Optimizing protein expression levels as a function of network topology minimizes nonfunctional complex formation

David O. Holland[1], Margaret E. Johnson[2]
[1]Johns Hopkins School of Medicine, Department of Biomedical Engineering, [2]Johns Hopkins University, Department of Biophysics

Cells are crowded by macromolecules, posing challenges for proteins to locate functional partners and avoid misinteractions. Overexpressed proteins may saturate partners, leaving leftovers for nonspecific binding. To avoid this, protein expression levels may be balanced according to the structure of their binding networks. We simulated several such networks under varying protein concentrations while allowing for nonspecific interactions. It was found that relative concentrations could be optimized to minimize misinteractions, and that network motifs determined how sensitive the networks were to non-optimal concentration levels. We conclude that there is evolutionary pressure on both protein abundance and network topology.

***Keywords* — Cellular crowding, dosage balance hypothesis, misbinding, protein-protein interaction network, network motif**

I. Introduction

To perform multiple functions reliably, cells have evolved a vast network of protein-protein interactions (PPI). Human cells alone contain about 20,000 genes encoding for at least 30,000 unique protein types[1]. One challenge that cells face is ensuring that their proteins bind to functional partners reliably as they diffuse through the cell. The cell interior is crowded: 5-40% of cell volume is occupied by macromolecules[2], posing challenges for proteins to both locate functional partners and avoid misbinding. Misbinding – nonspecific interactions that pose no benefit to the cell – can be hazardous to cell function, depleting resources and leading to pathogenic aggregations[3,4]. Highly abundant proteins are at particular risk for misbinding since they may saturate functional partners, leaving leftovers for nonfunctional binding. These "supersaturated" proteins have been linked to neurodegenerative diseases[5]. To avoid leftover proteins, cells may have evolved stoichiometrically balanced gene expression levels, a theory known as the "dosage balance hypothesis" (DBH)[6]. Indeed, copy number variations of genes have been linked to increased susceptibility to a number of diseases, including cancer and multiple sclerosis[7,8]. While the DBH has been explored for single protein complexes, one unexplored question is whether protein expression levels are balanced according to their overall binding networks.

II. Results

To study the effects of relative protein abundance on nonspecific complex formation, we first simulated five simple network motifs under varying protein concentrations using the Gillespie algorithm. While the motifs formed roughly the same proportion of nonspecific complexes under optimal conditions, they varied in sensitivity to initial concentrations (ICs), with the hub being the most sensitive and triangle being the least, and high sensitivity correlating with motifs that allow more ways to form nonspecific complexes. We then simulated 500 large networks of 90-200 nodes with varying topological properties under equal, random, and optimized ICs. Binding affinities for all specific and nonspecific interactions were determined using a coarse-grained protein sequence model. The proportion of protein in nonspecific complexes was recorded as a function of degree distribution, network density, average binding strength, local topology, and ICs. It was found that optimizing the local topology via introducing more hubs and less chains and flags; similar to real networks; decreased the number of nonspecific complexes under optimal ICs, but also increased sensitivity to ICs. Degree distribution, surprisingly, had little influence once local topology was optimized. A lower average binding strength resulted in a lower proportion of nonspecific complexes, in agreement with the hypothesis that abundant proteins are less sticky to avoid misinteractions[9].

III. Conclusion

We conclude that there is evolutionary pressure to both favor certain network motifs and to balance protein abundance to avoid misinteractions. Future work will add noncompetitive binding to the model and perform the analysis on real protein networks to compare with experimental expression level data.

References

[1] Kim M-S, et al. (2014) A draft map of the human proteome. *Nature* **509**, 575–81 (2014).
[2] Ellis RJ, Minton AP (2003) Cell biology: join the crowd. *Nature* **425**, 27–8 (2003).
[3] Zhang J, Maslov S, Shakhnovich EI (2008) Constraints imposed by non-functional protein-protein interactions on gene expression and proteome size. *Mol. Syst. Biol.* **4**, 210.
[4] Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE (2009) Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **78,** 959–91.
[5] Ciryam P, Tartaglia GG, Morimoto RI, Dobson CM, Vendruscolo M (2013) Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep.* **5**, 781–90.
[6] Veitia RA, Bottani S, Birchler JA (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet*. **24**, 390–7.
[7] Conrad DF, et al. (2010) Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–12.
[8] McElroy JP, et al. (2013) Copy number variation in pediatric multiple sclerosis. Mult. *Scler*. **19**, 1014–21.
[9] Levy ED, De S, Teichmann SA (2012) Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc. Natl. Acad. Sci*. **109**, 20461–6.