



ELSEVIER

## Proteomics/genomics and signaling in lymphocytes

Bernd Wollscheid<sup>1</sup>, Julian D Watts<sup>2</sup> and Ruedi Aebersold<sup>3</sup>

Recent technological advances in genomics, proteomics and bioinformatics have offered new insights into the molecular mechanisms that underlie lymphocyte signaling and function, and the development of new tools in these areas has opened up new avenues for biological investigation. By adding a quantitative dimension to lymphocyte proteome profiling, molecular machines and spatiotemporal regulatory processes can now be analyzed using such discovery-driven approaches. Biologists employing genomic and proteomic tools are gathering data at increasing speed and their struggle to extract maximal biological information is helped by new software tools that enable the detailed comparison of multiple datasets.

### Addresses

Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103, USA

<sup>1</sup>e-mail: bwollscheid@systemsbiology.org

<sup>2</sup>e-mail: jwatts@systemsbiology.org

<sup>3</sup>e-mail: raebersold@systemsbiology.org

### Current Opinion in Immunology 2004, 16:1–8

This review comes from a themed issue on Lymphocyte activation  
Edited by John C Cambier and Arthur Weiss

0952-7915/\$ – see front matter

© 2004 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.coi.2004.03.016

### Abbreviations

**ICAT** isotope coded affinity tag  
**RNAi** RNA interference  
**SILAC** stable isotope labeling with amino acids in cell culture  
**TAP** tandem affinity purification

### Introduction: the coming of a ‘new biology’

Over the past few decades, reductionist approaches to the study of biological systems have contributed greatly to our understanding of many basic principles of cellular biology. In lymphocyte biology, this has typically involved the study of one gene, receptor or protein at a time, to elucidate their role in lymphocyte biology and signaling pathways. Recent advances in experimental hardware and methodology have facilitated the emergence of the fields of genomics, proteomics and, through the data these generate, the field of bioinformatics. Furthermore, the recent and ongoing completion of whole genome sequences, especially those for human and mouse, have enabled the design of new types of experiments, whereby any previously unknown protein or gene, gene cluster or complex set of proteins can, poten-

tially, be identified via the integration of bioinformatics tools into the experimental design [1,2].

For the immunologist, such approaches promise the determination of all expressed genes and/or proteins in a lymphocyte system of choice, as well as the description of the interactions between all of the components of a targeted signaling network, within the confines of a single experiment. In turn, when integrated with existing research strategies, the ability to quantitatively measure large numbers of genes or gene products simultaneously, offers the opportunity to ask new questions about lymphocyte biology and signaling. These questions would address how and where different stimuli (and the signaling networks they impinge upon) operate within the cell, and how they interact with and potentially modulate the output of other signaling networks that are active in the same cell (i.e. the study of dynamic gene regulatory and signaling networks).

### Genomics

Even before the human and mouse genomic sequences were completed, the development of chip-based DNA microarray analyses allowed immunologists to look at global gene expression and induction within lymphocytes, either cultured *in vitro* or isolated *ex vivo* from suitable mouse model systems or human subjects. Such studies have provided new and invaluable insights into several long-standing and fundamental questions in immunology, enabling not only the identification of new key genetic signals, but also the revelation of functional gene clusters, activated during T- and B-cell differentiation, and regulating cell division, apoptosis and cytoskeletal reorganization [3–5]. Similarly, genetic mapping on a genome-wide scale has provided evidence that genes encoding for proteins that perform broadly related functions might also be physically clustered together in the genome sequence. In lymphocytes, for example, genomic regions surrounding the MHC and the antigen receptor genes for T, B and NK cells are rich in genes that are expressed in immune cells.

Gene microarray technology has also allowed for diagnostic classification and prediction of a sample, according to the gene expression profile. For example, the combination of gene expression profiling and bioinformatics was used to refine tumor classification, resulting in a more accurate prediction of disease course [6]. Other applications of gene expression profiling were used to diagnose clinically distinct subgroups of diffuse large-B-cell lymphoma [7], as well as distinguish between diffuse

## 2 Lymphocyte activation

large-B-cell lymphoma patient populations with high and low survival rates [8,9]. Similarly, the application of gene microarray technology to B-cell memory formation [10] and T-cell activation [11] have shown great promise for defining such processes, in terms of induced changes in gene expression profiles. Additionally, profiling of lymphomas using Lymphochip cDNA microarrays has been applied to the elucidation of molecular predictors of survival [12], where the gene expression signature for proliferation was shown to be a quantitative integrator of oncogenic events that predict survival in mantle cell lymphoma [13•].

Other recent genomic experiments have facilitated the identification of genes that are targeted by transcriptional repressors. For example, Blimp-1 is a transcriptional repressor that drives the terminal differentiation of B cells to plasma cells. DNA-microarray analysis of changes in B-cell gene expression induced by Blimp-1 showed that this differentiation is accomplished by extinguishing the mature B-cell gene expression program [14•]. Finally, a recently developed method for genome-wide location analysis, thus far only applied to the eukaryotic yeast *Saccharomyces cerevisiae*, combined chromatin immunoprecipitation with DNA microarrays to determine where on the DNA a specific protein was located [12,15]. The application of similar methodologies to immune cells may thus provide the foundation for a map of the many, overlapping regulatory networks that control lymphocyte gene expression.

