methods

RNA	Protein
Genes immobilized labeled RNA	Antibody arrays
RNAs immobilized labeled genes-	
Northern gel blot	Westerns
QRT-PCR	-none-
Reporter constructs	same
Fluorescent In Situ (Hybridization)	same (Antibodies)
Tag counting (SAGE)	-none-
Differential display	mass spec

Molecules per cell

E.coli/yeast	Human
Individual mRNAs:	
10^{-1} to 10^{3}	10^{-4} to 10^{5}
Proteins:	
10 to 10^{6}	10^{-1} to 10^{8}

MS Protein quantitation R=.84

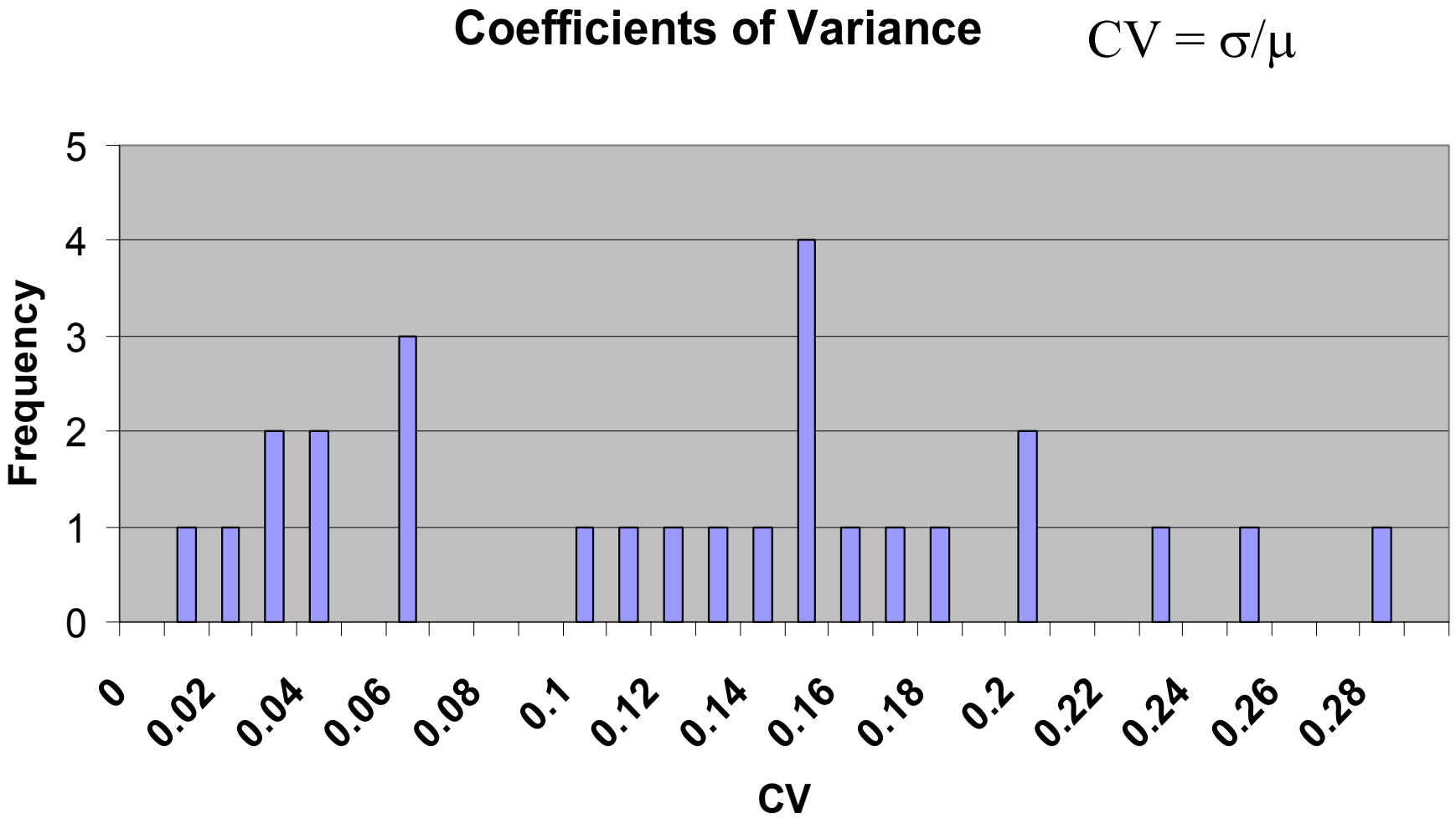


Link, et al

MS quantitation reproducibility

Sample: Angiotensin, Neurotensin, Bradykinin

Map: 600 – 700 m/z



Correlation between protein and mRNA abundance in yeast

See Gygi et al. 1999, **Mol. Cell Biol.** 19:1720-30 (Pub)

(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10022859&dopt=Abstract)

