

Meeting report

Shaping the future of interactome networks

Patrick Aloy

Addresses: Institute for Research in Biomedicine (IRB), c/ Josep Samitier 1-5, 08028 Barcelona, Spain, and Institutió Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

Published: 1 November 2007

Genome Biology 2007, **8**:316 (doi:10.1186/gb-2007-8-10-316)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/10/316>

© 2007 BioMed Central Ltd

A report of the third Interactome Networks Conference, Hinxton, UK, 29 August-1 September 2007.

Complex systems are often networked, and biology is no exception. Following on from the genome sequencing projects, experiments show that proteins in living organisms are highly connected, which helps to explain how such great complexity can be achieved by a comparatively small set of gene products. At a recent conference on interactome networks held outside Cambridge, UK, the most recent advances in research on cellular networks were discussed. At previous meetings in this series we heard much about the abstract properties of biological networks, often with little application to day-to-day biology, and the achievement of amazing milestones such as the first drafts of human interactomes or the completion of affinity-purification screens for protein complexes in yeast. This year's conference was more down to earth, focusing on identifying the strengths and weaknesses of currently resolved interaction networks and the techniques used to determine them - reflecting the fact that the field of mapping interaction networks is maturing.

Detecting challenging proteins and interaction types

One of the key points that kept popping up during the meeting was the need to identify and establish a reliable set of protein interactions, including binary pairs and larger assemblies. These could then be used to validate the results produced by each technique and, perhaps more importantly, to identify the advantages and drawbacks of each technique. Marc Vidal (Dana-Farber Cancer Institute, Boston, USA) presented the results of a thorough benchmarking of the yeast two-hybrid system. He convincingly showed that interactions discovered in high-throughput yeast two-hybrid screens were as reliable as those from individual experiments, and that their accuracy, in terms of the false-positive rate, was also comparable to that of affinity-purification

assays. He and Pascal Braun (Dana-Farber Cancer Institute) also showed that in an ideal scenario a two-hybrid experiment would be able to detect roughly 25% of the total number of possible interactions. However, the typical coverage of a single screen is only about 10%, and one would have to repeat the same experiment six times to reach the upper coverage limit of 25%. These criteria were used to estimate that there would be approximately 280,000 protein-protein binary interactions in humans, without including splice variants. Anne-Claude Gavin (European Molecular Biology Laboratory, Heidelberg, Germany) addressed similar questions for yeast protein interactions discovered by affinity purification coupled to mass spectrometry (MS). She showed that the reproducibility rate of purifications is about 69% and that, although she and her colleagues were able to see proteins from all functional classes and a wide range of copy numbers, there was a bias towards structural proteins and highly abundant proteins. Overall, they were able to detect roughly 60% of the proteins known to be expressed in exponentially growing yeast.

Both these studies highlighted the fact that no single technique will detect everything, and that to comprehensively chart an interactome network these methodologies and others will have to be combined. Moreover, some proteins are inherently underrepresented in all large-scale screens reported, mainly due to difficulty in their biochemical manipulation. This is the case for membrane and extracellular proteins, many of which have no binding partners reported so far. Gavin Wright (Wellcome Trust Sanger Institute, Hinxton, UK) presented a novel *in vitro* binding assay designed to detect low-affinity interactions between extracellular proteins. This protocol, called an avidity-based extracellular interaction screen (AVEXIS), enables the identification of hitherto unknown cell-surface receptor-ligand pairs and will help to reveal the systems that cells use to communicate with each other.

Igor Staljar (University of Toronto, Canada) introduced a modification of the yeast two-hybrid system designed to

detect interactions that include integral membrane proteins, with special emphasis on those of pharmacological interest. The new membrane yeast two-hybrid methodology (MYTH) involves constructs in which the two halves of a ubiquitin molecule are fused to two potentially interacting proteins, at least one of which is membrane bound, and a transcription factor is inserted after the ubiquitin. If the two proteins interact, a complete ubiquitin molecule is reconstituted and the transcription factor is cleaved by ubiquitin-specific proteases and released to switch on a reporter gene. Staljar also showed how this methodology has been applied to investigate complex cell signaling processes and membrane trafficking using collections of membrane proteins. In this context, Gavin showed how a slight variation in the protocol used in her large-scale affinity-purification screen in yeast enabled the retrieval of 340 membrane proteins out of 628 tagged, including some integral membrane complexes such as the Q/t-SNARE. All the methods mentioned above have been designed with the aim of using them in a high-throughput fashion and need very little modification to be fully automatable.

