Rationalizing the optimality of the gap gene system by ab-initio derivation of optimal ensembles of morphogenetic patterns

Thomas R. Sokolowski[1,2], Aleksandra Walczak[3], Thomas Gregor[4,5]
William Bialek[5], and Gašper Tkačik[2]

[1] Frankfurt Institute for Advanced Studies (FIAS), Frankfurt am Main, Germany
[2] Institute of Science and Technology Austria (IST Austria), Klosterneuburg, Austria
[3] École Normale Supérieure, Paris, France
[4] Institut Pasteur, Paris, France
[5] Princeton University, Princeton, New Jersey, U.S.A.

Early embryogenesis is driven by spatio-temporal patterns that specify distinct cell identities according to their locations in the embryo. This process is highly reproducible, although it results from regulatory interactions that are individually noisy. Despite intense study, we still lack comprehensive, biophysically realistic models that simultaneously reproduce quantitative data and rigorously explain high developmental precision. Moreover, traditional approaches fail to explain why particular patterning mechanisms evolved, and why they favor particular parameter values. We address both questions during early fly embryo development. In *Drosophila*, the gap gene expression patterns were shown to optimally encode positional information. We therefore asked whether one can mathematically derive the gap gene network—without any data fitting—by maximizing the encoded positional information. To this end we built a generic, biophysically accurate spatial-stochastic model of gene expression dynamics, where genes respond to morphogen input signals and mutually interact in an arbitrary way, and optimized its parameters for positional information. Firstly, our results show how the experimentally observed precision can be achieved by basic biochemical processes under known resource and time constraints. Secondly, we find that a rich ensemble of optimal solutions exists and systematically analyse its characteristics, finding that some optimal solutions closely correspond to the real gap gene pattern. Finally, we explore a broad range of "mutated" optimal ensembles in which relevant components of the wild-type system are altered or fully discarded, and systematically map out how this affects positional information and other pattern properties; this allows us to rationalize the design of the wild-type gap gene system and the possible roles of its specific components. To our knowledge our work provides the first successful ab-initio derivation of a nontrivial biological network in a biophysically realistic setting. Our results suggest that even though real biological networks are hard to intuit, they may represent optimal solutions to optimization problems which evolution can find.