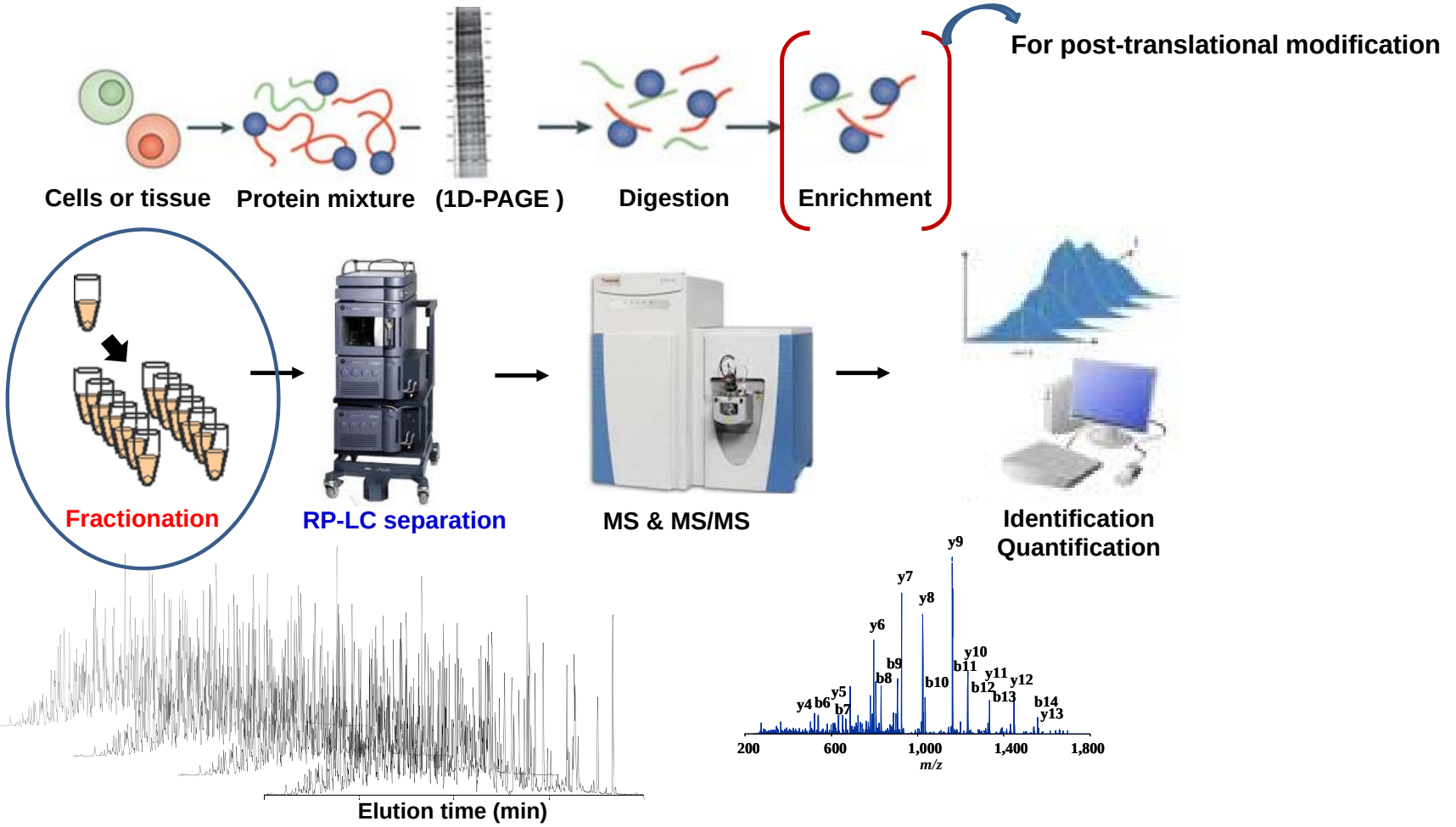
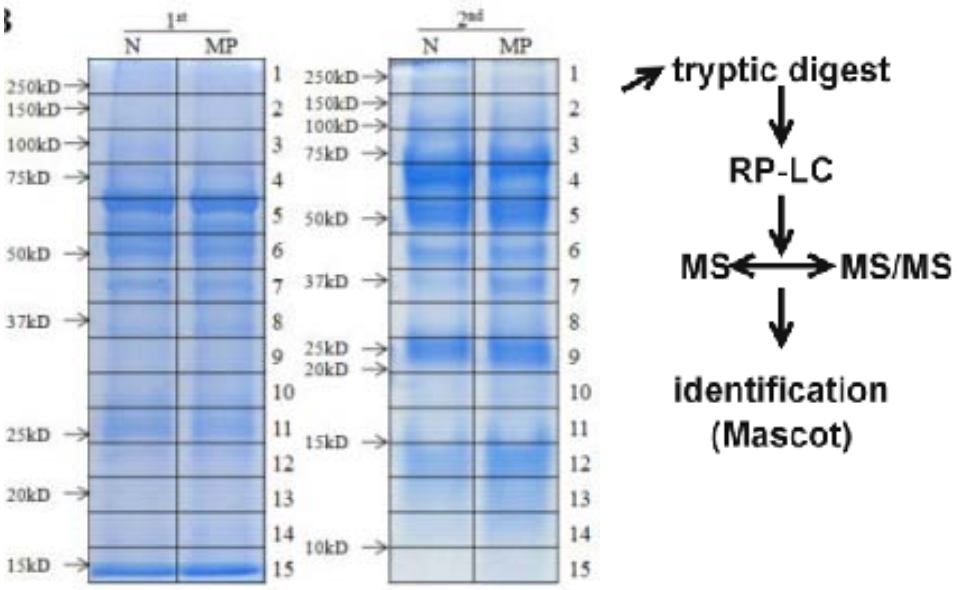
Bottom-up proteomics workflow



1D-PAGE / in-gel digestion
(molecular weight-based fractionation)



Pros

• Protein information (Mw)

• Easy to perform

• Biologist-friendly

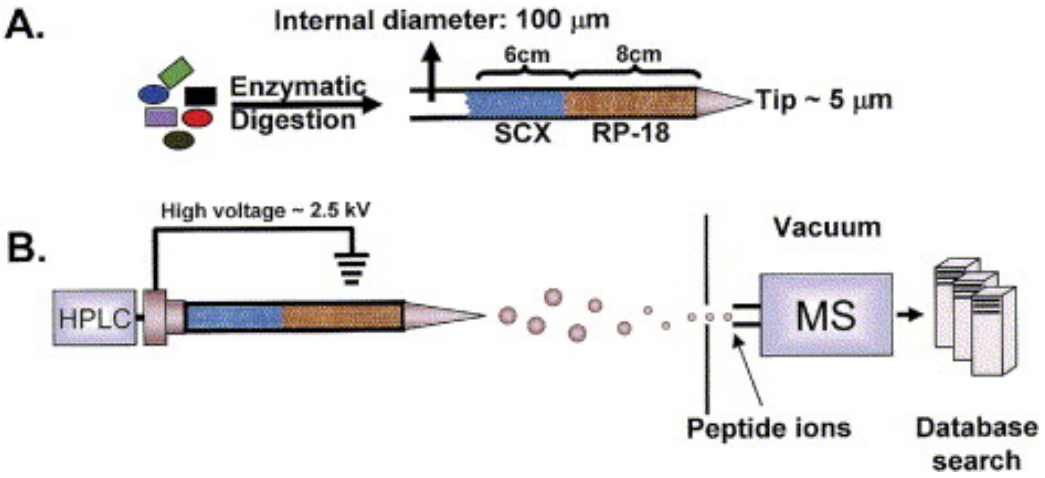
Cons

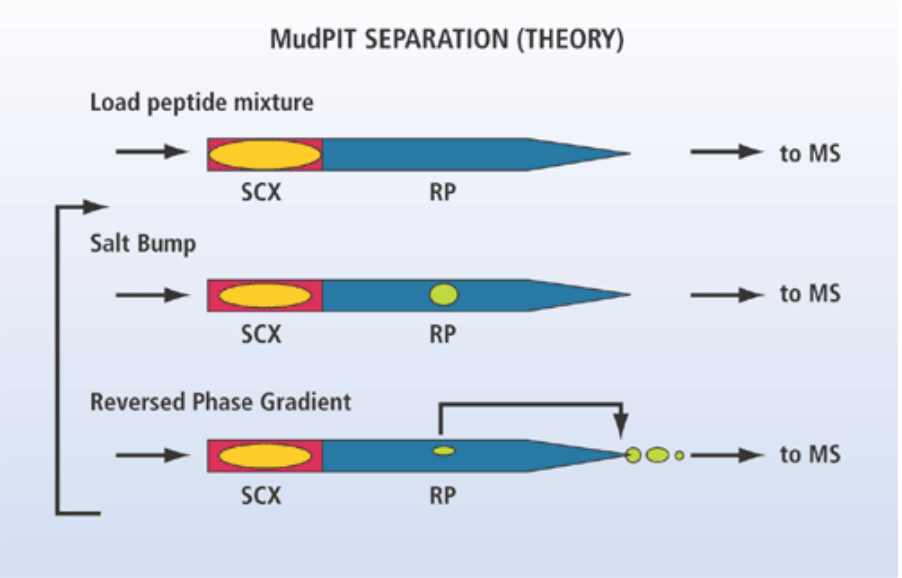
• Large amount of sample

• Low peptide recovery (in-gel digestion)

