

## Chapter 2

# Background

This chapter provides an overview of key topics that serve as background for the rest of the book. First, we discuss the roles of proteins in living organisms. This is followed by a brief discussion on protein-protein interactions and methods for analyzing them. Finally, we briefly highlight on the roles of databases, ontologies and annotations in proteins and their interactions.

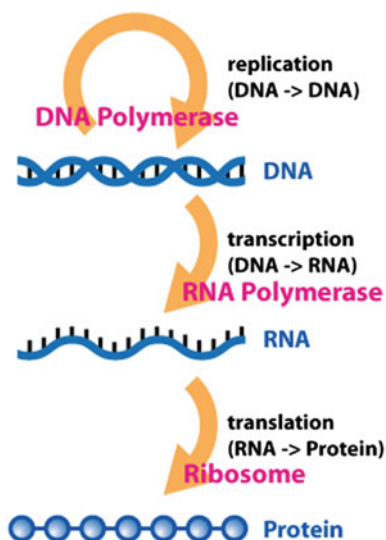
### 2.1 Proteins: The Building Block of Life

The basic building block of all living organisms is the cell. The cell itself is a complex machinery—within it a plethora of processes and components that govern the mechanisms of the cell [23]. Microtubules, tubular shaped scaffolds of the cell, provide not only shape and structure, but also act as tracks for transporting cellular cargoes. Mitochondrions are the molecular engines of the cell, generating fuel to power cellular machines. These are just a few examples of cellular components that regulate the cell machinery.

The various parts of the cell work in tandem to regulate *biological processes*—functionalities performed within the cell that control its behavior and state, depending on its internal and external environments. For example, the cell cycle is a biological process that controls the growth and replication of itself. Transport processes cargo cellular components within the cell, as well as exporting cargoes out of the cell and importing cargoes into it. Homeostatic processes regulate the equilibrium of chemical concentration in the cell to a desirable optimum.

Remarkably, the machines that run biological processes of cells are largely performed by one class of molecules called *proteins* [30]. A protein is composed of a

**Fig. 2.1** The central dogma of molecular biology. (Image by Dhorspool at en.wikipedia)



linear sequence of amino acids that are folded into a 3D structure. Informally, one can think of proteins as strings of words formed by an alphabet of amino acids. There are 20 “canonical” amino acids in eukaryotes [36]. Each amino acid exhibits distinct chemical properties (such as polarity and hydrophobicity) and also physical properties (such as mass), giving a 3D structured protein its character and behavior. The roles of proteins are many and varied. For instance, the protein *actin* lends structural integrity to cells. Enzymes are a special class of proteins that catalyze chemical reactions. Signaling proteins like *Ras* act as messengers that amplify and distribute signals from a stimuli.

Given the significance of proteins, this begs an important question: what directs their construction and regulation? *Genetic information* is the information required for construction of proteins. The central dogma of molecular biology [7] states that genetic information flows from deoxyribonucleic acid (DNA) to oxyribonucleic acid (RNA) to protein (Fig. 2.1). Essentially, the DNA (a sequential chain of polymers called nucleotides) serves as the blueprint for the construction of proteins. The sequence of nucleotides in DNA encodes the necessary information for protein construction, which is then transcribed into RNA before translated into proteins. Regions of the DNA that *directly* encode the construction of proteins are called *genes*. Beyond serving as the blueprint for protein construction, DNA and RNA also encode information that guides regulation of proteins. For instance, they regulate amount of proteins produced (expression level); signals to start or stop production (gene activation or suppression); and signals to modify proteins, affecting their behavior and interaction (protein modification).

