Protein1: Structure &Interactions (Last week)

- 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: Properties & Quantitation

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

2

Why purify?

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

3

(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

4

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.

5

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

Link et al. 1997 Electrophoresis 18:1259-313 (Pub)

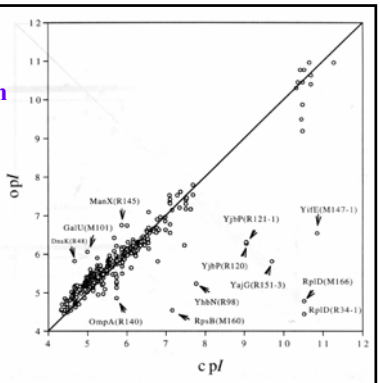


Figure 2. Comparison between the predicted and observed *pI* for identified proteins. For each protein, the deviation of op*I* from cp*I* was calculated. The arrows point to proteins for which this deviation is > 2

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)

7

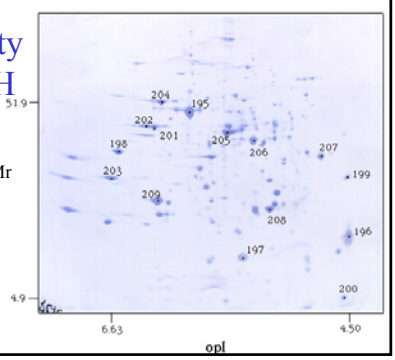
Protein2: Today's story & goals

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

8

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

Mr



Link et al. 1997
Electrophoresis
18:1259-313
(Pub)

Cell localization predictions

TargetP: using N-terminal sequence discriminates mitochondrion, chloroplast, secretion, & "other" localizations with a success rate of 85%. (pub)

Gromiha 1999, Protein Eng 12:557-61. A simple method for predicting transmembrane alpha helices with better accuracy. (pub)

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

10

Isotope calculations

Mass resolution 0.1% vs. 1 ppm

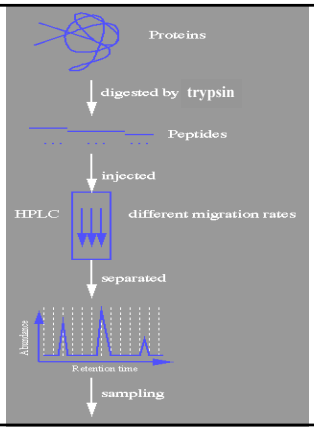
Symbol	Mass	Abund.	Symbol	Mass	Abund.
H(1)	1.007825	99.99	H(2)	2.014102	0.015
C(12)	12.000000	98.90	C(13)	13.003355	**1.10**
N(14)	14.003074	99.63	N(15)	15.000109	0.37
O(16)	15.994915	99.76	O(17)	16.999131	0.038
S(32)	31.972072	95.02	S(33)	32.971459	0.75

11

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)

12



High Performance Liquid Chromatography

13

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.

14

Stationary Phase of HPLC

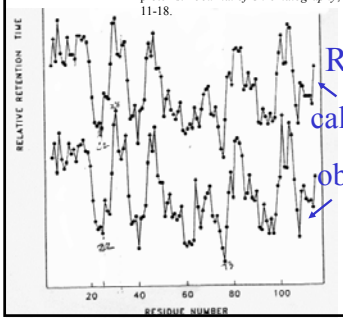
- The degree of interaction with samples determines the migration speed.
  - Liquid-Solid: polarity.
  - Liquid-Liquid: polarity.
  - Size-Exclusion: porous beads.
  - Normal Phase: hydrophilicity and lipophilicity.
  - Reverse Phase: hydrophilicity and lipophilicity.
  - Ion Exchange.
  - Affinity: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

15

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

(The calculated curve is displaced upward for clarity)

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



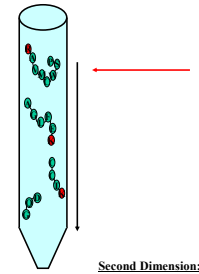
Empirical linear regression varies with type of LC-material

