ELSEVIER

Seminars in Immunology 16 (2004) 55–67

seminars in
IMMUNOLOGY

www.elsevier.com/locate/ysmim

A systems approach to dissecting immunity and inflammation

Alan Aderem [a,*], Kelly D. Smith [a,b]

[a] *Institute for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, USA*
[b] *Department of Pathology, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, USA*

Abstract

The immune and inflammatory responses are extraordinarily complex, involving the dynamic interaction of a wide array of tissues, cells, and molecules. Traditional approaches are by and large reductionist, shying away from complexity, but providing detailed knowledge of circumscribed physiologic, cellular and molecular entities. The sequencing of the human genome, in concert with emerging genomic and proteomic technologies permits the definition of a complete and dynamic parts list of the immune and inflammatory systems. When harnessed with powerful new computational approaches, this will for the first time provide a comprehensive description of these complex biological processes.

Keywords: Innate immunity; Inflammation; Systems biology

1. Innate immunity and the inflammatory response

The innate immune system is essential for host defense and is responsible for early detection and containment of pathogens [1,2]. Multiple cell types and tissues participate in the ensuing inflammatory response, which includes the recognition of microbes, the activation of anti-microbial defenses and the recruitment of circulating inflammatory cells. The nature of the invading pathogen specifies a response that provides optimal host defense, but this inflammatory response is a two edged sword that must be tightly regulated. The complex interactions initiated by the infection set off a wave of events that can lead to multiple outcomes: resolution of the infection with complete restoration of tissue architecture, resolution of the infection and destruction of tissue (scarring), control of the infection with ongoing inflammation (chronic inflammation), control of the infection with initiation of new inflammatory responses (autoimmunity), and failure to control the infection. The regulation of the inflammatory response is extraordinarily complicated and occurs on many levels. Indeed, it is this complexity that necessitates a systems approach to the problem [3]. In this article we will focus on one component of the inflammatory response, that of macrophage activation. In particular, the initial events that are triggered upon pathogen recognition will be discussed.

2. Recognition of pathogens and macrophage activation

2.1. Pattern recognition receptors (PRR)

The inflammatory response to infectious agents is activated when the phagocyte recognizes the foreign invaders using a battery of receptors including the Toll-like receptors (TLRs), scavenger receptors, complement receptors, members of the C-type lectin receptor family, and integrins. These germ line receptors have evolved to recognize conserved motifs on pathogens that are not found on higher eukaryotes; these structures have essential roles in the biology of the invader, and are therefore not subject to high mutation rates. These structural motifs include carbohydrates, glycolipids, proteolipids, glycoproteins and proteins; for example, TLR4 recognizes bacterial lipopolysaccharides (LPS), the mannose receptor binds mannosyl/fucosyl residues, dectin-1 binds β-glucans, and scavenger receptors bind negatively charged lipids. Pathogens are also opsonized by humoral components including complement and immunoglobulins, which are in turn recognized by complement- and Fc-receptors, respectively.

The molecular mechanism underlying the function of these receptors has received intense scrutiny, but cross-talk between them has received limited attention. When taken together with the enormous spectrum of pro- and anti-inflammatory responses induced when the host encounters microbial pathogens, a system of extraordinary complexity is revealed.

* Corresponding author. Tel.: +1-206-732-1200; fax: +1-206-732-1249.
E-mail address: aderem@systemsbiology.org (A. Aderem).

1044-5323/$ – see front matter © 2003 Published by Elsevier Ltd.
doi:10.1016/j.smim.2003.10.002

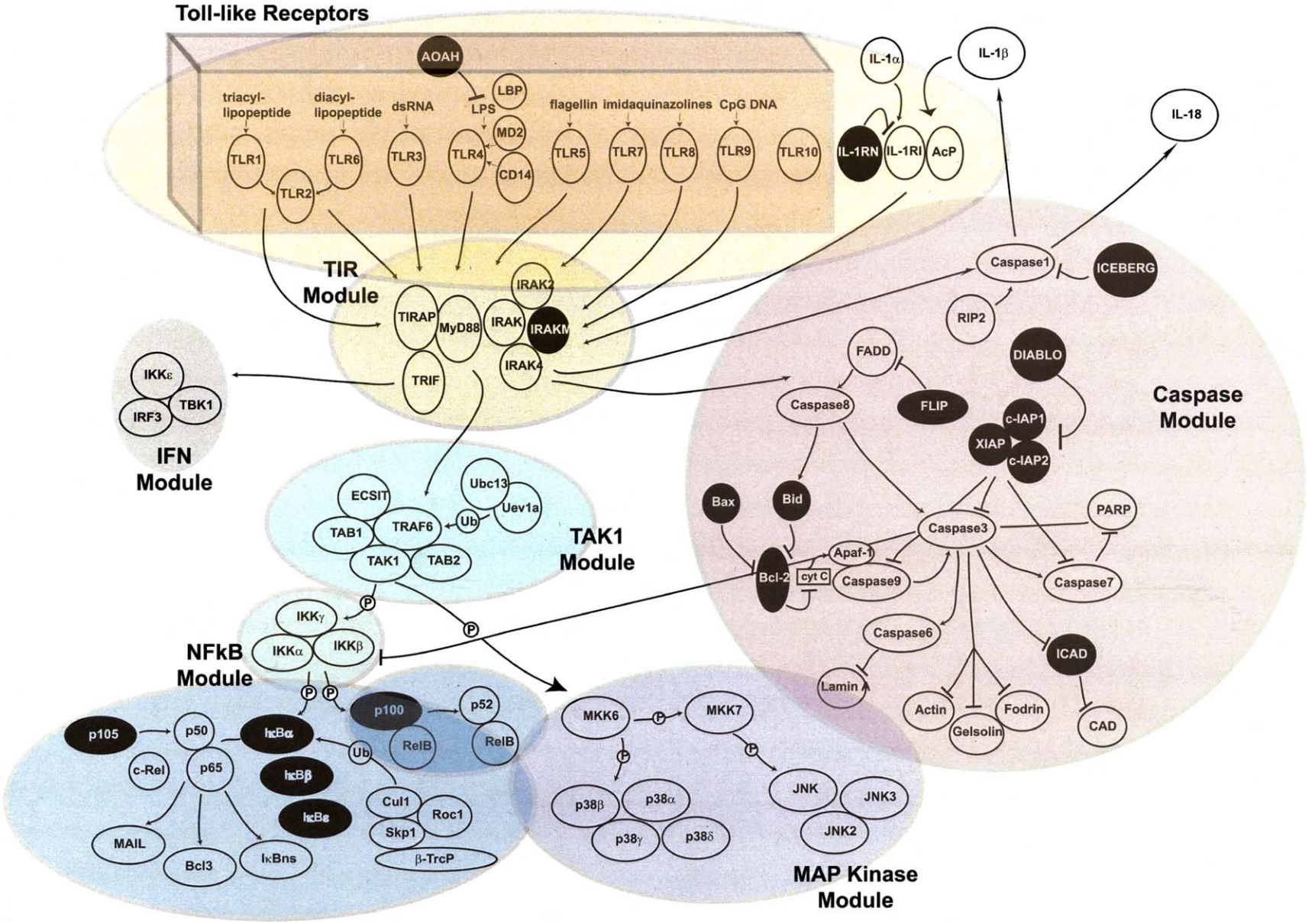

Fig. 1. A modular view of Toll-like receptor recognition and signaling. Shown are the 10 TLRs with their known agonists, as well as the IL-1 receptor system. The TIR, IFN, Tak1, NFκB, Map kinase, and caspase signaling modules are also shown. Inhibitory molecules are shown in black. Arrows demonstrate the flow of information.