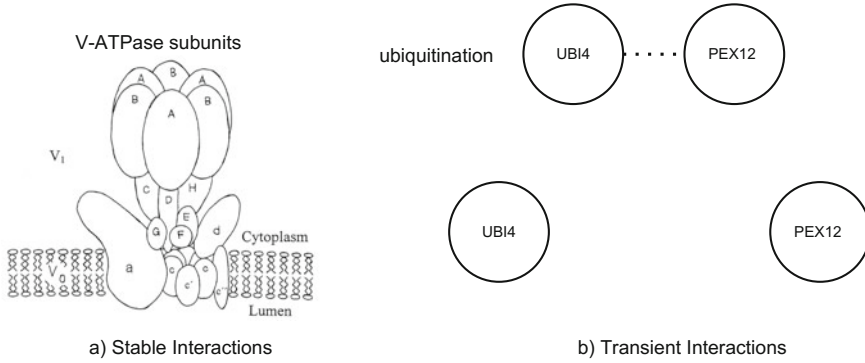
## 2.2 Protein-Protein Interaction (PPI)

Protein, DNA, RNA and other biological molecules do not work in isolation; they cooperate with other proteins to perform a particular biological activity. Two molecules that cooperate to perform a particular function are said to be *interacting*. It is the combination of these molecules and their interactions, and not the molecules alone, that characterize the mechanisms of a biological process. We wish to emphasize that although the rest of the book largely focuses on proteins, the concepts that we will discuss may extend to other molecules. Genes, DNAs, RNAs and other entities are also major drivers of a biological process. Interactions are typically grouped by their molecule types:

- **Protein-protein interactions**—cooperation between proteins to drive biological processes.
- **Gene regulatory interactions**—interplay of genetic information to regulate protein expression level.
- **Metabolic interactions**—cooperation between enzyme proteins to convert a substrate molecule into product molecule through several catalysis reactions.
- **RNA-DNA interactions**—cooperation between RNA-RNA or RNA-DNA interactions plays increasingly critical role in diseases.

In this book, major focus is placed on the class of protein-protein interactions, although most of the concepts covered here apply to other classes of interactions as well.

Protein-protein interactions can be *stable* or *transient* [25]. In *stable* protein-protein interactions, a group of proteins forms permanent protein-protein interactions to perform a biological role. A group of such stably interacting proteins is called a *protein complex*. An example of protein complexes is the V-ATPase (Fig. 2.2(a)). Multiple protein subunits combine to form the V-ATPase enzyme that transports protons across membranes [24]. In *transient* protein-protein interactions, two proteins associate with each other briefly to perform a biological activity before dissociating. These interactions regulate a significant portion of biological processes. The interactions occur when a region of one protein complements the region of another, forming non-covalent bonds like hydrogen bonds, Van der Waals forces and hydrophobic bondings. A common surface region is the *leucine zipper* [22], a 3D structural motif in proteins with hydrophobic regions that allow two proteins with complementing zipper motifs to “zip” together. Typically, transient interactions only occur under conditions that promote their interaction, for instance the phosphorylation state of the proteins involved, the protein conformation state or their localization. Figure 2.2(b) shows transient interaction between UBI4 and PEX12; physical interaction occurs only during ubiquitination.



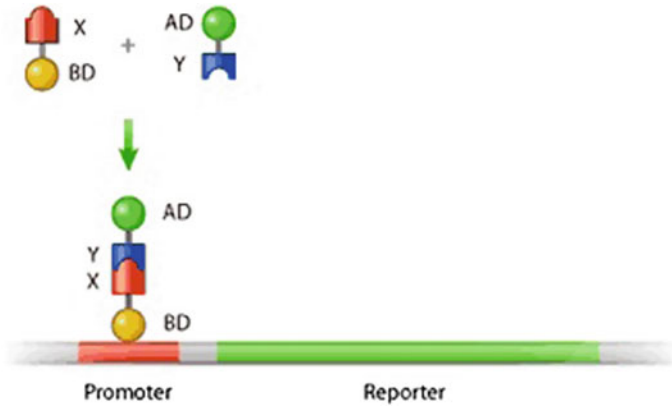
**Fig. 2.2** Stable vs. transient interactions

## 2.3 Methods to Analyze Protein-Protein Interactions

Given the importance of protein-protein interactions in characterizing the mechanisms of a biological process, biologists have developed a range of experimental methods to detect and predict interactions between proteins. We describe several pertinent ones below.

