



The interactome challenge

John D. Aitchison^{1,2} and Michael P. Rout³

¹Center for Infectious Disease Research (formerly Seattle Biomedical Research Institute), Seattle, WA 98109

²Institute for Systems Biology, Seattle, WA 98109

³Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065

The properties of living cells are mediated by a huge number of ever-changing interactions of their component macromolecules forming living machines; collectively, these are termed the interactome. Pathogenic alterations in interactomes mechanistically underlie diseases. Therefore, there exists an essential need for much better tools to reveal and dissect interactomes. This need is only now beginning to be met.

As a discipline or “ology,” the focus of cell biology is an unusually precise unitary one—the cell. The cell represents the fundamental unit of life, comprising the basic structural and functional unit of all living organisms. A cell’s living properties, in turn, emerge from the dynamic interactions of its millions of individual components, and particularly from the dynamic interactions of two major classes of information-rich polymers, proteins and nucleic acids. These interactions form macromolecular complexes that can function as discrete machines and dynamic molecular liaisons that transmit information and control cellular behaviors. Pathogenic alterations in these interaction networks underlie all diseases. Our understanding of cells, the biological systems they comprise, as well as their pathologies thus relies on the ability to elucidate and interpret these interactions and their dynamics, these interaction networks being loosely but collectively termed a cell’s interactome. Though our current understanding of interactomes remains woefully inadequate, methods and approaches to determine interactomes have improved greatly in the last few years, and particularly with the development of new and enhanced technologies to isolate and quantitatively assess complexes.

The nature of the interactome

Proteins are the central players in interactomes. Most eukaryotic cells express several thousand proteins, and multicellular organisms commonly express tens of thousands of proteins, exponentially increasing the probable total number of macromolecular interactions in these organisms, all passing information on pathways linking each other throughout the cell. For example, analyses suggest on the order of 50,000 different protein interaction pairs, and perhaps four times more in humans (Hart et al., 2006). This number is much higher if one considers the many possible combinations of interactions any protein may make at different times, or if one includes interactions between proteins and other macromolecules (e.g., DNA and RNA). A major

challenge in cell biology is thus to elucidate dynamic interactomes of consequence and to understand how these interactomes lead to cellular phenotypes. The impacts of such understanding are potentially vast, allowing us to create new medical therapeutics and diagnostics as well as to harness cells as factories in the biotechnology industry for medicines, pesticides, biofuels, and new foods and materials. Another major impact area for cell biology is in our understanding of pathogens, which act by directly influencing or hijacking cellular processes in host cells; understanding how leads both to therapeutic insights as well as to a better understanding of cellular processes. The promise of cell biology for its impacts on downstream translational applications thus remains extremely high. But at a time when scientists are increasingly challenged to focus on translational aspects of their research, it is perhaps worth reflecting on what we know about the cellular interactome and the extent to which our current understanding can lead to rational strategies with predictable outcomes for controlling biological systems.

The problem with proteins

For nucleic acids, recent advances have led to an explosion of the available genomic data. Genomes of entire organisms can now be fully sequenced in a day. In contrast, proteins are incredibly diverse in their abundance and their properties, making them highly versatile for the dynamic tasks at hand, but at the same time exceptionally difficult to analyze. Unfortunately, although we may sequence the genome of an organism quickly, we have yet to completely define the interactome of any organism at all! Worse, current technologies do not reveal dynamic interactions between macromolecules at sufficient scale, with sufficient reliability, or with sufficient sensitivity to keep pace with the genomic revolution brought about by sequencing technologies. New imaging technologies can place the components of an interactome in the cellular context and even study their normal dynamic behavior, but of themselves cannot provide all the information needed to elucidate and understand interactomes. It is for these reasons that the interactomic revolution lags badly behind the genomic revolution. At the current pace, it would take us hundreds of years to fully annotate the human interactome with respect to function and dynamic interactions.