2.1.1. The Toll-like family of receptors

Mammalian TLRs are a family of 10 pattern recognition receptors that are central to effective innate immunity [4]. TLRs recognize a broad spectrum of ligands, including modified lipids (LPS and bacterial lipoproteins), proteins (flagellin), and nucleic acids (DNA and double-stranded RNA). Each TLR recognizes specific components of the pathogen, and the specificity of this recognition is shown in Fig. 1. It is important to bear in mind that pathogens consist of a complicated cocktail of pathogen-associated molecular patterns (PAMPs) that stimulate the TLRs in concert, resulting in the activation of a number of cross-talking signaling pathways. The integration of this information ultimately gives rise to an appropriate, and measured, immune and inflammatory response.

2.1.2. C type lectins

The mannose receptor (MR) is the archetypal C type lectin. It binds mannosy/fucosyl or GlcNAc-glycoconjugate ligands on many bacteria, fungi, and protozoan parasites, and initiates a strong proinflammatory response to them. The MR also functions to clear glycosylated endogenous ligands, leading to the suggestion that it functions both to detect foreign pathogens and to mediate the clearance of injurious self molecules [5]. Dectin-1, another C-type lectin, is the major receptor for β-1,3-glucans; ligand binding induces phagocytosis, and reactive oxygen species production in macrophages. Interestingly, when macrophages encounter zymosan, dectin-1 cooperates with TLR2/6 to potentiate IL-12 production.

2.1.3. The scavenger receptors

The scavenger receptors (SR) are another example of an innate immune receptor doubling as a homeostatic receptor [6]. SR-A contributes to resistance to Gram-positive bacterial infections, and also appears to regulate LPS-induced inflammation. This is supported by the observation that SR-A knockout mice are more susceptible to LPS-induced shock, probably due to an imbalance between SR-A-dependent clearance of LPS and TLR4-dependent secretion of inflammatory mediators such as TNFα. SR-A also mediates the endocytosis of modified low-density lipoprotein (LDL) by macrophages, leading to inflammation and foam cell formation. It is not clear how the complex interplay between the immune and homeostatic functions of SR-A is orchestrated.

2.1.4. Integrins

Complement receptor 3 (CR3) is a myeloid cell phagocytic receptor for complement opsonised particles, and also for direct interactions with pathogens such as *Mycobacterium tuberculosis* and yeast cell wall [7,8]. It is a β2 integrin, also known as CD18/CD11b, and it plays a key role in myelomonocytic cell recruitment to sites of inflammation. It binds a wide range of ligands, including ICAM-1, selected clotting components, senescent platelets, and possibly, denatured proteins. The opsonic phagocytic mechanism differs from that mediated by Fc receptors and CR3-mediated uptake by macrophages, in that it does not trigger release of arachidonate or reactive oxygen metabolites.

2.1.5. Additional relevant receptors

While we have primarily focused on the pattern recognition receptors, it is important to bear in mind that a great many additional receptors participate in specifying a specific immune response. Thus, *cytokine* receptors influence the nature of the response, for example, a TH1 versus TH2 versus tolerogenic response. Chemokine receptors regulate, amongst other things the trafficking of cells to affected tissues and to the secondary lymphoid organs. Growth factor receptors maintain homeostasis and influences differentiation, and co-stimulatory receptors positively and negatively regulate the adaptive immune response. These receptors act in concert with the pattern recognition receptors to orchestrate an appropriate inflammatory response.

3. Signaling pathways regulating inflammation

3.1. Pro-inflammatory pathways

A vast number of signaling pathways regulate inflammatory responses, and the translation of information from extracellular signals to intracellular responses is the result of complex integration and interplay between signaling modules. A signaling module refers to an assembly of molecules that act in concert as a single functional unit. As an example, the stimulation of TLRs leads to the assembly of a Toll-interleukin-1 receptor (TIR) signaling module comprised of adaptors, such as Myd88, and kinases, such as IRAK-4 (Fig. 1). The TIR signaling module interfaces with other signaling modules, which include the Tak1 module, IFN module, and the caspase module. The Tak1 module activates both the NFκB and the Map kinase module [9].

The selected group of proximal signaling modules of the TLR pathway, shown in Fig. 1, provides a mere glimpse of the molecular complexity of TLR-induced responses. Additional complexity arises out of the differential use of individual components within a signaling module. For example, TLR4 uses the Myd88 adaptor of the TIR signaling module to stimulate the production of TNF, while it uses the TRIF adaptor, also within the TIR signaling module, to induce the secretion of IFNβ [O'Neill, 2003 #23]. It is also important to appreciate that the signaling pathways are dynamic, and that the specific components of a signaling module will change due to post-translational modifications, protein–protein interactions, subcellular compartmentalization, and differential gene regulation. The other receptor systems, described above, are equally complex. Since many of these pathways are activated concurrently when a macrophages encounters a single class of bacteria, the cross-talk and integration of information required for an appropriate host response is astounding.

Additional complexity in signaling is derived from the multifunctional nature of many PRRs [5]. As mentioned above, the MR both recognizes foreign pathogens and binds and removes glycosylated hormones from the blood. Similarly, SR-A functions as a PRR, and mediates the uptake of modified LDL. The CR3 receptor is an innate immune receptor, and has a role in cell motility and extravasation from the vascular. Finally, we should also bear in mind that all of these signaling pathways are subject to genetic variation and environmental influences.

3.2. Anti-inflammatory pathways

Although the initiation of the inflammation has been studied extensively, less is known about the mechanism by which the inflammatory response is dampened. A number of generic inhibitory mechanisms have been delineated. These include: (A) the secretion of inhibitory molecules; for example, phagocytosis of apoptotic cells is anti-inflammatory because the macrophage is induced to secrete inhibitory compounds such as prostaglandin E2, IL-10 and TGFβ [10]. (B) Varying the ratio of stimulatory and inhibitory molecules; for example, FcγRIII contains a stimulatory ITAM motif, whereas FcγRII contains an inhibitory ITIM motif [11]. (C) Competition for ligands amongst receptors; for example, SR-A knockout mice are hypersensitive to LPS because SR-A competes with TLR4 for LPS [6].

Negative regulation of the TLR pathway occurs at a number of levels. First, prolonged exposure with a TLR agonist results in tolerance to the agonist. Some aspects of tolerance can be attributed to down modulation of the TLRs, but evidence of cross-tolerance between different TLR agonists also indicates that TLR stimulation results in the modulation of signaling components [12]. Several signaling components that inhibit TLR signaling are up regulated after exposure to agonists. These include IRAK-M, SOCS1, a splice variant of Myd88, and SIGIRR, all of which have been implicated in negative feedback regulation of the TLR pathway [13].

4. Unraveling complexity using systems biology

Clearly host defense and the inflammatory response are overwhelmingly complex. Biochemical, cell biological and genetic approaches have been successful in unraveling, in broad brushstrokes, some of the functional components of these systems. However, the tools of systems biology will be essential in defining, in total, the interactions that underlie this complex biological system.