### 2.3.1 Yeast Two-Hybrid (Y2H)

The yeast-two-hybrid (Y2H) method relies on activating the transcription of a *reporter gene* to detect interaction between two proteins [17]. Reporter genes are typically genes with easily observable phenotype. Figure 2.3 summarizes the concept behind Y2H. In Y2H, biologists engineer the two tested proteins such that when these two proteins interact, transcription of the reporter gene is activated, and thus, if the reporter gene phenotype is sufficiently expressed, one can deduce that the two proteins interact. To this end, Y2H uses two types of protein domains: the DNA-binding domain (BD) and the activation domain (AD). The BD and AD domains must be brought together proximally to bind and form a transcription activator, which then activates reporter gene transcription. Given two proteins, the BD domain is fused to one protein (called the bait) and the AD domain is fused to the remaining protein (called the prey). If these two proteins interact, the two domains are brought together proximally and activates reporter gene transcription. Commonly used reporter genes (and their promoter) include *HIS3*, *URA3* and *lacZ*. For example, the *lacZ* reporter gene when activated causes the yeast cell to express  $\beta$ -galactosidase, which can be detected by the formation of blue colored yeast colonies. A strong advantage of this method is its scalability and Y2H can easily be used to screen thousands of proteins for interactions, giving rise to high-throughput experiment technologies.



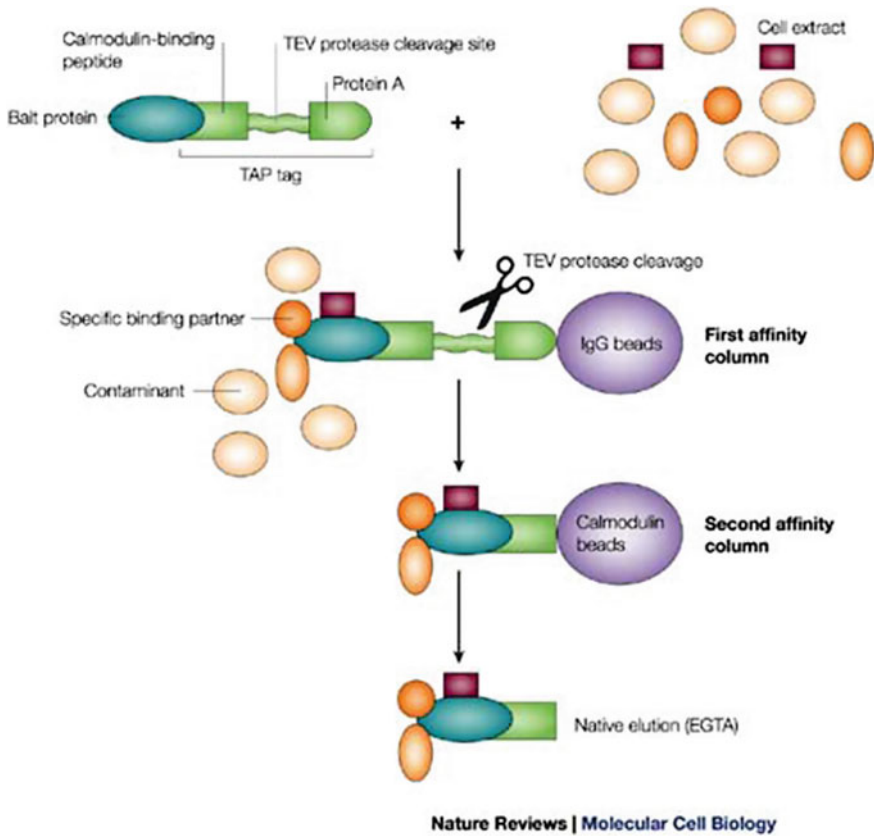
**Fig. 2.3** Yeast-two-hybrid to detect protein-protein interactions. (Adopted from *The Science Creative Quarterly*)

### 2.3.2 Tandem Affinity Purification (TAP)

The tandem affinity purification (TAP) method identifies protein-protein interaction by incorporating a TAP tag to the target protein, followed by fishing for other proteins that interact with the tagged protein [28]. Figure 2.4 illustrates the TAP method. The TAP tag comprises two Immunoglobulin G (IgG) binding domains and a Calmodulin-binding peptide (CBP). In TAP, the biologist engineers a fusion protein by fusing the TAP tag to the target protein. Next, the fusion protein, together with any other proteins attached to it, is isolated using beads coated with IgG. The biologist then applies the TEV cleavage enzyme to cleave the TAP tag from fusion, leaving behind the target protein plus the CBP domain bounded to the bead. A second isolation step is then applied using Calmodulin-coated beads. Here, the biologist obtains the final product of target protein, CBP and attached proteins that are interacting with the target protein. Finally, the end products are analyzed via mass spectrometry or SDS-PAGE [31]. The two step purification process minimizes the amount of contaminants obtained.