Normality tests

See Weiss 5th ed. Page 920.
Types of non-normality: kurtosis, skewness (www)
(log) transformations to normal.
(http://www.marketminer.com/prophet/statguide/n-dist_exam_res.html)

Futcher et al 1999, A sampling of the yeast proteome. **Mol.Cell.Biol.** 19:7357-7368. (Pub) **(http://mcb.asm.org/cgi/content/full/19/11/7357?view=full&pmid=10523624)**

Spearman correlation rank test

$r_s = 1 - \{6S/(n^3-n)\}$

Rank (from 1 to n, where n is the number of pairs of data) the numbers in each column. If there are ties within a column , then assign all the measurements that tie the same median rank. Note, avoids ties (which reduce the power of the test) by measuring with as fine a scale as possible. S= sum of the square differences in rank. (ref) **(http://www.wisc.edu/ecology/labs/phenology/spearman.html)**

	X	Y	Rx	Ry
	1	8	1	4
	6	2	3	1
	6	3	3	2
n=4	6	4	3	3

Correlation of (phosphorimager ^{35}S met) protein & mRNA

r_p = 0.76 for log(adjusted RNA) to log(protein)

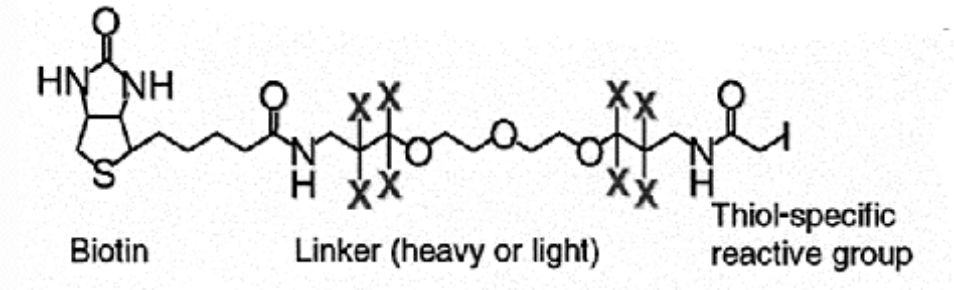
r_s = .74 overall; 0.62 for the top 33 proteins & 0.56 (not significantly different) for the bottom 33 proteins

Observed (Phosphorimage) protein levels vs. Codon Adaptation Index (CAI)

Codon Adaptation Index (CAI) Sharp and Li (1987); fi is the relative frequency of codon i in the coding sequence, and Wi the ratio of the frequency of codon i to the frequency of the major codon for the same amino-acid.

$$\ln(CAI) = \sum_{i=1,61} f_i \ln(W_i)$$

ICAT Strategy for Quantifying Differential Protein Expression.



X= H or D

See Gygi *et al*. Nature Biotechnology (1999)

Mass Spectrum and Reconstructed Ion Chromatograms.

See Gygi *et al*. Nature Biotechnology (1999)

Protein & mRNA Ratios +/- Galactose

See Ideker et al 2001

(http://arep.med.harvard.edu/pdf/Ideker01.pdf)

Protein2: Today's story & goals

- Separation of proteins & peptides
- Protein localization & complexes
- Peptide identification (MS/MS)
  - Database searching & sequencing.
- Protein quantitation
  - Absolute & relative
- Protein modifications & crosslinking
- Protein - metabolite quantitation

Post-synthetic modifications

- Radioisotopic labeling: PO_4 S,T,Y,H
- Affinity selection:

  Cys: ICAT biotin-avidin selection

  PO_4: immobilized metal Ga(III) affinity chromatography(IMAC)

  (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10424175&dopt=Abstract)

  Specific PO_4 Antibodies

  Lectins for carbohydrates
- Mass spectrometry

32P labeled phoshoproteomics

Low abundance cell cycle proteins not detected above background from abundant proteins

See Futcher et al 1999, A sampling of the yeast proteome. **Mol.Cell.Biol.** 19:7357-7368. (Pub)
(http://mcb.asm.org/cgi/content/full/19/11/7357?view=full&pmid=10523624)

Natural crosslinks

Disulfides	Cys-Cys
Collagen	Lys-Lys
Ubiquitin	C-term-Lys
Fibrin	Gln-Lys
Glycation	Glucose-Lys
Adeno primer proteins	dCMP-Ser

Crosslinked peptide Matrix-assisted laser desorption ionization Post-Source Decay (MALDI-PSD-MS)

