

# Describing Biological Protein Interactions in Terms of Protein States and State Transitions

THE LiveDIP DATABASE\*

Xiaoqun Joyce Duan, Ioannis Xenarios, and David Eisenberg‡

Biological protein-protein interactions differ from the more general class of physical interactions; in a biological interaction, both proteins must be in their proper states (e.g. covalently modified state, conformational state, cellular location state, etc.). Also in every biological interaction, one or both interacting molecules undergo a transition to a new state. This regulation of protein states through protein-protein interactions underlies many dynamic biological processes inside cells. Therefore, understanding biological interactions requires information on protein states. Toward this goal, DIP (the Database of Interacting Proteins) has been expanded to LiveDIP, which describes protein interactions by protein states and state transitions. This additional level of characterization permits a more complete picture of the protein-protein interaction networks and is crucial to an integrated understanding of genome-scale biology. The search tools provided by LiveDIP, Pathfinder, and Batch Search allow users to assemble biological pathways from all the protein-protein interactions collated from the scientific literature in LiveDIP. Tools have also been developed to integrate the protein-protein interaction networks of LiveDIP with large scale genomic data such as microarray data. An example of these tools applied to analyzing the pheromone response pathway in yeast suggests that the pathway functions in the context of a complex protein-protein interaction network. Seven of the eleven proteins involved in signal transduction are under negative or positive regulation of up to five other proteins through biological protein-protein interactions. During pheromone response, the mRNA expression levels of these signaling proteins exhibit different time course profiles. There is no simple correlation between changes in transcription levels and the signal intensity. This points to the importance of proteomic studies to understand how cells modulate and integrate signals. Integrating large scale, yeast two-hybrid data with mRNA expression data suggests biological interactions that may participate in pheromone response. These examples illustrate how LiveDIP provides data and tools for biological pathway discovery and pathway analysis. *Molecular & Cellular Proteomics* 1:104–116, 2002.

Understanding protein-protein interactions is crucial to integrated biology. The availability of the genome sequences of 100 species, ranging from bacteria to human, poses to biologists two major challenges of interpreting this blueprint of life, 1) understanding the function of each gene product, and 2) understanding phenotypes through the biological interactions of the gene products. Large scale and high throughput experimental techniques have been developed to address these questions by acquiring data on the whole genome, instead of just a few genes. One such example, an integrated genomic and proteomic analysis of a systematically perturbed yeast galactose utilization pathway, suggested the importance of analyzing both mRNAs and proteins for understanding biological systems (1). Protein-protein interactions are a crucial component of such integrated biology. The classical view of protein function focuses on the action of a single protein molecule, its biochemical activity or molecular function. An expanded view defines the function of a protein in the context of its network of interactions (2). Each protein interacts with several partners that also interact with other proteins. All these interactions connect proteins into an extensive and complex web. Every protein functions in the context of this web of interacting molecules, and its interactions with other molecules define how its biochemical activities are utilized and regulated in the related biological processes.

Genome-scale analysis of biological systems requires easy access to information on a large number of protein-protein interactions (1). Much of this information is embedded within many scientific articles and is hard to retrieve in an organized fashion. These facts call for new ways of representing such biological knowledge, as also noted by other researchers (3, 4). Several databases have been developed that provide web-accessible information on protein-protein interactions, as summarized in Table I.

A large proportion of known protein-protein interactions were detected by genome-scale yeast two-hybrid assays (5). These interactions indicate association of proteins for a certain period of time and can be termed physical protein interactions. Physical interactions may include nonspecific interactions, which do not have biological significance. A biologically important interaction requires the interacting partners to be in specific protein states. It also causes transitions in the states of one or both of the interacting partners (such as phosphorylation and/or activation). A biological interaction regulates the function of the interacting proteins or transmits a signal from one protein to another and underlies all the dynamic biological processes in cells. Such interactions can be termed biological protein interactions. DIP, the Database of Interacting Proteins

From the Howard Hughes Medical Institute, UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles, Los Angeles, California 90095-1570

Received, October 29, 2001

Published, MCP Papers in Press, January 9, 2002, DOI 10.1074/mcp.M100026-MCP200

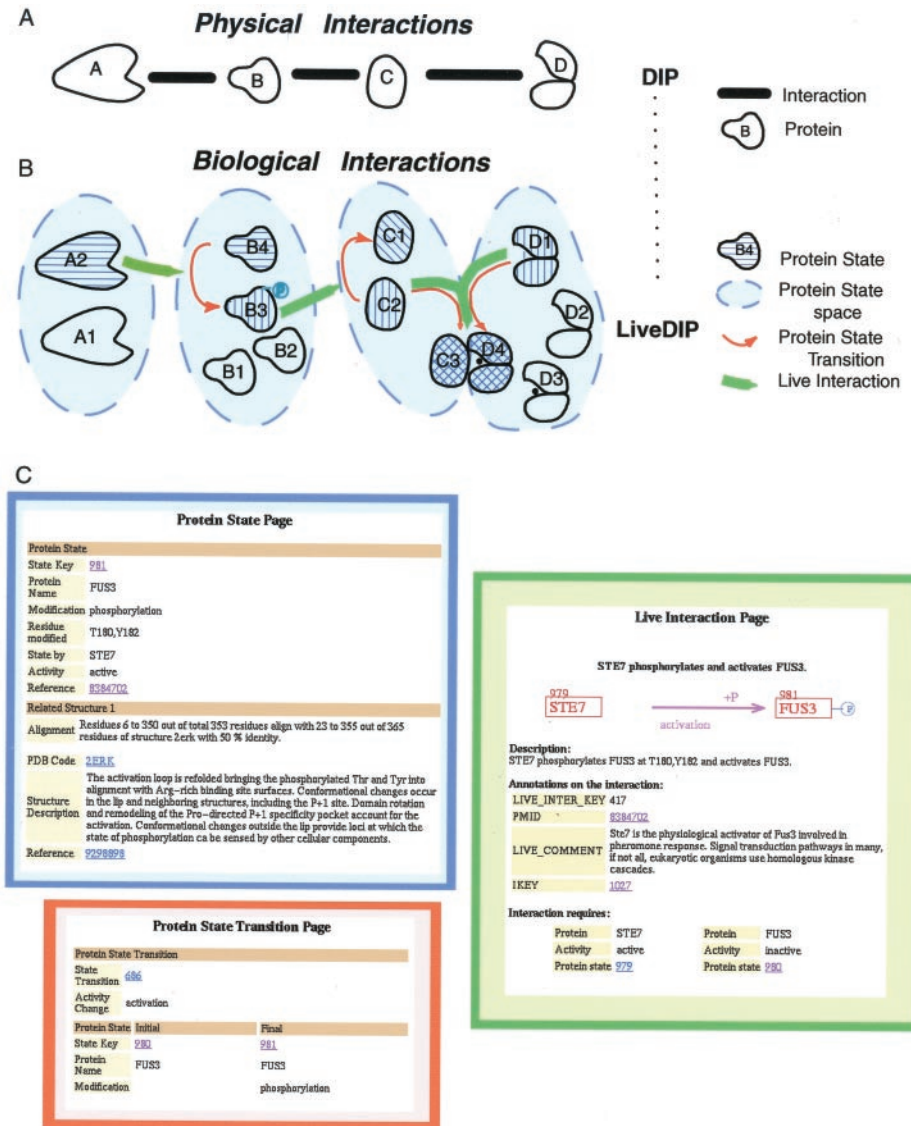
TABLE I  
A brief summary of web-accessible databases containing information on protein-protein interactions