### From genomics to proteomics

Array and other genomic technologies can tell us which genes are expressed in any cell or tissue, which genes are upregulated or downregulated in response to stimulus or environmental change, and the approximate absolute abundance of each message through serial analysis of gene expression (SAGE). Though powerful, these technologies are limited; they cannot tell us where expressed proteins are localized, identify the other molecules with which they interact, determine their state of post-translational modification, nor tell us how these properties will change in response to different stimuli. Indeed, it has been shown that there is not necessarily a good correlation between message level and protein expression level within cells [16,17]; as cellular signaling is regulated by alterations in precisely these properties of expressed proteins, it thus becomes necessary to study lymphocyte proteins to fully elucidate lymphocyte signaling mechanisms and function.

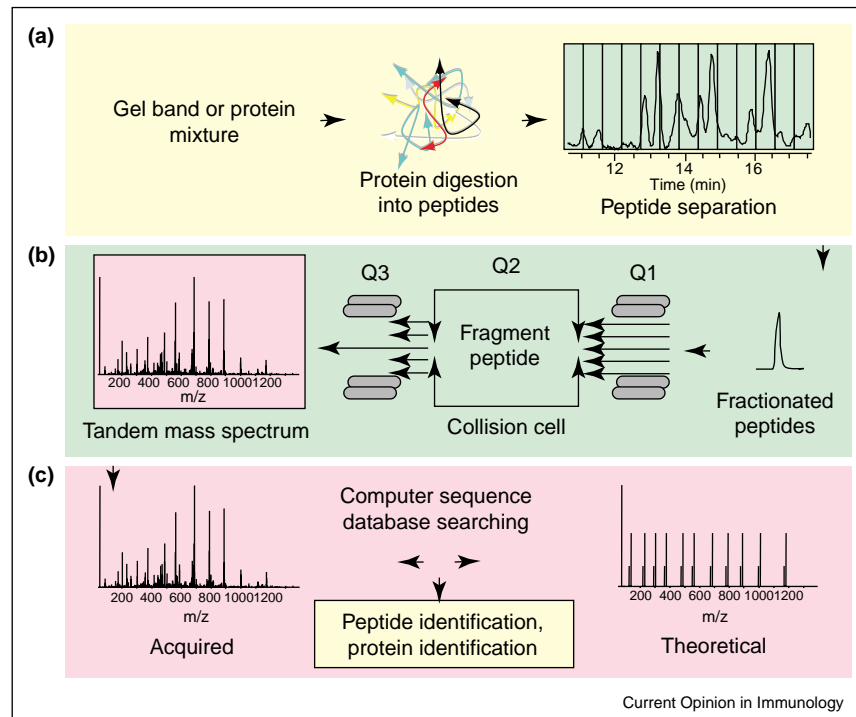
More recently than genomics, MS-based proteomic approaches have emerged, yielding the capability of identifying small amounts of proteins from increasingly complex mixtures [18]. Unlike genomics, however, at present there is no one single or superior proteomic platform. Rather, the best analytical platform depends on the spe-

cific biological question. Generally, proteins will be isolated from cells and then subjected to complexity reduction, either via chromatographic or electrophoretic fractionation, or via enrichment of a desired component, such as membrane proteins, a particular sub-cellular compartment, or specific protein complex isolation via affinity purification (for example, immunoprecipitation, DNA-binding and so on). Protein identification is achieved via proteolysis and analysis of resultant peptides in a mass spectrometer, followed by database-searching to identify the original protein(s) from which the peptides originated. The method for this has been well reviewed elsewhere [19] and the general strategy is summarized in Figure 1. If mixtures of proteins are analyzed, these peptides must also be prefractionated to maximize protein identification because the spectrometer can only identify one peptide at a time. This is most commonly achieved via HPLC separation, with on-line analysis by MS (LC-MS).

The application of proteomic technologies is already adding to our knowledge of lymphocyte signaling. For example, proteomic analysis of proteins co-purifying with lipid rafts identified raftlin ('raft-linking' protein), a new raft-binding protein in B cells [20]. Similarly, the extensive reorganization of the lipid raft proteome, following T-cell antigen receptor triggering has been studied successfully, using proteomics [21]. Another study of the relationship between organelle reshaping and Ig secretion in B lymphoma cells undergoing *in vitro* terminal differentiation, using 2D gels and MS, has determined that B cells anticipate their secretory role in a multistep process by clustering proteins according to temporal expression patterns [22•].

Another advancement is the introduction of 'tandem affinity purification' (TAP) methodology, which has substantially improved the purification and systematic genome-wide characterization of protein complexes in yeast [23,24]. Although the homozygous recombination used for yeast TAP-tag constructs cannot easily be applied to mammalian cells, a strategy has been developed in which the TAP approach can be combined with double-stranded RNA interference (RNAi) for the isolation and characterization of protein complexes from higher eukaryotic cells [25••]. In this modification of the TAP-tag approach, the expression of the endogenous mRNA is downregulated via RNAi, while the exogenously expressed TAG-tag construct is resistant to this treatment. This approach allows for the detection of individual protein-protein interactions and the subsequent construction of interaction networks. Multi-protein complexes can also be visualized and analyzed directly if stable and present in sufficient quantity. In fact, recent improvements in 2D native PAGE/SDS-PAGE permit the high-resolution separation of multi-protein complexes from whole cellular lysates, and subsequent component identification via LC-MS-MS [26•]. Not only

