

# Proteomics and Bioinformatics Soon to Resolve the Human Structural Interactome

Jean-Christophe Nebel\*

Faculty of Science, Engineering and Computing, Kingston University, London, UK

Following the completion of the Human Genome Project, came the humbling discovery that humans possess barely more protein-coding genes than the roundworm. Since then, humans' greater complexity has been substantiated by finely tuned regulatory and expression mechanisms, alternative splicing and post-translational modifications, and a dense protein interaction network resulting in a highly dynamic system. Indeed, the recent publication of the Encyclopedia of DNA Elements (ENCODE), which provides functional annotation for around 80% of the human genome, confirmed the intricate nature of gene expression control by revealing nearly 500,000 'promoter' and 'enhancer' regions [1]. In addition, studies of post-transcriptional alterations led to size estimates of the human proteome much higher than the number of genes, usually within the 100,000-1,000,000 protein range. Whatever the actual value, the number of possible protein-protein interactions is enormous.

To identify those interactions, high-throughput methods are required. Proteomics approaches, such as yeast two-hybrid screens (Y2H), combination of affinity purification and mass spectrometry (AP/MS) and protein microarrays, have been extensively applied to decipher the human protein interactome [2]. Despite those efforts, this network is still largely unknown. Therefore, bioinformatics has become the only practical way of revealing its full extend. Taking advantage of known interactions, computational methods are able to learn associated patterns and predict new interactions [3]. Moreover, as experimental techniques produce a substantial amount of false positives, software tools are also required to assess the validity of proposed interactions [4].

Protein interaction data are extremely valuable for biomedical research and drug design. However, knowledge of the existence of a given interaction is not sufficient to understand how a function is performed; the atomic structure of the protein complex is necessary. Low-throughput proteomics techniques, mainly X-ray crystallography, nuclear magnetic resonance and, to some extent, cryo-electron microscopy, provide detailed descriptions of those interactions. However, due to their high cost and technical limitations, there is no prospect of them resolving the human structural interactome in the foreseeing future. On the other hand, the amount of structural data currently available in the Protein Data Bank - above 80,000 entries 55 percent of which are complexes - may be sufficient to train bioinformatics tools aiming at predicting a significant portion of the human structural interactome.

Traditional docking approaches explore the space of configurations that proteins, involved in a complex, could adopt using energy based cost functions. Since those empirical functions are not optimal, docking tends to produce a set of possible models which need to be further evaluated [5]. Alternatively, additional constraints can be integrated within docking software to reduce the size of the search space [6]. Among them, locations of binding interfaces, hot spots or binding residues have proved particularly informative. In addition

to experimental mutagenesis studies, structure based bioinformatics approaches can provide such data. They can be divided into two categories: analysis of local surface patches according to their chemical and geometric properties [7] or usage of templates based on either local or global structural similarity [8,9], or homologous protein structures [10].

With the weekly release of hundreds of novel high resolution protein and complex structures, and continuous improvements in docking and protein structure prediction, the production of human structural protein-protein networks on a large scale is becoming a reality. Recently, essential elements of the MAP kinase cascades - the signalling pathway directing cellular responses to potentially harmful, abiotic stress stimuli - have been released [11]. A database (PrePPI) containing high confidence predictions of more than 300,000 structures of human protein complexes is now available [12]. Moreover, partial construction of the human structural interactome has already generated medically relevant observations [13]. Following the mapping of disease associated mutations on a structural network consisting of literature-curated binary protein-protein interactions, Wang et al. discovered that a significant number of those mutations is localised on protein interfaces.

Those very encouraging results should not mask the many remaining challenges such as experimental resolution of membrane protein structures, template-free protein structure prediction, docking involving large conformation changes and prediction of weak or transient interactions. However, while more than 10 years elapsed between release of the human genome and its near complete annotation, one can be confident that the next major milestone, i.e. the determination of the human structural interactome, will be reached by the end of the decade.

## References

1. Maher B (2012) ENCODE: The human encyclopaedia. *Nature* 489: 46-48.
2. Braun P, Gingras AC (2012) History of protein-protein interactions: from egg-white to complex networks. *Proteomics* 12: 1478-1498.
3. Gallone G, Simpson TI, Armstrong JD, Jarman AP (2011) Bio::Homology::InterologWalk--a Perl module to build putative protein-protein interaction networks through interolog mapping. *BMC Bioinformatics* 12: 289.

\*Corresponding author: Jean-Christophe Nebel, Faculty of Science, Engineering and Computing, Kingston University, London, UK, E-mail: [J.Nebel@kingston.ac.uk](mailto:J.Nebel@kingston.ac.uk)

Received September 24, 2012; Accepted September 24, 2012; Published September 26, 2012

Citation: Nebel JC (2012) Proteomics and Bioinformatics Soon to Resolve the Human Structural Interactome. *J Proteomics Bioinform* 5: xi-xii. doi:[10.4172/jpb.10000e15](https://doi.org/10.4172/jpb.10000e15)

Copyright: © 2012 Nebel JC. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

4. Kritikos GD, Moschopoulos C, Vazirgiannis M, Kossida S (2011) Noise reduction in protein-protein interaction graphs by the implementation of a novel weighting scheme. BMC Bioinformatics. 12: 239.
5. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20: 45-50.
6. Li B, Kihara D (2012) Protein docking prediction using predicted protein-protein interface. BMC Bioinformatics 13: 7.
7. Qin S, Zhou HX (2007) Meta-ppisp: a meta web server for protein-protein interaction site prediction. Bioinformatics 23: 3386-3387.
8. Jordan RA, El-Manzalawy Y, Dobbs D, Honavar V (2012) Predicting protein-protein interface residues using local surface structural similarity. BMC Bioinformatics 13: 41.
9. Zhang QC, Deng L, Fisher M, Guan J, Honig B, et al. (2011) Predus: a web server for predicting protein interfaces using structural neighbors. Nucleic Acids Res 39: W283-W287.
10. Esmailbeiki R, Nebel J-C (2012) Unbiased Protein Interface Prediction Based on Ligand Diversity Quantification. OASlcs 26: 119-130.
11. Kuzu G, Keskin O, GURSOY A, Nussinov R (2012) Constructing structural networks of signaling pathways on the proteome scale. Curr Opin Struct Biol 22: 367-377.
12. Honig B et al. (2012) Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature, in press.
13. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, et al. (2012) Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat Biotechnol 30: 159-164.

### Submit your next manuscript and get advantages of OMICS Group submissions

#### Unique features:

- User friendly/feasible website-translation of your paper to 50 world's leading languages
- Audio Version of published paper
- Digital articles to share and explore

#### Special features:

- 200 Open Access Journals
- 15,000 editorial team
- 21 days rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at PubMed (partial), Scopus, DOAJ, EBSCO, Index Copernicus and Google Scholar etc
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: <http://www.editorialmanager.com/proteomics>