Indeed, it is clear that most dynamic interactions relevant to both normal and disease-related cellular processes remain largely undescribed. The seriousness of this deficiency is further exacerbated by the fact that most data available in databases

Correspondence to Michael P. Rout: rout@rockefeller.edu; or John D. Aitchison: john.aitchison@cidresearch.org

© 2015 Rout and Aitchison This article is distributed under the terms of an Attribution-Noncommercial-Share Alike-No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms>). After six months it is available under a Creative Commons License (Attribution-Noncommercial-Share Alike 3.0 Unported license, as described at <http://creativecommons.org/licenses/by-nc-sa/3.0/>).

largely exclude dynamic interactions that change as cells progress through different states, such as those that occur through differentiation or accompany disease. Similarly, almost all interactions dependent on enzymes and their cognate substrates or on posttranslational modifications are most often ignored. As an example of the challenge, it is not uncommon to observe about half of a cell's transcriptome or protein abundance to change significantly during cellular transitions or infection. Moreover, proteins change partners and move; in one experiment in yeast, >400 proteins were detected to shift their localization between the cytosol and cytoplasmic organelles in response to carbon source (Jung et al., 2013). These findings underscore the kinds of extensive changes in interactomes that are normal to living systems and that remain largely unexplored.

There is therefore a desperate need for technologies that can quickly and reliably reveal the dynamic cellular interactome. The irony of the situation is that we are in the midst of a technological renaissance in biology, which has the potential to give the complete and accurate information necessary to go from bench to bedside. Today, the discovery process has been vastly accelerated by the advent of new "omics" and imaging approaches that have pushed the temporal and spatial resolution of cell biology studies to previously unimagined limits. As an example of how revolutions in approaches can transform our understanding of cell biology, let us consider the phenomenal improvements in cellular imaging during the past few decades. By the end of the 1970s, many considered electron and light microscopy to have reached their performance limitations. However, from the 1980s up to the present day, microscopic imaging has been in a constant state of revolution. This revolution is occurring mainly by building on existing platforms (the light and electron microscope), using established principles of physics and materials technology applied together in new ways. Similarly, we must build on the current developments in proteomic technologies, including affinity isolation of macromolecular complexes, mass spectrometry, and next generation DNA sequencing. By judiciously adding new technologies to address bottlenecks and limitations to the current techniques (throughput, speed, signal to noise, and data integration), these technologies are bringing proteomic and interactomic studies to a whole new level—i.e., the ability to produce enlightening dynamic pictures of how macromolecular assemblies form and function in the living cell (Russel et al., 2009; Mast et al., 2014).

Getting at the machinery

Any given macromolecule may make stable interactions with other macromolecules to form a tight complex, with which other macromolecules exchange rapidly in dynamic or transient interactions; and this whole network is surrounded by a macromolecular milieu of other complexes that jostle with it in vicinal interactions. Unfortunately, upon disruption of cells, macromolecular complexes tend to disintegrate and intermingle with components not normally exposed to one another, the resultant possibility of aberrant molecular interactions being a major source of nonspecific background. This problem is one of the most important facing biochemical approaches to the study of macromolecular interactions. Thus, ideally, we should aim to "freeze" a macromolecular complex in place within moments of visualizing its position in the cell and subsequently isolate the intact complex together with all its components and specific neighbors, including dynamic, transient, and vicinal interactors, no matter how fleeting.

One approach our laboratories have had some success with in this regard is cell breakage by cryolysis. The rapid freezing of cells almost instantly preserves their complexes as they were at the moment of freezing, preserving even dynamic and state-specific associations. Then, as the processes of cell breakage and dispersal occur in the solid phase, there can be no change in the relative distribution or association of component molecules, limiting the period during which such changes can occur to only the extraction and isolation stages. Alternatively, high pressure or high shear fluid processors can also break cells rapidly and efficiently while minimizing heating damage associated with other approaches (e.g., sonication).