• Low protein coverage

MudPIT
(Multidimensional Protein Identification Technology)
: era of shotgun proteomics





Nat Biotechnol. 2001 Mar;19(3):242-7

Nature Biotechnology **19**, 242 - 247 (2001)
doi:10.1038/85686

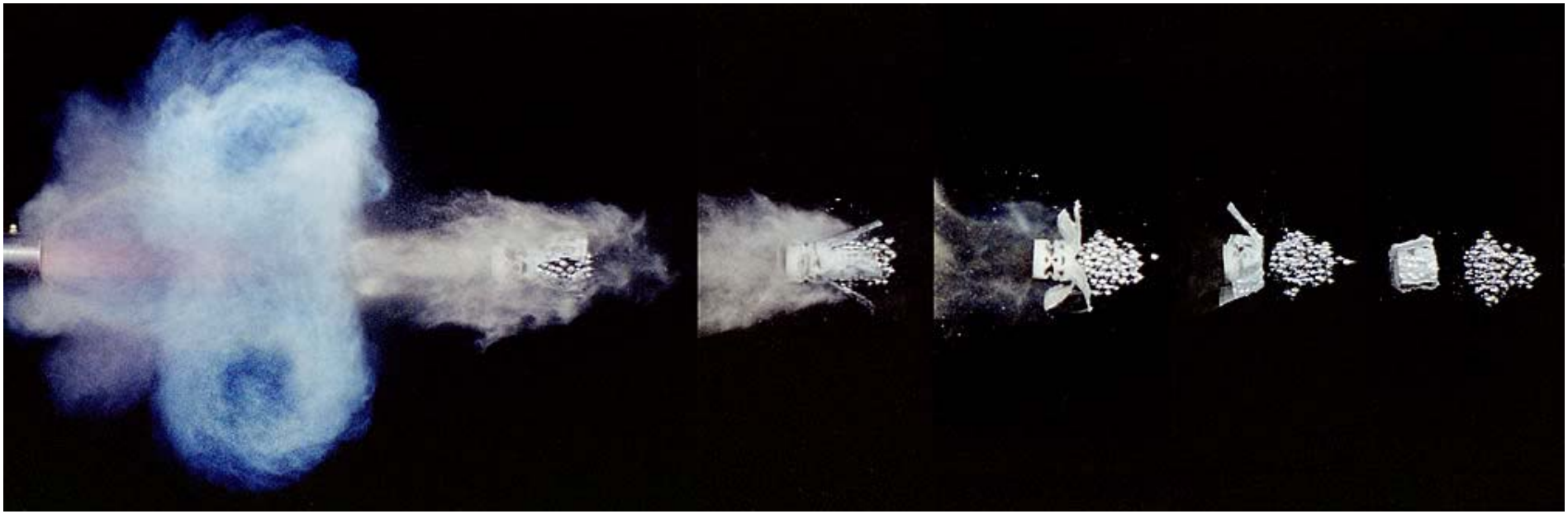
Large-scale analysis of the yeast proteome by multidimensional protein identification technology

Michael P. Washburn[1,3], Dirk Wolters[1,3] & John R. Yates, III[1,2]

We describe a largely unbiased method for rapid and large-scale proteome analysis by multidimensional liquid chromatography, tandem mass spectrometry, and database searching by the SEQUEST algorithm, named multidimensional protein identification technology (MudPIT). MudPIT was applied to the proteome of the *Saccharomyces cerevisiae* strain BJ5460 grown to mid-log phase and yielded the largest proteome analysis to date. A total of 1,484 proteins were detected and identified. Categorization of these hits demonstrated the ability of this technology to detect and identify proteins rarely seen in proteome analysis, including low-abundance proteins like transcription factors and protein kinases. Furthermore, we identified 131 proteins with three or more predicted transmembrane domains, which allowed us to map the soluble domains of many of the integral membrane proteins. MudPIT is useful for proteome analysis and may be specifically applied to integral membrane proteins to obtain detailed biochemical information on this unwieldy class of proteins.

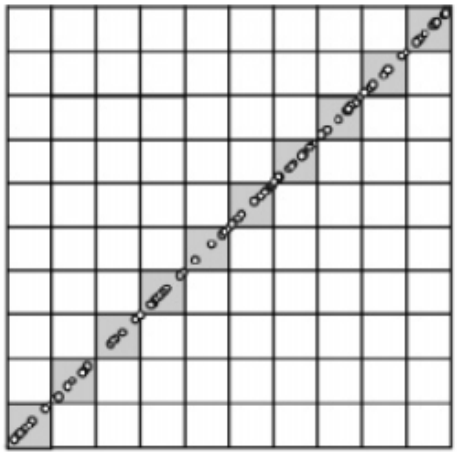
Shotgun ?? Proteomics ??