Figure 1



Schematic representation of protein identification strategy by MS. **(a)** A gel band or protein mixture (e.g. cell lysate, immune precipitated complex, isolated organelle) is proteolyzed to yield a complex peptide mixture. Peptides are separated by strong cationic exchange chromatography and reversed-phase HPLC. **(b)** Separated peptides are first detected by a survey scan (single-stage MS scan, for example, quadrupole1 (Q1) if a triple quadrupole mass spectrometer is used). Individual peptides are then sequentially and automatically selected for specific fragmentation in a collision cell, quadrupole 2 (Q2) in a triple quadrupole mass spectrometer, and the resultant fragments measured in a sequencing scan in quadrupole 3 (Q3) in a triple quadrupole instrument. The fragment ion spectrum that this MS-MS process creates is similarly known as a tandem mass spectrum. **(c)** This tandem mass spectrum contains sequence information for the original peptide ion. Bioinformatics software can then be used to correlate theoretical (i.e. computer-derived) MS-MS spectra from a sequence database with the acquired data, thus identifying the peptide sequence, and ultimately the protein(s) from which the peptides originated. In tandem mass spectrometers, in contrast to triple quadrupole instruments (such as ion traps; quadrupole time-of-flight; tandem time-of-flight), functionally analogous steps are performed using different physical principles.

does this technology possess the capacity to discriminate between complexes of different sizes [27], but it also enables the characterization of proteins as being members of a single signaling complex (or even multiple complexes) (Wolfgang W Schamel, personal communication).

Working with purified organelles, or other enriched sub-proteomes, offers the advantage of simplifying the complexity of crude cell extracts or tissue samples, thereby maximizing the probability of detecting lower abundance proteins in the mass spectrometer [28]. A particularly interesting sub-cellular compartment for immunologists is the plasma membrane, because this is where receptor and antigen complexes reside. Unfortunately, the physicochemical properties of the plasma membrane proteins render them refractory to systematic analysis by 2D electrophoresis (2DE). However, the application of MS to membrane proteomics has opened up new avenues of investigation [29]. The development of a method employing multidimensional protein identification technology (MudPIT) has resulted in the identification of

both the soluble and membrane proteins isolated from rat liver golgi [30<sup>\*</sup>] and could readily be applied to lymphocytes. Other strategies for the proteomic analysis of membrane proteins would include *in vivo* surface labeling with water-soluble biotinylation reagents, followed by cell lysis and membrane purification using affinity capture of biotinylated membrane proteins. Such an approach has been successfully applied to lymphocytes, leading to the identification of more than 40 plasma membranes from cultured lymphoma cells and more than 45 from primary murine splenocytes [31<sup>\*</sup>]. Similar approaches for cell-surface labeling have been applied to the plasma membrane sub-proteome in other cell systems, confirming the validity and general applicability of such methodologies [32–34].

### Quantitative proteomics

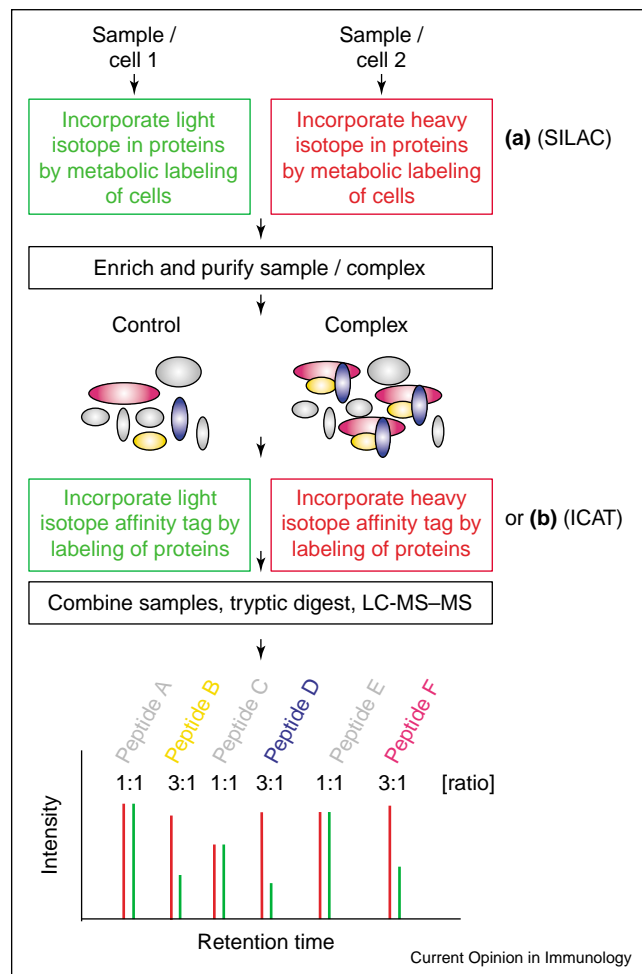
Although proteomics has the ability to identify large numbers of proteins in complex mixtures, to be able to study cell signaling at the proteomic level, it is necessary to have the capability to determine how proteomes or

## 4 Lymphocyte activation

sub-proteomes change, quantitatively, in response to stimuli, in a similar fashion to genomic analyses. This problem has been solved by the use of stable isotopic labeling of either proteins or peptides (following proteolysis). Through the introduction of either an isotopically 'normal' or 'heavy' signature in polypeptides, relative quantification of identified peptides derived from the proteins of two related samples, for example, resting and activated lymphocytes, can be determined. This is achieved by combining samples for MS analysis. The two versions of each peptide from these samples, being chemically identical will co-purify through any fractionation or enrichment protocol. In the spectrometer, however, these peptides can be distinguished via the introduced mass difference. The change in the relative abundance of the original proteins in the two samples is thus determined from the signal intensity ratio between these two isoforms of the peptides, derived from the original proteins.