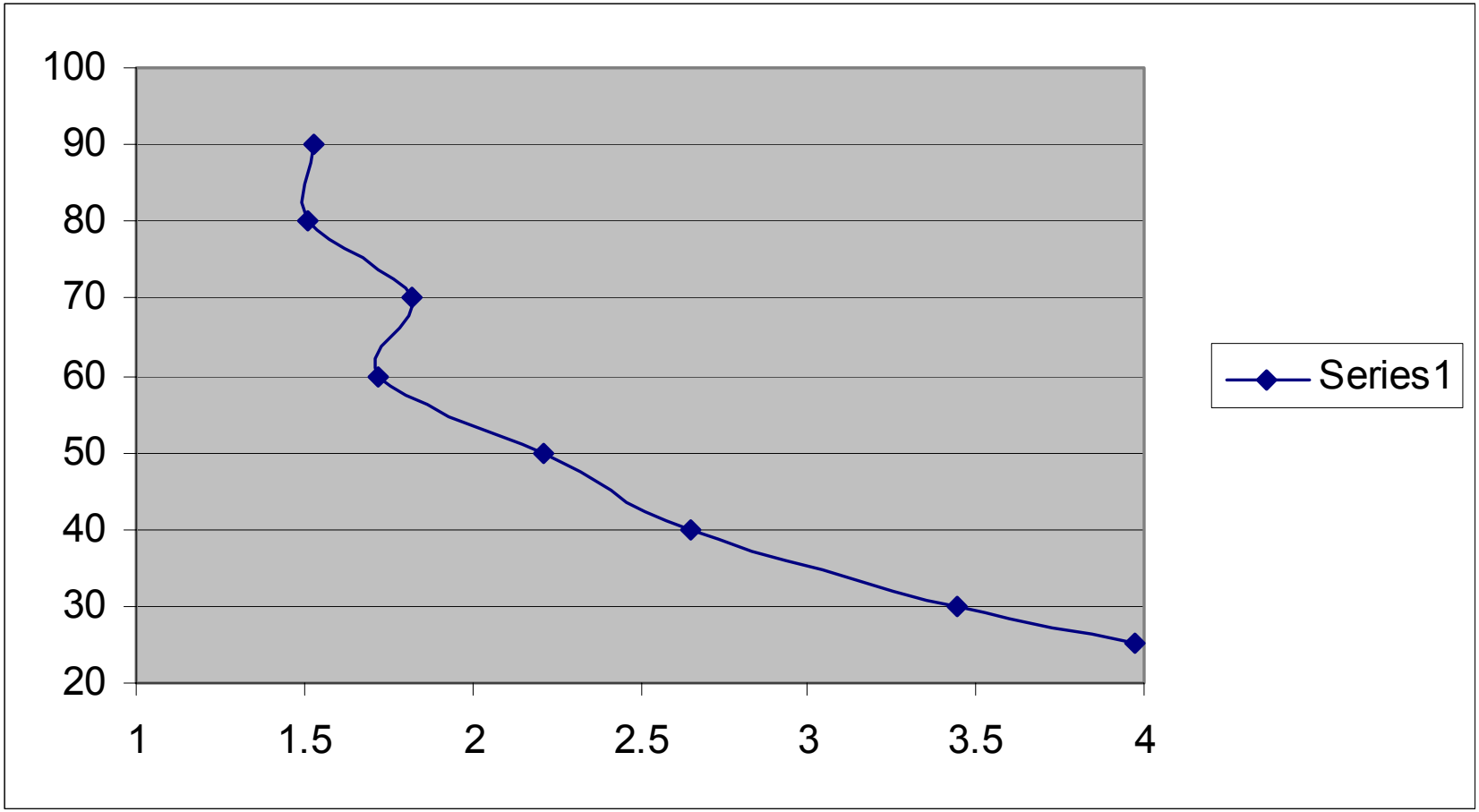
tryptic digest of BS3 cross-linked FGF-2. Cross-linked peptides are identified by using the program ASAP and are denoted with an asterisk (9). (B) MALDI-PSD spectrum of cross-linked peptide E45-R60 (M + H+ = m/z 2059.08).

Constraints for homology modeling based on MS crosslinking distances

The 15 nonlocal throughspace distance constraints generated by the chemical cross-links (yellow dashed lines) superimposed on the average NMR structure of FGF-2 (1BLA). The 14 lysines of FGF-2 are shown in red.

See Young et al 2000, PNAS 97: 5802 (Pub)
(http://www.pnas.org/cgi/content/full/97/11/5802)

Homology modeling accuracy



Swiss-model RMSD of the test set in Angstroms

(http://www.expasy.ch/swissmod/SM_LikelyPrecision.html)

Top 20 threading models for FGF ranked by crosslinking constraint error

See Young et al. PNAS | May 23, 2000 | vol. 97 | no. 11 | 5802-5806
http://www.pnas.org/cgi/content/full/97/11/5802

