Published in final edited form as:

Expert Rev Proteomics. 2016 May ; 13(5): 447–449. doi:10.1080/14789450.2016.1182023.

The potential impact of recent developments in three-dimensional quantitative interaction proteomics on structural biology

Nurcan Tuncbag[1], **Attila Gursoy**[2,3], **Ozlem Keskin**[2,3], and **Ruth Nussinov**[*,4,5]

[1] Middle East Technical University, Graduate School of Informatics, Department of Health Informatics, Ankara, Turkey

[2] Koc University, Center for Computational Biology and Bioinformatics, Istanbul, Turkey

[3] Koc University, Computer Engineering, College of Engineering, Istanbul, Turkey

[4] Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory, National Cancer Institute, Frederick, Maryland, United States

[5] Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Sackler Institute of Molecular Medicine, Tel Aviv University, Tel Aviv, Israel.

Keywords

integrative modeling; structural bioinformatics; assembly modeling; protein interactions

In the cell, proteins are not isolated entities. They interact with each other and with other molecules including DNA, RNA, small ligands, and membrane lipids to carry out their functions. They typically form multimeric assemblies through non-covalent interactions; examples include the nucleosome, ribosome, viral coats, the E2/E3 ubiquitination machinery, the Mediator, membrane-anchored nanoclusters and more. The strength of the non-covalent interactions and the life times of the assemblies vary in the cell and the spatial organization of their so-called ‘interactomes’ reflects a temporal and dynamic behavior. Some interactions are elicited by specific regulatory states; others take place in distinct cellular compartments or spatial locations. Some are long- others are short-lived regulating intracellular signaling bursts. Capturing these transient interactions with conventional techniques is challenging. In addition, some proteins interact with their partners following post-translational modification events such as phosphorylation, ubiquitination, methylation, or hydroxylation of specific residues. Abnormal alterations in protein interactions can cause cancer, neurodegenerative diseases and other pathologies. The ultimate goal in structural biology is to reveal the three dimensional interaction details of proteins which are essential in order to understand how cellular function is carried out. The characteristics of protein-protein interactions and their physicochemical properties is reviewed extensively in [1].

* nussinor@helix.nih.gov.

Declaration of interest

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy [2] are classical structural biology approaches to determine the structures of protein complexes at high-resolution (reviewed in [3]). Although we have access to a diverse set of protein complexes in atomic detail and the size of this set is increasing exponentially, protein complexes with low solubility and low stability or proteins whose conformations have low population times are difficult to detect with these techniques even though recently NMR has been able to capture rare conformations [4, 5]. Transient interactions are underrepresented because of their low stability. Further, although there are examples of large protein assemblies resolved with these techniques, the increase in their number is relatively slow. Due to inherent structural determination difficulties with these high resolution techniques, some important and populated protein groups are less represented in the structural database as well. For example, membrane proteins represent only ~1% of all structurally resolved proteins in the PDB (as of 2004), despite their functional and therapeutic significance [6]. Later, with the advances in experimental and comparative computational techniques, 10% of all known membrane proteins have been recovered [7]. To increase the structural coverage of protein assemblies, recent high resolution EM (even on the order of 2Å) [8] and conventional low-resolution techniques such as cryo-electron microscopy (cryo-EM) and small angle X-ray scattering (SAXS) (reviewed in [3]) are being used to determine the shape of the complex. However, in the absence of the crystal structures (or high resolution models) of the subunits, these techniques do not provide information about the organization of the subunits within the complex.