The first working method to achieve isotopic labeling for quantitative proteomic experiments was based on the introduction of the isotope signature via chemical reactions, using so-called 'isotope-coded affinity tag' (ICAT) reagents [35]. Since then, these reagents and other isotopic labeling strategies have been successfully applied to a wide range of biological systems, including lymphocyte signaling [36–38]. For example, a strategy employing stable isotopic amino acids in cell culture (SILAC) was developed to isotopically label two cell populations *in vivo* and, subsequently, elucidate functional protein–protein interactions that are regulated by epidermal growth factor signaling [39•]. In another study, ICAT reagents were used to study components of the yeast RNA polymerase II preinitiation complex [40•] (Figure 2). Applying the same strategy to study the change of transcriptional complexes controlling  $\beta$ -globin expression during erythroid differentiation, it was shown that the shift of the transcription factor NF-E2p18/MafK from an inactive to an active state is correlated with a distinctive change in the protein composition of the complex [41•]. ICAT labeling enabled genuine protein components of the complex to be distinguished from the large background of non-specific protein binders, identifying, in the process, a new, previously unknown component of the complex. It is feasible that a similar strategy could be successfully applied to the identification of the multiple components of lymphocyte receptor and signaling complexes. This ability of quantitative proteomics to distinguish between specific and unspecific associations of proteins on a larger scale was also demonstrated by applying the SILAC approach to HeLa cell lipid raft proteins, distinguishing those with a strong affinity for rafts from those with less affinity and those showing non-specific co-isolation [42•]. Another recent study using the ICAT approach identified 685 proteins co-purifying with lipid rafts isolated from cultured T lymphocytes [43].

Figure 2



The study of macromolecular complexes via quantitative proteomics. Two strategies for the incorporation of stable isotopes into proteins are typically used. (a) In a SILAC-based approach, two cell populations are grown in medium containing heavy and light isotopic forms of arginine, thus labeling all arginine-containing proteins in one cell population with a 6 kDa heavier amino acid. The protein complex or sub-proteome of interest is enriched and purified in the second step. (b) In an ICAT-based approach, two protein mixtures, representing different cell states, are enriched and purified first and labeled post isolation with light and heavy ICAT reagents, respectively. The ICAT reagent covalently reacts with (reduced) cysteinyl sidechains in every cysteine-containing protein, introducing a 9 kDa difference per cysteine. In both procedures, the protein mixtures are combined and proteolyzed to peptides. In contrast to SILAC, ICAT-labeled peptides can be isolated by use of the biotin-tag, to reduce sample complexity. Peptides are separated by microcapillary HPLC. An isotope-dependent mass difference of the co-eluting peptide isoforms is measured by a scanning mass spectrometer. Ratios of the original amounts of proteins from the two cell states are strictly maintained in the peptide fragments. Relative quantification is determined by the ratio of the ion currents for the light- and heavy-tagged peptide pairs, as computed by suitable software tools. Protein identification data are obtained via MS-MS of labeled peptides, as summarized in Figure 1.

One other area of proteomics that is rapidly emerging is the adaptation of chemical labeling, isotopic tagging and affinity isolation protocols to identify and quantify

sub-proteomes, on the basis of specific post-translational modifications [44]. Glycoproteins are of great biological and clinical importance, especially for the identification of diagnostic biomarkers and targets for immunotherapy or pharmacological intervention. Furthermore, because many cell surface proteins are glycosylated, such approaches can be used to characterize lymphocytes, possibly with a view to determining cell subtype, differentiation state and so on. Two methods for the identification of N-linked glycopeptides in complex biological samples have been reported to date. The first approach is based on lectin-column mediated affinity capture [45<sup>\*</sup>]. The second approach involves covalent conjugation of glycoproteins to a solid support [46<sup>\*</sup>]. Covalent conjugation of glycopeptides, in combination with an enzymatic release step, allows for a stringent purification procedure, which adds selectivity and permits the identification of the actual glycosylation site. Such approaches are readily adaptable to the characterization of lymphocyte surface markers, as well as induced shedding and/or secretion of glycoproteins by lymphocytes.

Another obvious target, relevant to cellular signaling in all cell types, is protein phosphorylation. The determination of the way that protein phosphorylation is regulated and the identification of the sites of phosphorylation on signaling molecules is a prerequisite for understanding the regulation of signaling pathways and networks. To this end, several strategies for the quantitative analysis of phosphorylation have been developed. In one approach, a variation of the ICAT strategy was used to introduce a biotin tag into phosphoserine and phosphothreonine residues by  $\beta$ -elimination and Michael reaction, enabling enrichment and simultaneous quantification of phosphoproteins [47]. The other approach used an N-terminal peptide-labeling strategy for incorporation of isotopic tags without the need for specific internal standards, phosphorylation-specific antibodies or radioactivity [48]. However, despite these recent developments, it remains difficult to identify the phosphorylation sites of low-abundance phosphoproteins on a larger scale using existing technology, which might still be too immature for widespread application to models of lymphocyte signaling. The progress that has been made, to date, in phosphoprotein analysis by MS has been explored in more detail in other recent reviews [49,50]. A referenced summary of currently available proteomic tools and strategies that are most applicable to lymphocyte signaling and biology is given in Table 1.