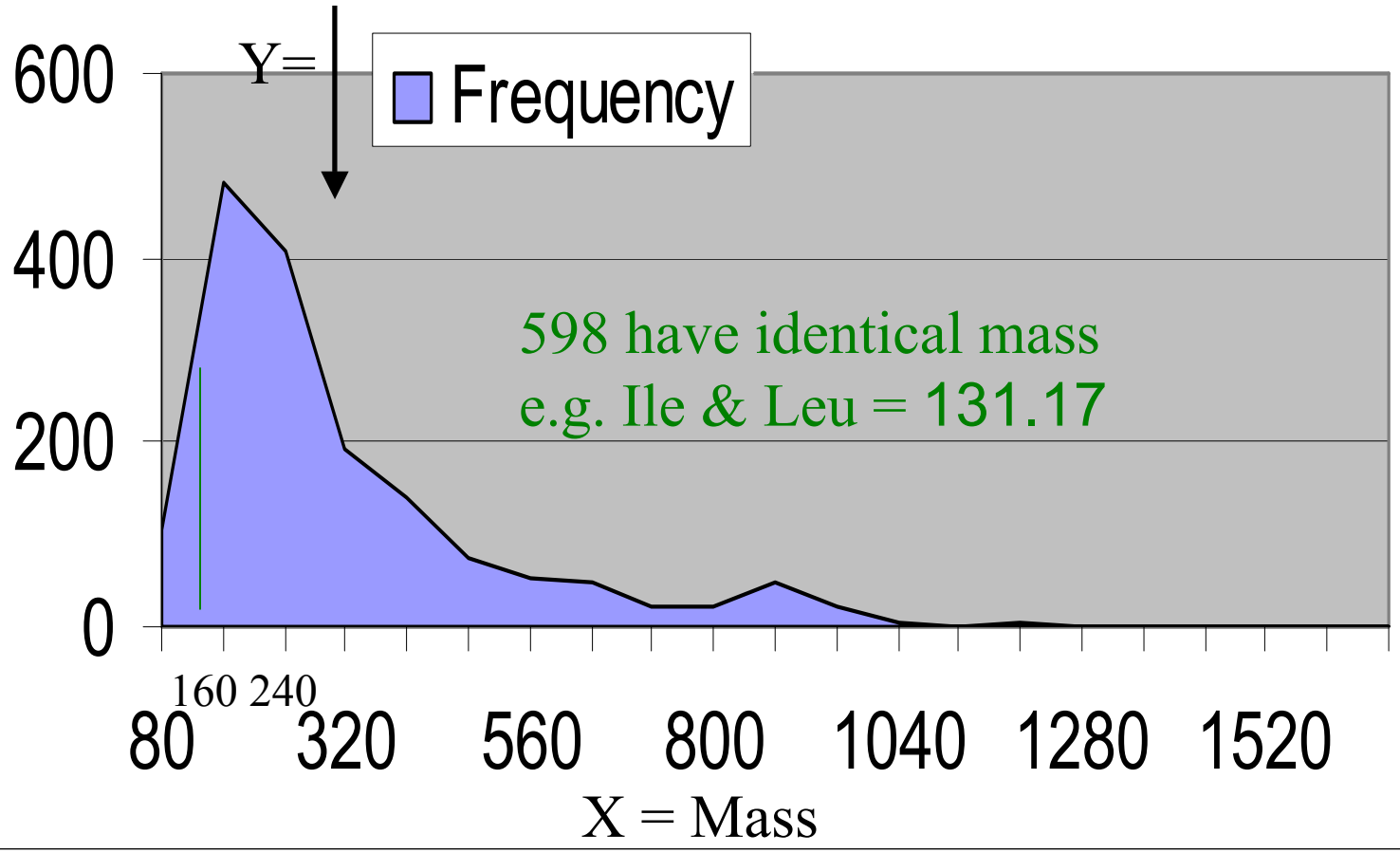
Protein2: Today's story & goals

- Separation of proteins & peptides
- Protein localization & complexes
- Peptide identification (MS/MS)
  - Database searching & sequencing.
- Protein quantitation
  - Absolute & relative
- Protein modifications & crosslinking
- Protein - metabolite quantitation