→ Bottom-up proteomics



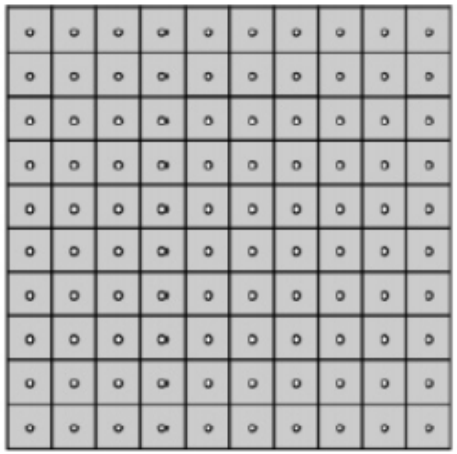
Orthogonality Concept

A



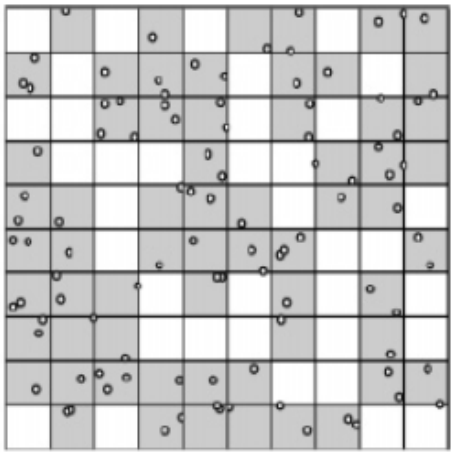
0% orthogonality

B



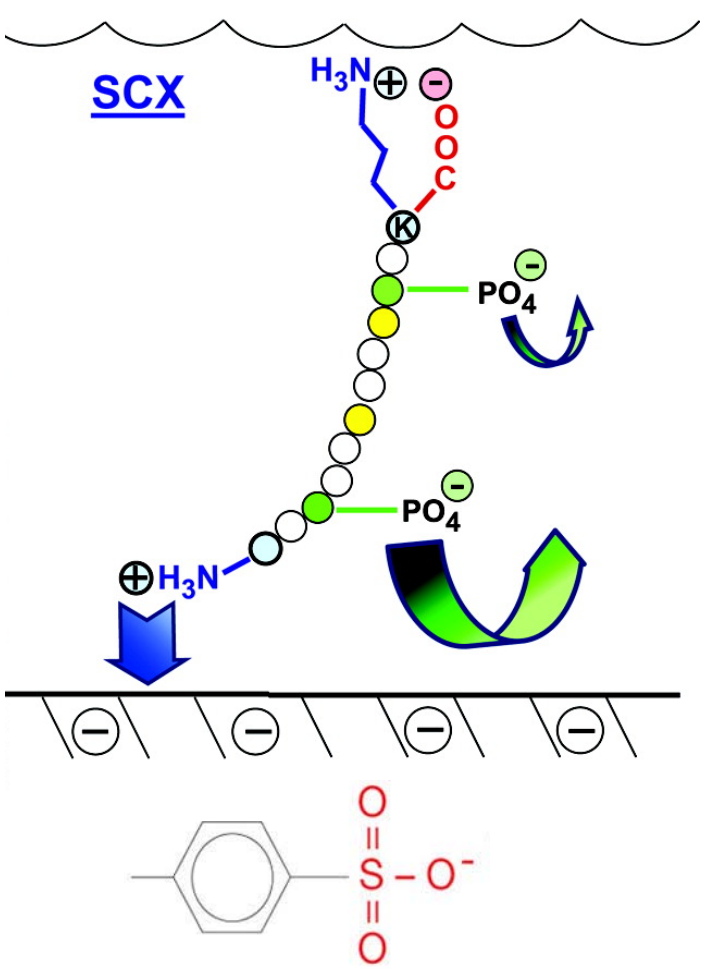
Hypothetical

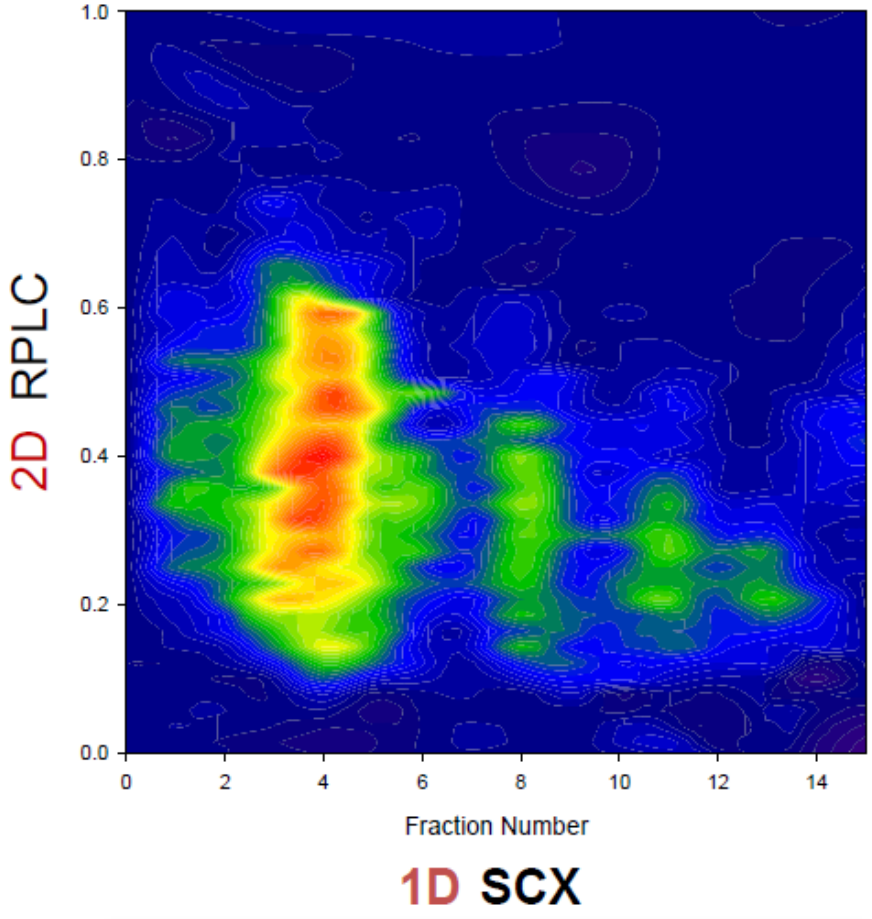
C



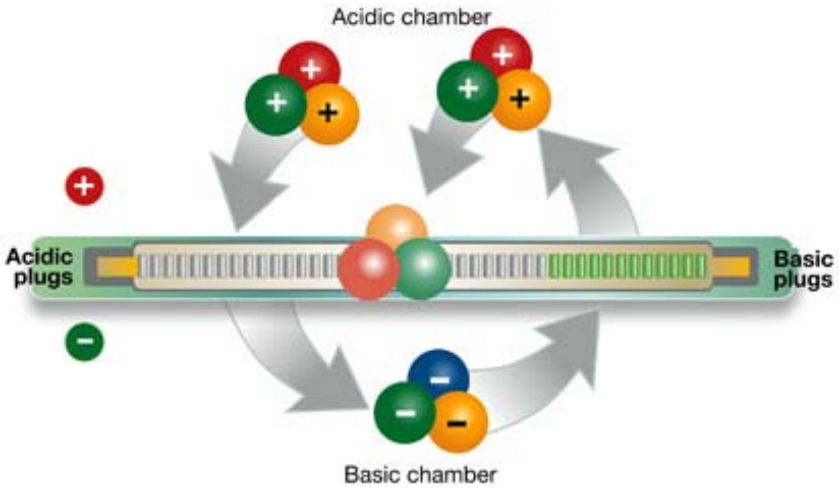
100% orthogonality

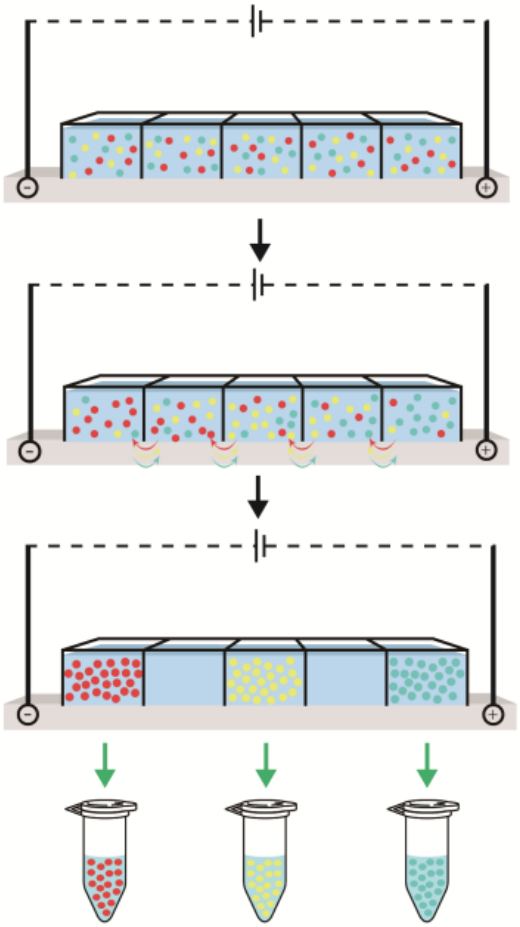
SCX (strong cation exchange) - LC





Isoelectric focusing (IEF): Off-gel fractionation





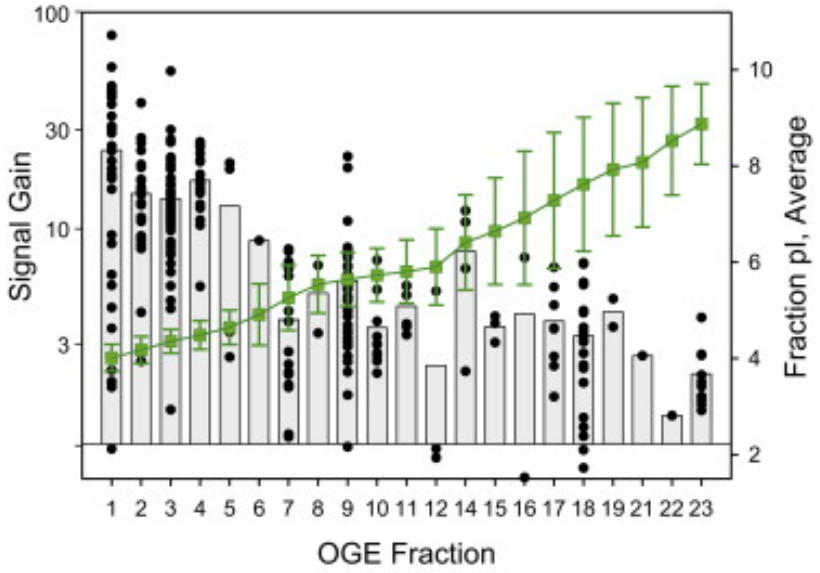
Fractionation by isoelectric point

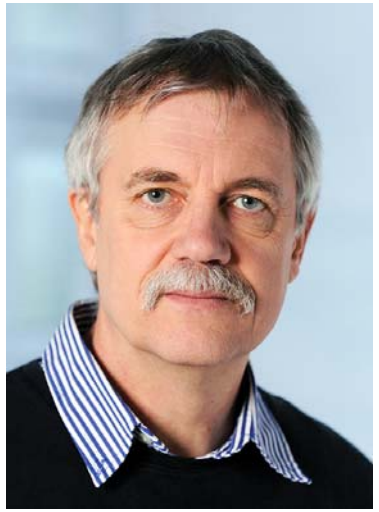
Resource

Cell

Full Dynamic Range Proteome Analysis of *S. cerevisiae* by Targeted Proteomics

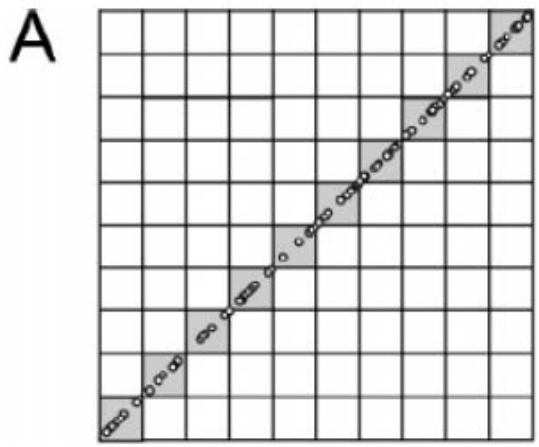
Paola Picotti,[1] Bernd Bodenmiller,[1] Lukas N. Mueller,[1] Bruno Domon,[1] and Ruedi Aebersold[1,2,3,4,*]
[1]Institute of Molecular Systems Biology, ETH Zurich, Zurich CH 8093, Switzerland
[2]Competence Center for Systems Physiology and Metabolic Diseases, Zurich CH 8093, Switzerland
[3]Institute for Systems Biology, Seattle, WA 98103, USA
[4]Faculty of Science, University of Zurich, Zurich CH 8057, Switzerland
*Correspondence: aebersold@imsb.biol.ethz.ch
DOI 10.1016/j.cell.2009.05.051