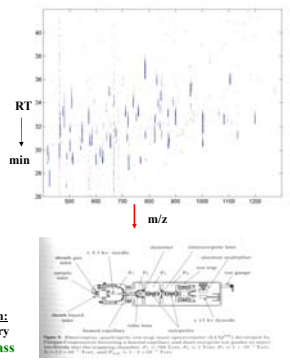
	α-NH3+ no	α-NH3+ yes	C18 no
W	10.1	9.3	9.8
P	**8.8**	**5.5**	8.8
L	7.5	4.6	9.5
I	5.8	3.0	8.4
M	4.8	3.0	2.6
Y	4.5	3.1	6.1
V	3.5	1.3	4.9
C	3.4	2.9	0.5
P	2.7	0.7	2.8
E	0.3	0.5	0.8
A	0.2	0.1	1.7
D	0.0	0.6	1.1
G	0.0	0.0	0.4
T	-0.1	1.0	1.8
S	-0.8	-0.1	0.3
Q	-0.9	0.0	-0.7
N	-3.0	-2.1	0.0
R	-3.1	-2.1	2.4
H	-3.3	-1.5	0.6
K	**-3.5**	**-1.6**	0.0

16

A Map is Like a 2D Peptide Gel

First Dimension:
Reverse Phase Chromatography
Separation By Hydrophobicity

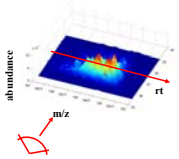




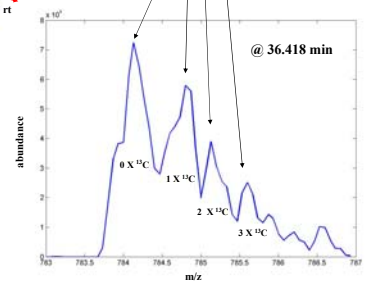
Second Dimension:
Mass Spectrometry
Separation by Mass

17

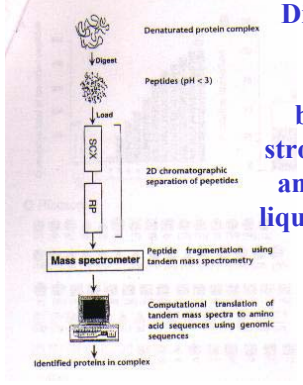
What Information Can Be Extracted From A Single Peptide Peak



Isotopic Variants of DAFLGSFLYEYSR



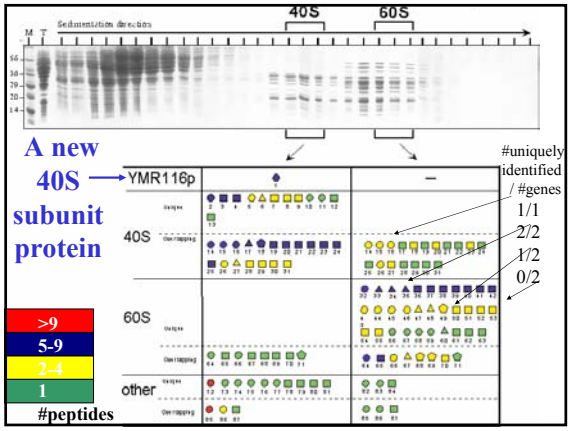
K.Leptos 2001



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)

19



Protein2: Properties & Quantitation

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

21

The Finnigan LCQ: An ESI-QIT Mass Spectrometer

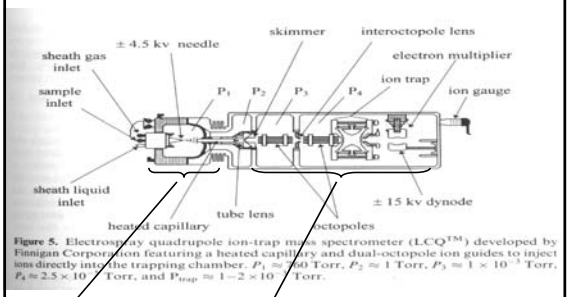


Figure 5. Electrospray quadrupole ion-trap mass spectrometer (LCQ™) developed by Finnigan Corporation featuring a heated capillary and dual-octopole ion guides to inject ions directly into the trapping chamber. $P_1 \approx 760$ Torr, $P_2 \approx 1$ Torr, $P_3 \approx 1 \times 10^{-3}$ Torr, $P_4 \approx 2.5 \times 10^{-5}$ Torr, and $P_{trap} \approx 1-2 \times 10^{-3}$ Torr.