The most remarkable advances in quantitative interaction proteomic approaches in the last decades, especially in mass spectrometry (MS), are in structural biology [9]. MS-based methods can be used to obtain the stoichiometry of protein interactions [10]. MS-SWATH combined with affinity purification (AP-SWATH) is another approach which quantifies the temporal changes in protein interaction accurately and efficiently upon perturbation. One such example is the 14-3-3β mediated interactome after IGF1 stimulation [11]. In addition to identifying pairwise protein interactions, MS-based approaches also offer more details about the interaction. The most significant advantage of MS-based approaches is its general applicability -- which is independent of the size, mass, flexibility and conformation of the complex. Several MS-based approaches can reveal protein structures and their interactions. Native MS is a technique that allows structural and functional analysis of intact protein complexes [12]. Although native MS does not identify protein complexes at atomic resolution, it provides information about the stoichiometry, composition, stability and organization of the subunits in the assemblies. It is also important to note that native MS keeps protein complexes intact in the spectrometer by conserving the noncovalent interactions across the protein subunits [13]. The first step in MS-based 3D structure determination is finding the mass of each component in the protein complex which is not an easy task. In general, the mass of the first product is not always equal to the mass derived from the protein sequence. The MS analysis of the denatured complex can help in determining the mass of each subunit. The mass of the protein assembly with the masses of each subunit can give the composition and stoichiometry of the complex. At this point, algorithms (e.g. SUMMIT) that consider all possible combination of subunits within an error range can be used to determine the composition [14]. The ion-mobility MS (IM-MS)

technique discriminates among packing possibilities of the subunits in the complexes, determining if the subunit is ring-shaped, linear or densely packed [15]; combining native MS with IM-MS gives further information on the shape and topology of the complex. In addition, IM-MS can be combined with EM data to find out the subunit arrangements in the complex. Although native MS combined with IM-MS is great for determining the stoichiometry, topology and the overall shape of the complex, it cannot detect the binding residues. Another MS-based method, the chemical crosslinking (XL-MS) technique, resolves this problem by covalently connecting the modules across subunits or within subunits [16, 17]. Inter-protein crosslinks provide information about interacting partners while intra-protein crosslinks provide information about protein topology. With MS techniques, it is also possible to monitor the dynamics of protein complexes which can capture residue level fluctuations in real time. All these techniques give spatial restraints for three-dimensional modeling of protein assemblies and provide the composition, topology, shape, subunit contacts and interaction sites. To have high-resolution structures of the complexes, integrative approaches which combine these techniques with conventional techniques and computational approaches are crucial [17].

Besides the experimental approaches, many computational techniques have been developed to structurally model protein complexes. A detailed review and a list of tools or resources for computational techniques can be found in [1]. These can be grouped into *ab initio* docking, homology modeling and template-based docking as well as predictive models using machine learning and network based approaches (reviewed in [1, 18]). In addition, molecular dynamics simulations can be used to extend the coverage of conformational ensembles for docking approaches to better consider the flexibility of proteins [19]. These approaches can be also considered as high-resolution techniques. As in all predictive approaches, there are false positives as well as true positives. Prior knowledge relating to the interaction of the subunits in the protein complex is expected to improve the prediction accuracy. For example, chemical crosslinking data are especially important to filter out the false positives in the output of these computational approaches. In a recent work, the architecture of the TIR domain signalosome has been computationally modelled [20] using the PRISM algorithm [21], a template-based docking approach, and the modelled assembly has been validated with biochemical data [20]. In another work, homology modeling techniques have been integrated with EM maps to model a high resolution molecular model of the auxillin-clathrin complex. Further, the PRISM algorithm integrated with conformational ensembles of protein monomers and EM maps performed well in modeling large protein assemblies [22]. Integrative modeling has been also applied to the yeast RNA exosome where the spatial restraints from MS data and computational docking of the subunits have been used (as reviewed in [3]).

Each technique, experimental or computational, has its strengths and weaknesses. A single approach is not enough to properly model protein complexes. For more reliable results hybrid approaches are emerging, especially those integrating experimental and computational techniques which complement each other really well [20, 23-26]. It is apparent that the field is moving toward integrative approaches to accurately identify the three-dimensional protein interactions. Spatial restraints obtained by MS techniques can be combined with other structural techniques to model protein assemblies in atomic detail.