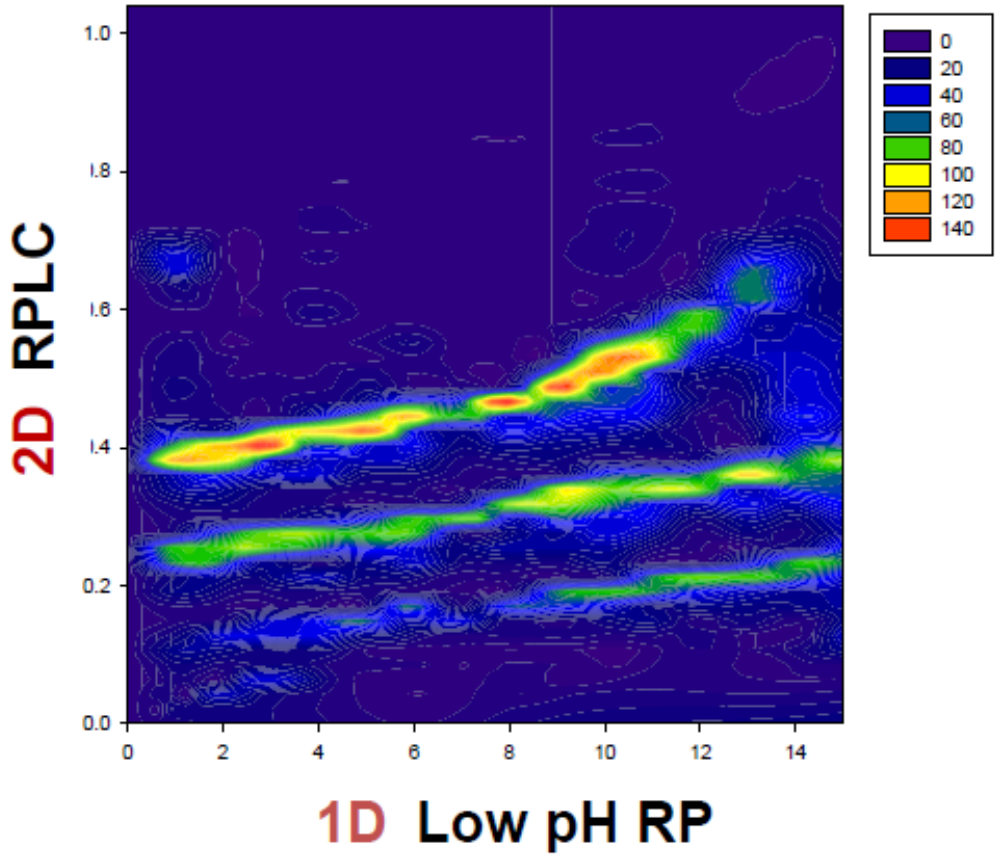


Moving forward to “clean & orthogonal” fractionation approaches

Reverse phase LC: low pH

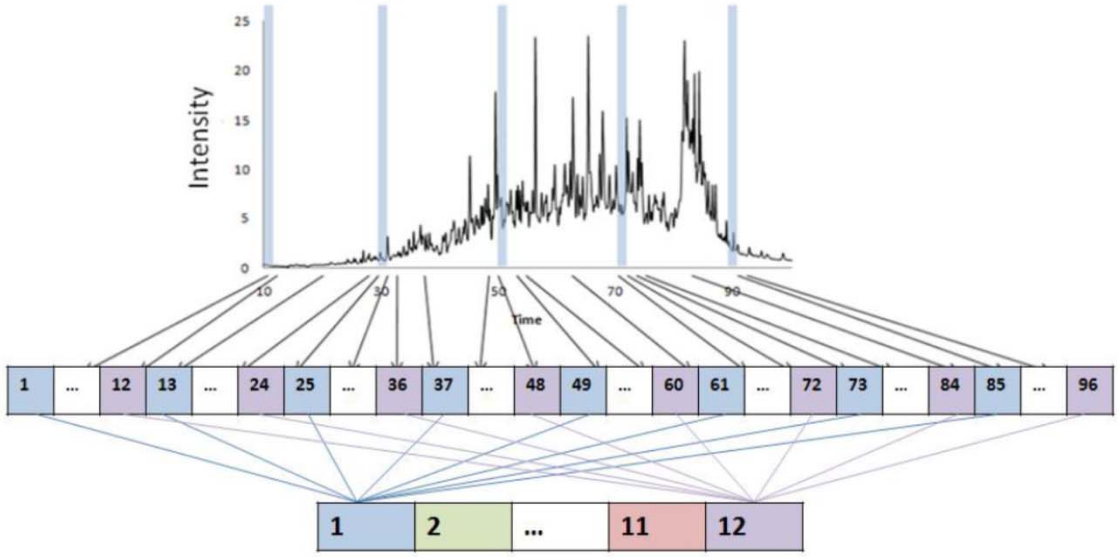


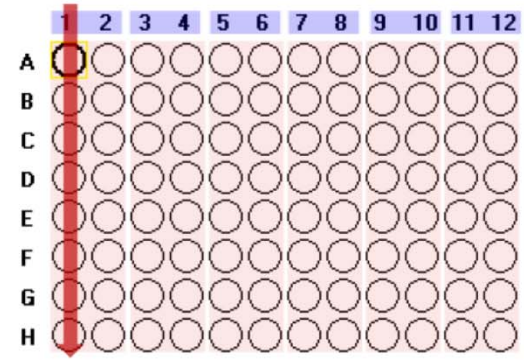
0% orthogonality



Reverse phase LC: high pH

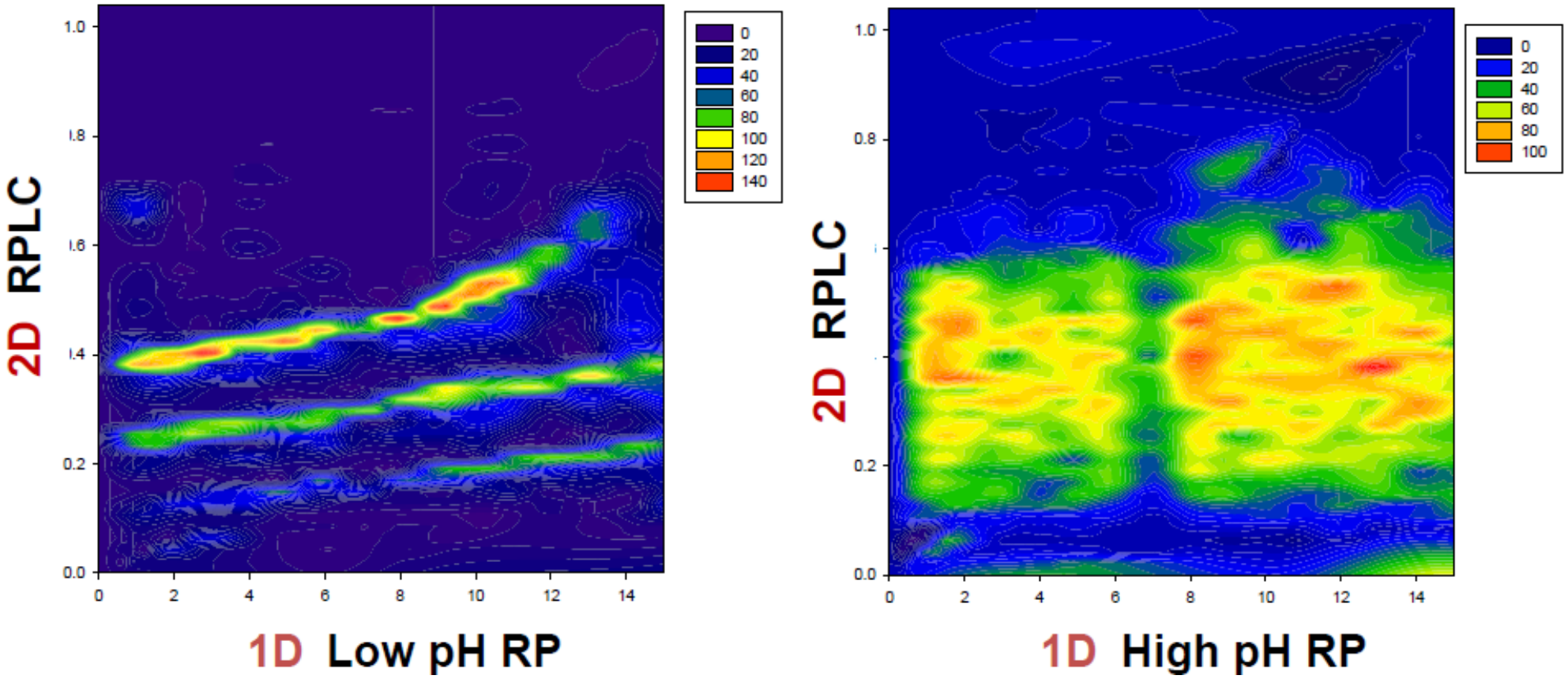
Concatenated high-pH RPLC-RPLC separation



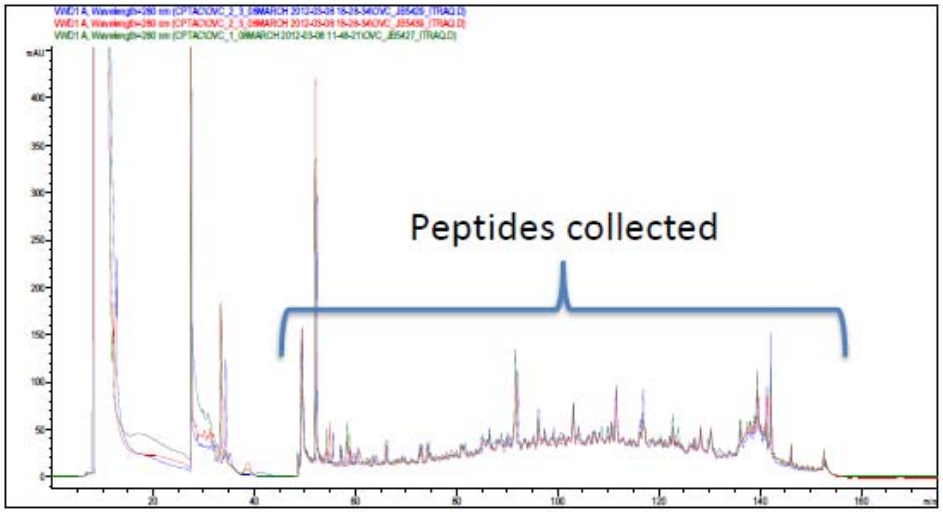


Fraction 1

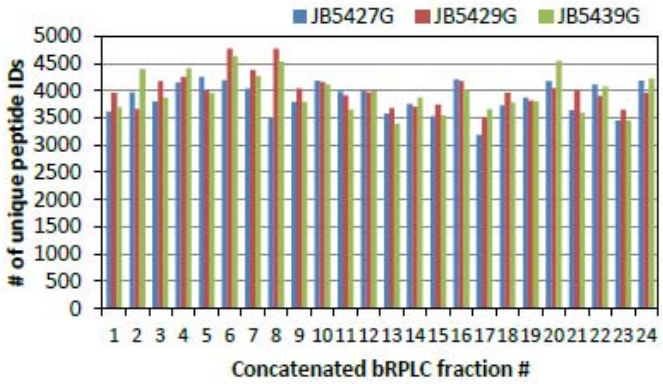
Power of off-line concatenation



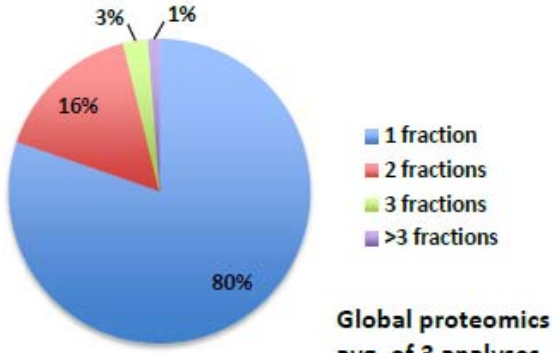
Effectiveness of bRPLC fractionation with concatenation