Database and URL	Protein-Protein interaction-related contents	Unique features	Comments	Data queries
Alliance for Cellular Signaling (AFCS) www.cellularsignaling.org/	"Mini Molecule Pages" contain literature-derived information about signaling molecules including interactions.	Details about signaling pathways studied by the alliance; recruit members from scientific community to author individual ACFC molecule page.	Part of a concerted multidisciplinary research alliance to understand cell signaling using two selected model systems. Only molecules and pathways of interest to the Alliance are included in the database. The signaling maps are static and drawn individually and may not be complete.	Browse or search protein list; browse pathway maps.
Amaze www.ebi.ac.uk/research/amaze/	Clickable signaling maps and legends provide information about signaling pathways. Pathway graphs.	Prototype pathway analysis tools.	A data model to deal with information on metabolic pathways, gene regulation, sub- or supracellular locations, and transport.	Browse pathway graphs; prototype pathway analysis tools.
Biomolecular Interaction Network Database (BIND) www.biondb.org	The three main data types are interactions, molecular complexes, and pathways.	BIND BLAST; blast search against molecules in BIND. PreBIND, a data mining tool locating papers that may contain information about interactions related to a query protein. Visualize-interaction, Java applet drawing schematic diagram interactions.	A complete data model designed to store full descriptions of interactions or chemical reactions between any two molecules of any type.	Search interactions, complexes, or pathways by protein names or text queries.
Cell Signaling Networks Database (CSNDB) geo.nih.gov/jp/csndb	Database models consist of cell signaling, paper, and signal molecule.	Pathway finder, find the pathway around a target or between two target molecules.	Compiles the information on biological molecules, sequences, structures, functions, and biological reactions that transfer the cellular signals in human cells.	Query by keyword; browse by molecules. Pathway finder.
Database of Interacting Proteins (DIP and LiveDIP) dip.doe-mbi.ucla.edu/	DIP contains data on proteins, interactions, and experimental methods. The major components of LiveDIP are Protein State Pages, Protein State Transition Pages, and Live Interaction Pages.	Pathfinder and batch search, find path connecting query molecules, and automatically draw the Interaction Map. Map expression data, display expression data schematically on the Interaction Map. JDIP, a stand-alone Java application that provides a graphical, browser-independent interface to the DIP database.	DIP and its extensions emphasize collecting experimentally determined protein-protein interactions and providing tools to explore and visualize the network of interactions, the supporting experimental evidence, and tools to integrate functional genomic data with protein-protein interactions.	Browse protein state; search interactions; Pathfinder; batch search.
Kyoto Encyclopedia of Genes and Genomes (KEGG) kegg.genome.ad.jp/kegg/	Gene annotation and pathway information.	Ortholog group tables contain information about conserved subpathways (pathway motifs).	PATHWAY database contains graphical representations of cellular processes, such as metabolism, membrane transport, signal transduction, and cell cycle.	Search or browse.

TABLE 1—Continued

Database and URL	Protein-Protein interaction-related contents	Unique features	Comments	Data queries
Proteome www.proteome.com	Protein Report pages contain detailed annotation on proteins including protein-protein interactions.	Extensive and well-referenced information on proteins collated from scientific literature.	Extensive information on proteins.	Search by keywords, browse by various categories.
Signaling pathway database (SPAD) www.grt.kyushu-u.ac.jp/spad/	Clickable pathway map and protein information page.	Browse by extracellular signal molecules; signaling pathway map.	SPAD is compiled to describe information on interaction between protein and protein and protein and DNA, as well as information on sequences of DNA and proteins.	Browse by extracellular signal molecules.
Science's signal transduction knowledge environment (STKE) stke.sciencemag.org	Connection maps, connections, and component pages.	Canonical pathways document basic properties of signaling modules, and specific pathways describe specific components and relations in a particular organism, tissue, or cell type.	Emphasizes information supplied by authorities in the field, provides information management tools to organize all kinds of knowledge related to signaling transduction including protein-protein interactions.	Search by query string; browse by connections map.
Transpath transpath.gbf.de/index.html	Molecule, reaction, literature reference pages.	Browseable map of signaling pathways.	An extension module to the TRANSFAC data base on transcription factors and their binding sites. TRANSPATH focuses on pathways involved in the regulation of transcription factors in different species, mainly human, mouse, and rat.	Search by query string; browse pathway map; explore cross-talks.

**FIG. 1. Biological (A) versus physical (B) interactions.** Shown is a schematic illustration of the concepts of the DIP and LiveDIP databases and their relationship to layered information about protein-protein interactions. **A**, in DIP, physical interactions are modeled as binary relationships between pairs of proteins. **B**, LiveDIP is an extension of DIP in which biological interactions are modeled as transitions between protein states caused by the protein-protein interactions. The single protein entity in DIP corresponds to a collection of states in the protein state space of LiveDIP. For example, protein C can exist in any of the protein states C1, C2, or C3; B can exist in B1, B2, B3, or B4. A given protein state (e.g. A2) interacts only with a given state of its interacting partner (B4), represented by pairs of protein states shaded with the same pattern (A2 and B4; B3 and D1). The interaction (green arrow) causes a state transition (red arrow) in one or both of the interacting pairs (B to B3; C2 to C1), and the resulting new state either gains or loses its ability to interact with its other interacting partners: protein B in state B3 interacts with protein C in state C2, and protein C in state C1 can no longer interact with protein D. **C**, the three major types of data in LiveDIP and the interconnections among them. LiveDIP presents information about protein states in the Protein State Page, information about the transition between protein states in the Protein State Transition Page, and information about protein-protein interactions in the Live Interaction Page. Each page is shown here enclosed in a rectangular box with the same color as the corresponding symbols in the B.



that documents experimentally determined protein-protein interactions (5), has now been expanded to LiveDIP to describe biological interactions in terms of protein states and state transitions.

## MATERIALS AND METHODS

LiveDIP is a relational database consisting of eleven tables. Protein-protein interactions and their related transitions among the protein states are described by the three main tables, the protein state table, the state transition table, and the live interaction table. The protein state table stores the detailed information on protein states. It includes description of chemical modification, cellular localization, whether it is part of a complex, the presence of ligands, and what triggers the protein state. Information on transitions between protein states is stored in the state transition table, which stores the keys for the initial state and the final state, and the resultant changes. The live interaction table describes the two protein state transitions of an interacting pair, factors that affect the interaction, and the article reporting this interaction. This database schema defines the prerequisite in the protein states for a given interaction to occur and the

transitions in the protein states resulted from the interaction.

All data in LiveDIP are collated from published literature. Medline abstracts are rated for information on protein-protein interactions (6). Abstracts with high scores are read by curators, and data are entered manually to assure the quality of the data, through HTML forms.