### 2.3.3 Bimolecular Fluorescence Complementation (BIFC)

Bimolecular fluorescence complementation (BIFC) is another protein-protein interaction screening strategy that relies on a reporter protein [16]. In this method, the reporter protein is fluorescent, allowing it to be easily detected and located using tools such as flow cytometry. A reporter protein, the yellow fluorescent protein (YFP) for instance, is designed as two complementary fragments (YN and YC). Given two candidate proteins, one can separately fuse each fragment to the candidate proteins.



**Fig. 2.4** Tandem affinity purification. (Adopted from [15])

When these two proteins interact, the two fragments will be brought to close proximity, encouraging them to re-attach and re-assemble into the YFP reporter protein. The fluorescent reporter protein can then be screened through a variety of techniques including flow cytometry.

### 2.3.4 Noise in High-Throughput Screening Methods

Rapid high-throughput protein-protein interaction screening methods, however, suffer from significant noise and coverage issues. For instance, the *false negative* rate, defined as the probability of interacting protein detected as negative, could be as high as 70–90% with Y2H data [5]. This implies that there is a significant *coverage* gap (coverage here refers to the ratio between the number of detected interactions and the number of actual interactions in the network). Moreover, high-throughput

protein-protein interaction screening methods also suffer from relatively high *false positives* [13], which is defined as the probability of non-interacting protein detected as positive.

## 2.4 Protein-Protein Interaction Databases

Advancements in protein-protein interaction screening methods have enabled the capability of generating large scale interaction data. Therefore, it is important to catalog and store these datasets to allow rapid and convenient access. We discuss several public databases that catalog key protein-protein interaction datasets. Table 2.1 lists several well known knowledge-bases with significant protein-protein interaction datasets. The STRING database [34] hosts a large collection of predicted and known protein-protein interactions. In addition, the STRING database links key information about the gene that codes for the interactor proteins, including their DNA sequence, biological annotations, co-occurrence, and co-expression data. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database [19] is a resource of manually curated pathway datasets. The KEGG database is especially notable for its large collection of metabolic pathways for bacterial microbes. Important signaling pathways for a variety of organisms are also hosted in the KEGG database. The REACTOME database [18] hosts detailed biological pathways specifically for the human species. As is the KEGG database, pathways in the REACTOME database are manually curated and handcrafted. The IntAct database [21] stores a large

**Table 2.1** Selected protein-protein interaction databases

Database	Reference
Human Protein Reference Database (HPRD)	[27]
Biological General Repository for Interaction Datasets (BioGRID)	[32]
Database of Interacting Proteins (DIP)	[35]
Kyoto Encyclopedia of Genes and Genomes (KEGG)	[19]
Biomolecular Interaction Network Database (BIND)	[2]
The MIPS Mammalian Protein-Protein Database	[26]
STRING: functional protein association networks	[34]
REACTOME	[18]
IntAct	[21]
BioCyc	[20]
BioCarta Pathways	[4]
PHOSIDA	[10]
Phospho-ELM	[8]
DOMINE: a database of protein domain interactions	[29]

amount of protein-protein interaction datasets submitted by individual labs. The datasets can range for a several protein-protein interactions per dataset to several hundred thousands of interactions per dataset. The Munich Information Center for Protein Sequences (MIPS) database [26] is noted for its repository of protein complexes. Other significant databases hosting protein-protein interaction datasets are the Human Protein Reference Database (HPRD) [27], Biological General Repository for Interaction Datasets (BioGRID) [32], Database of Interacting Proteins (DIP) [35].

Apart from general protein-protein interaction resources, several web resources host context-specific datasets that focus on a particular biological topic of interest. For example, the PHOSPIDA [10] and Phospho-ELM [8] knowledge-bases contain protein phosphorylation sites information, which can be used to deduce their interacting partners. DOMINE [29] is a database of protein domain-domain interactions. Apart from molecular function specific datasets, disease specific datasets are also abundant. The IntAct database contains a number of disease-related protein-protein interaction datasets that include Alzheimer's, cancer and cerebellar ataxia.