Distribution of peptide identifications in each concatenated bRPLC fraction

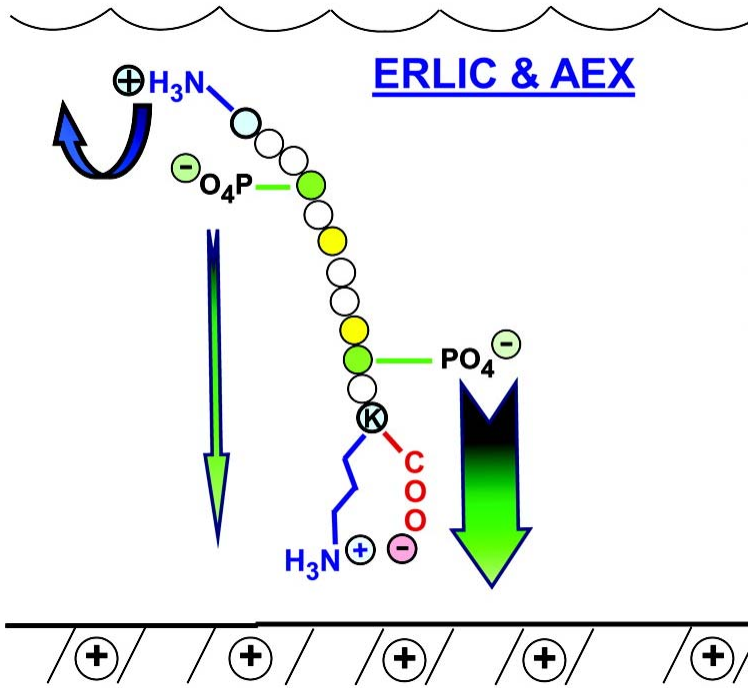


Uniqueness of peptides in the concatenated bRPLC fractions



Global proteomics avg. of 3 analyses

ERLIC (electrostatic repulsion hydrophilic interaction chromatography)



Mix mode of anion exchange and hydrophilic interaction chromatography

ERLIC provides:

• More peptide identifications than SCX

65% more unique peptide identifications

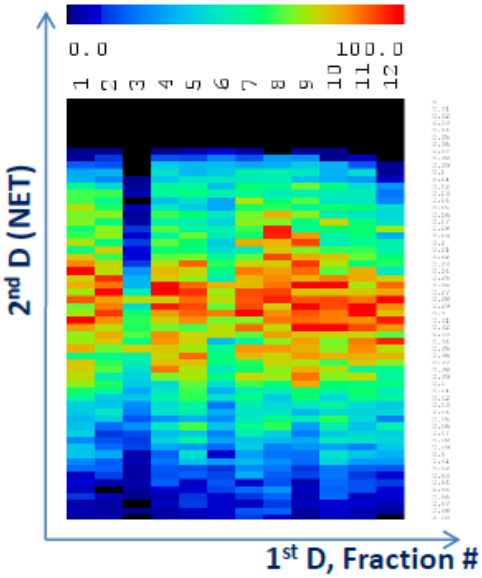
40% more protein identifications

• Clean fractionated sample

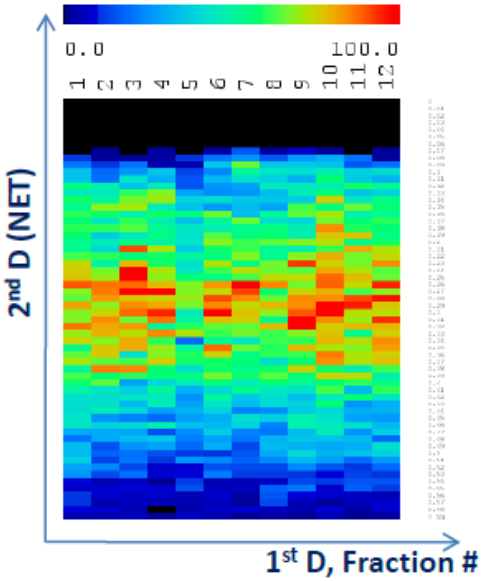
Sol A (90% ACN/0.1% acetic acid, pH 3.6)