It also became clear at the meeting that current methods not only miss certain types of proteins but also miss specific types of interactions. All the techniques currently used in large-scale studies are poor at detecting very transient interactions or those that depend on posttranslational modifications, and efforts to remedy this were reported. In these difficult cases computational methods seem to be useful. Rune Linding (Mount Sinai Hospital, Toronto, Canada) and colleagues have exploited the modularity observed in signaling networks to predict specific phosphorylation patterns in DNA-damage responses, thus deciphering some of the most elusive regulatory networks *in vivo*. Linding also reported the experimental validation of some of the predicted relationships by immunoprecipitation and MS analyses. Richard Edwards (University College Dublin, Ireland) showed how computational approaches can be used to discover new motifs in peptide-mediated transient protein interactions.

Proteins are not the only molecules in living organisms and so it makes little sense to study protein interactions in isolation. We now have the experimental tools to investigate protein interactions with other molecules in a systematic way. Martha Bulyk (Harvard Medical School, Boston, USA) and Marian Walkout (University of Massachusetts Medical School, Boston, USA) presented two different systems for studying protein-DNA interactions to reveal the mechanisms underlying regulatory transcriptional networks. Bulyk introduced a DNA microarray-based *in vitro* assay that enables high-throughput characterization of the binding sites for specific transcription factors in DNA and identifies the combinatorial co-regulation of certain genes. Walkout reported the development of an *in vivo* yeast one-hybrid system for the high-throughput identification of interactions

between transcription factors and their target genes in *Caenorhabditis elegans*.

Quantitative proteomics and mass spectrometry

The amount and quality of the data yielded by high-throughput protein-interaction experiments are also being extended. For example, Etienne Formstecher (Hybrigenics, Paris, France) presented a *Drosophila* interaction-mapping project using domain-based yeast two-hybrid technology, which is designed to throw light on cell signaling in human cancer. This technique enables identification of the specific domains in the interacting proteins that are responsible for the binding and extends the scope of the classic yeast two-hybrid experiment, which is only able to detect whether two full-length proteins interact.

Probably the most spectacular advances in this area are related to MS. Hitherto, MS has been used in combination with pull-down assays to identify which proteins are purified together. The field has now advanced to a point where MS can confidently provide information about the composition of functional sub-complexes, protein stoichiometry and even dissociation constants. Albert Heck (Utrecht University, Utrecht, The Netherlands) reported innovative MS-based approaches to disentangle the three-dimensional assembly of components and the dynamic composition of several complexes (for example, RNA polymerases or the exosome, a protein complex involved in RNA processing and degradation). He also showed how the gradual dissociation of complex components can be used to estimate dissociation constants and cooperative effects between proteins.

Matthias Mann (Max Planck Institute for Biochemistry, Martinsreid, Germany) demonstrated the power of his newly developed quantitative proteomics technique for MS - stable-isotope labeling with amino acids in cell culture (SILAC) - to remove false positives in protein-interaction networks and to reveal kinetic aspects of the control of signaling by protein phosphorylation. The abundance of high-quality data confirms that quantitative MS is here to stay and is already making key contributions to most areas in proteomics. The improved methods and new data should allow the field to move on from the static representation of interaction networks to the more realistic and dynamic models necessary for a systems-biology approach.