The science of systems biology has grown directly out of the Human Genome Project. For the first time, the entire parts list of the inflammatory and immune responses have been defined and annotated. This, in turn, has permitted the quantitative analysis of all the mRNAs (transcriptome) or proteins (proteome) present in a particular cell type. The sequencing effort also includes the introns containing the regulatory elements, essential for the eventual deciphering of the regulatory code. In addition, the availability of the genomic sequence will reveal polymorphisms within the population leading to a deeper understanding of the genetic factors influencing disease susceptibility.

Importantly, the genome has catalyzed fundamental changes in how we view and practice biology. These changes in paradigm can be summarized as follows. (A) Biology is an informational science. There are two major types of biological information: the information of our genes which encode the molecular machines composed of protein or RNA, and the information of the regulatory networks controlling the behavior of the molecular machines. (B) High-throughput biological tools are essential for following the flow of information in biological systems. The Human Genome Project has catalyzed the development of high-throughput DNA sequencing, DNA arrays, genotyping, and proteomics. These tools have permitted global studies, the study of behavior of all, or most, of the elements in a system—an essential component of systems biology. Many other high-throughput tools will also be required for systems approaches; these include the visualization of biological information in cells, tissues, and even organisms. (C) Computer science and applied mathematics are critical tools for deciphering biological information and for modeling complex biological systems. The dataset generated during the analysis of a biological system is so vast that the development of advanced computational and graphical tools is necessary in order to integrate the data into informational pathways and networks. It is almost certain that new types of mathematics will be required for these challenges. (D) Model organisms can be manipulated to provide insights into complex biological systems. In order to study the coordinate behavior of elements of a biological system, genetic or pharmacological perturbations of the system must be carried out in model organisms or cell lines, and the flow of information through the various hierarchical levels must be captured. For immunologists, the mouse has been the model organism of choice, although tantalizing insights into the innate immune system have been and will be gleaned by studying other organisms, including insects and sea urchins. (E) Comparative genomics permits powerful analyses of development, physiology, and evolution. Much of the complexity of living organisms stems from complex regulatory networks, rather than gene diversity. Comparing orthologous chromosomal regions in different species provides powerful tools for identifying coding regions and regulatory elements.

Each of these paradigm changes contributes to the modern concept of systems biology. Thus, all the components of the system are defined, the informational pathways within the system are elucidated, and mathematical models must be developed that accurately represent the system.

5. Global technologies for dissecting immunity and inflammation

5.1. Genomics

5.1.1. Genotyping

The understanding of complex biological processes is facilitated by global analyses of genetic variation and gene regulation. The mouse is the best characterized mammalian model organism. Several phenotypes have arisen spontaneously, and have been characterized through breeding, but the majority of phenotypes have been engineered using gene deletion and transgenic strategies. The sequencing of the mouse genome, together with advances in genotyping, permits the creation of novel genetic variants and phenotypes, akin to the strategies pioneered with *Drosophila*. These approaches have recently led to the discovery of the *lps2*/*trif* gene, which encodes a novel adaptor belonging to the Myd88 family [14,15]. Human genetic variation results in altered susceptibility to infectious and inflammatory diseases. Occasionally, genetic variation in single genes results in immune and inflammatory disorders, but more commonly these diseases are caused by a complex interplay between a number of genes, and exacerbated by environmental influences. Classical genetic studies make use of variable elements within the genome that are stably inherited to map phenotypes to genetic loci.

Identification of genes with known biological relevance also provides the opportunity to address the contribution of genetic variation to disease in the reverse direction. Candidate genes can be interrogated for sequence variation within human populations. Many variants detected by this strategy constitute single nucleotide polymorphisms (SNPs), which can be tested for their prevalence and associations with disease phenotypes. Such candidate gene analysis has led to the association of NOD2 with familial and sporadic forms of Crohn's disease [16], as well as TLR4 with atherosclerosis and bacterial meningitis [17,18]. We recently demonstrated that a common variant of TLR5 is unable to mediate flagellin signaling, acts in a dominant negative fashion, and is associated with susceptibility to pneumonia caused by *Legionella pneumophila*, a flagellated bacterium. The association of these genes with disease is significant, but, in general, the odds ratios of association are relatively low. This is to be expected, since human diseases are complex and multi-factorial, and mutations in single genes would only partially contribute to the overall risk. More drastic mutations, resulting in significant susceptibility to disease, are exceedingly rare, since they have been selected against. Clearly, an understanding of multi-factorial diseases requires the integration of a vast amount of variables, and can only be attempted using systems approaches. This will result in a compendium of potential susceptibility loci of genetic disease association, and will the pave the way for predictive and personalized medicine.

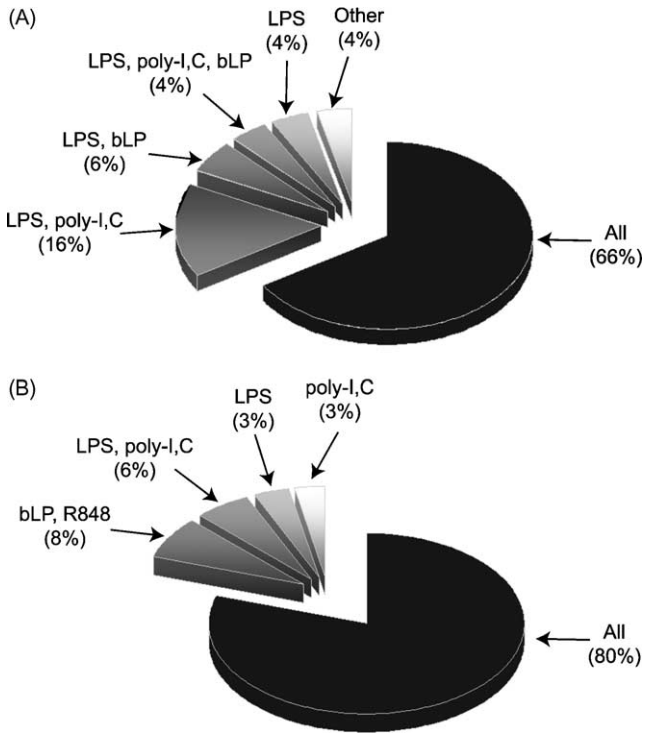

Fig. 2. Regulation of gene expression by Toll-like receptors. Macrophages were stimulated with the indicated TLR agonists and the transcriptome was measured. (A) The proportion of genes induced by individual stimuli, compared with those that were common to all TLR agonists. (B) The down-regulated genes.

5.1.2. Transcriptional regulation

The transcriptome refers to the sum of all the expressed genes in a system, and is commonly measured by microarray technologies.

Comparative gene expression studies define global differences between biological perturbations, which can be any one of a vast number of situations, including stimulus, time course, drug treatments, or genetic manipulations. We have defined the transcriptome induced in macrophages by a variety of TLR agonists (Fig. 2). The data clearly demonstrates that individual TLRs induce both common and individual regulatory networks within the cell. This suggests a mechanism whereby TLRs translate pathogen identity into an appropriate host inflammatory response.

Transcriptional regulatory networks can be defined by integrating gene expression data, obtained using microarrays, with transcription factor binding data, obtained using chromatin immunoprecipitation and *cis*-regulatory element microarrays (ChIP-chip technology) [19]. These global approaches have to date been most successful in yeast. However, the availability of the genomic sequences of mouse and human permits the extension of these methods to the definition of the gene networks that regulate immune and inflammatory responses.