## RESULTS

### The Scheme of LiveDIP and Its Relationship to DIP

**DIP Describes Physical Protein-Protein Interactions**—Protein-protein interactions have different attributes. A protein-protein interaction can be described as a physical association between two proteins, as represented by the black lines linking a pair of symbols in Fig. 1A. A protein may interact with more than one protein. For example, protein B interacts with both protein A and protein C; protein C also interacts with protein D. These proteins may be different in properties such as size or domain structure. For example, proteins A, B, and C are single-domain proteins, and protein D has multiple

TABLE II  
Description of protein states, state transitions, and protein-protein interactions in LiveDIP

Attributes that define protein states. Characterization of a protein state in LiveDIP includes one or several of these attributes, which describe the difference between the protein state and the base state of the protein.

Type of attribute	List of descriptions
Post-translational modification	Autophosphorylation, unphosphorylated, phosphorylation, ubiquitination, multiubiquitination, monoubiquitination, <i>N</i> -acetylation, acetylation, farnesylation, <i>O</i> -glycosylation, <i>N</i> -glycosylation, glycosylation, unglycosylated, methylation, carboxymethylation, <i>N</i> -myristoylation, myristoylation, palmitoylation, unknown
Presence of ligand	+Ca, +glucose, +GTP, +GDP, +heme, +ATP, +ADP
Oligomeric state	Complex, homodimer, homo-oligomer
Cellular localization	Nucleus, cytosol, ER, Schmoo tip, plasma membrane, bud
Proteolytic form	N-terminal cleavage, C-terminal cleavage, cleavage
Protein structure	Open form, close form, conformation 1, conformation 2

Type of changes and their descriptions associated with protein state transitions as categorized in LiveDIP. These changes describe the differences between two protein states.

Type of change	List of descriptions
Molecular activity	Activation, inhibition, stimulation, suppression, regulation
Chemical modification	Autophosphorylation, unphosphorylated, phosphorylation, ubiquitination, multiubiquitination, monoubiquitination, <i>N</i> -acetylation, acetylation, farnesylation, <i>O</i> -glycosylation, <i>N</i> -glycosylation, glycosylation, unglycosylated, methylation, carboxymethylation, <i>N</i> -myristoylation, myristoylation, palmitoylation, unknown
Protein stability	Stabilization, destabilization
Subcellular localization	Localization, transport, translocalization
Oligomeric state	Complex formation, complex dissociation
Structure	Conformational change
Proteolytic form	Cleavage
Others	Targeting

Example description of biological interactions causing transitions in the protein states as represented in LiveDIP. A protein state transition may be characterized by one or several types of changes listed above.

Example interaction	Transition in the protein states	Description of change
Ste7 phosphorylates and activates Fus3.	Unphosphorylated, inactive Fus3 → phosphorylated, active Fus3	Activation, phosphorylation
Cln1 forms complex with Cdc28.	Cln1 → Cln1 in complex with Cdc28	Complex formation
Msn5 transports Far1 across the nuclear membrane.	Far1 in the nucleus → Far1 in cytoplasm	Transport
Ste4 stimulates activity of Ste5 by inducing conformational change.	Ste5 active, conformation 1 → Ste5 more active, conformation 2: association of NC termini	Conformational change, stimulation

domains. These interactions can be termed physical interactions, and they may or may not have biological significance. DIP contains data on physical interactions including the identities of the interacting proteins, their domain structures, the ranges of amino acids or domains involved, the binding affinity, and the experimental techniques used to detect these interactions (5). DIP also provides tools to help assess the significance of these interactions.

*LiveDIP Describes Protein-Protein Interactions in Terms of Protein States and State Transitions*—Each protein can exist in different protein states as shown in Fig. 1B. The collection of all states for a given protein forms its protein state space (protein C can be in any of the protein states C1, C2, or C3; B can be in B1, B2, B3, or B4). A protein state is defined by one or a combination of several attributes including post-translational modification (protein state B3), presence or absence of ligand (protein state D3), oligomeric state (part of a complex,

protein states C3 and D4). Other attributes include cellular localization, alternative splicing, proteolytic form, etc. Table II lists the attributes currently included in LiveDIP. The state of a protein determines its activity. For example, some proteins are active when phosphorylated and inactive when dephosphorylated. Post-translational modifications may also stabilize or destabilize the protein or target the protein to certain subcellular location. One example is Gpa1, the G- $\alpha$  subunit of the receptor-coupled G protein, which is targeted to the plasma membrane through dual lipid modification (7). At a given time, the pool of molecules of a protein inside a cell may exist in one or several of its protein states depending on the cellular context.

Another type of information related to protein states is protein three-dimensional structures. Structural information is important for understanding the behaviors of different protein states; e.g. why one state is active and another is not. For example, the phosphorylated state of Erk2, a homologue of

Fus3, causes refolding of the activation lip in the protein and makes the active site accessible. It also causes conformation changes in regions outside the activation lip, through which the phosphorylation state can be sensed by other proteins (8). This type of information about the relationship between protein structures and protein states, including alignments and brief annotations on the structures, is stored in LiveDIP. Other related information in LiveDIP includes what triggers a protein state and the article reporting this state.

Proteins inside a cell are not static; they undergo state transitions (*green arrow* in Fig. 1B) in response to environmental signals or change in cellular context. For example, Ste20 is a kinase of the pheromone response pathway in yeast. In response to pheromone treatment, yeast cells arrest as unbudded, G<sub>1</sub> phase cells. Ste20 from cells of pheromone-arrested cultures is post-translationally modified by phosphorylation. Removal of pheromone results in the appearance of the unphosphorylated form within 30 min (9).

Transition between protein states regulates biological processes including cell signaling. One example is the use of cellular localization as a common mechanism to regulate activity of signaling proteins (10). Proper signaling requires co-localization at the right time of the proteins involved in succeeding steps of a signaling pathway. It was reported that Ste5 shuttles through the nucleus (10). In the presence of pheromone, Ste5 undergoes enhanced export from the nucleus and is recruited to plasma membrane by the G $\beta$  subunit of the G protein receptor and triggers activation of downstream kinase Fus3. This function of the nucleus to sequester proteins destined for the plasma membrane may prevent activation of downstream targets in the absence of signal. Thus, cells can regulate biological processes by modifying protein states. A protein state transition in LiveDIP is characterized by one or several of the changes listed in Table II.