### Translating data into biological knowledge

The distillation of biological significance from the numerous volumes of data that have been generated in both large-scale genomic and proteomic experiments perhaps presents the greatest challenge in these fields right now. To be of use, bioinformatic tools created to deal with such experimental data must be capable of automated

**Table 1**

#### Proteomic toolbox for immunologists

Proteomic toolbox for immunologists	References
1. Glycoprotein analysis	[44,45 <sup>*</sup> ,46 <sup>*</sup> ]
2. Phosphoprotein analysis	[47–49]
3. Membrane protein analysis	[29,30 <sup>*</sup> , 31 <sup>*</sup> ,32–34]
4. Single protein analysis	[20,21,22 <sup>*</sup> ,26 <sup>*</sup> ]
5. Quantitative protein analysis	[35,37,38,39 <sup>**</sup> ,40 <sup>**</sup> ,41 <sup>*</sup> ,42 <sup>**</sup> ]
6. Organelle analysis	[28]

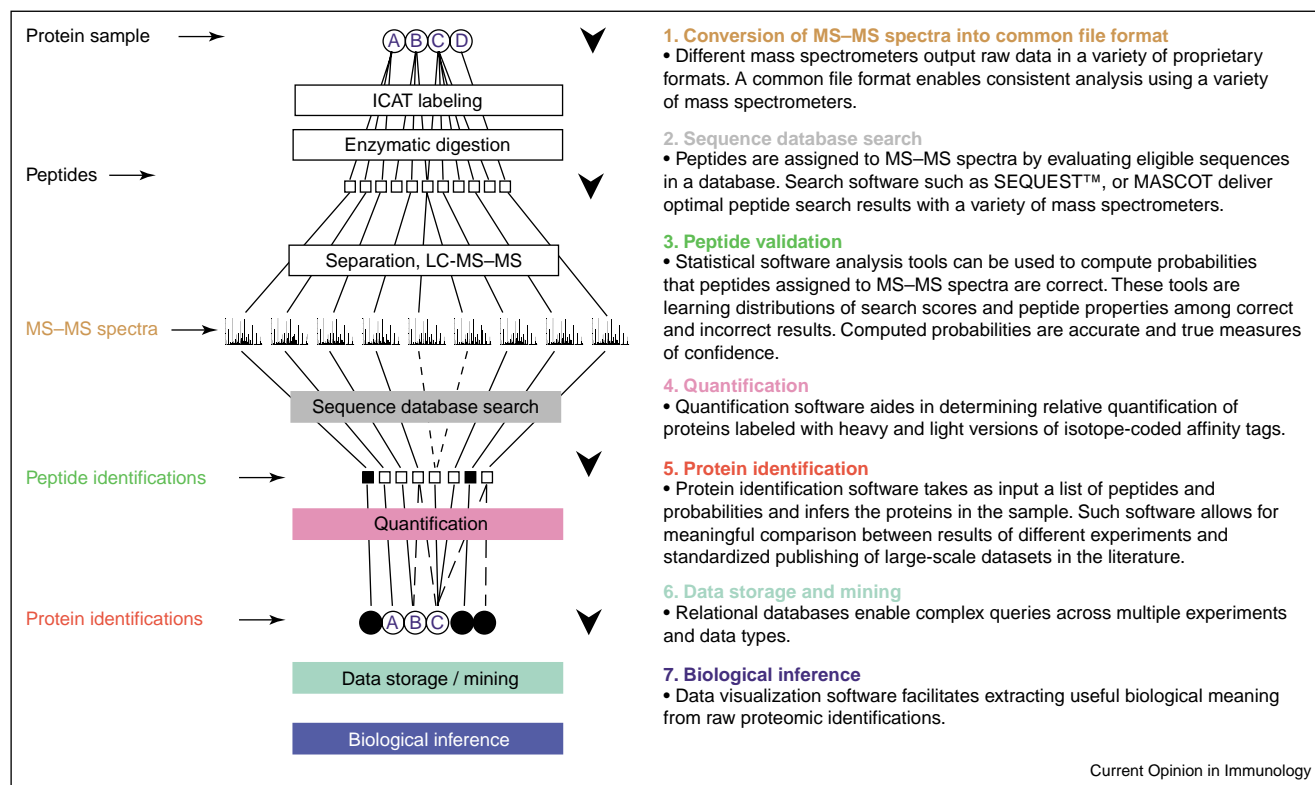
and systematic processing, analysis, validation and presentation of such data. Tools and guidelines for modeling, capturing and disseminating genomic data are more advanced than those for proteomic data, and are examined elsewhere [51]. By contrast, bioinformatic tools for the interpretation of proteomic data are still in their infancy (summarized in Figure 3).

As a prerequisite for efficient, portable and transparent data interpretation, it is first necessary to develop algorithms that assign scores to every observation, as a form of statistical data evaluation. New software tools have recently been published that attempt to address this need. Scores are assigned to peptide and protein identifications, or to the quantitative component of the data ratio — in effect, calculating the likelihood, as a function of probability, that each observation is correct, independent of variables such as the type of mass spectrometer used to generate the data, the database search tool used or the quality of the sample, and independent of subjective, expert interpretation [52<sup>\*</sup>,53–55]. The application of such tools to large proteomic datasets has been shown to allow for the meaningful comparison of experiments, based on consistent and verifiable parameters [43,56].