6. Proteomics

Proteomics can be defined as the systematic analysis of all the proteins expressed by a cell or a tissue. Quantitative measurements are particularly useful for the study of biological systems and pathways, since they reveal dynamic changes in the proteome. The traditional method for quantitative proteome analysis combines protein separation by high resolution two-dimensional gel electrophoresis with mass spectrometric or tandem mass spectrometric analysis (2DE/MS). In this method protein quantification is achieved by recording the staining intensities of the separated protein species.

Ruedi Aebersold and colleagues from the Institute for Systems Biology recently developed a new experimental strategy for quantitative proteomics that alleviates most of the limitations inherent in the 2DE/MS method [20]. It quantifies the relative abundance and identifies every protein present in two or more samples, even if the proteins are of low abundance. This method is based on a class of new chemical reagents termed isotope coded affinity tags (ICAT), MS/MS, and a suite of software tools for data analysis.

The ICAT strategy is schematically illustrated in Fig. 3. Protein mixtures from macrophages, either stimulated with LPS or unstimulated, are treated with the ICAT^{TM} reagent (Fig. 2A). The reagent consists of three functional groups. The first is a protein reactive group, which is used to covalently attach the reagent at a specific site in the protein. The second is a linker group that exists in two isotopic forms, light (d0) and heavy (d8). The third is an affinity tag, which is used to selectively extract the reagent–peptide conjugates from the sample mixture. The protein samples to be compared are derivatized with either the heavy or light form of the ICAT reagent, proteolyzed, and the resulting peptides enriched on an affinity column (Fig. 2B). These relative abundance and identity of each of the peptides are then measured by mass spectrometry. The procedure can be automated and multiplexed. A number of ICAT^{TM} reagents with different reactive group specificities have been developed. This permits the comparative quantitation of all proteins, phosphoproteins, glycoproteins, and specific proteases (Fig. 2C).

The ICAT^{TM} reagents can be applied broadly to analyze immunity and inflammation. We are using this method to identify TLR-dependent changes in secreted proteins, membrane proteins, phagosomal proteins, nuclear proteins, signaling proteins, and signaling complexes.

6.1. Membrane and secreted proteins

The glyco-ICAT strategy specifically quantifies and identifies secreted and membrane glycoproteins [21]. The carbohydrate groups of proteins are oxidized and covalently attached to hydrazide groups on a solid support. The immobilized, purified, glycoproteins are then trypsinized, and the resulting glycopeptides are labeled with either of the

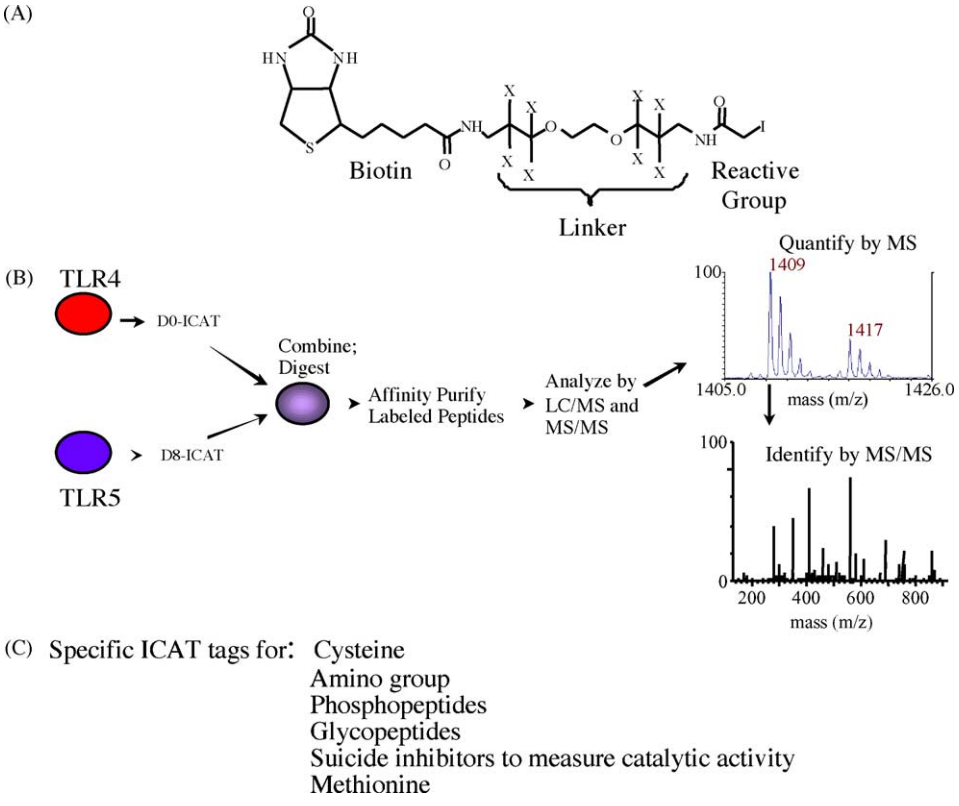

Fig. 3. Quantitative analysis of proteomes, using isotope-coded affinity tag (ICAT^{TM}) technology. (A) The ICAT reagent structure. (B) The ICAT procedure. (C) Protein reactive group specific reagents.

heavy or light ICAT reagent. These peptides are released enzymatically and analyzed using tandem mass spectrometry (MS/MS). This method is particularly important when analyzing secreted proteins since it permits the removal of albumin, and other high abundance, non-glycosylated protein, thereby facilitating deeper protein coverage.

6.2. Phosphoproteins and signaling complexes

The phospho-ICAT reagent is particularly useful in delineating signaling pathways; the procedure permits the quantitative comparison and identification of dynamic protein phosphorylation during macrophage activation [22].

We have used tandem affinity purification together with ICAT reagents and mass spectrometry to identify novel interacting partners within signaling complexes. This strategy is appealing since it circumvents the artifacts associated with the yeast 2-hybrid system, and permits dynamic quantification and identification of proteins without need for gel separation. The experimental approach involves expressing epitope-tagged versions of the bait protein in macrophages, and affinity purifying the bait and associated proteins after stimulating the cells with an appropriate TLR ligand.

We have found that tandem affinity purification (TAP) pioneered by Bertrand Seraphin and colleagues, is required to minimize the noise derived from molecules associating non-specifically. The bait and its associated molecules are eluted and the peptides identified and quantitated by their tandem mass spectra (Fig. 4). We have validated the method by demonstrating the dynamic association of Myd88 and Tirap upon activation of TLR2 (Fig. 4). Studies are underway to comprehensively catalog protein signaling complexes that are assembled in response to TLR activation.

7. Computational approaches to defining complex interactions in the immune response

As discussed, macrophage activation is a result of complex dynamic behavior. This includes positive and negative feedback loops, cross-connections between pathways and modules, kinetic effects such as competitive binding, and genetic variation. Powerful computational approaches are required in order to make sense of this complexity. In particular, it is important to display the multiple signaling modules, and the transcriptional regulatory networks, that mediate common and distinct TLR-dependent pathways.