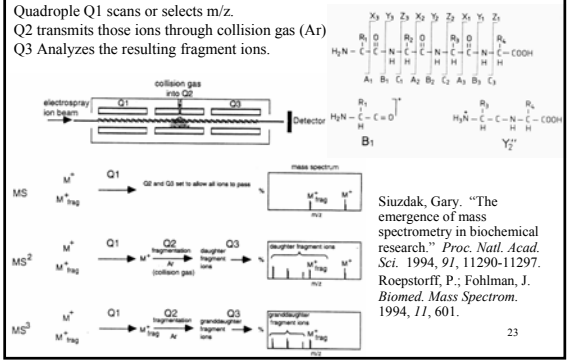
Electro-Spray Ionization chamber

Mass Analyzer/Detector

22

Tandem Mass Spectrometry

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



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.

23

Ions

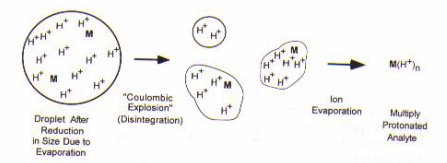
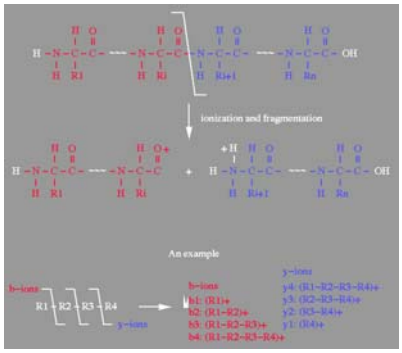


FIG. 11.2. Sketch showing macroscopic droplets containing the analyte and an excess of positive charges (protons). After reduction in size due to evaporation, a given droplet disintegrates due to coulombic repulsion of like charges into smaller droplets, until finally the analyte molecule *(M)* is ejected together with some of the excess charge.

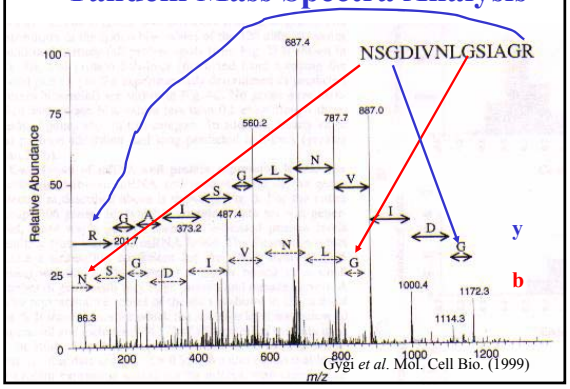
24

Peptide Fragmentation and Ionization



25

Tandem Mass Spectra Analysis



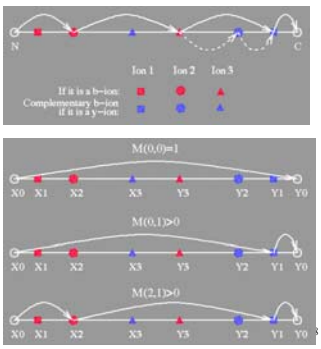
Gygi *et al*. Mol. Cell Bio. (1999)

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.

27

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



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)