Once collected and scored, proteomic data must be stored, annotated and integrated with genomic data and other information, to extract meaningful biological information, and ultimately enable the construction of protein interaction and signaling networks. Data repositories and standard representations of both the methods used and the data generated in proteomic experiments have yet to emerge [57]. Attempts to store, manage and access the plethora of data in relational database management systems for further biological interrogation on a higher level include the Proteomics Experiment Data Repository (PEDRo; <http://pedro.man.ac.uk>) [58] or the Systems Biology Experiment Analysis Management System (SBEAMS; <http://db.systemsbio.org/projects/sbeams/>). New software tools are also emerging that make possible *in silico* prediction and visualization of protein networks, derived from large-scale proteomic datasets, either individually or from multiple sources; an example of such bioinformatic software is Cytoscape (<http://www.cytoscape.org/>). The construction of such networks

## 6 Lymphocyte activation

Figure 3



Computational tools for high-throughput proteomics. Software tools are needed to automate the analysis of large amounts of proteomic data generated by a variety of types of mass spectrometers. A complete pipeline contains tools for peptide and protein identifications, validation, and quantification. A protein sample is enzymatically digested to peptides that are subjected to LC-MS-MS. Resulting spectra are assigned computationally to their best amino acid sequence matches and are quantified according to the signal size of corresponding isotopic peaks. Validation is necessary to distinguish correct (white) from incorrect (black) results. From assigned peptides, the original protein contents of the sample are inferred. This process is complicated by peptides that correspond to more than one protein in the database (dashed lines), and by redundancies in sequence databases. Some standardization of bioinformatics software and statistical data analysis are essential to allow the comparison of multiple datasets from related experiments, produced in different laboratories and on different spectrometers. A suite of software tools that constitute such an analysis pipeline are described at, and are available from, <http://www.proteomecenter.org/software.php>.

will help to address new questions and make predictions that can then be tested *in vivo*, and the model can then be reformulated, according to the observed experimental results. Indeed, at least in yeast, it has been demonstrated that simple protein networks can be measured via genomic and proteomic methodologies, a model constructed *in silico*, and predictions can be made about new network connections, which were subsequently verified experimentally [59]. It is thus reasonable to expect that, with sufficiently mature analytical and bioinformatics tools and quantitative proteomic data in hand, it will be possible to perform similar experiments on lymphocyte signaling networks and to predict (and test) how they react to specific perturbations.

## Conclusions

The emergence of the global analytical fields of genomics and proteomics are clearly opening up new avenues of investigation in many areas of biological inquiry. Perhaps

for the first time, the description of entire biological systems, in terms of quantitative measurements of their set of expressed genes and translated protein products is becoming a reality. The development of new bioinformatic tools to manage, display, model and disseminate datasets of such magnitude will also allow for the construction of dynamic gene-protein interaction networks that will subsequently enable the prediction of the outcome of any new perturbation of the system in question; predictions that can then be tested. Although in many cases, particularly in proteomics, these new methodologies are still being designed and validated using simpler model systems, the implications for such systematic, global analyses of gene-protein signaling networks in all cell types, including lymphocytes, is clearly highly significant. Ultimately, they will not only provide a much more profound understanding of how individual signaling pathways act, and the other biological processes upon which they impinge, but will also enable accurate

diagnostic and predictive evaluation of cell type, differentiation state and physiological outcome, equally applicable to laboratory experimental models of human biology as well as human disease itself.