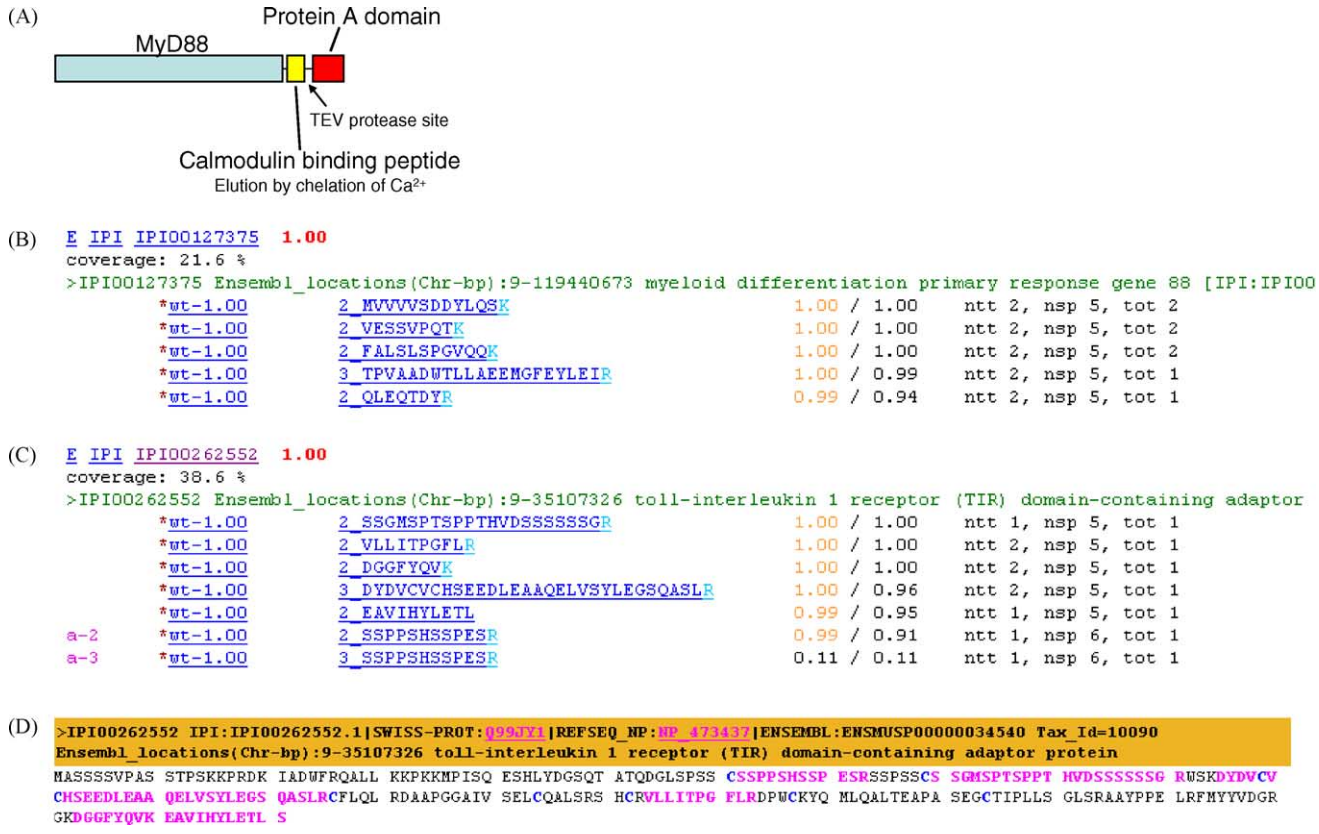

Fig. 4. Tandem affinity purification (TAP) procedure. (A) TAP-epitope tagged Myd88 was expressed in RAW 264.7 macrophages. (B) After immunoprecipitation, peptides derived from Myd88-TAP were identified by mass spectrometry. (C) After stimulating macrophages with lipopetide for 10 min, Tirap was identified to co-precipitate with Myd88. (D) Six peptides, corresponding to 38% of the Tirap protein were identified by mass spectrometry. The interaction of Myd88 with Tirap was not detected in unstimulated macrophages.

This integrative network analysis defines regulatory constriction points within the pathways and predicts the effect of perturbation on the network.

To reconstruct and understand both signal transduction and transcriptional regulatory networks, it is necessary to develop:

- A list of all the components of the system. This step is typically carried out with high-throughput global technologies and the resulting information is stored in a local database; in the case of the ISB it is called SBEAMS.
- A map of the network of interactions among the components. The information stored within SBEAMS is then analyzed, and the network structure can be visualized with graph handling and analysis software such as Cytoscape (http://www.cytoscape.org).
- An understanding of the nature of interactions among the parts. Such a description can be at various levels of abstraction, depending on the data available. For instance, one may describe interactions in terms of Boolean logic of the type: "if ligand L is present, then receptor R is activated" (summarized as "if L then R" in the Boolean formalism). In enzymatic networks, the kinetics of reactions frequently affect the behavior of the network as a whole. In such cases, it is necessary to describe the average behavior of chemical reactions using mass action kinetics. Often, there is good reason to suspect considerable variability between cells. In such cases, average behavior models can be misleading and it may be desirable and necessary to model the interactions of interest in terms of individual, stochastic molecular events.
- A model of how the interactions specified in the above three steps, result in overall system behaviors experimentally observed in wild-type and perturbed cells.

8. SBEAMS: a systems biology database

SBEAMS is a software and database framework for collecting, storing, and accessing different types of experimental data (http://www.sbeams.org/). This system combines a relational database management system (RDBMS) back end, a collection of tools to store, manage, and query experiment information and results in the RDBMS, a web front end for querying the database and providing integrated access to remote data sources, and an interface to other data processing and analysis programs. All data are organized in a modular schema in the RDBMS using similar designs to simplify quality control, data analysis, and data integration. Investigators may use web-based tools or custom scripts to correlate, explore, and annotate the experimental results.

SBEAMS is a modular framework wherein each module can operate independently of the others. The current major modules provide support for microarrays, proteomics, molecular interactions, histology, phenotyping, genotyping, and EST clustering. A single-user project, as managed by the core, may include data from one or several of the SBEAMS modules; for example, Fig. 5 shows the SBEAMS core and the microarray and proteomics modules.

To manage molecular-interaction information contributing to a systems-level understanding of macrophage activation, we developed the module SBEAMS-Interactions. This module provides a database and interface for curating protein–protein and other types of interactions obtained from the literature, other databases, ISB experiments, and algorithms for inferring interactions. The web interface also facilitates queries, such as commands to display only those interactions that are of interest at a particular time. Query results can be automatically piped to Cytoscape for data integration, graphical visualization and exploration (Fig. 5).

9. Data visualization using Cytoscape

Visual data integration plays an important role in the following components of systems biology. First, it displays complex information as an organized representation that is intelligible. Second, the displayed data is interpreted and integrated with the existing world knowledge base derived from global databases and the literature. This suggests additional testable hypotheses that generate more data, which, when integrated into the model generates more testable hypotheses. This leads to an iterative process, in which the choice of successive experiments is driven by the reevaluation of the previous model (Fig. 6). In this way the model is continuously refined.

We will use the TLR pathway to demonstrate some of the functions and capabilities of Cytoscape (http://www.cytoscape.org). Fig. 7A shows a partial interaction network for macrophage activation (approximately 600 interactions are shown). The squares are nodes, and represent individual molecules, and the lines are edges that represent the interactions between them. In this example, we are integrating LPS and CpG induced gene expression data with the interaction network. We selected the nearest neighbors to TNF; this brings up the nodes and edges in a new window; Fig. 7B demonstrates the view for LPS induction, whereas Fig. 7C shows CpG induction. The data shown are relative to unstimulated macrophages. The colors of the nodes designate relative changes in gene expression; green designates induced genes whereas red designated repressed genes. The thickness of the edges corresponds to the level of confidence in the interaction (not shown). The edges can be depicted in many different forms to convey additional information; for example arrows can show directionality of the interaction. Pathways and molecules that are differentially modulated are easy to discern. Imbedded within each node and edge is an enormous amount of additional data and links all of which are available with a click of a mouse, but demonstrating this is beyond the scope of this review.