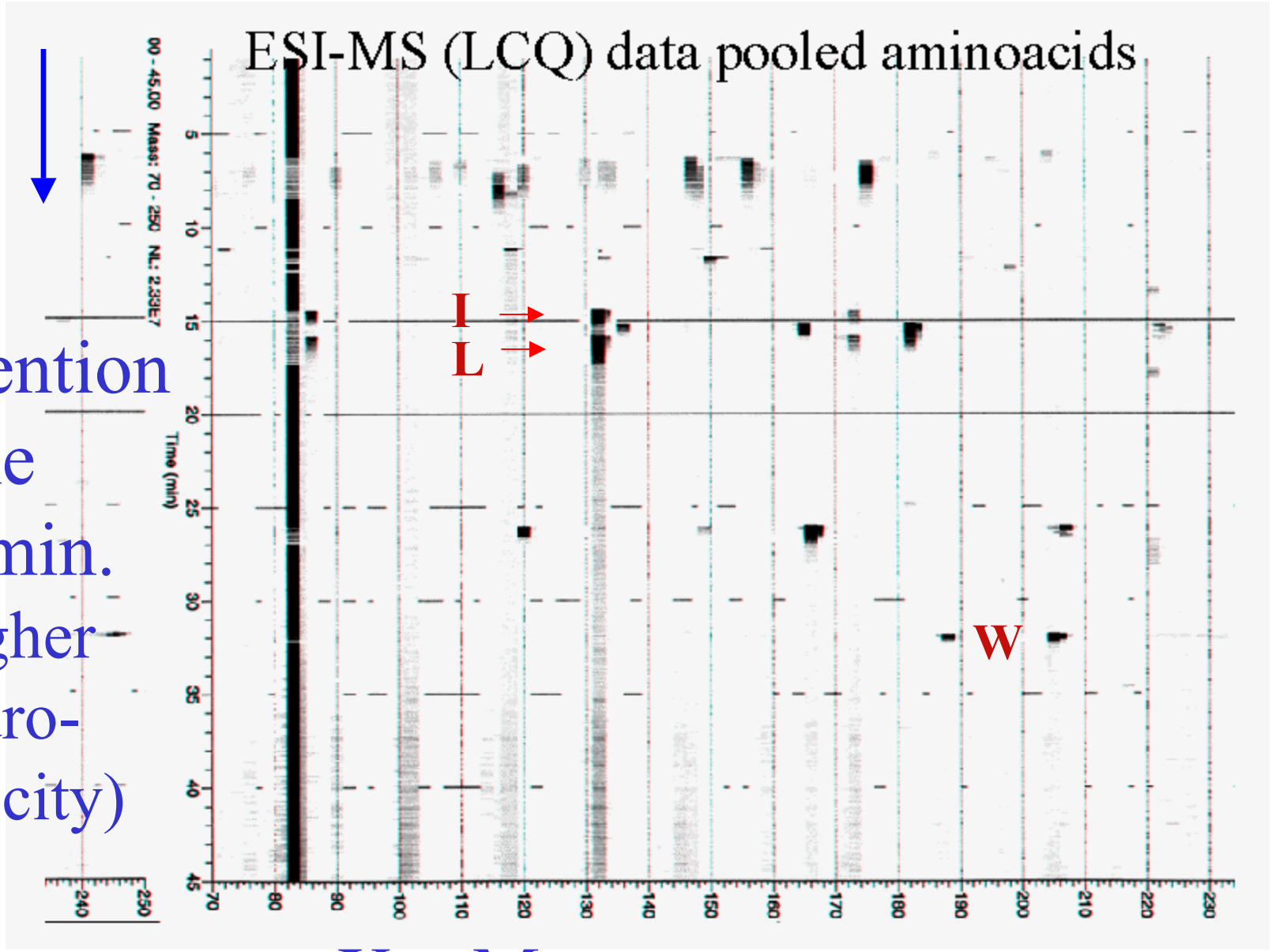
Challenges for accurately measuring metabolites

- Rapid kinetics
- Rapid changes during isolation
- Idiosyncratic detection methods:
  enzyme-linked, GC, LC, NMR
  (albeit fewer molecular types than RNA& protein)

Databases



Karp et al. (1998) *NAR* 26:50. EcoCyc; Selkov, et al. (1997) *NAR* 25:37. WIT
Ogata et al. (1998) Biosystems 47:119-128 KEGG


ESI-MS (LCQ) data pooled aminoacids
Y= RP LC retention time in min. (higher hydro-phocity)
X = Mass
L
L
W
Time (min)

Metabolite fragmentation & stable isotope labeling

See Wunschel J Chromatogr A 1997, 776:205-19 Quantitative analysis of neutral & acidic sugars in whole bacterial cell hydrolysates using high-performance anion-exchange LC-ESI-MS2.
(Pub)
(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9291597&dopt=Abstract)

Isotopomers

See Klapa et al. Biotechnol Bioeng 1999; 62:375. Metabolite and isotopomer balancing in the analysis of metabolic cycles: I. Theory. (Pub) "accounting for the contribution of all pathways to label distribution is required, especially ... multiple turns of metabolic cycles... 13C (or 14C) labeled substrates."

(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10099550&dopt=Abstract)

MetaFoR: Metabolic Flux Ratios

Fractional 13C labeling > Quantitative 2D NMR
Why use amino acids from proteins rather than metabolites directly?