Sol B (30% ACN/0.1% formic acid, pH 3.0)

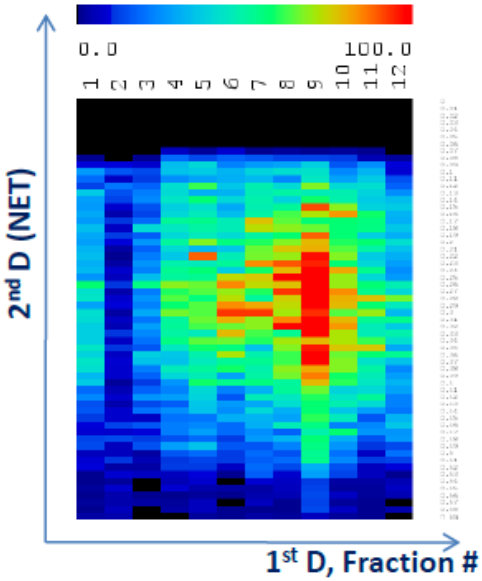
Clean fractionation methods comparison



pH 10 RP



pH 7.5 RP

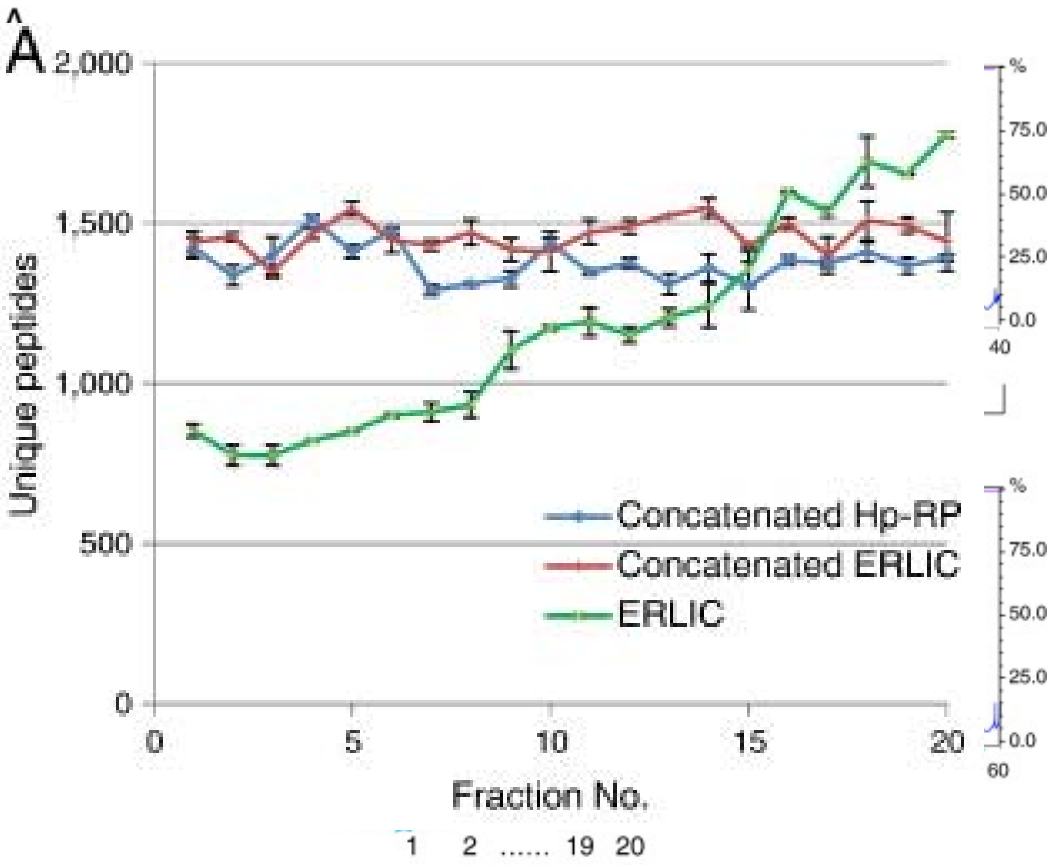


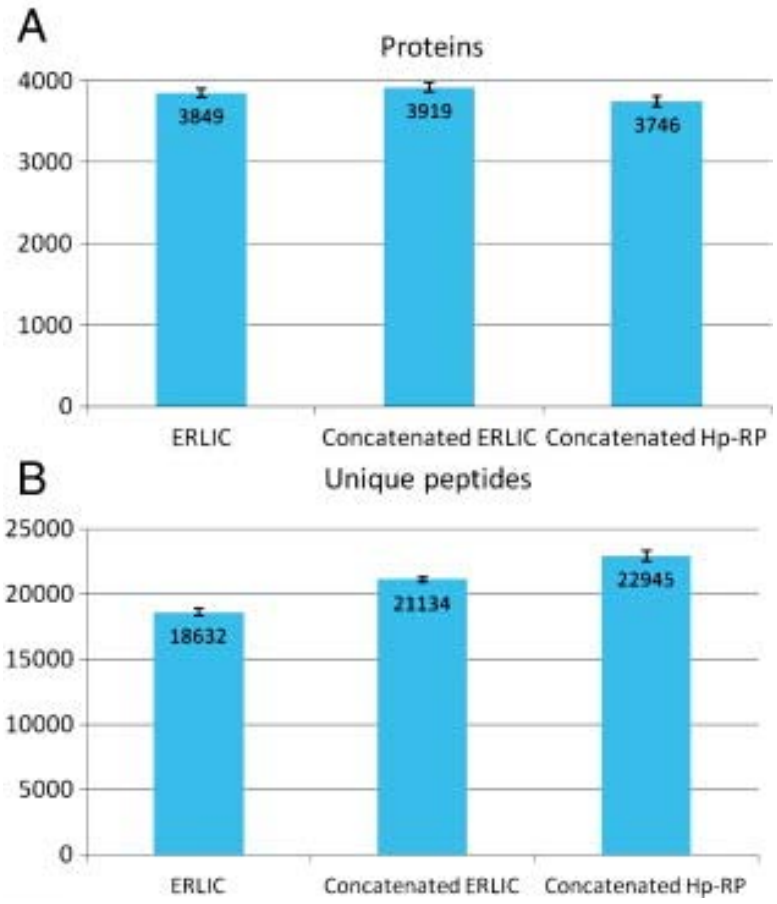
ERLIC

12 analysis

	pH 10 RP	pH 7.5 RP	ERLIC
phosphopeptides	21529	20901	17020
phosphoproteins	4519	4441	4132

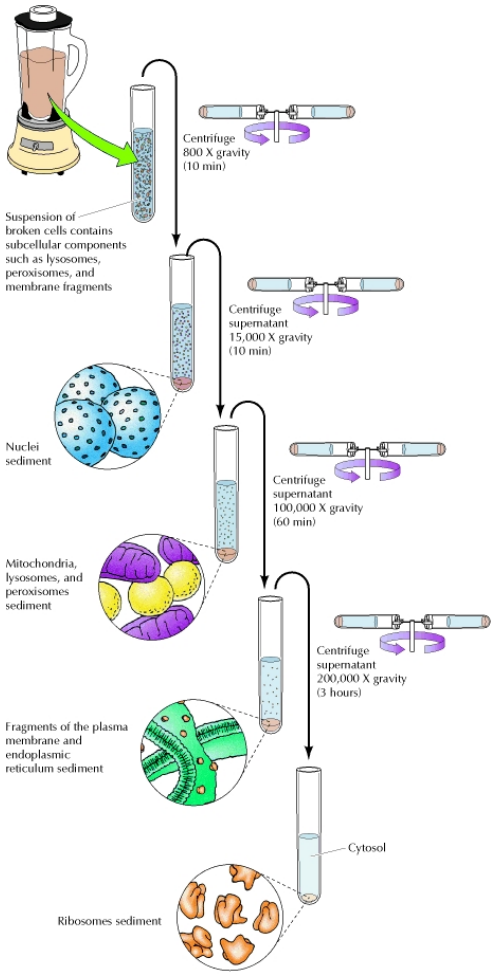
Off-line concatenation of ERLIC separation