When is a physical interaction (*black lines* in Fig. 1A) a biological interaction (*green arrows* in Fig. 1B)? This requires the interacting partners to be in certain protein states for the interaction to occur and that the interaction causes a transition in the state of one or both of the interacting partners as shown schematically in Fig. 1B. The interaction (*green arrow*) between two proteins can occur only when each protein is in its specific state, represented, respectively, by the two states shaded with the same pattern (A2 interacts with B4 and B3 with D1). Each protein state can participate only in a subset of all the interactions of the protein. The interaction (*green arrow*) causes a state transition (*red arrow*) in one or both members of the interacting pair. The transition can be formation or dissociation of a protein complex (the complex of proteins C and D), changing the states of both proteins (C2 into C3, D1 into D4). Alternatively, one protein can modify its interacting partner and convert its partner into a new state by chemically modifying it (A phosphorylates B and changes its state from B4 into B3) or by other possibly unknown modification (B modifies C and changes its state from C2 to C1). The protein

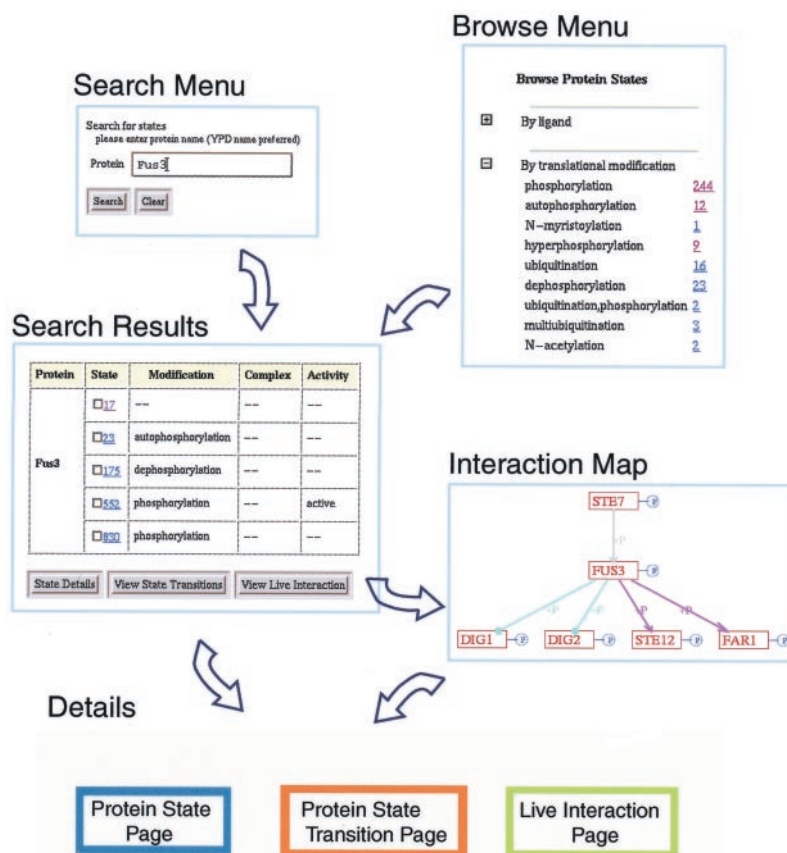
in its new state may gain or lose its ability to interact with its other interacting partners, e.g. A2 interacts with B4 and changes it into B3, which then modifies C1; B3 modifies C2 into C1, which can no longer form a complex with protein D. Through coupling of chemical events and mechanical actions, protein-protein interactions cause transitions in the protein states, regulate the function of the proteins, or transmit a signal from one protein to another. For example, pheromone treatment activates Fus3 by changing it into its phosphorylated state, which subsequently activates transcription factor Ste12 and induces the pheromone response. Msg5, on the other hand, can dephosphorylate Fus3 and change it into the inactive state and is suggested to function in attenuation of the mating signal (11). A Live Interaction entry in LiveDIP describes the state transitions of the interacting partners. Other information includes references, factors that affect the interaction, the cell stage or subcellular localization wherein the interaction was detected, and cellular function of the interaction such as what signaling pathway it belongs to.

In summary, DIP and LiveDIP provide layered descriptions of protein-protein interactions. DIP documents physical interactions as binary relationships, and LiveDIP describes biological interactions as transitions in the protein states of the interacting pairs caused by the interactions. A single protein entity (protein C) in DIP may correspond to several protein states (protein states C1, C2, or C3) in LiveDIP (Fig. 1A). Each interaction in DIP may correspond to several LiveDIP interactions, because different aspects of the same binary interaction may be reported by different research articles.

#### *Presentation of Interactions in LiveDIP*

Information on biological interactions collected by LiveDIP is presented by three major web pages, the Protein State Pages, the Protein State Transition Pages, and the Live Interaction Pages, as shown in Fig. 1B. The Protein State Page describes the different attributes that define a protein state, including chemical modification, presence/absence of ligand, and subcellular location, etc. Other information includes related three-dimensional structures and factors triggering the protein state. A link to the reference in PubMed is provided. Each Protein State Transition Page offers a short description of the beginning and ending protein states and the resultant changes. All protein states in the state transition are linked to their related Protein State Pages, respectively. The Live Interaction Page uses schematic drawings and tables to describe the interaction. The Live Interaction Page is linked to the related Protein State Transition Pages, Protein State Pages, and references in PubMed. Links are also provided to a related entry in DIP, which presents the experimental evidence, interacting domains, and binding affinity. Our aim is to present information about biologically important protein-protein interactions in a form familiar to biologists.

FIG. 2. **Basic Search and Browse in LiveDIP.** Users can find information about protein-protein interactions either by browsing the protein states (Browse Menu) or by performing a query on protein states with a protein name (Search Menu). The Search Results page lists all the protein states found for the query protein in a table, each row providing a short description of one protein state. There are also buttons linking to relevant interacting proteins and protein state transitions. The schematic Interaction Map shows all proteins interacting with a selected protein state, with *rectangular boxes* representing protein states, *cyan lines ending on circles* representing inhibition, and *magenta lines with arrows* representing activation. All the entries in these pages are linked to the relevant Protein State pages, Protein State Transition pages, and Live Interaction pages as denoted by the boxes following the color scheme of Fig. 1.



#### Database Statistics

The September 2001 release of LiveDIP contains 304 unique proteins, 35 types of chemical modifications, and 408 interactions. All these data were collected from 341 papers.

