A geometric analysis of multiple-scale models for stochastic gene expression

Applicants: Dr. Nikola Popovic (School of Mathematics) and Prof. Peter Swain (Centre for Systems Biology at Edinburgh)

Gene expression is a fundamentally stochastic process, which makes the modelling of genetic networks challenging. In recent work, Shahrezaei and Swain [1] have investigated a class of two and three-stage models under the assumption that the decay of protein is substantially slower than that of mRNA. Introducing the ratio of the degradation probabilities of mRNA and protein as a natural large parameter in the corresponding chemical master equation, formulating an equivalent partial differential equation and performing a perturbation analysis, they obtained a leading-order approximation for the resulting probability distribution of protein.

Preliminary analysis by Popovic – prompted by an Applied and Computational Mathematics seminar presentation by Swain – indicated that dynamical systems techniques and, in particular, geometric singular perturbation theory [2] can provide a more systematic and rigorous understanding of the fast-slow (multiple-scale) characteristic system of ordinary differential equations that is equivalent to the original master equation. In collaboration with Swain and Dr. Carsten Marr (Helmholtz Center Munich and Institute for Bioinformatics and Systems Biology, who was visiting Edinburgh at the time), Popovic has since completely characterized the dynamics of the two-stage model of gene expression proposed in [1]: in the process, we have obtained a systematic expansion procedure (in terms of the large perturbation parameter) for the probability distribution function that can in principle be taken to any order, both in the time-dependent and the stationary cases, thus improving on the results of [1].

In future work, we will develop an analogous procedure for the three-stage model formulated in [1], and we will investigate the parameter dependence of the resulting distributions. Moreover, we will consider generalised multi-stage models that will account for "refractory" states and transcriptional bursting [3]. Finally, we will perform a systematic and comprehensive numerical verification of our results. (This verification has already been initiated, and will be completed by Marr in collaboration with both Popovic and Swain.) The co-investigators' contributions to the programme can be summarised as follows: the necessary analytical skills will be provided by Popovic, the numerical expertise by Marr, and the essential biological inputs by Swain.

Our results will be highly relevant to biologists: recent technological advances in microscopy and microfluidics have allowed for the generation of long-term time series data of gene expression in single cells, which is usually fitted either via extensive numerical simulation or via so-called likelihood-free Markov chain Monte Carlo methods. (In particular, such time series are being generated in the laboratory of Swain in budding yeast, and by Marr and his collaborators in murine embryonic stem cells.) By contrast, our approach will yield analytical (closed-form) expressions for the propagators that describe the probability of observing a particular level of protein given a previous observation at some earlier point in time. These expressions will enable us to calculate the likelihoods of experimentally observed time series – resulting in a substantial increase of the accuracy and usability of the available fitting algorithms – and, ultimately, to distinguish between competing models of gene expression.

[1] Shahrezaei and Swain, Proc. Acad. Sci. USA 105(45), 17256-17261, 2008.
[2] Jones, Lecture Notes in Mathematics 1609, 44-118, 1995.
[3] Suter, Molina, Gatfield, Schneider, Schiebler, and Naef, Science 332(6028), 472-474, 2011.