29

Protein2: Properties & Quantitation

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

30

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

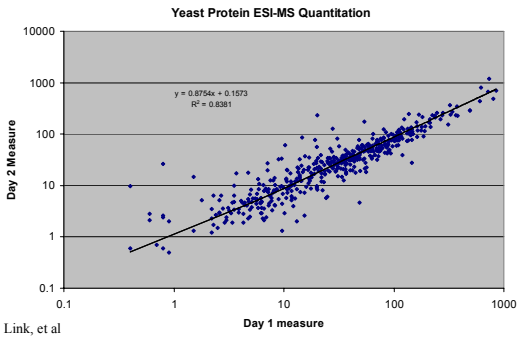
31

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}

32

MS Protein quantitation R=.84



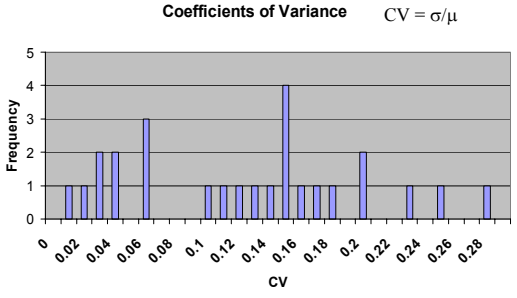
Link, et al

33

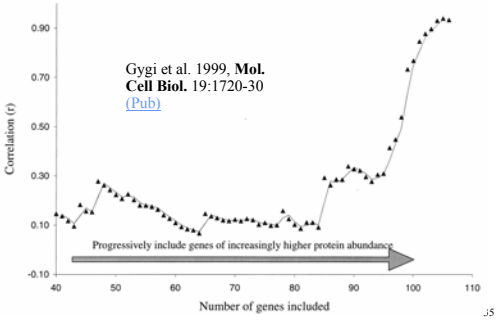
MS quantitation reproducibility

Sample: Angiotensin, Neurotensin, Bradykinin

Map: 600 – 700 m/z



Correlation between protein and mRNA abundance in yeast



35

Normality tests

See Weiss 5th ed. Page 920.
Types of non-normality: kurtosis, skewness (www)
(log) transformations to normal.

Futcher et al 1999, A sampling of the yeast proteome. **Mol.Cell.Biol.** 19:7357-7368. (Pub)

36

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)

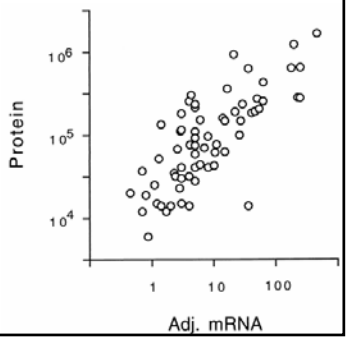
	X	Y	Rx	Ry
	1	8	1	4
	6	2	3	1
	6	3	3	2
n=4	6	4	3	3

37

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

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

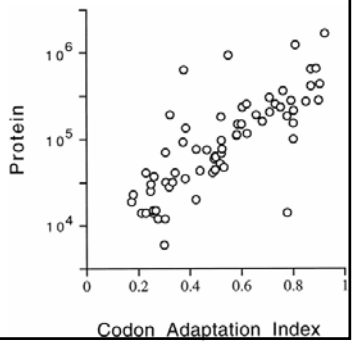
r_s = .74 overall; 0.62 for the top 33 proteins & 0.56 (not significantly different) for the bottom 33 proteins

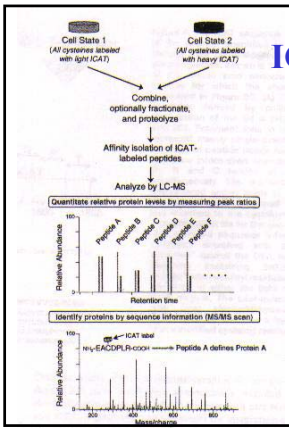


Observed (Phosphorimage) protein levels vs. Codon Adaptation Index (CAI)

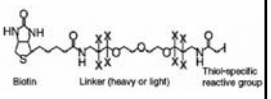
Codon Adaptation Index (CAI) Sharp and Li (1987); fi is the relative frequency of codon i in the coding sequence, and Wi the ratio of the frequency of codon i to the frequency of the major codon for the same amino-acid.

$$\ln(CAI) = \sum_{i=1,61} fi \ln(Wi)$$



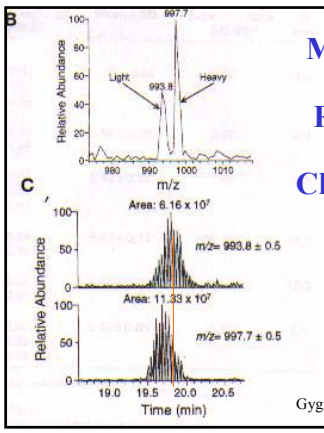


ICAT Strategy for Quantifying Differential Protein



X= H or D

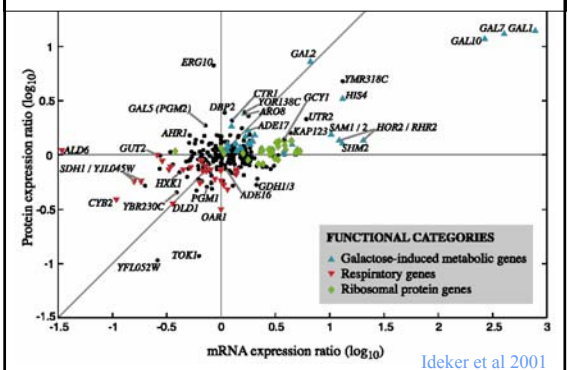
Gygi *et al*. Nature Biotechnology (1999)



Mass Spectrum and Reconstructed Ion Chromatograms.