#### Data Queries

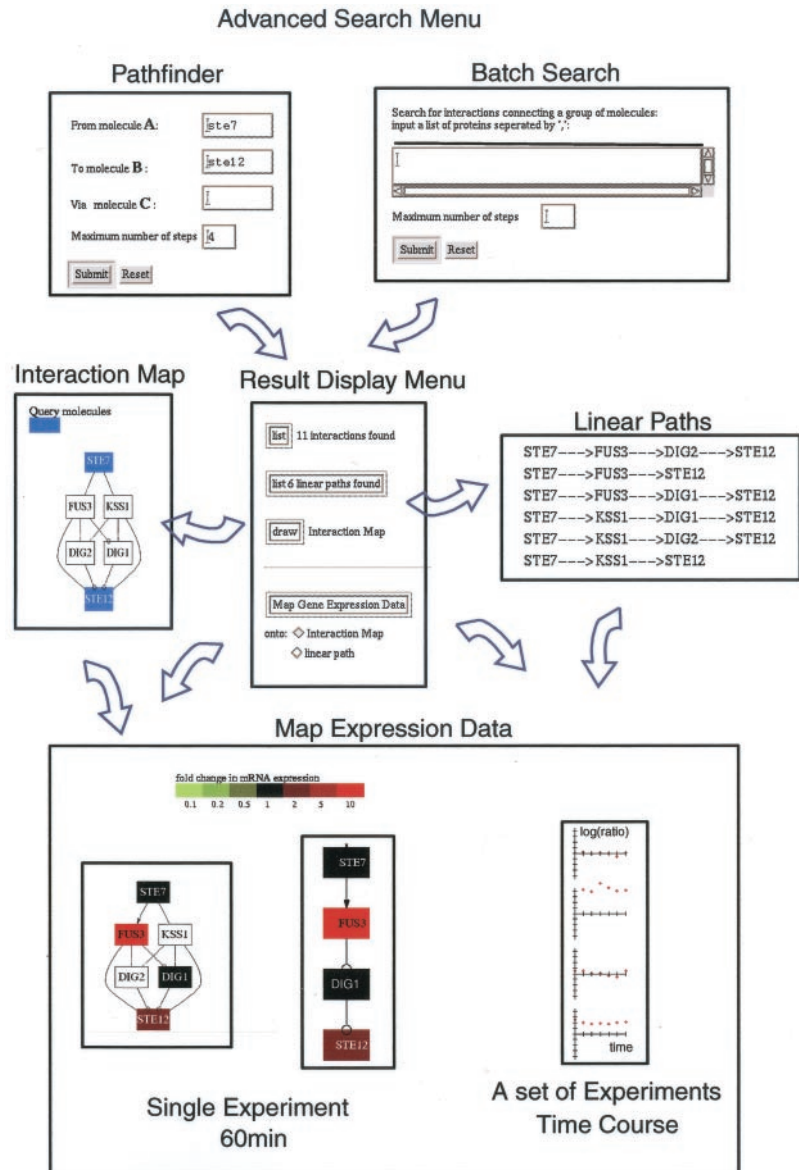
Users can enter the database by browsing protein states by the types of post-translational modifications or ligands bound. Alternatively, one may query the protein states with a protein name (Fig. 2). The Search Results page lists all the protein states found in a table, each row displaying a short description of one protein state. The View Live Interaction button leads to the Interaction Map. It shows interactions related to the selected protein state, with *rectangular boxes* representing protein states, *cyan lines ending on circles* representing inhibition, and *magenta lines with arrows* representing activation. All the entries in these pages are linked to the relevant Protein State Pages, State Transition pages, and Live Interaction Pages as denoted by the boxes following the color scheme of Fig. 1. Advanced search tools have been developed to facilitate pathway analysis and pathway discovery (Fig. 3). The Pathfinder feature searches for paths going from molecule A to molecule B via molecule C with molecules A, B, and C supplied by the user. Users can leave one or two of these molecules unspecified. For example, a query with molecule A unspecified will search for all the paths going from any molecule through molecule B and ending in molecule C. A query

where only molecule B is specified will find all the paths going through molecule B. Also available is the Batch Search feature that searches for all interactions connecting a group of proteins within a given number of steps. Results can be explored in the following several ways as shown in the Result Display Menu: a list of all interactions with links to Live Interaction Pages, a list of all the linear paths, or an Interaction Map automatically drawn and schematically showing all the interactions. Tools are also provided to integrate mRNA gene expression data with protein-protein interaction networks. Users can choose to map data from a single experiment onto the Interaction Map found by advanced search tools or onto a specific path extracted from the search results (see example below). Data from a set of expression experiments such as time course can also be plotted as a line chart for each component of a selected path to analyze temporal regulation of gene expression.

#### Example Applications

**Pathways Function within the Network**—The advanced search features of LiveDIP can help researchers analyze signaling pathways in the context of protein interaction networks as illustrated by the example shown in Fig. 4. Using proteins involved in the pheromone signaling transduction pathway as query molecules (Ste2, Ste18, Ste4, Ste20, Ste5, Ste11, Ste7, Fus3, Dig1, Dig2, Ste12), results of the Batch Search for

**FIG. 3. Tools for analyzing pathways by integrating microarray data with protein-protein interactions.** Two Advanced Search features are provided. The Pathfinder feature searches for protein interactions going from molecule A to molecule B via molecule C (top left box), and the Batch Search feature looks for interactions that connect a group of proteins within a given number of steps (top right box). Users can choose different ways to explore the interactions found through the Result Display Menu (center box, middle row). Results can be represented by listing all the linear paths (middle right box) or by a schematic, automatically drawn Interaction Map, with rectangular boxes representing proteins and lines between them representing interactions (middle left box). The bottom of the figure illustrates how mRNA expression data of a selected experiment can be displayed in the Interaction Map. Boxes representing the corresponding proteins are colored to indicate changes in mRNA expression level, red for increase and green for decrease. Data from a set of experiments, such as time course data, can also be plotted for each component of a selected linear path.



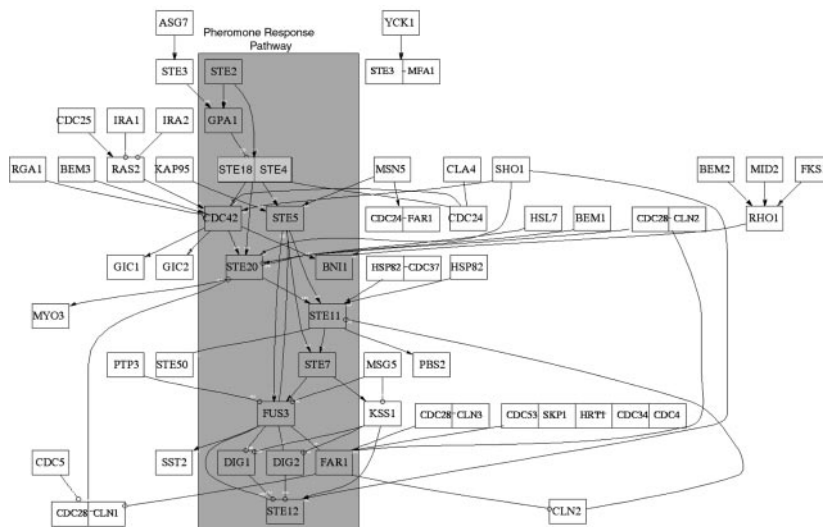
interactions within two steps away from the query molecules are presented by the Interaction Map of Fig. 4. The *small rectangular boxes* represent proteins and complexes. The *lines connecting boxes* represent distinct kinds of interactions; *arrowheads* indicate activating interactions, and *empty circles* indicate inhibiting interactions. The pheromone response pathway is enclosed in a *shaded rectangular box*. One observation from this map is that the pheromone signaling pathway, like so many other pathways in higher organisms, does not work in isolation (12). Seven of the eleven components of the signaling transduction pathway are regulated by other proteins in the current version of LiveDIP, with one interaction each for Dig1 and Dig2 and up to five interactions for Ste20. These interactions change the states of the signaling proteins mainly through phosphorylation or dephosphorylation and modulate their activity, and consequently, inten-

sity of the signal. Phosphatase Msg5, for example, causes transition in the protein state of Fus3 by dephosphorylating and inactivating it and is suggested to be involved in adaptation and recovery of cells from pheromone treatment (11).