See:

Sauer J et al. Bacteriol 1999;181:6679-88 (Pub)
(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10542169&dopt=Abstract)

Szyperski et al 1999 Metab. Eng. 1:189.
(http://www.biotech.biol.ethz.ch/sauer/pdf/1999 - Uwe Sauer Metabol Eng.pdf)

Dauner et al. 2001 Biotec Bioeng 76:144
(http\www.biotech.biol.ethz.ch\sauer\pdf\2001 - Michael Dauner et al B&B.pdf)

A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations

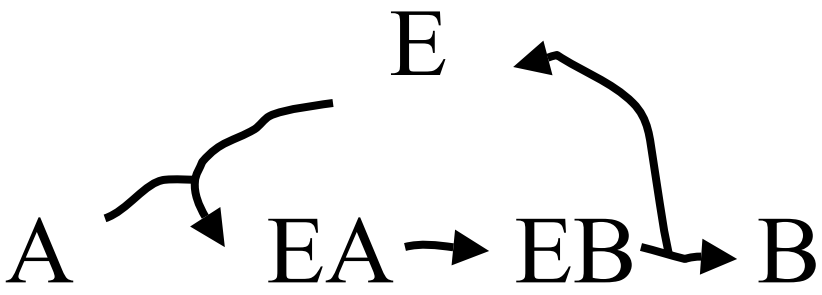
See Raamsdonk et al. 2001 Nature Biotech 19:45.
(http://arep.med.harvard.edu/pdf/Raamsdonk01.pdf)

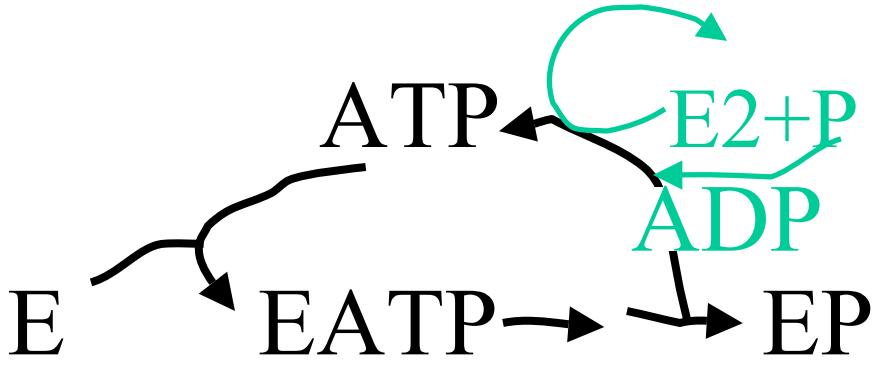
Types of interaction models

Quantum Electrodynamics	subatomic
Quantum mechanics	electron clouds
Molecular mechanics	spherical atoms (101Pro1)
Master equations	stochastic single molecules (Net1)
Phenomenological rates ODE	**Concentration & time (C,t)**
Flux Balance	dC_{ik}/dt optima steady state (Net1)
Thermodynamic models	$dC_{ik}/dt = 0$ k reversible reactions
Steady State	$\Sigma dC_{ik}/dt = 0$ (sum k reactions)
Metabolic Control Analysis	$d(dC_{ik}/dt)/dC_j$ (i = chem.species)
Spatially inhomogenous models	dCi/dx

Increasing scope, decreasing resolution

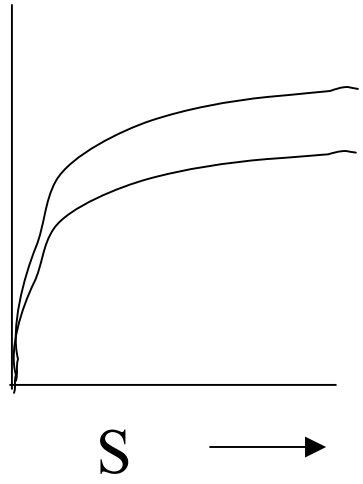
How do enzymes & substrates formally differ?





Catalysts increase the rate (&specificity) without being consumed.

Enzyme rate equations with one Substrate & one Product



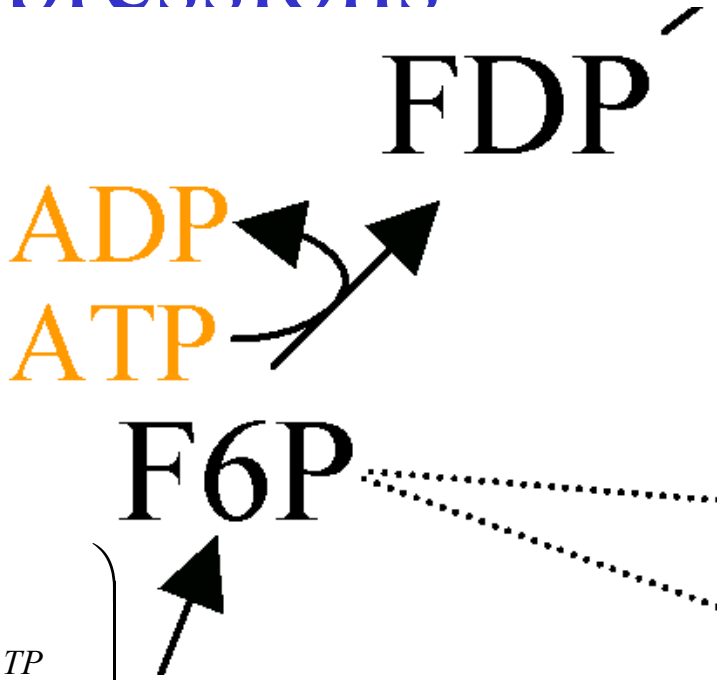
$$\frac{dP}{dt} = \frac{V\,(S/Ks - P/Kp)}{1 + S/Ks + P/Kp}$$

