Functional Genomics of Protein-Protein Interactions

Takashi Ito

Division of Genome Biology, Cancer Research Institute, Kanazawa University, 13-1 Takaramachi, Kanazawa 920-0934, Japan

It is well recognized that protein-protein interactions play key roles in structural and functional organization of the cell. Uncovering of protein interaction schemes often sheds light on molecular mechanisms underlying biological processes. Hence increasing attention is being paid to protein-protein interactions in both structural and functional studies. Although such studies are intensively conducted on the proteins of interest to a wide audience of researchers, genomes harbor a number of novel proteins currently lacking any hint as to their specific functions. For such novel proteins, interactions with known proteins serve as invaluable clues to their functions or biological roles. These interactions also suggest directions one may take for more detailed phenotypic examination of mutants for the novel genes. Organisms with completely sequenced genomes are quite suitable for such studies, because all of the proteins can be predicted and used for the comprehensive examination of protein-protein interactions. Such genome-wide interaction mapping would be a novel type of genomic data and strongly accelerate the comprehensive understanding of the cell as a molecular machinery. Therefore, the study on protein-protein interactions is one of the most important issues in functional genomics.

It should also be emphasized that recent studies of protein-protein interactions, in particular, those involved in signal transduction, uncovered a number of protein-binding domains or motifs, which are evolutionarily conserved and used in various signaling pathways. Thus, the identification of novel protein interaction modules contributes not only to the studies on the

particular pathway but also to much wider fields of biomedical research. A large set of protein-protein interaction data would lay a foundation for the search of such modules by both experimental and computational means.

Bearing these in our mind, we have launched a systematic identification of mutually interacting protein pairs in the budding yeast *Saccharomyces cerevisiae.* We use the budding yeast, because its genome is completely sequenced and all the open reading frames (ORFs) have been predicted. In addition, a huge amount of knowledge in molecular biology, cell biology, genetics and biochemistry is accumulated for this simple eukaryotic organism to help evaluate the biological relevance of the identified interactions. It should also be noted that the ease of genome modification in this yeast enables one to create mutants defective for each interaction in order to examine its biological role.

First, we cloned almost all of the yeast ORFs individually as a DNA-binding domain fusion ("bait") in a *MAT***a** strain PJ69-2A (*MAT***a**, *trp1, leu2, ura3, his3, gal4, gal80, GAL2::ADE2, GAL1::HIS3*). Similarly, each ORF was cloned as an activation domain fusion ("prey") in a *MAT*α strain MaV204K (*MAT*α, *trp1, leu2, his3, ade2::kanMX, gal4, gal80, SPAL10::URA3, UASGAL1::HIS3, GAL1::lacZ*). Subsequently, these clones were divided into pools, each containing 96 clones. These "bait" and "prey" clone pools were then systematically mated with each other, and the transformants were subjected to strict selection for the activation of three reporter genes driven by different GAL4-repsonsive promoters (*ADE2*, *HIS3*, and *URA3*). The positive clones were subjected to colony PCR or plasmid isolation followed by sequence tagging. Our pilot project covering ~10% of the total combinations to be tested has demonstrated the feasibility and effectiveness of the project (Ito et al., PNAS 97, 1143-1147, 2000).

Now we have almost finished the examination of all possible combinations between the pools, and obtained over 10,000 interacting sequence tags. We are expecting that some 3,000 independent interactions will be finally obtained, thereby providing the largest two-hybrid interaction data sets. Currently, we are constructing a database for these interactions, which will also integrate previously known interactions as well as those by CuraGen and others. Bioinformatic analysis of these huge data allows us to extract more complex protein-protein interaction networks. These networks include both proteins of well-known functions and those of totally unknown functions, and hence provide invaluable hints for the functions of the latter.

Following the cataloging of two-hybrid interactions, we are preceding to the second step of our project, that is, systematic mapping of the protein modules (domains, motifs, etc.) that mediate each interaction. We put particular emphasis on such interaction modules, because each of them can be used as a "perturbagen" to interrupt the cognate interaction *in vivo* and the induced phenotypes would tell its biological role. Furthermore, identification of such interaction modules would help define a general rule for interaction, such as "Domain A interacts motif B", that can be used for the prediction of interactions occurring not only in yeast but also in other species. To facilitate the mapping of protein regions responsible for each protein-protein interaction, we are currently developing a novel method termed the "two-hybrid footprinting", which would open a way for the systematic hunting of protein-interacting modules.