Gygi *et al*. Nature Biotechnology (1999)

Protein & mRNA Ratios +/- Galactose



Ideker et al 2001

Protein2: Properties & Quantitation

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

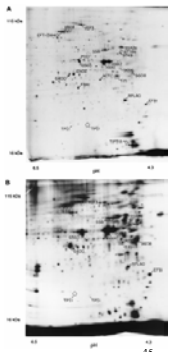
43

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)
  Specific PO_4 Antibodies
  Lectins for carbohydrates
- Mass spectrometry

44

32P labeled phoshoproteomics



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

Futcher et al 1999, A sampling of the yeast proteome. **Mol.Cell.Biol.** 19:7357-7368. (Pub)

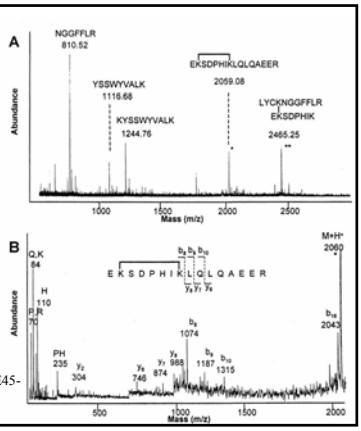
45

Natural crosslinks

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

46

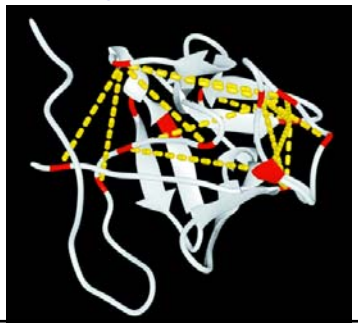
Crosslinked peptide Matrix-assisted laser desorption ionization Post-Source Decay (MALDI-PSD-MS)



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).

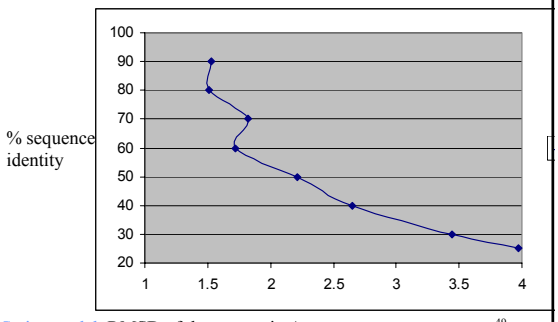
Constraints for homology modeling based on MS crosslinking distances

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.



Young et al 2000, PNAS 97: 5802 (Pub)

Homology modeling accuracy



Swiss-model RMSD of the test set in Angstroms

49

Top 20 threading models for FGF ranked by crosslinking constraint error

Name	Fold family	% Sequence identity	Threading rank	Constraint error, Å	violatio
FGF-2	β-Trefoil	98.6	1	0.0	0
IL-1β	β-Trefoil	12.7	5	0.0	0
Gastrotropin	Lipocalin	7.1	8	2.9	1
Hisactophilin	β-Trefoil	8.6	12	5.5	2
Guanylate kinase	P-loop	12.4	9	7.4	4
NTP pyrophosphohydrolase	NTP pyrophosphohydrolase	9.3	6	14.5	3
Glutathione peroxidase	Thioredoxin	11.1	14	16.6	5
Retinol-binding protein	Lipocalin	9.1	18	17.1	3
Nucleoside diphospholkinase	Ferridoxin-like	8.8	20	18.6	2
Cytochrome c_4	Cytochrome c	12.6	11	21.4	5
Aspartate carbamoyltransferase	Ferridoxin-like	9.8	13	22.6	4
D-UTPase	β-Clip	7.8	2	27.5	7
Disulfide bond formation protein	Thioredoxin	8.4	15	28.1	8
ASV integrase	Ribonuclease H-like	7.8	19	28.6	5
Endoglucanase C	Galactose binding	11.6	4	33.8	6
TATA-box-binding protein	TATA-box-binding protein-like	10.3	7	40.0	8