The Interaction Map also provides visualization of the cross-talk between different pathways. Signaling pathways rarely operate in isolation; instead the action of an individual signaling pathway often affects and/or is integrated with other pathways to control cell response to outside stimuli. One such example is the cross-talk between the pheromone response pathway and the filamentous growth pathway. Previous studies showed that low concentrations of mating pheromones increases agar-invasive growth of haploid yeast cells and that filamentation reporter genes are activated in fus3 mutant haploid cells (13). The molecular basis of this cross-talk is visualized in the Interaction Map in Fig. 4 where MAPKK Ste7



FIG. 4. **Pathway analysis using the Advance Search feature of LiveDIP, applied to the pheromone signaling pathway of yeast.** The interactions were generated by the Batch Search feature of LiveDIP, using proteins known to be involved in pheromone signaling transduction (enclosed in the *shaded rectangular box*) as query molecules (Ste2, Ste18, Ste4, Ste20, Ste5, Ste11, Ste7, Fus3, Dig1, Dig2, and Ste12). The results are represented by the automatically drawn, schematic Interaction Map. The *small rectangle boxes* depict proteins and complexes, and the *lines* connecting pairs of *boxed* proteins represent different kinds of interactions; *arrowheads* indicate activating interactions, and *empty circles* indicate inhibiting interactions.



activates both Fus3 in the pheromone response pathway and Kss1 in the filamentous growth pathway. During pheromone response, scaffold protein Ste5 is required for activation of Ste7 by Ste11 and activation of Fus3 by Ste7; interactions among Ste11, Ste7, and Fus3 require these proteins to be in the protein states where they are part of a complex with Ste5; activation of Kss1, on the other hand, does not require Ste5 (14). Changes in the protein states of Ste11, Ste7, and Fus3, because of either low dose of pheromone or by mutation of Fus3, may cause misactivation of Kss1 and the cross-talk. LiveDIP provides a tool to map an individual pathway in the protein-protein interaction network. This visualization helps understanding of how signals are modulated by regulating the protein states of the pathway components.

**Understanding Transcriptional Regulation of the Signaling Network**—LiveDIP also provides tools to integrate gene expression data with interaction networks as an aid to understanding how a signaling pathway is regulated to generate the proper amount of signal intensity at the right time in response to an outside stimulus. As an example, we analyzed data on the time course of gene expression upon pheromone treatment; these data represent changes in mRNA levels between wild-type cells treated with  $\alpha$  factor at different time points from 0 to 120 min *versus* untreated cells (13). The number of genes induced significantly is about 30 at 0 min, going up to 112 after 30 min, remaining approximately constant for a while, and then jumping up to 658 after 120 min. Fig. 5 displays the changes in mRNA expression level of the proteins in the Interaction Map generated in Example I (confidence level higher than 90%, with  $p$  value less than 0.1) at three representative time points (initial, 0 min; mid-point, 45 min; and late point, 120 min) after pheromone treatment. The proteins in the interaction map are colored according to changes in their expression level, with *red* representing increase and *green* representing decrease as shown by the *color scale*. Uncolored proteins do not have high confidence data available ( $p > 0.1$ ).

There is no apparent correlation among changes in transcription levels of proteins involved in pheromone signaling. Although participating in the same signaling process, components of the pheromone pathway (proteins enclosed in the *cyan box* in Fig. 5) exhibit different time course profiles. Some show increases in their mRNA levels immediately (Ste12, Fus3) or 30 min after (Ste4, Ste2, Gpa1); half of them do not show significant change throughout the time course. The only evident observation is the lack of down-regulation. This seems to be in contrast with the consistent up-regulation of the glycerol biosynthetic pathway during salinity stress, which was more strong as time progressed (15). Assuming a general correlation between mRNA level and protein level, the above analysis suggests that cells use different mechanisms to regulate metabolic pathways and signaling pathways. Metabolic pathways are regulated more on the transcription level, and signaling pathways are regulated by changing the signaling proteins from one state to another rather than by changing the amounts of these proteins. All these observations point to the importance of future proteomic studies for understanding cell signaling.

Roberts *et al.* (13) showed that the entire transcriptional response to pheromone is derived from pathway-dependent activation of the transcription factor Ste12. In LiveDIP, there are 31 proteins in the protein network around the pheromone signaling pathway that affect activity of Ste12, directly or indirectly through protein-protein interactions, 19 by activation and 12 by inhibition. More interactions may be discovered as the database grows. Among the regulators that have high quality expression data during the time course, most activators (7 of 9) reach their maximum change before 60 min whereas all (7) inhibitors reach their maximum change after 45 min. The activator Fus3 shows the largest increase of about 12-fold after 45 min. The increased mRNA level of these activators may function as positive feedback, leading to stronger signals at the early stage of signaling. Dig2, which inhibits transcription factor Ste12, shows an increase of 2.5-fold only after 90 min. Because

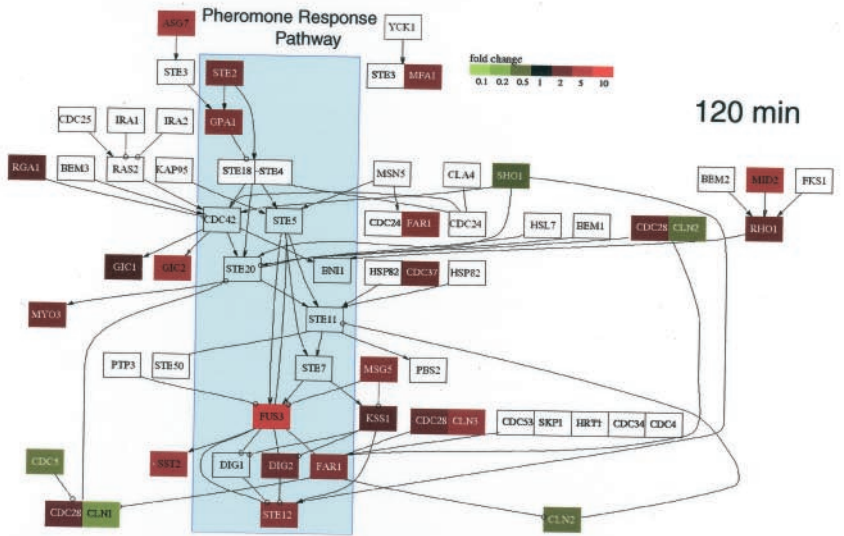
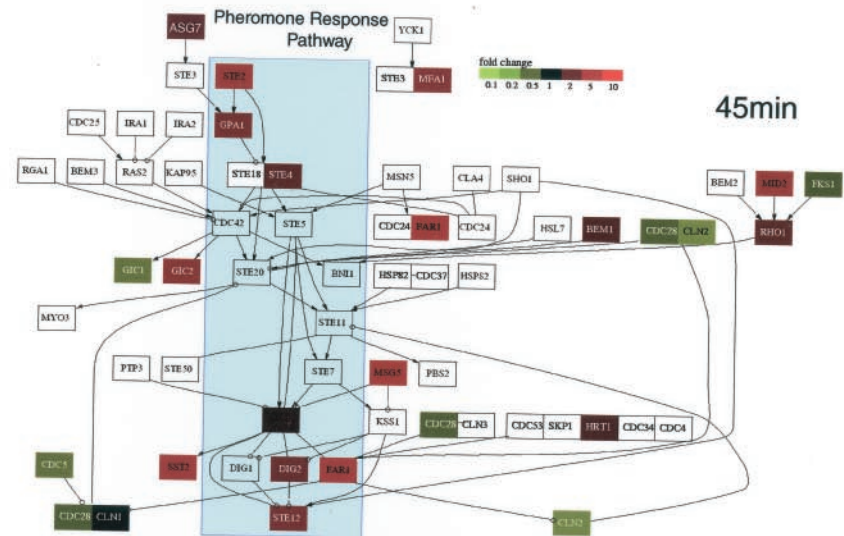
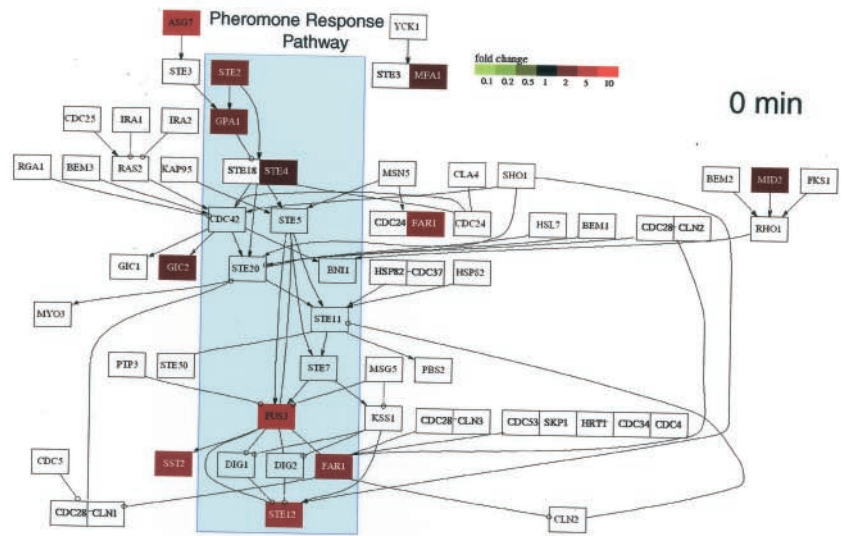