As P approaches 0:

$$\frac{dP}{dt} = \frac{V}{1 + Ks/S}$$

Enzyme Kinetic Expressions

Phosphofructokinase

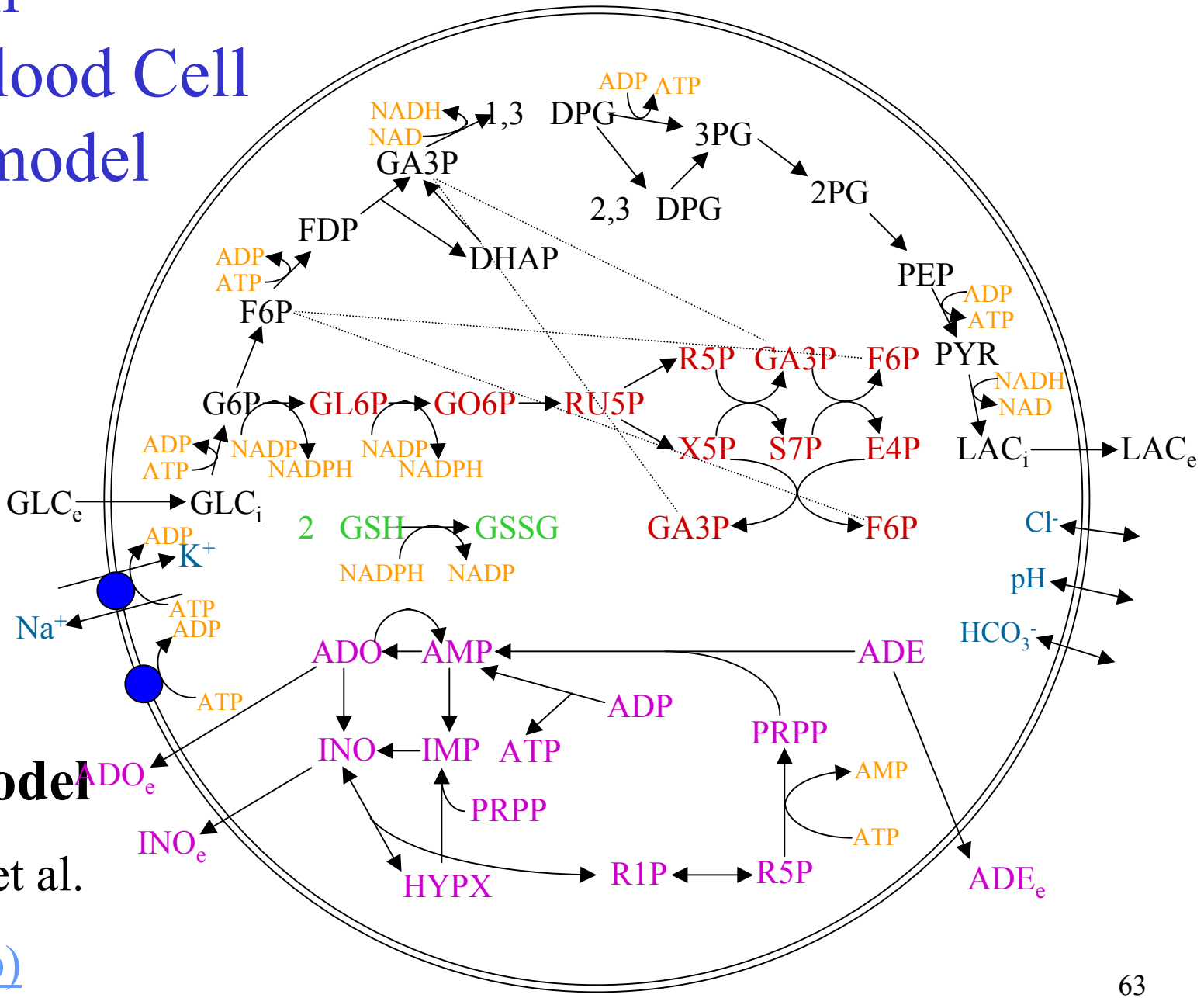


$$v_{PFK} = \frac{v_{mx}^{PFK}}{N_{PFK}} \left(\frac{F6P / K_{F6P}^{PFK}}{1 + F6P / K_{F6P}^{PFK}} \right) \left(\frac{Mg \bullet ATP / K_{Mg \bullet ATP}^{PFK}}{1 + Mg \bullet ATP / K_{Mg \bullet ATP}^{PFK}} \right)$$

$$N_{PFK} = 1 + L_0^{PFK} \frac{\left(1 + ATP_{free} / K_{ATP}^{PFK}\right)^4 \left(1 + Mg / K_{Mg}^{PFK}\right)^4}{\left(1 + AMP / K_{AMP}^{PFK}\right)^4 \left(1 + F6P / K_{F6P}^{PFK}\right)^4}$$

Allosteric kinetic parameters for AMP, etc.