## 2.5 Annotating the Roles of Proteins and Their Interactions

With the growth of biological literature on the roles of proteins, groups of proteins, as well as their interactions, the need to annotate these information in a structured manner becomes pertinent. The Gene Ontology (GO) [12] is developed as a standard for providing a structured *ontology* describing attributes of genes and gene products (including proteins). An ontology is a set of controlled concepts (GO *terms*) and their relationships that models the domain. In GO, the concepts describe the roles of the genes and their products, while the concept relationships connect the various concepts in GO. For example, the activation of protein kinase activity concept can be used to annotate the MAPK protein, giving it that particular function. Now the concept relationships in GO may provide additional inferences to this concept. If suppose GO states that activation of protein kinase activity is a type of regulation of protein phosphorylation, then one can reason that MAPK protein also has the attribute of regulation of protein phosphorylation.

The role of GO as controlled vocabulary also resolves ambiguity in word descriptions. Functional descriptors that describe the role and function of proteins in the literature can be ambiguous, redundant and domain specific [1]. For instance, the gene names CDC28, Cdc28p or cdc-28 all refer to the same biological entity. With a controlled vocabulary, computation methods can infer functional roles of proteins in a consistent manner.

Gene Ontology Annotation (GOA) database [6] stores associations of genes and proteins to GO terms. GO term annotation can be undertaken either manually or automatically. In manual annotation, a domain expert or curator who is aware of the functional description of the gene or protein annotate that protein with the relevant GO terms. The automatic approach, on the other hand, predicts and infers the GO terms



relevant of the protein via a multitude of machine learning techniques including literature mining and graph-based inferencing tools. The Online Mendelian Inheritance in Man (OMIM) database [11] supplies important annotations regarding diseases associated with human proteins.

### 2.5.1 The Structure of Gene Ontology

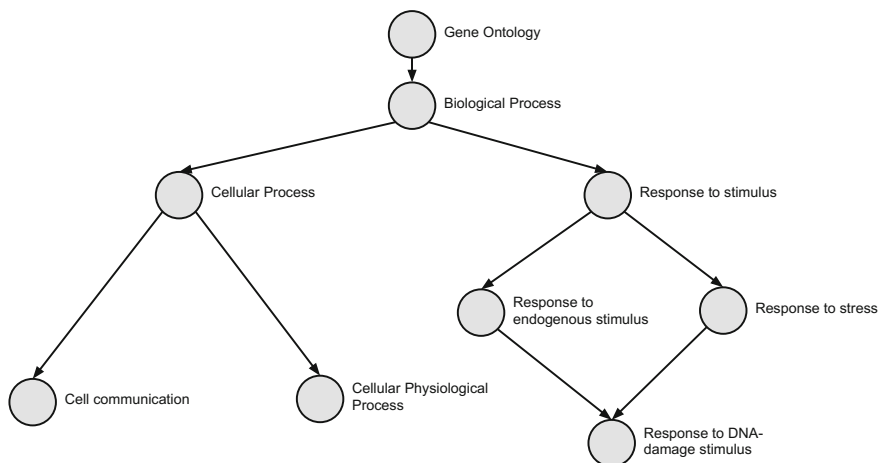
The Gene Ontology is modeled as a directed acyclic graph (DAG) and is divided into three major domains: *biological process*, *cellular component* and *molecular function*. The total number of GO terms in the GO DAG exceeds 30,000.

The biological process domain contains GO terms describing the functional processes in cells, tissues, organs and organisms that proteins may take part in. The Gene Ontology defines a biological process as “a recognized series of events or molecular functions” with a defined beginning and end. A biological process GO term may describe the process itself, or it may describe an encompassing process that is made up of subprocesses. For instance, the biological process term *apoptosis* describes cell apoptosis pathways in the cell. Thus, if the p53 protein is annotated with *apoptosis* GO term, then one can infer that p53 protein participates in cell apoptosis. The GO term *cell cycle* may describe the cell cycle process which itself is made up of several subprocesses, such as M-phase cell cycle and G-phase cell cycle. In Gene Ontology, a process term may be connected to its parent via *is\_a* and *part\_of* relationships; the former describes that the process is an instance of the parent process, while the latter describes that the process is only a part of the parent process.