FIG. 5. **Pathway discovery by combining interaction and microarray data.** Mapping changes in mRNA expression level (13) onto the interaction network at three representative time points after phormone treatment (initial, 0 min; middle, 45 min; late, 120 min) are shown. The interaction network is the same as that in Fig. 4. The gene expression data represent changes in mRNA level of cells treated with  $\alpha$  factor *versus* untreated cells at various time points from 0 to 120 min (13). Changes in mRNA expression level with confidence levels higher than 90% ( $p$  value less than 0.1) are represented by *red* for increase and *green* for decrease. Proteins that are not colored correspond to those with expression data of lower quality ( $p > 0.1$ ).

prolonged activation of a signaling pathway may lead to desensitization, Dig2 may provide negative feedback to down-regulate the signal. These observations may indicate that activators of the signal are induced in the early stage to maximize the intensity whereas inhibitors are generally induced later to dampen the signal and promote recovery.

*Integrating Yeast Two Hybrid Data with Functional Genomic Data for Pathway Discovery*—It is possible to integrate mRNA expression with large scale protein-protein interaction data. Among proteins shown to interact with one or more components of the yeast pheromone signaling pathway in the large scale yeast two-hybrid experiments, *Sph1* interacts with *Ste7* and *Ste11*, and *Spa2* interacts with *Ste7*, *Ste11*, and *Bni1*. Both show more than 3-fold decreases in their mRNA level during the time course (13), suggesting these interactions may be relevant to cell response to pheromone treatment, consistent with previous observations that both proteins are involved in shmoo formation (16). Integration of mRNA expression data with large scale protein-protein interaction data is one way to validate yeast two-hybrid data and discover new proteins that are involved in signaling processes. It is also possible to use other data, such as protein expression levels and predicted phosphorylation sites, to help elucidate the relevance of protein-protein interactions.

#### DISCUSSION

##### *Protein-Protein Interactions and Biological Pathways*

As illustrated by our analysis of the pheromone response pathway in yeast, biological pathways are not isolated; they exist in the context of complex protein-protein interaction networks. Consequently, LiveDIP does not feature static pathways. Furthermore, our knowledge of biological pathways is growing rapidly. In many cases it is still not well understood whether and/or when a particular interaction is part of a given pathway. Instead of storing static pathways, we provide query tools such as Pathfinder and Batch Search to assemble pathways from currently available knowledge of protein interactions in LiveDIP. With sufficient annotation, such as when and where an interaction occurs, and by applying different types of filters, we may be able to reconstruct which subset of interactions take place inside the cell under specific conditions. The goal of LiveDIP is to reflect objectively what has been reported on the subject of a particular signaling pathway in the scientific literature.

##### *Application of DIP and LiveDIP*

Information about protein-protein interactions provided by DIP and LiveDIP can be used in three ways by biologists. Based on the functions of its interacting partners, possible general functions can be assigned to an unannotated protein, or new functions can be discovered for a previously characterized protein (17). The large amounts of binary interactions generated by two-hybrid studies are useful for this purpose. On the other hand, annotations on protein-protein interac-

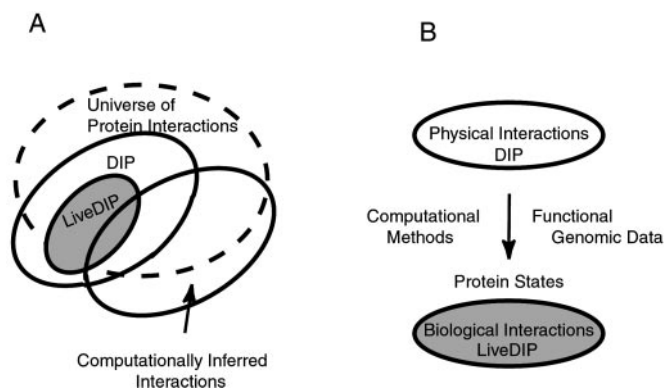
tions, as provided by LiveDIP, including the effect of interactions on protein states and state transitions, provide the molecular details on how these proteins carry out their functions. This type of information is important for understanding the mechanisms of biological processes and how misregulation leads to disease. Lastly, easy access to large quantities of experimental observations of protein-protein interactions facilitates computational manipulation for validating and assessing the quality of these data. It provides benchmarks for developing computational methods to predict protein functions or protein-protein interactions.

##### *From DIP to LiveDIP: from Physical Interactions to Biological Interactions*

LiveDIP contains the subset of interactions in DIP about which we have detailed information on related protein states. These interactions in LiveDIP (*green arrows* in Fig. 1B) can be termed biological interactions to distinguish them from physical interactions (*black lines* in Fig. 1A) in DIP. Some physical interactions are not biological interactions, because the interacting proteins are not in the right states. For example *Bni1* regulates polarized growth within the bud by perhaps establishing an anchoring site for *Kar9p-Bim1p* complex, which is involved in capturing microtubules at the bud cortex (16). This interaction can contribute to bud formation only when both proteins are in the protein states defined by localization to the bud. Other physical interactions do not belong to biological interactions, because they are nonspecific interactions and do not change the protein states. Data from large scale two-hybrid experiments are physical interactions; their biological relevance remains to be tested.