However a cell is broken open, upon its breakage, the normal microenvironment and larger cellular context surrounding the macromolecular complex of interest is replaced by an artificial one consisting of buffers, salts, and stabilizing agents. Ideally, these are carefully selected to mimic the natural milieu. Even so, we cannot hope to exactly replicate the conditions found inside the cell. In the absence of constant replenishment from a living cell, macromolecular complexes and their microenvironments will rapidly disperse. Moreover, during cell lysis extraction, there is usually a dilution step into the extraction buffer. Dilution favors macromolecular complex dissociation by making reassociation less likely. Disruption of the cell and dissociation of the complexes also leads to time-dependent intermingling of components not normally exposed to one another and the resultant possibility of aberrant molecular interactions, a major source of nonspecific background. One obvious way to address this issue is to isolate complexes rapidly, thereby minimizing time-dependent decay. Such speedy capture can be achieved through high specificity and high affinity capture agents. Fortunately, the need for such agents—nanobodies, ScFvs (single-chain variable fragments), monoclonal antibodies, aptamers, and the like—is appreciated by many and has become a major push in many laboratories and corporations.

Another way to address the problem is to optimize the affinity capture solvent so that it helps to preserve the complex, slowing or preventing the decay process. Unfortunately, such optimization remains empirical and time consuming, such that affinity capture practices often adopt a one-size-fits-all approach to protein isolation that cannot account for the diverse physicochemical properties of protein complexes and their constituents. Hence, there exists a pressing need to expedite the affinity capture optimization process through multiparameter searches of extraction solvents to identify those highly optimized for affinity capture of the protein of interest in order to enable the facile exploration of a broad extraction solvent space.

Finally, chemical stabilizers or cross-linkers can be used to rivet a complex together; chemical cross-linking irreversibly captures binding partners so that even the most transient interactions can, in principle, be detected. Clearly, this approach is highly promising at two levels: as chemical stabilizers to preserve the structure and interactions surrounding a given tagged protein, and as chemical rulers to measure interatomic distances to determine the high resolution structure of a complex. Recently published studies have underscored the tremendous potential that this stabilizer and chemical ruler technology has to revolutionize the elucidation of endogenous protein complexes (Rappsilber, 2011).

Mapping the machinery

It is the dynamic and regulated interactions of macromolecular interaction hierarchies that breathe life into a cell, and it is