Protein2: Properties & Quantitation

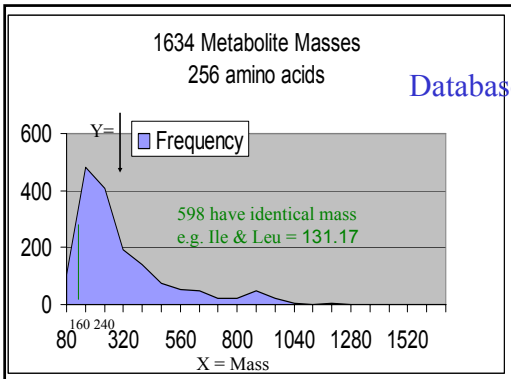
- Separation of proteins & peptides
- Protein localization & complexes
- Peptide identification (MS/MS)
  - Database searching & sequencing.
- Protein quantitation
  - Absolute & relative
- Protein modifications & crosslinking
- Protein - metabolite quantitation

51

Challenges for accurately measuring metabolites

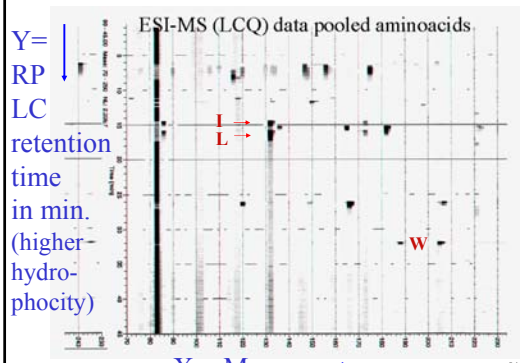
- Rapid kinetics
- Rapid changes during isolation
- Idiosyncratic detection methods:
  enzyme-linked, GC, LC, NMR
  (albeit fewer molecular types than RNA& protein)

52

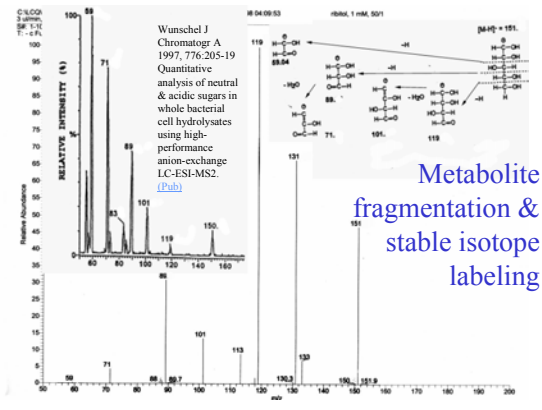


Karp et al. (1998) *NAR* 26:50. EcoCyc; Selkov, et al. (1997) *NAR* 25:37. WIT
Ogata et al. (1998) Biosystems 47:119-128 KEGG

53



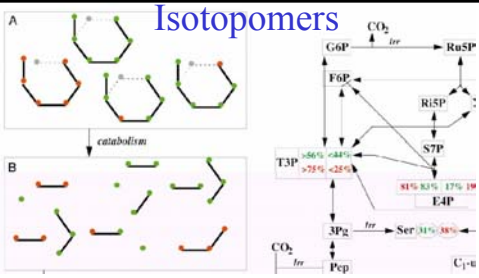
54



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)

Metabolite fragmentation & stable isotope labeling

Isotopomers



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."

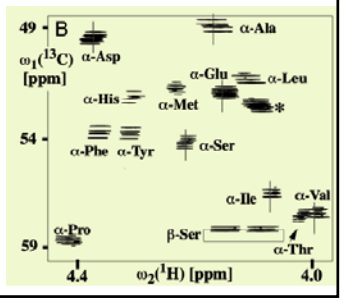
MetaFoR: Metabolic Flux Ratios

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

Sauer J et al. Bacteriol 1999;181:6679-88 (Pub)

Szyperski et al 1999
Metab. Eng. 1:189.