The cellular component domain contains GO terms describing the components of the cell and its extracellular environment. Cellular components may be anatomical structures or macromolecular complexes. In GO, a protein annotated with a cellular component GO term is said to be *located in* or is a *subcomponent of* the component described by the term. For example, the GO term *mitochondrial ribosome* describes the mitochondrial ribosomal components. Proteins like ribosomal protein L41 may be annotated with such GO terms.

Finally, the molecular process contains GO terms pertaining to an elemental activity of a protein. Activities here include any function performed by proteins like catalysis, binding, phosphorylation, and other enzymatic roles. For example, the GO term *phosphorylation* describes the molecular activity that a protein may perform, which in this case is phosphorylation activity. A protein may be annotated with multiple activities. This is because proteins itself may participate in multiple functions. Protein kinases like PKC are known to have such capabilities and could be annotated with these terms.

Figure 2.5 depicts a part of the GO DAG. Formally, the Gene Ontology for each domain is modeled as a directed acyclic graph  $D = (V_{go}, E_{go})$  where  $V_{go}$  denotes



**Fig. 2.5** Subset of the Gene Ontology directed acyclic graph

the set of GO terms and  $E_{go}$ —the set of pair relationships between GO terms in  $V_{go}$ —denotes the set of GO term relationships. Here, an edge  $(v_1, v_2) \in E$  represents a parent-child connection between two GO terms  $v_1 \in V_{go}$  and  $v_2 \in V_{go}$ . The ordered set  $\Delta = \langle \Delta_1, \Delta_2, \dots, \Delta_d \rangle$  is a topological sort of  $D$ . Each  $\Delta_i$  represents a single GO term. We assume that a protein node  $v \in V_i$  is annotated with a set of GO terms  $D_v \subset \Delta$ . The indicator function of terms annotated in node  $v$  is given by  $I_{\{x \in D_v\}} : \Delta \rightarrow \{0, 1\}$  such that  $I_{\{x \in D_v\}} = 1$  if  $x \in D_v$  and 0 if otherwise.

The root node absorbs all GO terms of its descendants, i.e., each descendant GO term is a or is part of the root node. The root nodes of biological process, cellular component and molecular function domains are biological process, cellular component and molecular function, respectively. As the GO DAG branches from the root node, the specificity of the functional description increases. Thus, one can utilize GO DAG and its associated annotations to group proteins by their function or parts in a hierarchical manner. For example, in Fig. 2.5, if proteins MAPK, MAPKK, and MAPKKK are annotated with the intracellular signaling process term, then these proteins are also part of the signal transduction, cell communication, cellular process and biological process.

Gene Ontology and its annotations has been applied to a large number of bioinformatics approaches [14]. A pertinent usage is in gene expression analysis studies [3]. Typically, groups of genes which are either significantly up-regulated or de-regulated are identified using techniques such as gene clustering and enrichment analysis [33]. Then, the GO annotations are utilized to identify over-expressed func-

tional roles of these groups of proteins. An example of algorithms of such nature is the `MAPPFinder` [9], which looks for genes that are significantly deregulated using the GO annotations.

## 2.6 Summary

This chapter can be summarized as follows:

- Proteins, DNAs, RNAs and other biological molecules work in tandem to regulate biological processes. Cooperating molecules that perform a particular function are said to be interacting, and their interactions can be either transient or stable.
- A range of experimental methods have been developed to detect and predict interactions between proteins in a high-throughput manner. Among them are Y2H, TAP and BIFC.
- Advancement in protein-protein interaction screening methods has led to large scale interaction datasets. Several public databases serve as important repositories of such datasets, including STRING, KEGG and REACTOME.
- The Gene Ontology (GO) is developed as a standard for providing a structured ontology describing attributes of genes and gene products (including proteins). Gene Ontology Annotation (GOA) database stores associations of genes and proteins to GO terms. GO term annotations are useful as functional descriptions of a gene or protein.

