Abstract

Mathematical models of biochemical processes are essential tools to understanding the dynamics of intercellular events in living organisms. The copy number of species in such a system fluctuates in time due to the random occurrence of chemical reactions, leading to variability in the population of living cells. Therefore, characterising the number of species at a certain time requires stochastic models of particular interest. To this end, numerous modelling approaches have been proposed in both deterministic and stochastic settings to elucidate the dynamic behaviour of biochemical reaction networks.

Gene expression, the production process of gene products such as proteins, is one of such biological mechanisms intensively studied over the last decades. The (basic) two-stage model gives the essential description of gene expression, which entails the dynamics of transcription and translation processes involving mRNA and protein species. In the stochastic context, the dynamics of species in gene expression is given by the chemical master equation (CME). Despite its simple structure, finding a solution to the CME is often challenging, and analytical solutions are inaccessible for most systems. Consequently, numerical methods, including stochastic simulations, emerge as a remedy to the underlying problem.

On the other hand, with the recent advances in technology, experimental studies open a new window into the more advanced models capable of capturing the detailed dynamics of molecular systems. Accordingly, the present biochemical reaction models need to be revisited. In particular, gene expression is far more complex than the basic two-stage model. The mRNA molecules, for example, can switch between their active and inactive states.

In this thesis, we consider and study various structured chemical reaction systems tailored for gene expression, generalising the results of the basic two-stage model. Specifically, we first begin with a stochastic gene-expression model that accounts for a self-regulating protein molecule with exponential and phase-type lifetimes. We show that the one-dimensional and multiclass-multistage models exhibit the same stationary distribution in the case of non-bursty production of protein regime.

Second, we focus on extending the basic two-stage model to that involving an mRNA inactivation loop, where mRNA molecules are allowed to transition between

active and inactive states. For this model, we present a systematic mathematical derivation of the generating function of the stationary distribution of the species, providing the marginal protein distribution given in terms of special functions.

Next, we characterise the protein noise in terms of the two metrics, Fano factor and the squared coefficient of variation, concluding that the extended model exhibits less protein noise than the basic two-stage model. Importantly, we demonstrate how the models studied here play an important role in modelling the formation of stem–loops, thus controlling noise.

Finally, we present a generalisation of the two-stage and the extended model, which involves multiple mRNA states. We give a detailed mathematical analysis of the model, obtain marginal molecular distributions, and provide an additional gene expression example, which can be obtained from the generalised model.

Overall, in this thesis, we develop and study various gene expression models that may contribute to understanding the stochastic dynamics of biochemical reaction networks arising in relevant research fields.

Keywords: biochemical processes • Master equation • stochastic simulation • gene expression