Alternatively, low-resolution structural maps can be integrated with the computational methods. The Integrative Modeling Platform [26] serves to provide an integrative framework with several tools for the analysis of different experiments including SAXS and cryoEM data, and integrating these with computational pairwise docking. In a recent integrative approach using this platform, the molecular architecture of the yeast Mediator complex has been determined by combining chemical-crosslinking and MS, X-ray crystallography, homology modeling and cryo-EM data [25].

Hybrid approaches integrating these techniques have been applied to several other cases too. MS of the intact protein complexes has been used to analyze the eukaryotic initiation factor (eIF) and the total mass, masses of all subunits, the stoichiometry and the subunit interaction network have been determined [24]. The model has been further analyzed with phosphorylation proteomics of the complex and 29 phosphorylation sites have been identified. The phospho-sites were found to be located within the core part of the complex which implies that dissociation and association of the subunits are controlled through these events.

In another example, chemical crosslinking and MS have been used to structurally analyze the p53 tetramer [23]. The XL-MS data of the p53 tetramer was found to be consistent with the p53 model obtained by SAXS experiments; however, the regulatory domains of p53 are closer to each other than in the model obtained with SAXS. Additionally, the EM structure of the p53 tetramer is in disagreement with the crosslinking data. These results illustrate that although conventional structural biology techniques fail to obtain high resolution structures of proteins having many disordered regions, chemical crosslinking MS might be a way to overcome this issue. It may also suggest different preferred organizations under different experimental conditions and protocols.

Understanding the functional complexes or assemblies in the cell in three-dimensional space is crucial. Integrative approaches allow advanced modeling of these complexes by combining multi-level data from multiple experiments. This is useful in two ways. First, hybrid approaches consider multiple aspects and spatial restraints of the protein complex which improves its structural model. Second, using multiple data from different experiments can help in cross-validating the results which can limit the false positives in the final list of putative protein models. Although many protein assemblies were successfully determined by hybrid approaches, a large-scale application is still missing. In the near future, these technologies will likely evolve to accurately model intact protein complexes and their dynamics at the global level. Toward this aim, data from multiple experiments accumulate, including EM maps, state of-the-art MS-based data and more and are organized in databases. In line with these experimental advances, the performance of computational approaches to docking and modeling keeps improving. We expect that in the future structural biology will be able to put together a proteome scale map of protein complexes in high resolution obtained by such integrative approaches. The next step will involve putting these in the cellular milieu, consider their interactions with the membrane and with the cytoskeleton, and merge these with data on their stability, dynamics and eventually exploit these to gain further insight into cell life.

Acknowledgments

N Tuncbag acknowledges support from the TUBITAK-Marie Curie Co-funded Brain Circulation Scheme (114C026) and the Young Scientist Award Program of the Science Academy (Turkey). O Keskin and A Gursoy are members of the Science Academy (Turkey). The authors acknowledge the partial funding from TUBITAK projects (114M196 and 113E164).