## Acknowledgements

We thank Jeff Ranish and Anne-Claude Gingras for comments on the manuscript. This work was supported, in part, through funding from NIH RO1-AI-41109-01 (to JW) and RO1-AI-51344-01 (to RA) and by federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, under contract no. N01-HV-28179.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Patterson SD, Aebersold RH: **Proteomics: the first decade and beyond.** *Nat Genet* 2003, **33**(Suppl):311-323.
  2. Tyers M, Mann M: **From genomics to proteomics.** *Nature* 2003, **422**:193-197.
  3. Glynn R, Ghandour G, Rayner J, Mack DH, Goodnow CC: **B-lymphocyte quiescence, tolerance and activation as viewed by global gene expression profiling on microarrays.** *Immunol Rev* 2000, **176**:216-246.
  4. Shaffer AL, Rosenwald A, Hurt EM, Giltne JM, Lam LT, Pickeral OK, Staudt LM: **Signatures of the immune response.** *Immunity* 2001, **15**:375-385.
  5. Glynn RJ, Ghandour G, Goodnow CC: **Genomic-scale gene expression analysis of lymphocyte growth.** *Curr Opin Immunol* 2000, **12**:210-214.
  6. Schmidt U, Begley CG: **Cancer diagnosis and microarrays.** *Int J Biochem Cell Biol* 2003, **35**:119-124.
  7. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM: **A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma.** *Proc Natl Acad Sci USA* 2003, **100**:9991-9996.
  8. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS *et al.*: **Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning.** *Nat Med* 2002, **8**:68-74.
  9. Staudt LM, Wilson WH: **Focus on lymphomas.** *Cancer Cell* 2002, **2**:363-366.
  10. Vinuesa CG, Cook MC, Cooke MP, Maclennan IC, Goodnow CC: **Analysis of B cell memory formation using DNA microarrays.** *Ann N Y Acad Sci* 2002, **975**:33-45.
  11. Marrack P, Mitchell T, Hildeman D, Kedl R, Teague TK, Bender J, Rees W, Schaefer BC, Kappler J: **Genomic-scale analysis of gene expression in resting and activated T cells.** *Curr Opin Immunol* 2000, **12**:206-209.
  12. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltne JM *et al.*: **The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma.** *N Engl J Med* 2002, **346**:1937-1947.
  13. Rosenwald A, Wright G, Wiestner A, Chan WC, Connors JM, Campo E, Gascoyne RD, Grogan TM, Muller-Hermelink HK, Smeland EB *et al.*: **The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma.** *Cancer Cell* 2003, **3**:185-197.
- A good example of the use of gene-expression profiling to establish a molecular diagnosis of a specific cancer.
14. Shaffer AL, Lin KI, Kuo TC, Yu X, Hurt EM, Rosenwald A, Giltne JM, Yang L, Zhao H, Calame K *et al.*: **Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program.** *Immunity* 2002, **17**:51-62.
- This paper demonstrates the way that DNA microarrays can be used to analyze a cascade of gene expression changes in lymphocytes.
15. Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakkola TS *et al.*: **Serial regulation of transcriptional regulators in the yeast cell cycle.** *Cell* 2001, **106**:697-708.
  16. Greenbaum D, Luscombe NM, Jansen R, Qian J, Gerstein M: **Interrelating different types of genomic data, from proteome to secretome: 'oming in on function.** *Genome Res* 2001, **11**:1463-1468.
  17. Greenbaum D, Colangelo C, Williams K, Gerstein M: **Comparing protein abundance and mRNA expression levels on a genomic scale.** *Genome Biol* 2003, **4**:117.
  18. Aebersold R, Mann M: **Mass spectrometry-based proteomics.** *Nature* 2003, **422**:198-207.
  19. Peng J, Gygi SP: **Proteomics: the move to mixtures.** *J Mass Spectrom* 2001, **36**:1083-1091.
  20. Saeki K, Miura Y, Aki D, Kurosaki T, Yoshimura A: **The B cell-specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction.** *EMBO J* 2003, **22**:3015-3026.
  21. Bini L, Pacini S, Liberatori S, Valensin S, Pellegrini M, Raggiaschi R, Pallini V, Baldari CT: **Extensive temporally regulated reorganization of the lipid raft proteome following T-cell antigen receptor triggering.** *Biochem J* 2003, **369**:301-309.
  22. van Anken E, Romijn EP, Maggioni C, Mezghrani A, Sitia R, Braakman I, Heck AJ: **Sequential waves of functionally related proteins are expressed when B cells prepare for antibody secretion.** *Immunity* 2003, **18**:243-253.
- This paper describes a dynamic proteomics study, using 2D Gels of B lymphoma cells undergoing *in vitro* terminal differentiation.
23. Puig O, Caspari F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B: **The tandem affinity purification (TAP) method: a general procedure of protein complex purification.** *Methods* 2001, **24**:218-229.
  24. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B: **A generic protein purification method for protein complex characterization and proteome exploration.** *Nat Biotechnol* 1999, **17**:1030-1032.
  25. Forler D, Kocher T, Rode M, Gentzel M, Izaurralde E, Wilm M: **An efficient protein complex purification method for functional proteomics in higher eukaryotes.** *Nat Biotechnol* 2003, **21**:89-92.
- This paper describes the combined application of RNAi and TAP technology to avoid competition from corresponding endogenous proteins, while isolating and characterizing protein complexes from higher eukaryotic cells.
26. Camacho-Carvajal MM, Wollscheid B, Aebersold R, Steimle V, Schamel WW: **Two-dimensional Blue Native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates - a proteomics approach.** *Mol Cell Proteomics* 2003, in press.
- This paper describes how Blue Native PAGE (BN-PAGE) can be used for the high-resolution separation of multi-protein complexes under native conditions in combination with MS-based identification of complex members.
27. Schamel WW, Reth M: **Monomeric and oligomeric complexes of the B cell antigen receptor.** *Immunity* 2000, **13**:5-14.
  28. Brunet S, Thibault P, Gagnon E, Kearney P, Bergeron JJ, Desjardins M: **Organelle proteomics: looking at less to see more.** *Trends Cell Biol* 2003, **13**:629-638.
  29. Wu CC, Yates JR III: **The application of mass spectrometry to membrane proteomics.** *Nat Biotechnol* 2003, **21**:262-267.
  30. Wu CC, MacCoss MJ, Howell KE, Yates JR III: **A method for the comprehensive proteomic analysis of membrane proteins.** *Nat Biotechnol* 2003, **21**:532-538.
- This paper describes a strategy that facilitates the identification of large numbers of membrane proteins and modifications. This strategy has the potential to provide insights into protein topology and orientation in membranes.

