



Identifying protein interactions

Computational approaches

Ettore Appella and Carl W. Anderson

Macromolecular interactions, including protein–protein interactions, are crucial for most biological processes. A typical cell expresses tens of thousands of proteins that may participate in hundreds of thousands of physical interactions at any one instant. Although experimental methods are beginning to produce meaningful protein interaction network maps, they have significant limitations even for moderately complex systems. However, rapid progress in whole genome sequencing and annotation is providing a rapidly growing database of predicted proteins from many different organisms throughout the extant evolutionary tree of life. Computational approaches and methods take advantage of this wealth of information for identifying and understanding the functional and physical interactions that create networks, thus complementing experimental approaches.

The four minireviews in this issue address recent progress in computational approaches to identifying and characterizing protein networks. In the first, Altschul and colleagues describe when and how to use compositionally adjusted substitution matrices for searching protein databases for related protein sequences with programs such as BLAST. Adjusting the substitution matrix when comparing sequences with biased amino acid compositions improves alignments and reduces false-positive identifications for a relatively small cost in speed. Criteria for invoking compositional adjustments are given, and a compositional substitution matrix compatible version of BLAST is available at NCBI's website. In the second review, Bowers and colleagues describe a natural extension of pairwise analysis of genomic data, in which proteins triplets are considered for inferences about biological associations. Given a sufficient number of fully sequenced genomes, the additional power of the derived logic relationships combining two binary states to match a third permits novel biological associations to be inferred computationally. These inferences may then be used to generate

hypotheses explaining cellular functions that can be tested experimentally. In the third review, William Noble and colleagues describe RANKPROP, a program that, like PSI-BLAST, produces a ranking of all proteins in a network with respect to a given query sequence. RANKPROP relies not only upon the similarities identified by PSI-BLAST but also uses information from the global network topology, thereby improving upon the rankings produced by PSI-BLAST. RANKPROP is available through the UC Santa Cruz Gene Sorter (www.genome.ucsc.edu). Protein–protein interactions are organized into well-coordinated networks. While many proteins have only one or a few connections, a few hub proteins possess tens or even hundreds of links. The connectivity of these hub proteins must be reflected in their structure. The fourth minireview by Dunker and colleagues explores the possibility that regions of intrinsic disorder contribute to protein network architecture in two general ways: the intrinsic disorder can serve as the structural basis of hub proteins, or intrinsically disordered proteins can bind well-ordered hub proteins. Several examples are given.

The minireviews in this series were derived from presentations given at the 15th Methods in Protein Structure Analysis Conference (MPSA2004) which took place in Seattle, WA in early September, 2004. MPSA meetings are biannual, international meetings held alternately in the US, Europe, and occasionally other countries. The purpose of MPSA meetings is to promote the discovery and exchange of new methods and techniques for the analysis of protein structure, and to facilitate the application of methods in protein structure analysis in the pursuit of solutions to biological problems. MPSA2006 will be held in Lille, France, 29 August–2 September 2006 (<http://www.iapsap.bnl.gov>). A subsequent issue of *FEBS Journal* will feature additional minireviews derived from MPSA2004 presentations that explore further the themes of protein–protein interactions and protein families.

doi:10.1111/j.1742-4658.2005.04944.x



Ettore Appella is Chief of the Chemical Immunology Section, Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD. Raised in Italy, he received his MD from the University Medical School, Rome, Italy in 1959 and then undertook postdoctoral studies in protein chemistry and immunology at Johns Hopkins University with Clement Markert and subsequently at the NIH with Bruce Ames, Gordon Tomkins and Michael Potter. In 1966, he started his group at the National Cancer Institute. His research interests encompass protein chemistry, immunology, tumor suppressors, and the synthesis of small organic molecules that target the HIV nucleocapsid and the p53 tumor suppressor.



Carl W. Anderson is a Senior Geneticist and Chair of the Biology Department at the Brookhaven National Laboratory (BNL) on Long Island, NY. He received his BA in chemistry and physics from Harvard College in 1966 and his PhD in Microbiology from Washington University, St. Louis, in 1970. After postdoctoral work in the tumor virus group at the Cold Spring Harbor Laboratory, he joined the Biology Department at BNL in 1975 and became department chair in 1999. His research interests encompass the roles of post-translational modifications in regulating cellular stress responses including activation of the tumor suppressor p53 and the mechanism of DNA repair by nonhomologous end-joining.

FEBS Journal MiniReview Series
Identifying Protein Interactions



SERIES I: Computational Approaches

Volume **272**, Issue 20, October, 2005

Cover: Combinations of gene expression predict clinical outcomes (P. M. Bowers *et al.*, pp. 5110-5118).

- Appella, E., and C.W. Anderson. 2005. Identifying protein interactions. *FEBS J* **272**: 5099-100.
- Altschul, S.F., J.C. Wootton, E.M. Gertz, R. Agarwala, A. Morgulis, A.A. Schäffer, and Y.-K. Yu. 2005. Protein database searches using compositionally adjusted substitution matrices. *FEBS J* **272**: 5101-9.
- Bowers, P.M., D. O'Connor B, S.J. Cokus, E. Sprinzak, T.O. Yeates, and D. Eisenberg. 2005. Utilizing logical relationships in genomic data to decipher cellular processes. *FEBS J* **272**: 5110-8.
- Noble, W.S., R. Kuang, C. Leslie, and J. Weston. 2005. Identifying remote protein homologs by network propagation. *FEBS J* **272**: 5119-28.
- Dunker, A.K., M.S. Cortese, P. Romero, L.M. Iakoucheva, and V.N. Uversky. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* **272**: 5129-5148.



SERIES II. Experimental Approaches

Volume **272**, Issue 21, November, 2005

Cover: Nucleic acid programmable protein microarray (N. D. Ramachandran *et al.*, pp 5412-5425.)

- Appella, E., and C. W. Anderson. 2005. Identifying protein interactions. *FEBS J* **272**: 5389-90.
- Fields, S. 2005. High-throughput two-hybrid analysis. The promise and the peril. *FEBS J* **272**: 5391-99.
- Bartone, P., and M. Snyder. 2005. Advances in functional protein microarray technology. *FEBS J* **272**: 5400-5411.
- Ramachandran, N., D. N. Larson, P. R. H., E. Hainsworth, and J. LaBaer. 2005. Emerging tools for real-time label-free detection of interactions on functional protein microarrays. *FEBS J* **272**: 5412-25.
- Houtman, J. C. D., M. Barda0Saad, and L. E. Samelson. Examining multiprotein signaling complexes from all angles. The use of complementary techniques to characterize complex formation at the adapter protein, linker for activation of T cells. *FEBS J* **272**: 5426-36.