A number of additional programs have been written that interact with Cytoscape or that serve as plug-ins.

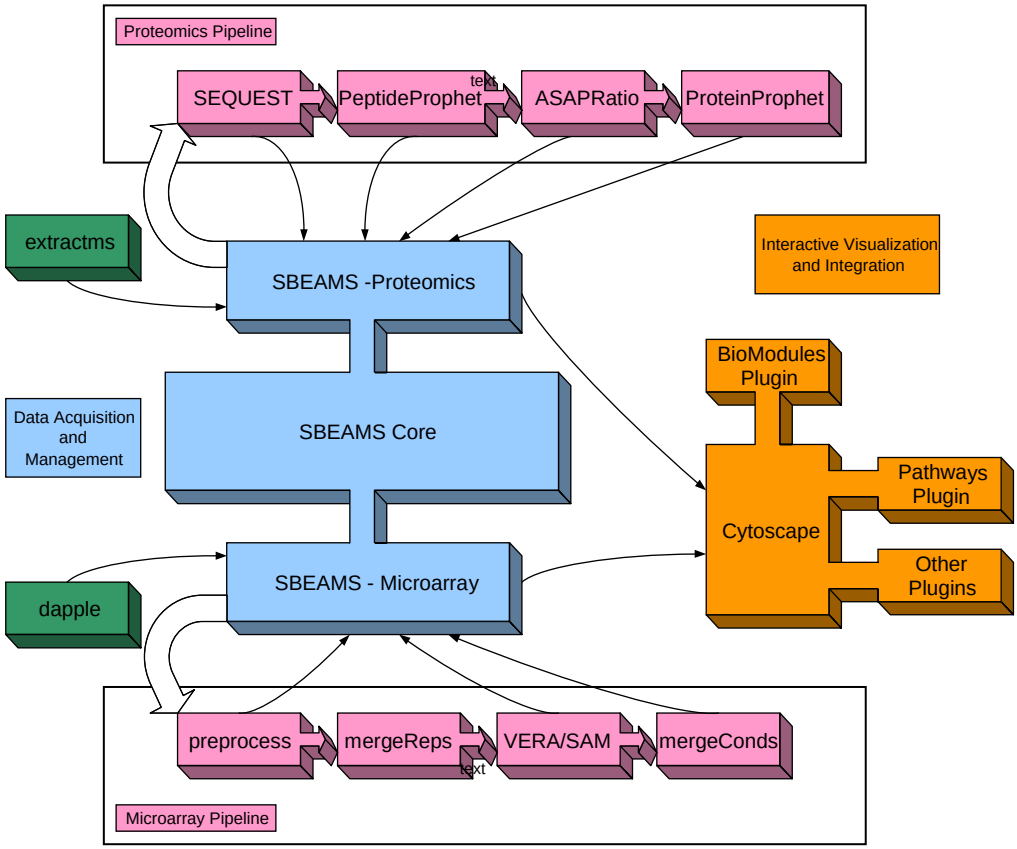

Fig. 5. Architecture of SBEAMS and Cytoscape; integration of data acquisition, management and analysis tools. Two modules are demonstrated; the proteomic pipeline and the microarray pipeline. In the proteomics pipeline, Sequest is a high throughput, scalable, customizable sequence database search engine for tandem mass spectrometry data. Peptide prophet is a statistical program that validates peptide identifications made by tandem mass spectrometry. Protein prophet is a statistical program that validates identification at the protein level. SBEAMS data is the piped into Cytoscape for visualization.

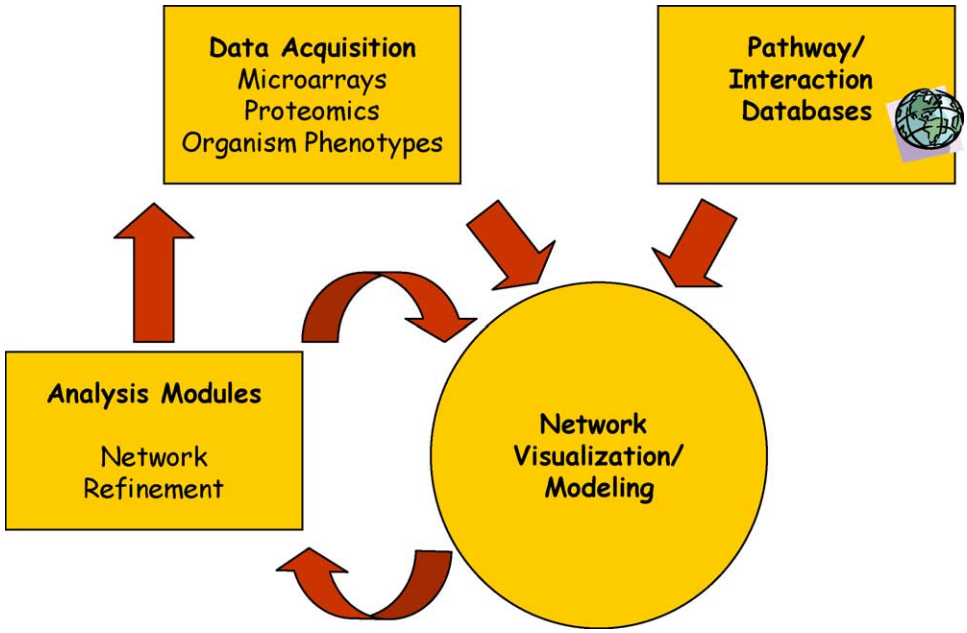

Fig. 6. Network modeling by iterative refinement.