## 8 Lymphocyte activation

31. Peirce MJ, Wait R, Begum S, Saklatvala J, Cope AP:  
 • **Expression profiling of lymphocyte plasma membrane proteins.** *Mol Cell Proteomics* 2003, in press.  
 An *in vivo* cell-surface labeling technique with a water-soluble biotinylation reagent is presented for investigating differential expression of surface proteins.
32. Shin BK, Wang H, Yim AM, Le Naour F, Brichory F, Jang JH, Zhao R, Puravs E, Tra J, Michael CW *et al.*: **Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function.** *J Biol Chem* 2003, **278**:7607-7616.
33. Jang JH, Hanash S: **Profiling of the cell surface proteome.** *Proteomics* 2003, **3**:1947-1954.
34. Chen WN, Yu LR, Strittmatter EF, Thrall BD, Camp DG II, Smith RD: **Detection of *in situ* labeled cell surface proteins by mass spectrometry: application to the membrane subproteome of human mammary epithelial cells.** *Proteomics* 2003, **3**:1647-1651.
35. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* 1999, **17**:994-999.
36. Patton WF, Beechem JM: **Rainbow's end: the quest for multiplexed fluorescence quantitative analysis in proteomics.** *Curr Opin Chem Biol* 2002, **6**:63-69.
37. Sechi S, Oda Y: **Quantitative proteomics using mass spectrometry.** *Curr Opin Chem Biol* 2003, **7**:70-77.
38. Goshe MB, Smith RD: **Stable isotope-coded proteomic mass spectrometry.** *Curr Opin Biotechnol* 2003, **14**:101-109.
39. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M:  
 •• **A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling.** *Nat Biotechnol* 2003, **21**:315-318.  
 This paper reports an MS-based approach to identify protein complexes from background binding using quantitative proteomics, by differentially labeling proteins in epidermal growth factor-stimulated versus unstimulated cells, using SILAC.
40. Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J,  
 •• **Aebersold R: The study of macromolecular complexes by quantitative proteomics.** *Nat Genet* 2003, **33**:349-355.  
 This is a good example of the identification of functional protein complexes and discrimination of real components of a complex from contaminant proteins using ICAT-based quantitative proteomics.
41. Brand M, Ranish JA, Kummer NT, Hamilton J, Igarashi K,  
 • **Francastel C, Chi TH, Crabtree GR, Aebersold R, Groudine M: Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics.** *Nat Struct Mol Biol* 2004, **11**:73-80.  
 This paper shows how the ICAT technique can be used to reveal quantitative changes in proteins interacting with NF-E2p18/MafK during differentiation. The results define MafK as a 'dual-function' molecule that shifts from a repressive to an activating mode during erythroid differentiation.
42. Foster LJ, De Hoog CL, Mann M: **Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors.** *Proc Natl Acad Sci USA* 2003, **100**:5813-5818.  
 This is a good example of identification of a sub-proteome and discrimination of real components of the sub-proteome from contaminant proteins, using SILAC-based quantitative proteomics.
43. Von Haller PD, Yi E, Donohoe S, Vaughn K, Keller A, Nesvizhskii AI, Eng J, Li XJ, Goodlett DR, Aebersold R *et al.*: **The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: i. statistically annotated datasets for peptide sequences and proteins identified via the application of ICAT and tandem mass spectrometry to proteins copurifying with T cell lipid rafts.** *Mol Cell Proteomics* 2003, **2**:426-427.
44. Zhang H, Yan W, Aebersold R: **Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes.** *Curr Opin Chem Biol* 2004, **8**:1-10.
45. Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J,  
 • Kasai K, Takahashi N, Isobe T: **Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins.** *Nat Biotechnol* 2003, **21**:667-672.  
 This paper describes MS analysis of protein glycosylations using the combined techniques of lectin affinity purification and isotope-coded tagging.
46. Zhang H, Li XJ, Martin DB, Aebersold R: **Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry.** *Nat Biotechnol* 2003, **21**:660-666.  
 This paper describes a method to identify and quantify N-linked glycoproteins. It also applies the method to the enrichment of cell surface proteins and the efficient profiling of serum proteins via removal of the albumin from serum samples.
47. Goshe MB, Veenstra TD, Panisko EA, Conrads TP, Angell NH, Smith RD: **Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins.** *Anal Chem* 2002, **74**:607-616.
48. Zhang X, Jin QK, Carr SA, Annan RS: **N-Terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry.** *Rapid Commun Mass Spectrom* 2002, **16**:2325-2332.
49. Kalume DE, Molina H, Pandey A: **Tackling the phosphoproteome: tools and strategies.** *Curr Opin Chem Biol* 2003, **7**:64-69.
50. Shou W, Verma R, Annan RS, Huddleston MJ, Chen SL, Carr SA, Deshaies RJ: **Mapping phosphorylation sites in proteins by mass spectrometry.** *Methods Enzymol* 2002, **351**:279-296.
51. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC *et al.*: **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.** *Nat Genet* 2001, **29**:365-371.
52. Keller A, Nesvizhskii AI, Kolker E, Aebersold R:  
 • **Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.** *Anal Chem* 2002, **74**:5383-5392.  
 This paper describes the problems that are involved in estimating the accuracy of peptide identifications and one approach on how to automate the analysis of large volumes of MS-MS database search results.
53. Nesvizhskii AI, Keller A, Kolker E, Aebersold R: **A statistical model for identifying proteins by tandem mass spectrometry.** *Anal Chem* 2003, **75**:4646-4658.
54. MacCoss MJ, Wu CC, Liu H, Sadygov R, Yates JR III: **A correlation algorithm for the automated quantitative analysis of shotgun proteomics data.** *Anal Chem* 2003, **75**:6912-6921.
55. Li XJ, Zhang H, Ranish JA, Aebersold R: **Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry.** *Anal Chem* 2003, **75**:6648-6657.
56. Von Haller PD, Yi E, Donohoe S, Vaughn K, Keller A, Nesvizhskii AI, Eng J, Li XJ, Goodlett DR, Aebersold R *et al.*: **The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: ii. evaluation of tandem mass spectrometry methodologies for large-scale protein analysis.** *Mol Cell Proteomics* 2003, **2**:428-442.
57. Fenyó D, Beavis RC: **Informatics and data management in proteomics.** *Trends Biotechnol* 2002, **20**:S35-S38.
58. Taylor CF, Paton NW, Garwood KL, Kirby PD, Stead DA, Yin Z, Deutsch EW, Selway L, Walker J, Riba-Garcia I *et al.*: **A systematic approach to modeling, capturing, and disseminating proteomics experimental data.** *Nat Biotechnol* 2003, **21**:247-254.
59. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** *Science* 2001, **292**:929-934.