Dauner et al. 2001
Biotec Bioeng 76:144



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

Table 2. Internal metabolite concentrations[a]

Strain	Growth rate (h⁻¹)	G6P	F6P	ATP	Pyr	ADP	AMP	ATP/ ADP
Wild type	0.31	2.05 ±0.11	0.40 ±0.03	2.80 ±0.32	3.39 ±0.54	0.42 ±0.11	0.20 ±0.02	6.67
hxk1	0.31	2.22 ±0.10	0.39 ±0.03	2.71 ±0.07	4.35 ±0.40	0.43 ±0.06	0.20 ±0.03	6.30
pfk26Δ	0.30	2.24 ±0.06	0.57 ±0.04	2.45 ±0.13	5.45 ±1.01	0.71±0.10	0.17 ±0.00	3.45
pfk27Δ	0.30	2.71 ±0.02	0.67 ±0.02	2.35 ±0.02	4.56 ±0.38	0.67 ±0.07	0.15 ±0.02	3.51
cox5aΔ	0.30	1.81 ±0.08	0.53 ±0.09	1.70 ±0.17	5.51 ±1.91	0.76 ±0.14	0.16 ±0.01	2.24
pet191Δ	0.31	1.88 ±0.11	0.46 ±0.07	1.75 ±0.06	5.97 ±1.04	0.84 ±0.09	0.15 ±0.03	2.08

[a]Given in mM, assuming that 3.75 ml cytosol is equivalent to 1 g of total protein. Each metabolite concentration is the average of the values obtained from three independently grown cultures (± s.e.m.).

-40C MeOH> 80C EtOH > Cobas Enzymatic BioAutoanalyser & Quantitative ^{1}H NMR 0 to 4.4 ppm (1300 measures)

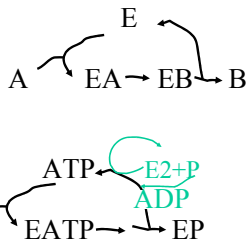
Raamsdonk et al. 2001 Nature Biotech 19:45.

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

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

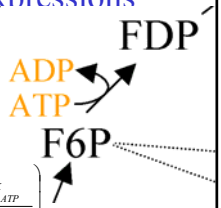
As P approaches 0:

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

61

Enzyme Kinetic Expressions

Phosphofructokinase



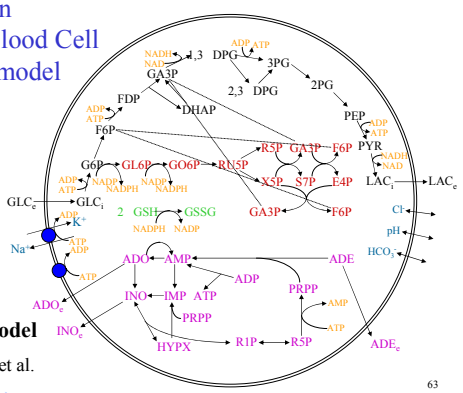
$$v_{PFK} = \frac{v_{mx}^{PFK}}{N_{PFK}} \left(\frac{F6P / K_{F6P}^{PFK}}{1 + F6P / K_{F6P}^{PFK}} \right) \left(\frac{Mg \bullet ATP / K_{Mg \bullet ATP}^{PFK}}{1 + Mg \bullet ATP / K_{Mg \bullet ATP}^{PFK}} \right)$$

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

Allosteric kinetic parameters for AMP, etc.

62

Human Red Blood Cell ODE model



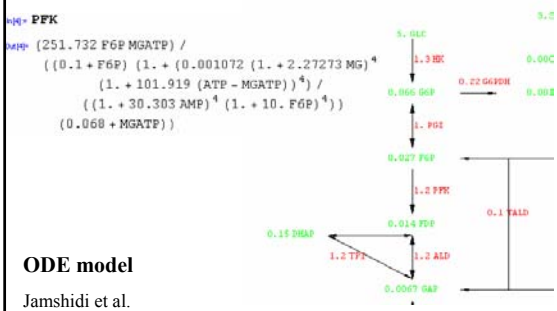
ODE model

Jamshidi et al.

2000 (Pub)

63

Red Blood Cell in Mathematica



ODE model

Jamshidi et al.

2000 (Pub)

64

Protein2: Properties & Quantitation

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

65