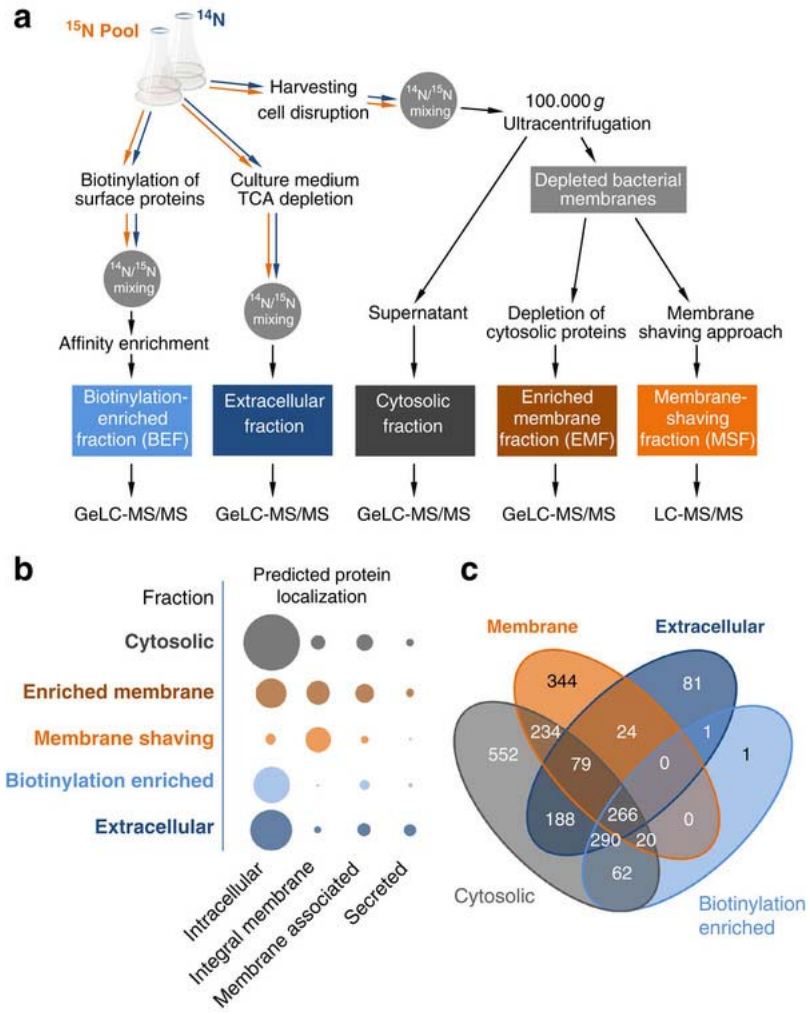




J Proteomics. 2013 Apr 26;82:254-62.

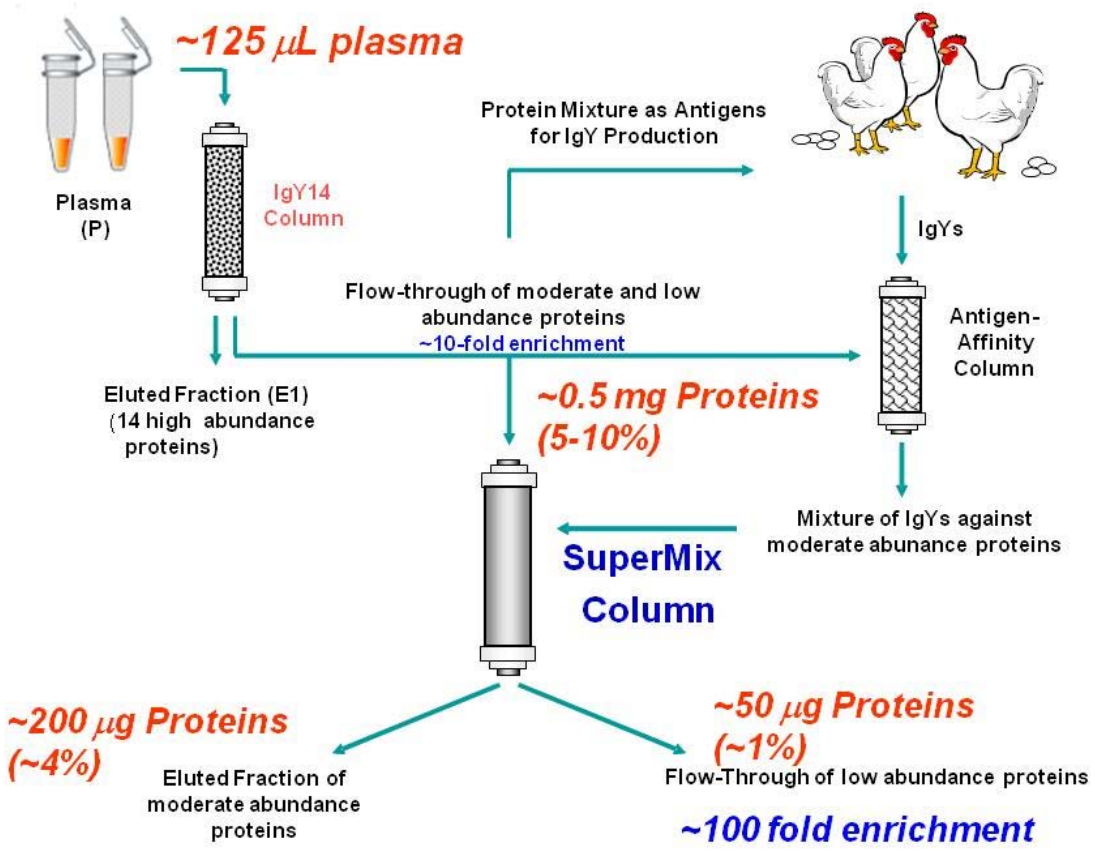
What else to decrease sample complexity?

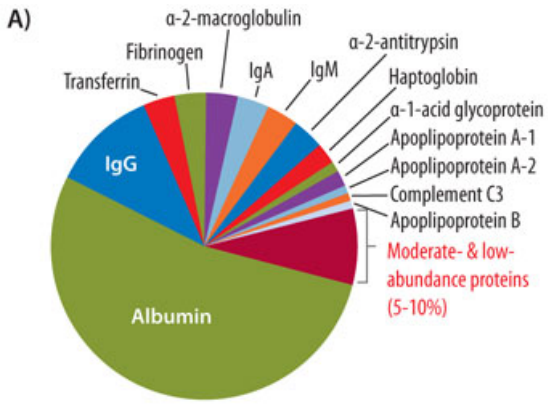




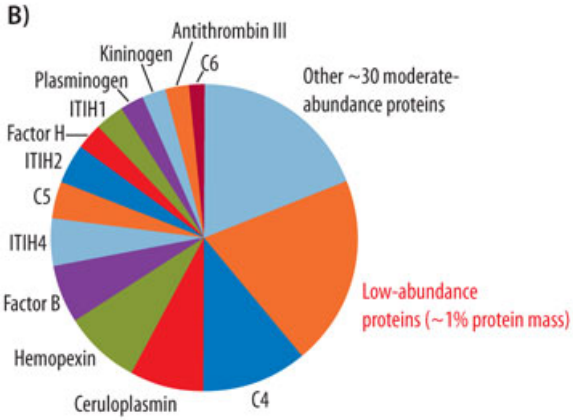
What about depleting of abundant proteins?

IgY14-SuperMix tandem immuno-depletion





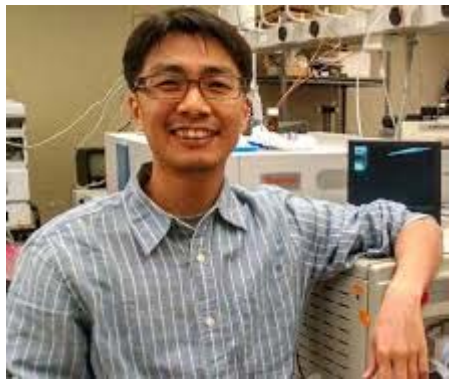
IgY14 column
(Mixture of antibodies against 14 abundant proteins)



SuperMix column
(Mixture of antibodies against 50 or more moderate abundant proteins)

Qian WJ et al. MCP 2008

How many proteins now, 2015?