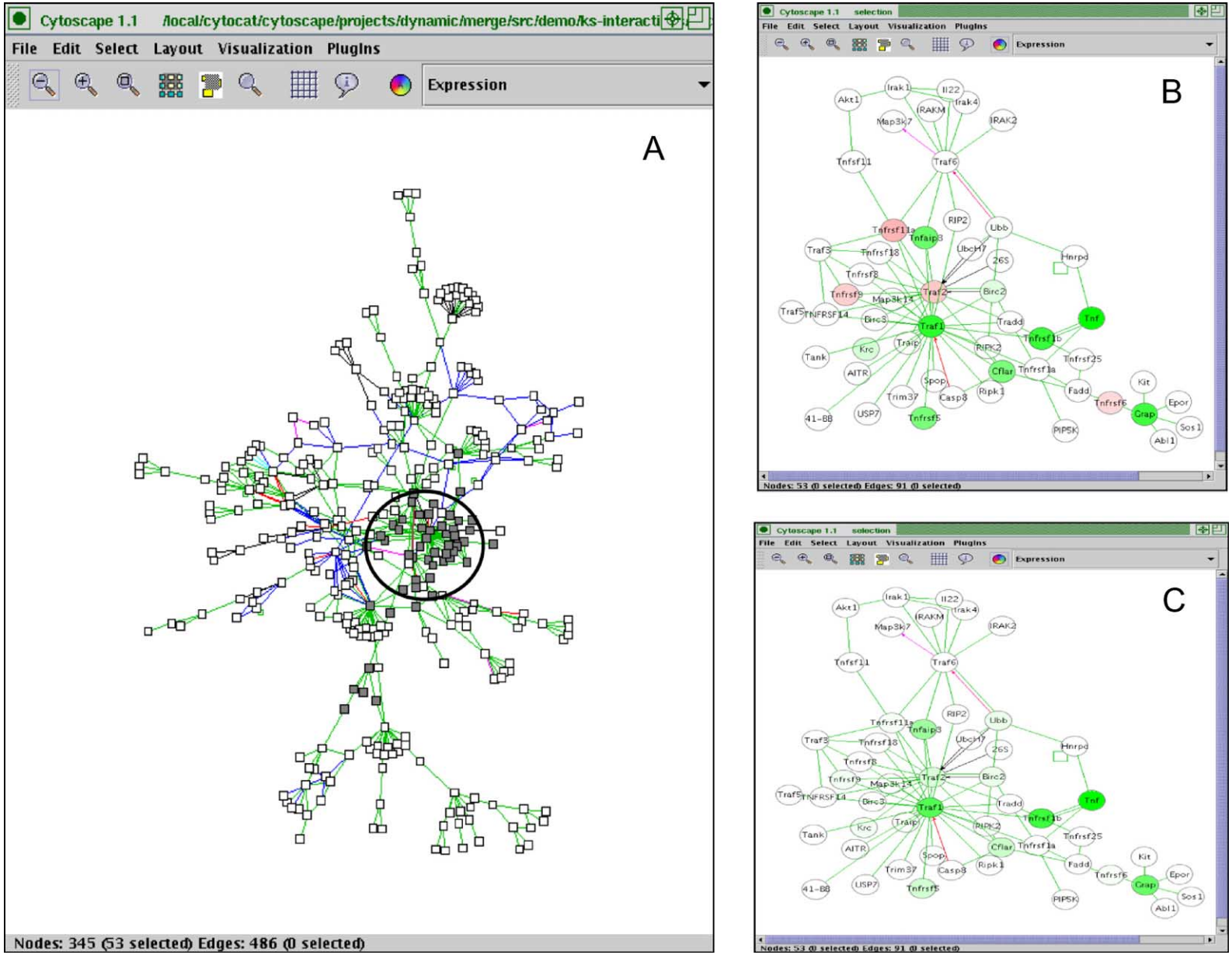

Fig. 7. Integration of molecular interactions with gene expression data using Cytoscape. (A) Cytoscape view of the macrophage activation interaction network. A close-up of the network (circled region), shows LPS (B) and CpG (C) regulated genes (green: induced, and red: repressed) within the context of the molecular interaction network.

9.1. Modular structure of bionetworks

Our genome-scale understanding of immunity suggests a hierarchical view of the cell in which groups of interacting molecules form biological modules, and biological modules interact in complex networks that control the properties of a cell. Biological hypothesis generation is an inherently integrative process in which insight is derived from the global context of interacting biological processes, i.e. modules. Scientists use this approach all the time, but their analysis is subjective and reflects their own personal biases. In addition, the human mind is not capable of processing thousand of variables. The loose molecular associations in networks can be identified computationally, thus abstracting the molecular network into a modular network [23]. The "modular structure" plug-in of Cytoscape uses algorithms to extract from the data a simplified module that is both unbiased and useful in that it leads to the generation of hypotheses.

9.2. Gene regulatory network discovery

The enzymatic networks underlying signal transduction and cellular physiology commonly operate on time scales of the order of seconds. In contrast, the transcriptional regulatory networks that are responsible for changing the state of a cell, for example during LPS-dependent macrophage activation, typically operate on time scales that are two to three orders of magnitude slower.

As a result, viewed from the perspective of genes, signaling events often appear instantaneous, while viewed from the perspective of enzymatic reactions, changes in gene transcription levels happen so slowly as to be insignificant. This observation simplifies modeling of cellular processes by uncoupling the behaviors of enzymatic and genetic networks.

A gene regulatory network is derived from multiple data sources [24]. Microarrays are used to define co-expressed genes; and candidate transcription factors are used for ChIP-chip analysis. A protein–protein interaction map is used to augment the protein–DNA map. To verify DNA–protein interactions determined with global technologies, algorithms have been developed to identify and compare putative transcription factor binding sites in co-regulated genes. This structural information is used to drive "*cis*-regulatory analysis," leading to a model of transcriptional regulation of inflammatory responses.

9.3. Comparative genomics

Comparative genomics is a powerful tool in defining gene function. When complete, this plug-in will compare sequences across species in order to identify orthologs, homologs, conserved domains and genomic regulatory sequences. These analyses provide three basic insights into the biology of the system. First is the identification of evolutionarily conserved genes and modules within the system. Second, protein function or domain function can be inuited from studies performed in model organisms leading to hypothetical assignment of function. Third, these gene comparisons will aid in the identification of genomic regulatory elements that comprise transcriptional regulatory networks.

9.4. Structure explorer

As discussed above, many protein functions and motifs can be assigned by primary sequence analysis, using a variety of different algorithms. However, when the primary sequence does not suggest functional motifs, we have found that predicted structure can lead to the identification of structural homologues that suggest function. Structure explorer uses a combination of a vast library of short known structural motifs and energy minimization principles to predict protein folds.

9.5. Intelligent integration of global knowledge into SBEAMS/Cytoscape

It is important to mine relevant data that exists within the world literature and in the public databases. For example, when a particular gene is identified, this plug-in executes searches that extract all known information about the gene from public databases, and then integrates the information into the working model generated by Cytoscape. The acquisition of information from database is relatively straightforward and can be automated. Much more difficult is the extraction of the text based information found within PubMed. Importantly, this data must be evaluated before curation and annotation. Currently this is hand curated, but with the advent of sophisticated artificial intelligence algorithms this task can be partially automated. Critical to this effort will be the involvement of the Journals and scientific community. A prerequisite for publication might be the extraction of the relevant data by the authors, editors and reviewers for submission to a central database. For example, the molecule database managed by the Alliance for Cell Signaling and hosted by Nature could serve as a model, which when expanded and modified might meet these needs.

10. Concluding remarks

Systems biology approaches are necessary for a complete understanding of the innate immune system and the inflammatory response. These systems wide approaches are now feasible due to availability of complete genomic sequences and high throughput global technologies. Most importantly, new computational strategies need to be developed to make sense of the mountains of data that are generated using these technologies. At the ISB we have developed a suite of programs (Fig. 8) that serve as a starting point to meet this need. These approaches generate complete molecular description of complex biological events and predict the behavior of the system. The models lay the foundation for predictive,

