Rationalizing the optimality of the *Drosophila* gap gene system by ab-initio derivation of optimal solutions for morphogenetic patterns

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

***Abstract*— Early fruit fly development is outstandingly precise in spite of the high level of stochasticity in the underlying biochemical processes. While the gap gene system driving fly embryo patterning has been shown to encode positional information optimally, the precise mechanisms that enable this remain elusive. We show that optimal solutions for the gap gene regulatory network can be obtained by ab-initio optimization of a spatial-stochastic embryo model, without inferring from data. Firstly, our predictions mechanistically explain how the observed developmental precision can be attained. Secondly, by exploring rich sets of optimal solutions, we elucidate the role of key components controlling early fly patterning.**

***Index Terms*— Embryo development, Drosophila, gap genes, morphogens, patterning, gene regulation, biochemical networks, spatial-stochastic modeling, information theory, optimization**

I. Background

Early embryogenesis is driven by complex spatio-temporal patterns that specify distinct cell identities according to their locations in the embryo. This process is remarkably reproducible, even though it results from regulatory interactions that are individually noisy. Despite intense study, we still lack a comprehensive, biophysically realistic model for at least one biological system that could simultaneously reproduce quantitative data and rigorously explain the emergence of developmental precision. Moreover, traditional approaches fail to provide any insight as to why certain patterning mechanisms (and not others) evolved, and why they favor particular sets of parameter values. We address both questions during early fly embryo development.

II. Scientific Approach

Previous work has shown that the gap gene expression patterns in *Drosophila* optimally encode positional information [1], [2]. We therefore asked whether one can mathematically derive the gap gene network—without any fitting to data—by maximizing the encoded positional information. To this end, we extended our previous models [3], [4], [5] into a generic, biophysically accurate spatial-stochastic model of gene expression dynamics, where gap genes respond to morphogen input signals and mutually interact in an arbitrary fashion, and optimized its parameters for positional information.

[1]Institute of Science and Technology (IST) Austria, Am Campus 1, A-4300 Klosterneuburg, Austria.

[2]Present Address: Frankfurt Institute for Advanced Studies (FIAS), Ruth-Moufang-Str. 1, D-60438 Frankfurt, Germany.
Email: sokolowski@fias.uni-frankfurt.de

[3]Department of Physics / Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08540, U.S.A.

[4]Insitut Pasteur, Department of Developmental and Stem Cell Biology, 25 Rue du Dr. Roux, F-75015, Paris, France

III. Results

Firstly, our results show how the experimentally observed precision can be achieved with basic biochemical processes and within known resource and time constraints. Secondly, we show that a rich ensemble of optimal solutions exists and systematically analyse its characteristics, finding that some of the optimal solutions closely correspond to the real gap gene expression pattern. While numerous, optimal solutions still constitute a small subset of all possible solutions, and feature common properties. Finally, we explore a broad range of "mutated" optimal ensembles in which relevant components of the wild-type setting are altered or completely discarded, and systematically map out how this affects the encoded positional information and other relevant pattern properties; this allows us to rationalize the design of the wild-type gap gene system and the possible roles of its specific components.

IV. Conclusion

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.

References

[1] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, "Positional information, in bits," *Proceedings of the National Academy of Sciences*, vol. 110, no. 41, pp. 16 301–16 308, Oct. 2013.

[2] M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, and T. Gregor, "Optimal decoding of cellular identities in a genetic network," *Cell*, vol. 176, no. 4, pp. 844–855.e15, Feb. 2019.

[3] T. R. Sokolowski and G. Tkačik, "Optimizing information flow in small genetic networks. IV. spatial coupling," *Physical Review E*, vol. 91, no. 6, Jun. 2015.

[4] T. R. Sokolowski, A. M. Walczak, W. Bialek, and G. Tkačik, "Extending the dynamic range of transcription factor action by translational regulation," *Physical Review E*, vol. 93, no. 2, Feb. 2016.

[5] G. Tkačik and A. M. Walczak, "Information transmission in genetic regulatory networks: a review," *Journal of Physics: Condensed Matter*, vol. 23, no. 15, p. 153102, Apr. 2011.