All the biological protein-protein interactions in a given organism may be considered to form the universe of protein-protein interactions as shown in Fig. 6A. The actual boundary of the universe of interactions is unknown and is represented by the *oval in a dashed line*. Because of false positive interactions, some fraction of DIP and LiveDIP may lie outside of the universe of protein interactions. Advances in large scale experimental techniques, such as yeast two-hybrid assay and mass spectrometry, lead to accumulation of data in DIP. Another source for protein-protein interactions is through computational methods, and these inferred interactions include both physical interactions and functional interactions such as participation in the same pathway (18). It is not clear how much these interactions from different sources overlap with the interaction universe.

Many new interactions detected by large scale methods are physical interactions between two protein entities with unknown biological relevance. They belong only to DIP and not LiveDIP. It is clear that to further our understanding about protein-protein interactions and thus to increase the overlap between DIP and the biological interaction universe, and the overlap between LiveDIP and the interaction universe, we are faced with the following two tasks: 1) To evaluate the quality



**FIG. 6. Relationship of DIP and LiveDIP to the universe of protein-protein interactions.** *A*, all protein-protein interactions important for biological processes in a given organism form the universe of protein-protein interactions. The actual boundary of the universe of protein-protein interactions is unknown and is represented by the area enclosed by the oval in a dashed line. Advances in large scale experimental techniques such as the yeast two-hybrid assay (19) and mass spectrometry (20) lead to the accumulation of protein-protein interaction data in DIP. LiveDIP contains the subset of protein-protein interactions in DIP, about which we have detailed information related to the protein states of the interacting pairs. Because of false positive interactions, some fraction of DIP and LiveDIP may lie outside of the universe of protein interactions. Computationally inferred interactions include both physical interactions and functional interactions such as participation in the same pathway (18). *B*, understanding protein-protein interactions. To understand the network of protein interactions documented in DIP requires us to make these interactions go Live; that is, the details of protein states and state transitions and other biologically relevant information need to be specified as in LiveDIP. Deepened understanding can come by supplementing the interaction data in DIP with data from functional genomics and computational methods.

of the large scale interaction data in DIP. It is important to compare and/or integrate interaction data from different sources to estimate the confidence for all the interacting protein pairs. DIP annotates each interaction with all the experiments used and can be used as a standard benchmark for large scale interaction maps. 2) To understand the functions and dynamics of these interactions. Information about protein-protein interactions, such as the protein states required for the interactions to occur and the transition in the protein states of the interacting pairs caused by the interactions, must be specified, such as in LiveDIP. Functional genomics data, such as mRNA expression array and computational methods, can be used to supplement large scale interaction data and to help elucidate their biological meaning (Fig. 6B).

#### Unique Features of LiveDIP

LiveDIP has several unique features. It aims to store experimentally determined biological protein-protein interactions instead of all kinds of biological relations. The design of LiveDIP captures the essence of biological interactions, their prerequisite on the states of the interacting partners and the resultant transition in the states of these proteins. One of the major goals of LiveDIP is to provide the biological community

with convenient access to information about protein-protein interactions. We emphasize providing easy data access through such tools as Browse and Basic Search and on presenting the information in a format and language familiar to biologists. The Batch Search and Pathfinder features can be used to assemble biological pathways based on knowledge about protein-protein interactions on the fly, and the results can be explored in various ways including automatically drawn schematic Interaction Maps. We provide tools to integrate large scale functional genomics data with protein-protein interaction networks. These features of LiveDIP render it a useful tool for biological pathway analysis and pathway discovery.

*Acknowledgments*—We thank Lukasz Salwinski, Parag Mallick, and Michael Thompson for suggestions.

\* This work was supported by Department of Energy and National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Howard Hughes Medical Inst., UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles, P. O. Box 951570, Los Angeles, CA 90095-1570. Tel.: 310-825-3754; Fax: 310-206-3914; E-mail: david@mbi.ucla.edu.

#### REFERENCES

- Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934
- Eisenberg, D., Marcotte, E. M., Xenarios, I., and Yeates, T. O. (2000) Protein function in the post-genomic era. *Nature* **405**, 823–826
- Bader, G. D., and Hogue, C. W. (2000) BIND—a data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics* **16**, 465–477
- van Helden, J., Naim, A., Mancuso, R., Eldridge, M., Wernisch, L., Gilbert, D., and Wodak, S. J. (2000) Representing and analysing molecular and cellular function using the computer. *Biol. Chem.* **381**, 921–935
- Xenarios, I., Fernandez, E., Salwinski, L., Duan, X. J., Thompson, M. J., Marcotte, E. M., and Eisenberg, D. (2001) DIP: the database of interacting proteins: 2001 update. *Nucleic Acids Res.* **29**, 239–241
- Marcotte, E. M., Xenarios, I., and Eisenberg, D. (2001) Mining literature for protein-protein interactions. *Bioinformatics* **17**, 359–363
- Manahan, C. L., Patnana, M., Blumer, K. J., and Linder, M. E. (2000) Dual lipid modification motifs in G(alpha) and G(gamma) subunits are required for full activity of the pheromone response pathway in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **11**, 957–968
- Canagarajah, B. J., Khokhlatchev, A., Cobb, M. H., and Goldsmith, E. J. (1997) Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell* **90**, 859–869
- Wu, C., Leeuw, T., Leberer, E., Thomas, D. Y., and Whiteway, M. (1998) Cell cycle- and Cln 2p-Cdc28p-dependent phosphorylation of the yeast Ste20p protein kinase. *J. Biol. Chem.* **273**, 28107–28115
- Elion, E. A. (2000) Pheromone response, mating and cell biology. *Curr. Opin. Microbiol.* **3**, 573–581
- Zhan, X. L., Deschenes, R. J., and Guan, K. L. (1997) Differential regulation of FUS3 MAP kinase by tyrosine-specific phosphatases PT, 2/PTP3 and dual-specificity phosphatase MSG5 in *Saccharomyces cerevisiae*. *Genes Dev.* **11**, 1690–1702
- Chambers, D. (2001) Scientists signal the way forward. *Trends Genet.* **17**, 309–310
- Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker, W. L., Hughes, T. R., Tyers, M.,

- Boone, C., and Friend, S. H. (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**, 873–880
14. Gustin, M. C., Albertyn, J., Alexander, M., and Davenport, K. (1998) MAP kinase pathways in the yeast *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **62**, 1264–1300
15. Posas, F., Chambers, J. R., Heyman, J. A., Hoeffler, J. P., de Nadal, E., and Arino, J. (2000) The transcriptional response of yeast to saline stress. *J. Biol. Chem.* **275**, 17249–17255
16. Arkowitz, R. A., and Lowe, N. (1997) A small conserved domain in the yeast Spa 2p is necessary and sufficient for its polarized localization. *J. Cell Biol.* **138**, 17–36
17. Schwikowski, B., Uetz, P., and Fields, S. (2000) A network of protein-protein interactions in yeast. *Nat Biotechnol.* **18**, 1257–1261
18. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O., and Eisenberg, D. (1999) A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86
19. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4569–4574
20. Gygi, S. P., Rist, B., and Aebersold, R. (2000) Measuring gene expression by quantitative proteome analysis. *Curr. Opin. Biotechnol.* **11**, 396–401