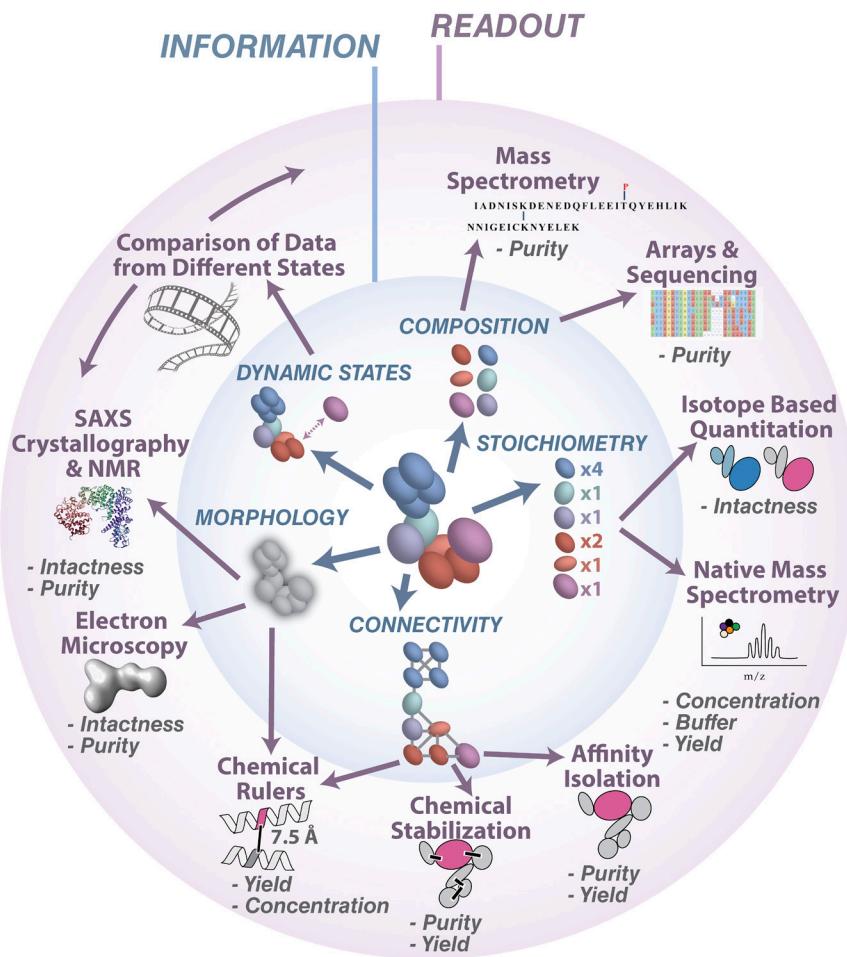


Figure 1. Diagram illustrating analyses of interactomes and complexes, showing the information gathered and the analytical readout. Blue, information gathered; purple, analytical readout (NMR, nuclear magnetic resonance; SAXS, small-angle x-ray scattering). Below each readout are provided the key factors that must be addressed with the kind of technology and data.

these kinds of dynamic data that we must gather and interpret to elucidate cellular and complex functions. As well as providing a highly detailed, albeit static, picture of macromolecular hierarchies, we must gather two kinds of data that inform on the dynamics of macromolecular complexes: snapshots of the dynamic process mediated by a complex, obtained by freezing that complex in sequences in space and time; and comparisons of ensembles of complexes in different states, where the states are defined by differences in the composition, connectivity, and morphology.

The kinds of analyses that can be performed to gain data on the organization and dynamics of an interactome and its component machineries are multitudinous and highly varied, and so are beyond the scope of this essay. However, mass spectrometry is a mainstay of proteomics approaches to determine the protein composition of complexes, but quantitative mass spectrometry approaches, combined with clever biochemical mixing and enrichment analyses, have been designed to detect contaminants and to determine the purity and stoichiometry of complexes. If nucleic acids form a part of a complex, several microarray or high throughput sequencing approaches can be used. If the complex is homogenous enough, it can be morphologically mapped by the ever-improving techniques of electron microscopy. Nevertheless, each kind of analysis requires a macromolecular complex to be presented to it in a way that most efficiently optimizes the analysis, so as to maximize the amount and quality of information obtained (Fig. 1). Indeed, in many cases, if a sample is not optimized, the analysis becomes

impaired or impossible. First, each analytical technique requires an appropriate degree of purity; e.g., electron microscopy requires that virtually every complex that is visualized is identical, and although mass spectrometry is less demanding, high levels of contaminants can be limiting. Second, adequate yield is crucial for techniques where sensitivity is an issue. Third, high concentration is currently an absolute requirement for some approaches (e.g., native mass spectrometry). Fourth, the appropriate buffer is key for some applications such as native mass spectrometry, where only volatile buffers can be used. Finally, morphological intactness (i.e., a low degree of damage) is important in, for example, electron microscopy studies.

Integration into a dynamic and interpretable picture of the interactome

We will then need to be able to integrate these data into models that represent in unprecedented detail the changing interactions of the macromolecular players in almost any dynamic subcellular assembly. The integration of collected data into meaningful representations presents four key challenges. The first is how to extract the maximum amount of information from noisy data obtained from heterogeneous samples. The second is how to find static structural models that satisfy all the data points within their uncertainties. The third challenge is how to extend our techniques for building models of static structures to the modeling of individual snapshot states in the dynamic processes, followed by connecting these snapshots to capture the entire dynamic process. The fourth challenge is to reveal key dynamics

of complex networks without exhaustive measurements of all biochemical parameters. These challenges are coupled, for example when the heterogeneous samples come from many stages of the process measured simultaneously. Moreover, our cellular maps need to map data at the right level of granularity to reveal sufficient detail of dynamic systems and to provide the necessary conceptual framework to navigate and understand the biology without including superfluous or misleading information.