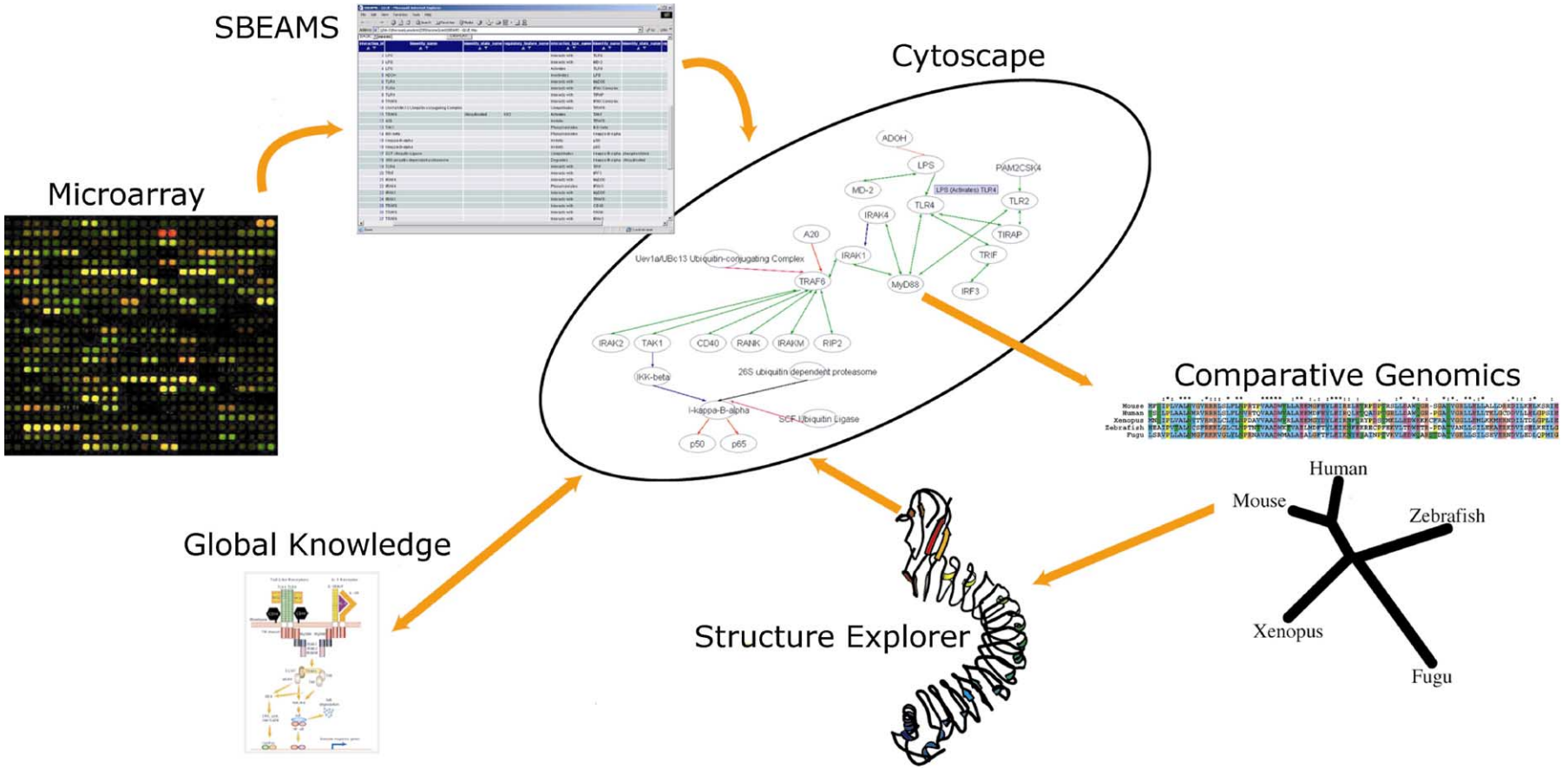

Fig. 8. Computational integration, data analysis, visualization, and model refinement. High-throughput data is processed and transcribed into a structured database (SBEAMS). Database information is translated into graphical forms suitable for input to human intuition (Cytoscape). Computational analysis of each node within a network adds interpretative value and aids iterative refinement of the network. Comparative genomics and structure explorer enrich functional understanding, as described in the text. Cytoscape, and SBEAMS extract global knowledge to integrate with experimental results.

preventive, and personalized medicine. Thus, human genetic variation predicts susceptibility to disease, and a complete knowledge of the pathways leading to disease will serve to design diagnostics for disease progression and rationally define drug targets.

Acknowledgements

The article relies on a large number of scientific contributions and papers. Because of page constraints we have referenced review articles, which hopefully, will lead the reader to the appropriate citations. We apologize to those whose work have been referred to, but have not been cited. We thank our colleagues at the Institute for Systems Biology for their helpful discussions and contributions, especially Drs. Hamid Bolouri, Eric Deutsch, Tim Galitski, Adrian Ozinsky, Javed Roach, Paul Shannon and Vesteinn Thorsson.

References

[1] Fearo DT, Locksley RM. The instructive role of innate immunity in the acquired immune response. Science 1996;272:50.
[2] Janeway Jr CA, Medzhitov R. Innate immune recognition. Annu Rev Immunol 2002;20:197.
[3] Aderem A, Hood L. Immunology in the post-genomic era. Nat Immunol 2001;2:373.
[4] Takeda K, Kaisho T, Akira S. Toll-like receptors. Annu Rev Immunol 2003;21:335.
[5] Gordon S. Pattern recognition receptors: doubling up for the innate immune response. Cell 2002;111:927.
[6] Platt N, Haworth R, Darley L, Gordon S. The many roles of the class A macrophage scavenger receptor. Int Rev Cytol 2002;212:1.
[7] Aderem A, Underhill DM. Mechanisms of phagocytosis in macrophages. Annu Rev Immunol 1999;17:593.
[8] Ernst JD. Macrophage receptors for *Mycobacterium tuberculosis*. Infect Immun 1998;66:1277.
[9] Akira S. Toll-like receptor signaling. J Biol Chem 2003;30:30.
[10] Savill J, Dransfield I, Gregory C, Haslett C. A blast from the past: clearance of apoptotic cells regulates immune responses. Nat Rev Immunol 2002;2:965.
[11] Ravetch JV, Lanier LL. Immune inhibitory receptors. Science 2000;290:84.
[12] Sato S, Takeuchi O, Fujita T, Tomizawa H, Takeda K, Akira S. A variety of microbial components induce tolerance to lipopolysaccharide by differentially affecting MyD88-dependent and -independent pathways. Int Immunol 2002;14:783.
[13] O'Neill LA. SIGIRR puts the brakes on Toll-like receptors. Nat Immunol 2003;4:823.
[14] Hoebe K, Du X, Georgel P, Janssen E, Tabeta K, Kim SO, et al. Identification of Lps2 as a key transducer of MyD88-independent TIR signalling. Nature 2003;424:743.
[15] Yamamoto M, Sato S, Hemmi H, Hoshino K, Kaisho T, Sanjo H, et al. Role of adaptor TRIF in the MyD88-independent Toll-like receptor signaling pathway. Science 2003;301:640.
[16] Inohara N, Nunez G. NODs: intracellular proteins involved in inflammation and apoptosis. Nat Rev Immunol 2003;3:371.
[17] Kiechl S, Lorenz E, Reindl M, Wiedermann CJ, Oberhollenzer F, Bonora E, et al. Toll-like receptor 4 polymorphisms and atherogenesis. N Engl J Med 2002;347:185.
[18] Smirnova I, Mann N, Dols A, Derkx HH, Hibberd ML, Levin M, et al. Assay of locus-specific genetic load implicates rare Toll-like receptor 4 mutations in meningococcal susceptibility. Proc Natl Acad Sci USA 2003;100:6075.
[19] Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 2002;298:799.
[20] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;422:198.
[21] 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.
[22] Zhou H, Watts JD, Aebersold R. A systematic approach to the analysis of protein phosphorylation. Nat Biotechnol 2001;19:375.
[23] Rives AW, Galitski T. Modular organization of cellular networks. Proc Natl Acad Sci USA 2003;100:1128.
[24] Bolouri H, Davidson EH. Modeling transcriptional regulatory networks. Bioessays 2002;24:1118.