ARTICLE

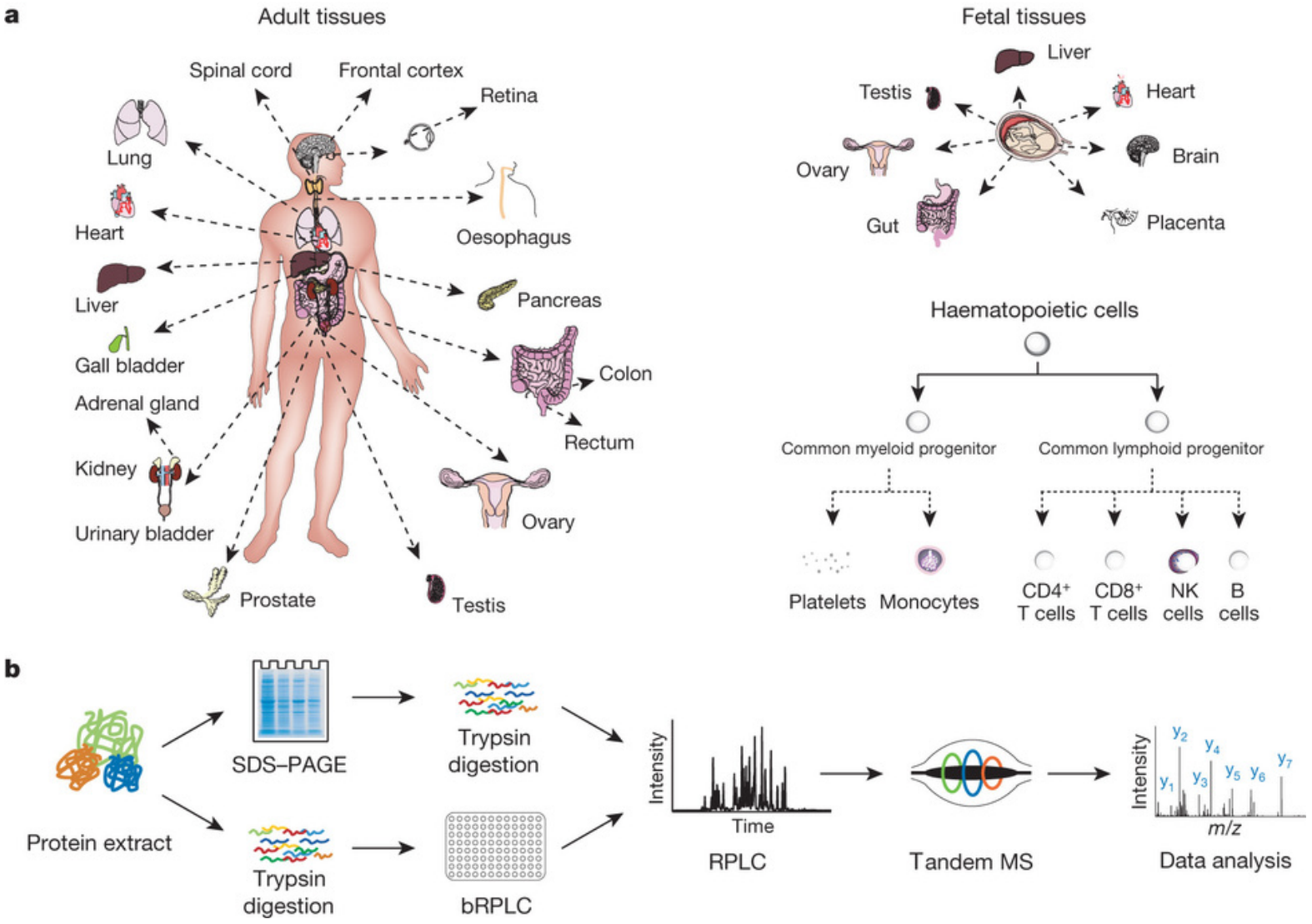
doi:10.1038/nature13302

A draft map of the human proteome

Min-Sik Kim[1,2], Sneha M. Pinto[3], Derese Getnet[1,4], Raja Sekhar Nirujogi[3], Srikanth S. Manda[3], Raghothama Chaerkady[1,2], Anil K. Madugundu[3], Dhanashree S. Kelkar[3], Ruth Isserlin[5], Shobhit Jain[5], Joji K. Thomas[3], Babylakshmi Muthusamy[3], Pamela Leal-Rojas[1,6], Praveen Kumar[3], Nandini A. Sahasrabuddhe[3], Lavanya Balakrishnan[3], Jayshree Advani[3], Bijesh George[3], Santosh Renuse[3], Lakshmi Dhevi N. Selvan[3], Arun H. Patil[3], Vishalakshi Nanjappa[3], Aneesha Radhakrishnan[3], Samarjeet Prasad[1], Tejaswini Subbannayya[3], Rajesh Raju[3], Manish Kumar[3], Sreelakshmi K. Sreenivasamurthy[3], Arivusudar Marimuthu[3], Gajanan J. Sathe[3], Sandip Chavan[3], Keshava K. Datta[3], Yashwanth Subbannayya[3], Apeksha Sahu[3], Soujanya D. Yelamanchi[3], Savita Jayaram[3], Pavithra Rajagopalan[3], Jyoti Sharma[3], Krishna R. Murthy[3], Nazia Syed[3], Renu Goel[3], Aafaque A. Khan[3], Sartaj Ahmad[3], Gourav Dey[3], Keshav Mudgal[7], Aditi Chatterjee[3], Tai-Chung Huang[1], Jun Zhong[1], Xinyan Wu[1,2], Patrick G. Shaw[1], Donald Freed[1], Muhammad S. Zahari[2], Kanchan K. Mukherjee[8], Subramanian Shankar[9], Anita Mahadevan[10,11], Henry Lam[12], Christopher J. Mitchell[1], Susarla Krishna Shankar[10,11], Parthasarathy Satishchandra[13], John T. Schroeder[14], Ravi Sirdeshmukh[3], Anirban Maitra[15,16], Steven D. Leach[1,17], Charles G. Drake[16,18], Marc K. Halushka[15], T. S. Keshava Prasad[3], Ralph H. Hruban[15,16], Candace L. Kerr[19]†, Gary D. Bader[5], Christine A. Iacobuzio-Donahue[15,16,17], Harsha Gowda[3] & Akhilesh Pandey[1,2,3,4,15,16,20]

The availability of human genome sequence has transformed biomedical research over the past decade. However, an equivalent map for the human proteome with direct measurements of proteins and peptides does not exist yet. Here we present a draft map of the human proteome using high-resolution Fourier-transform mass spectrometry. In-depth proteomic profiling of 30 histologically normal human samples, including 17 adult tissues, 7 fetal tissues and 6 purified primary haematopoietic cells, resulted in identification of proteins encoded by 17,294 genes accounting for approximately 84% of the total annotated protein-coding genes in humans. A unique and comprehensive strategy for proteogenomic analysis enabled us to discover a number of novel protein-coding regions, which includes translated pseudogenes, non-coding RNAs and upstream open reading frames. This large human proteome catalogue (available as an interactive web-based resource at http://www.humanproteomemap.org) will complement available human genome and transcriptome data to accelerate biomedical research in health and disease.

	Number of genes involved or annotation confirmed /added/altered
Genes whose products were detected	17,294
Confirmed exons	223,385
Confirmed N termini	4,105
Confirmed exon–exon junctions	66,947
Signal peptide cleavage site	329
Confirmation of annotated sites	128
Unannotated cleavage sites	201
Novel protein-coding regions	193
Pseudogenes	140
Non-coding RNAs	9
Upstream ORFs	29
Other ORFs	15
Novel coding regions/exons	106
Novel N termini	198
Gene/protein extensions	70
N-terminal extension	58
C-terminal extension	12
Exon extension	40


a
Adult tissues
Spinal cord
Frontal cortex
Retina
Lung
Heart
Oesophagus
Liver
Pancreas
Gall bladder
Colon
Adrenal gland
Rectum
Kidney
Ovary
Urinary bladder
Prostate
Testis
Fetal tissues
Liver
Testis
Heart
Ovary
Brain
Gut
Placenta
Haematopoietic cells
Common myeloid progenitor
Common lymphoid progenitor
Platelets
Monocytes
CD4+ T cells
CD8+ T cells
NK cells
B cells
b
Protein extract
SDS–PAGE
Trypsin digestion
Trypsin digestion
bRPLC
Intensity
Time
RPLC
Tandem MS
Intensity
m/z
Data analysis