EMBO reports

opinion

Toward a role model

Roded Sharan

With ever-growing amounts of omics data, the next challenge in biological research is the integration and interpretation of these data to gain mechanistic insights about cellular function. Protein–protein and protein–DNA interaction networks served for over a decade as the workhorses behind those interpretation efforts, but their use was limited to topologically based observations. Specifically, work in this domain aimed to characterize a given phenotype by the network locations or attributes of its affected (signature) proteins [1] or to distil the specific subnetwork that underlies the phenotype [2,3]. While these methods were successful in associating novel proteins with investigated phenotypes, the resulting models had limited predictive value due to their topological nature.

In contrast to signalling–regulatory networks, metabolic networks are amenable to a rich modelling framework, which is termed constraint-based modelling [4]. This framework quantitatively describes the network's fluxes under a steady-state assumption, which allows the *in silico* simulation of any process of interest under different genetic and environmental perturbations. Borrowing this richness to apply it to signalling and regulation has proven challenging [5] because of the discrete nature of signalling–regulatory circuits.

Although a variety of other models can be employed to enrich the topological description, most prior work focused on logical (boolean) models or variations thereof [6] because of the rich history of these models in the biological domain, which goes back to the 1960s and 1970s. Examples are boolean modelling of fundamental signalling systems, algorithms to optimize boolean models against experimental data and generalizations of these models to multivalued and probabilistic ones.

Despite their predictive power, methods for logical modelling suffer from stark computational and experimental bottlenecks. On the computational side, they rely on the availability of a signalling network for the process at hand, whose interactions are annotated with directions of signal flow and activation or repression effects. This requirement greatly limits the applicability of these methods, as very few processes have been characterized in such fine detail. For most processes, only partial information on key proteins is available, which calls for methods to reconstruct the underlying network and annotate it with data from large-scale screens. On the experimental side, most prior works required the measurement of protein-activation levels under different perturbations, a task for which no standardized large-scale technology yet exists. Phosphoproteomic screens that try to address this challenge are further complicated by the huge size of the phosphoproteome and the need to infer the activity level of a protein from the phosphorylation status of its potentially multiple phosphosites.

Thus, the building of computational pipelines that can exploit other types of data to learn a logical model from scratch for any process of interest is desirable. An alternative to phosphoproteomics could be the use of gene expression data. While not directly reflecting protein-level modifications, such data have the advantage of being much more abundant, larger in scale and gene-centred. Reassuringly, gene expression data have been successfully used for network reconstruction, annotation [7] and model learning [8], but these pieces of the puzzle have never been glued together. The arising challenge is to develop an automatic modelling pipeline that receives as input large-scale data pertaining to the system of interest and outputs a complete logical model that maximizes the fit to the given data. In particular, such a pipeline requires piecing together the learning of topology, annotation and logic to optimize a common objective. Going beyond isolated systems, one could envision a cellular-level model that integrates multiple systems and reveals their crosstalk. Having such a model could greatly enhance our understanding of the cell by allowing the simulation of the effects of specific internal or external perturbations on its workings.

CONFLICT OF INTEREST

The author declares that he has no conflict of interest.

REFERENCES

1. Said MR *et al* (2004) *PNAS* **101:** 18006–18011
2. Shachar R *et al* (2008) *Mol Syst Biol* **4:** 172
3. Yeger-Lotem E *et al* (2009) *Nat Genet* **41:** 316–323
4. Reed JL, Palsson BO (2003) *J Bacteriol* **185:** 2692–2699
5. Klamt S *et al* (2006) *BMC Bioinformatics* **7:** 56
6. Morris MK *et al* (2010) *Biochemistry* **49:** 3216–3224
7. Yeang CH, Ideker T, Jaakkola T (2004) *J Comput Biol* **11:** 243–262
8. Gat-Viks I *et al* (2006) *J Comput Biol* **13:** 165–181

Roded Sharan is at the School of Computer Science, Tel Aviv University, Tel Aviv, Israel. E-mail: roded@post.tau.ac.il

EMBO *reports* (2013) **14,** 948; published online 15 October 2013; doi:10.1038/embor.2013.161