## References

1. M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene Ontology: tool for the unification of biology. the Gene Ontology consortium. *Nat. Genet.* **25**, 25–29 (2000)
2. G.D. Bader, I. Donaldson, C. Wolting, B.F. Ouellette, T. Pawson, C.W. Hogue, BIND-the biomolecular interaction network database. *Nucl. Acids Res.* **29**, 242–245 (2001)
3. T. Beissbarth, T.P. Speed, GStat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics (Oxford, England)* **20**, 1464–1465 (2004)
4. Biocarta, BioCarta
5. P. Braun, M. Tasan, M. Dreze, M. Barrios-Rodiles, I. Lemmens, H. Yu, J.M. Sahalie, R.R. Murray, L. Roncari, A.-S. de Smet, K. Venkatesan, J.-F. Rual, J. Vandenhaute, M.E. Cusick, T. Pawson, D.E. Hill, J. Tavernier, J.L. Wrana, F.P. Roth, M. Vidal, An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6**, 91–97 (2009)
6. E. Camon, M. Magrane, D. Barrell, V. Lee, E. Dimmer, J. Maslen, D. Binns, N. Harte, R. Lopez, R. Apweiler, The Gene Ontology annotation (GOA) database: sharing knowledge in uniprot with Gene Ontology. *Nucl. Acids Res.* **32**, D262–D266 (2004)
7. F. Crick, Central dogma of molecular biology. *Nature* **227**, 561–563 (1970)
8. H. Dinkel, C. Chica, A. Via, C.M. Gould, L.J. Jensen, T.J. Gibson, F. Diella, Phospho.ELM: a database of phosphorylation sites-update 2011. *Nucl. Acids Res.* **39**, D261–D267 (2011)

9. S.W. Doniger, N. Salomonis, K.D. Dahlquist, K. Vranizan, S.C. Lawlor, B.R. Conklin, MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Gen. Biol.* **4**, R7–R7 (2003)
10. F. Gnad, S. Ren, J. Cox, J.V. Olsen, B. Macek, M. Oroschi, M. Mann, PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Gen. Biol.* **8**, R250–D250 (2007)
11. A. Hamosh, A.F. Scott, J.S. Amberger, C.A. Bocchini, V.A. McKusick, Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucl. Acids Res.* **33**, D514–D517 (2005)
12. M.A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G.M. Rubin, J.A. Blake, C. Bult, M. Dolan, H. Drabkin, J.T. Eppig, D.P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J.M. Cherry, K.R. Christie, M.C. Costanzo, S.S. Dwight, S. Engel, D.G. Fisk, J.E. Hirschman, E.L. Hong, R.S. Nash, A. Sethuraman, C.L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S.Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E.M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, R. White, The Gene Ontology (GO) database and informatics resource. *Nucl. Acids Res.* **32**, D258–D261 (2004)
13. H. Huang, J.S. Bader, Precision and recall estimates for two-hybrid screens. *Bioinformatics (Oxford, England)* **25**, 372–378 (2009)
14. D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009)
15. L.A. Huber, Is proteomics heading in the wrong direction? *Nat. Rev. Mol. Cell Biol.* **4**, 74–80 (2003)
16. C.-D. Hu, Y. Chinenov, T.K. Kerppola, Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **9**, 789–798 (2002)
17. T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, Y. Sakaki, A comprehensive two-hybrid analysis to explore the yeast protein interactome. in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98 (2001), pp. 4569–4574
18. G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G.R. Gopinath, G.R. Wu, L. Matthews, S. Lewis, E. Birney, L. Stein, Reactome: a knowledgebase of biological pathways. *Nucl. Acids Res.* **33**, D428–D432 (2005)
19. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* **28**, 27–30 (2000)
20. P.D. Karp, C.A. Ouzounis, C. Moore-Kochlacs, L. Goldovsky, P. Kaipa, D. Ahrén, S. Tsoka, N. Darzentas, V. Kunin, N. López-Bigas, Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucl. Acids Res.* **33**, 6083–6089 (2005)
21. S. Kerrien, Y. Alam-Farouque, B. Aranda, I. Bancarz, A. Bridge, C. Derow, E. Dimmer, M. Feuermann, A. Friedrichsen, R. Huntley, C. Kohler, J. Khadake, C. Leroy, A. Liban, C. Lieftink, L. Montecchi-Palazzi, S. Orchard, J. Risse, K. Robbe, B. Roechert, D. Thorncroft, Y. Zhang, R. Apweiler, H. Hermjakob, IntAct-open source resource for molecular interaction data. *Nucl. Acids Res.* **35**, D561–D565 (2007)
22. W.H. Landschulz, P.F. Johnson, S.L. McKnight, The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science (New York, N.Y.)* **240**, 1759–1764 (1988)
23. J. Lodish, H.F. Baltimore, D. Berk, A. Darnell, *Molecular Cell Biology* (WH Freeman, New York, 1995)
24. V. Marshansky, M. Futai, The V-type H<sup>+</sup>-ATPase in vesicular trafficking: targeting, regulation and function. *Curr. Opin. Cell Biol.* **20**, 415–426 (2008)
25. I.M.A. Nooren, J.M. Thornton, Diversity of protein-protein interactions. *EMBO J.* **22**, 3486–3492 (2003)