Improving data gathering and dissemination

Being able to trace, verify and clarify the experiments that generate interaction network data is as important as the data themselves. Sandra Orchard (European Bioinformatics Institute, Hinxton, UK) presented the MIMIX initiative (minimum information requirement to report a molecular interaction experiment) as an attempt to standardize the data that any interaction discovery experiment should

report. These guidelines are supported by most of the main data producers, which guarantees their wide acceptance, and will hopefully result in publications of increased clarity and a rapid, systematic capture of molecular-interaction data in public databases. Advances and updates in the data repositories were reported by Jyoti Khadake (European Bioinformatics Institute) and Andrew Chatr-aryamontri (University of Rome Tor Vergata, Rome, Italy), who presented the IntAct [<http://www.ebi.ac.uk/intact>] and MINT [<http://mint.bio.uniroma2.it/mint>] databases, respectively, two of the world's largest warehouses of protein-interaction data. It was good news indeed to hear that they have agreed to cooperate within the International Molecular Exchange (IMEX) consortium, together with the Database of Interacting Proteins (DIP) and the Munich Information Center for Protein Sequences protein-interaction data resource (MPact), and to cover more journals with manual curation - a tremendous amount of work.

Biophysically possible versus biologically relevant

Virtually all the high-throughput attempts to chart interactome networks detect interactions between macromolecules that are biophysically possible - which does not necessarily mean that they occur in the living cell. Nature has many control mechanisms that can prevent biophysically plausible interactions, such as subcellular compartmentalization, different times of expression and tight control of specificity via competition. Christopher Sanderson (University of Liverpool, UK) addressed this question in an analysis of more than 8,500 putative interactions between E2 ubiquitin-conjugating enzymes and E3 ubiquitin ligases within the human 'ubiquitome'. The resulting gene-family-specific high-density protein-interaction map was combined with information from mutants that perturb true E2-E3 interactions and with bioinformatics analyses, which revealed that, although many spurious interactions are possible, proteins show clear preferences for specific partners. Linding also emphasized the importance of biological control mechanisms of interaction specificity. He showed that, for instance, to consider the biological scenario surrounding an interaction increases the computational ability to assign *in vivo* substrate specificity in phosphorylation events to around 60-80%.

A very encouraging message from the meeting is that, although being error-prone and incomplete, the data and models generated so far have proved useful in understanding biological processes and have triggered innovative biomedical applications. Andrea Califano (Columbia University, New York, USA) showed how the existing data, combined with complex Bayesian integration approaches and a few biochemical validations, has enabled a first draft of the protein-interaction network in the human B

lymphocyte. This has led to the identification of deregulated interactions in specific pathological or physiological phenotypes and helped to identify some key effectors in normal physiology and the causal lesions in several well-studied B-cell malignancies.

Towards a visual proteomics

"We know about molecules; we know about cells and organelles; but the stuff in between is messy and mysterious." In his keynote lecture on how to bridge the resolution gap between single molecules and whole cells, Wolfgang Baumeister (Max Planck Institute for Biochemistry) was quoting from an article by the writer Philip Ball. Classical structural biology techniques, such as X-ray crystallography or single-particle electron microscopy, can provide atomic-level information in the angstrom range about small proteins and large macromolecular complexes. Cell biology has the necessary tools to study cellular organization with a resolution approaching 150 nanometers, but the nanometre range is completely uncharted territory. Baumeister discussed electron tomography as a tool for visualizing large molecular machines and their associations in supramolecular structures in their functional environment. His exciting talk was very well received by an audience that saw the power of combining interaction discovery and structural biology, in what he calls "visual proteomics", to construct a pseudo-atomic atlas of the cellular inner space. Starting from another viewpoint - abstract representations of interaction networks - Ewan Birney (European Bioinformatics Institute) presented a Java-based navigation tool for moving across the biological pathways defined in the Reactome database. The tool is easily adapted to work on any network, and it is easy to imagine how to combine this technology with high- and medium-resolution three-dimensional structures of macromolecular complexes and whole-cell tomograms to create 'Google maps'. By highlighting blurry regions where data are lacking, these navigable models will help to identify the needed research.

It is fascinating to see how the interactome networks research community is evolving in response to new scientific and technological advances. We are clearly on a journey analogous to the one that started 15 years ago and ended with the first draft of the human genome. At last year's meeting, Richard Gibbs, one of the fathers of the human genome project, suggested that we focus on methods development, automation, data gathering and quality checks - which we have done to a large extent. This year, Ed Harlow (Harvard Medical School, Boston, USA) pointed out the need to team up and to cooperate as a real consortium to tackle a model system to completion. Who knows, if we follow his advice, this might be the beginning of the interactome networks era. I look forward to seeing where we have got to at the next meeting in 2008.