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

1. Keskin O, Tuncbag N, Gursoy A. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. Chemical reviews. 2016
2. Takeuchi K, Wagner G. NMR studies of protein interactions. Current opinion in structural biology. 2006; 16(1):109–117. [PubMed: 16427776]
3. Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. Nature. 2007; 450(7172): 973–982. [PubMed: 18075576]
4. Clore GM. Visualizing lowly-populated regions of the free energy landscape of macromolecular complexes by paramagnetic relaxation enhancement. Molecular bioSystems. 2008; 4(11):1058–1069. [PubMed: 18931781]
5. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T. Hidden alternative structures of proline isomerase essential for catalysis. Nature. 2009; 462(7273):669–673. [PubMed: 19956261]
6. Walian P, Cross TA, Jap BK. Structural genomics of membrane proteins. Genome biology. 2004; 5(4):215. [PubMed: 15059248]
7. Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS. Three-dimensional structures of membrane proteins from genomic sequencing. Cell. 2012; 149(7):1607–1621. [PubMed: 22579045]
8. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S. 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science. 2015; 348(6239):1147–1151. [PubMed: 25953817]
9. Benesch JL, Ruotolo BT. Mass spectrometry: come of age for structural and dynamical biology. Current opinion in structural biology. 2011; 21(5):641–649. [PubMed: 21880480]
10. Wepf A, Glatter T, Schmidt A, Aebersold R, Gstaiger M. Quantitative interaction proteomics using mass spectrometry. Nature methods. 2009; 6(3):203–205. [PubMed: 19198594]
11. Collins BC, Gillet LC, Rosenberger G, Rost HL, Vichalkovski A, Gstaiger M, Aebersold R. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. Nature methods. 2013; 10(12):1246–1253. [PubMed: 24162925]
12. Heck AJ. Native mass spectrometry: a bridge between interactomics and structural biology. Nature methods. 2008; 5(11):927–933. [PubMed: 18974734]
13. Belov ME, Damoc E, Denisov E, Compton PD, Horning S, Makarov AA, Kelleher NL. From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry. Analytical chemistry. 2013; 85(23):11163–11173. [PubMed: 24237199]
14. Taverner T, Hernandez H, Sharon M, Ruotolo BT, Matak-Vinkovic D, Devos D, Russell RB, Robinson CV. Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. Accounts of chemical research. 2008; 41(5):617–627. [PubMed: 18314965]
15. Uetrecht C, Rose RJ, van Duijn E, Lorenzen K, Heck AJ. Ion mobility mass spectrometry of proteins and protein assemblies. Chemical Society reviews. 2010; 39(5):1633–1655. [PubMed: 20419213]
16. Sinz A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. Expert review of proteomics. 2014; 11(6):733–743. [PubMed: 25227871]

17. Leitner A, Faini M, Stengel F, Aebersold R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. Trends in biochemical sciences. 2016; 41(1):20–32. [PubMed: 26654279]
18. Tuncbag N, Gursoy A, Keskin O. Prediction of protein-protein interactions: unifying evolution and structure at protein interfaces. Physical biology. 2011; 8(3):035006. [PubMed: 21572173]
19. Hospital A, Goni JR, Orozco M, Gelpi JL. Molecular dynamics simulations: advances and applications. Advances and applications in bioinformatics and chemistry : AABC. 2015; 10:37–47. [PubMed: 26604800]
20. Guven-Maiorov E, Keskin O, Gursoy A, VanWaes C, Chen Z, Tsai CJ, Nussinov R. The Architecture of the TIR Domain Signalosome in the Toll-like Receptor-4 Signaling Pathway. Scientific reports. 2015; 5:13128. [PubMed: 26293885]
21. Tuncbag N, Gursoy A, Nussinov R, Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. Nature protocols. 2011; 6(9):1341–1354. [PubMed: 21886100]
22. Kuzu G, Keskin O, Nussinov R, Gursoy A. Modeling protein assemblies in the proteome. Molecular & cellular proteomics : MCP. 2014; 13(3):887–896. [PubMed: 24445405]
23. Arlt C, Ihling CH, Sinz A. Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. Proteomics. 2015; 15(16):2746–2755. [PubMed: 25728495]
24. Damoc E, Fraser CS, Zhou M, Videler H, Mayeur GL, Hershey JW, Doudna JA, Robinson CV, Leary JA. Structural characterization of the human eukaryotic initiation factor 3 protein complex by mass spectrometry. Molecular & cellular proteomics : MCP. 2007; 6(7):1135–1146. [PubMed: 17322308]
25. Robinson PJ, Trnka MJ, Pellarin R, Greenberg CH, Bushnell DA, Davis R, Burlingame AL, Sali A, Kornberg RD. Molecular architecture of the yeast Mediator complex. eLife. 2015; 4
26. Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS biology. 2012; 10(1):e1001244. [PubMed: 22272186]