26. P. Pagel, S. Kovac, M. Oesterheld, B. Brauner, I. Dunger-Kaltenbach, G. Frishman, C. Montrone, P. Mark, V. Stümpflen, H.-W. Mewes, A. Ruepp, D. Frishman, The MIPS mammalian protein-protein interaction database. *Bioinformatics* (Oxford, England) **21**, 832–834 (2005)
27. S. Peri, J.D. Navarro, R. Amanchy, T.Z. Kristiansen, C.K. Jonnalagadda, V. Surendranath, V. Niranjan, B. Muthusamy, T.K.B. Gandhi, M. Gronborg, N. Ibarrola, N. Deshpande, K. Shanker, H.N. Shivashankar, B.P. Rashmi, M.A. Ramya, Z. Zhao, K.N. Chandrika, N. Padma, H.C. Harsha, A.J. Yatish, M.P. Kavitha, M. Menezes, D.R. Choudhury, S. Suresh, N. Ghosh, R. Saravana, S. Chandran, S. Krishna, M. Joy, S.K. Anand, V. Madavan, A. Joseph, G.W. Wong, W.P. Schiemann, S.N. Constantinescu, L. Huang, R. Khosravi-Far, H. Steen, M. Tewari, S. Ghaffari, G.C. Blobel, C.V. Dang, J.G.N. Garcia, J. Pevsner, O.N. Jensen, P. Roepstorff, K.S. Deshpande, A.M. Chinnaiyan, A. Hamosh, A. Chakravarti, A. Pandey, Development of human protein reference database as an initial platform for approaching systems biology in humans. *Gen. Res.* **13**, 2363–2371 (2003)
28. O. Puig, F. Caspari, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, B. Séraphin, The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* (San Diego, Calif.) **24**, 218–229 (2001)
29. B. Raghavachari, A. Tasneem, T.M. Przytycka, R. Jothi, DOMINE: a database of protein domain interactions. *Nucl. Acids Res.* **36**, D656–D661 (2008)
30. G. Scatchard, The attractions of proteins for small molecules and ions. *Ann. New York Acad. Sci.* **51**, 660–672 (1949)
31. H. Schägger, Tricine-SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006)
32. C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, BioGRID: a general repository for interaction datasets. *Nucl. Acids Res.* **34**, D535–D539 (2006)
33. A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102 (2005), pp. 15545–15550
34. D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguéz, T. Doerks, M. Stark, J. Müller, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucl. Acids Res.* **39**, D561–D568 (2011)
35. I. Xenarios, D.W. Rice, L. Salwinski, M.K. Baron, E.M. Marcotte, D. Eisenberg, DIP: the database of interacting proteins. *Nucl. Acids Res.* **28**, 289–291 (2000)
36. T.S. Young, P.G. Schultz, Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J. Biol. Chem.* **285**, 11039–11044 (2010)

Summarizing Biological Networks

Bhowmick, S.S.; Seah, B.-S.

2017, XV, 146 p. 54 illus., 31 illus. in color., Hardcover

ISBN: 978-3-319-54620-9