Optimizing information flow in small genetic networks

Gašper Tkačik[1], Aleksandra M Walczak[2], and William Bialek[2]

Short Abstract — We explore how cells can maximize their control power in the presence of genetic regulatory noise; formally, we solve the problem of maximizing the information transferred from transcription factors (TFs) to target genes. In the simplest version of the problem, a single TF controls the readout of one or more genes by binding to DNA. We find a rich set of optimal input/output relations for the target genes that are completely determined once the constraints on the number of molecules (and therefore noise) have been specified. Even with our simplifying assumptions, we see parallels between the structure of solutions and the behavior of actual genetic regulatory networks.

***Keywords* — Gene regulation, networks, noise, information theory**

THE he ability of cells to control the expression levels of their genes is central to growth, development and survival. In this work (see Ref [1] for the preprint and additional references) we have explored perhaps the simplest model for this control process, in which changes in the concentration of a single transcription factor (TF) protein modulate the expression of one or more genes by binding to specific sites along the DNA. Such models have many parameters, notably the binding energies of the transcription factor to the different target sites and the interactions or cooperativity among factors bound to nearby sites that contribute to the control of the same gene. This rapid descent from relatively simple physical pictures into highly parameterized models is common to most modern attempts at quantitative analysis of biological systems. Our goal in this work is to understand whether these many parameters can be determined by appeal to theoretical principles, rather than solely by fitting to data.

The theoretical principle to which we appeal is the optimization of information transmission [2,3]. In the context of genetic control systems, if the system can transmit I bits of information, then adjustment of the inputs allows the cell to access, reliably, 2^I distinguishable states of gene expression (analysis of experiments on real control elements suggests a limit of 1-3 bits on the capacity of a TF to control the expression level of one gene; in at least in one case, the system can achieve ~90% of its capacity [4]).

[1]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia PA 19104. E-mail: gtkacik@sas.upenn.edu

[2]Joseph Henry Laboratories of Physics, Lewis-Sigler Institute for Integrative Genomics, and Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544. E-mails: awalczak@princeton.edu, wbialek@princeton.edu

The strategy needed to optimize information transmission depends on the structure of the noise in the system. In the case of transcriptional control, the interplay between the input and the output noise [5,6] sets a characteristic scale for the concentration of transcription factors, c_0 ~ 15 - 150 nM [1]. If the maximum available concentration is too much larger or smaller than this scale, then the optimization of information transmission becomes degenerate, and we lose predictive power. Further, c_0 sets the scale for diminishing returns on information. With any reasonable cost for producing TF proteins, we find that the optimal tradeoff between bits and cost will set the typical TF concentration in the range of c_0.

The optimization of information transmission is largely a competition between the desire to use the full dynamic range of outputs and the preference for outputs that can be generated at low noise. Because of the combination of noise sources, this competition has non-trivial consequences, even for a single transcription factor controlling one gene. As we consider the control of multiple genes, the structure of the solutions becomes richer. Activators and repressors are both possible, and can achieve nearly identical information capacities. At low input dynamic range, the optimal arrangement of several output genes is redundant, with overlapping input/output relations; at high input dynamic range, the activation curves for output genes "tile" the TF concentration range. We compute exact numerical results, find analytical approximations for capacity and optimal parameters, and analyze the robustness to parameter variations. Extensions to this work include the computation of optimal solutions with (self-) interacting output genes, thereby generating a diversity of (locally) optimal solutions which can be compared to real regulatory networks, e.g. in early fly embryogenesis [5].

REFERENCES

[1] Tkačik G, Walczak AM, Bialek W (2009) Optimizing information flow in small genetic networks. *arXiv:0903.4491*.

[2] Ziv E, Nemenman I, Wiggins CH (2007) Optimal signal processing in small biochemical networks. *PLoS One* **2**, e1077.

[3] Tkačik G, Callan CG Jr, Bialek W (2008) Information flow and optimization in transcriptional regulation. *Phys Rev E* **78**, 011910.

[4] Tkačik G, Callan CG Jr, Bialek W (2008) Information flow and optimization in transcriptional regulation. *PNAS* **105,** 12265.

[5] Gregor T, Tank DW, Wieschaus EF, Bialek W (2007) Probing the limits to positional information. *Cell* **130,** 153.

[6] Tkačik G, Gregor T, Bialek W (2008) The role of input noise in transcriptional regulation. *PLoS One* **3,** e2774.