The ultimate goal is to use computational methods for building data-derived models of static and dynamic macromolecular structures as well as molecular networks representing cellular processes. Modeling approaches must be relevant to and tuned for the data types we seek to generate, ideally emphasizing data on molecular interactions that favor quality over quantity and mechanism over scale. It is of course important to construct and advance each model in parallel with experimental biology. This close juxtaposition between modeling and experiment generates a cycle in which experiments set the initial parameters for a model that is then refined based on further experiments inspired by the model, optimizing the completeness, precision, accuracy, and efficiency of the determination of the structural or network models despite noise, sparseness, and ambiguity of the data, even when collected from heterogeneous samples. The earlier a model can be generated, the more effective are the experiments and thus the overall process. The hope is to generate structural and network models resulting in nontrivial insights and hypotheses that can be tested experimentally. Such models are predictive, actionable, and prioritize experiments that are most critical for advancing our understanding, yielding insights into how the macromolecular assemblies and networks operate, how they evolved, how they can be controlled, and how similar functionality can be designed. In particular, they hold promise for rational target-based intervention and drug design strategies.

Interactomics: From bottlenecks to bench to bedside

We envision that current and emerging technologies can be assembled into a benchtop pipeline that can reveal part or all of an interactome under study. One might think of this approach as a multiscale molecular microscope; the first goal prepares the samples for observation, the second enables detailed observation and analysis of the sample, and the last enables integration of the data into a dynamic and interpretable picture of the sample that enables an understanding of the function and dynamic properties of the system. The majority of successful research today is based around individual, small to medium-sized laboratories investigating a particular area of research. This is a tremendous advantage to the field of cell biology. Although it

is clear that high throughput interactome studies have suffered from data quality issues, the high level of expertise, focus, and thoroughness of cell biologists ensures that these issues will be minimized. The molecular microscope, therefore, is of a scale and scope that can enormously empower any individual cell biology group. Furthermore, given that any area of research is just a segment of an interactome, it is possible to cover an entire interactome via the overlapping research of many such groups. We must continue to evolve the most quantitative and robust approaches that seek to preserve complexes in their native states. These approaches, when combined with interpretation through structural and dynamic modeling, will begin to reveal the dynamic molecular architecture of cells and their components.

Such discoveries will impact medical research in several ways. We will uncover novel interactions that contribute to pathophysiological states in the areas of infectious disease and cancer. Advancing methods to reveal high quality, high confidence protein interactions will enable the construction of accurate and comprehensive complex networks that form the basis of physiological and pathophysiological states. These networks are the basis of systems medicine, through which scientists are taking a network view of disease to enable drug repurposing and drug discovery. Finally, discovering macromolecular structures and interfaces between macromolecules at increasing structural resolution will open new opportunities for discovering new classes of druggable targets that disrupt or potentiate complexes and information flow.

Acknowledgments

J.D. Aitchison and M.P. Rout were supported by grants P50 GM076547, P41 GM109824, and R01 GM112108 from the National Institutes of Health.

References

- Hart, G.T., A.K. Ramani, and E.M. Marcotte. 2006. How complete are current yeast and human protein-interaction networks? *Genome Biol.* 7:120. <http://dx.doi.org/10.1186/gb-2006-7-11-120>
- Jung, S., J.J. Smith, P.D. von Haller, D.J. Dilworth, K.A. Sitko, L.R. Miller, R.A. Saleem, D.R. Goodlett, and J.D. Aitchison. 2013. Global analysis of condition-specific subcellular protein distribution and abundance. *Mol. Cell. Proteomics.* 12:1421–1435. <http://dx.doi.org/10.1074/mcp.O112.019166>
- Mast, F.D., A.V. Ratushny, and J.D. Aitchison. 2014. Systems cell biology. *J. Cell Biol.* 206:695–706. <http://dx.doi.org/10.1083/jcb.201405027>
- Rappoport, J. 2011. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* 173:530–540. <http://dx.doi.org/10.1016/j.jsb.2010.10.014>
- Russel, D., K. Lasker, J. Phillips, D. Schneidman-Duhovny, J.A. Velázquez-Muriel, and A. Sali. 2009. The structural dynamics of macromolecular processes. *Curr. Opin. Cell Biol.* 21:97–108. <http://dx.doi.org/10.1016/j.ceb.2009.01.022>