Human Red Blood Cell ODE model



ODE model

Jamshidi et al.

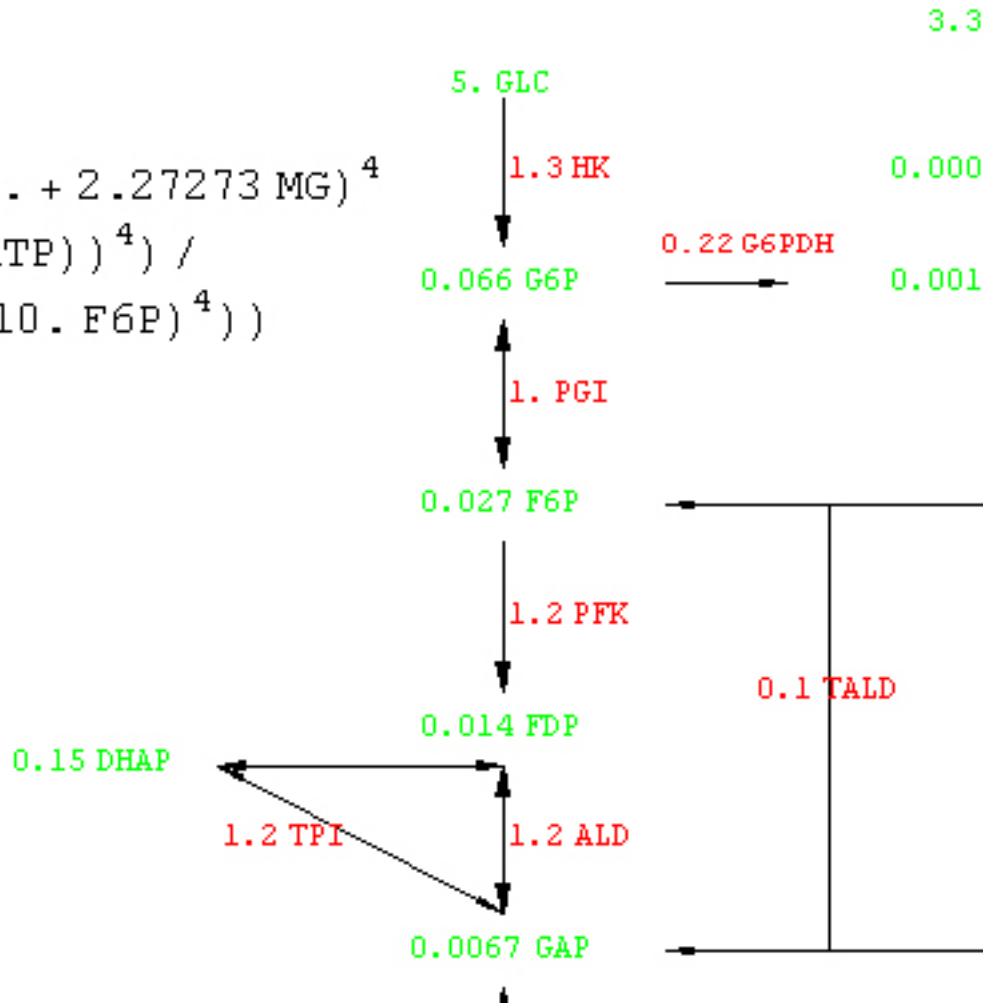
2000 (Pub)

(http://atlas.med.harvard.edu/gmc/rbc.html)

Red Blood Cell in Mathematica

```
In[4]:= PFK

Out[4]= (251.732 F6P MGATP) /
   ((0.1 + F6P) (1. + (0.001072 (1. + 2.27273 MG)^4
          (1. + 101.919 (ATP - MGATP))^4) /
       ((1. + 30.303 AMP)^4 (1. + 10. F6P)^4))
    (0.068 + MGATP))
```



ODE model

Jamshidi et al.

2000 (Pub)

(http://atlas.med.harvard.edu/gmc/rbc.html)

Protein2: Today's story & goals

- Separation of proteins & peptides
- Protein localization & complexes
- Peptide identification (MS/MS)
  - Database searching & sequencing.
- Protein quantitation
  - Absolute & relative
- Protein modifications & crosslinking
- Protein - metabolite quantitation