Chapter 5

Conclusion

Synthetic biology has come a long way since Wacław Szybalski first articulated it as a possibility. Yet, we are still only scratching the surface of the vast potential of this field. We believe that the use of mathematics and modeling can be just as rewarding in our field as it has been in other engineering disciplines. It will not only help us gain deeper understanding of the biological systems we are working with, but also enable more nuanced and carefully designed applications, and ultimately complex architectures such as Szybalski's "new better mouse".

In this thesis, we have developed and demonstrated new applications, tools and mathematical frameworks associated with the use of modeling in synthetic biology.

We began in Chapter 2 with an application of modeling approaches to reduce variability across extracts. We showed that not only could this be achieved, but the classical problem of parameter non-identifiability, first articulated in the control systems literature by Bellman and Astrom in their paper on structural identifiability [4], is not necessarily a death knell for modeling. Indeed, we show in that chapter that for our modeling objective, the sets of values that parameters can take due to non-identifiability may be treated as equivalent with some restrictions. At a fundamental level, this happens because of an exact alignment between two facts. Firstly, the only data we are attempting to transform using the models is the output data. Secondly, the parameters that are non-identifiable are so because their value cannot be discerned from the outputs, precisely because the output is insensitive to variations in these parameters within the sets of output-indistinguishable values. This equivalence of the values the non-identifiable parameters can take allows the

use of *arbitrary* values within these sets, greatly easing the modeling task. Said differently, we showed that with some restrictions, the sets of output indistinguishable parameters can be treated as equivalence classes with respect to the modeling objectives, and therefore any element can be used as a representative of its respective class. Indeed, these results hint at a somewhat deeper fact about the nature of parameter spaces and their use in model-based prediction in a reductionist framework. Specifically, we think that it might be possible to prove and further develop the following statement: Given system component models containing non-identifiable parameters, and the composition of the component models into a whole, the set of parameter values identified using the whole model must be a superset of the Cartesian product of the sets of the corresponding parameters' values identified from the component models. This result is exactly analogous to Condition 2.14 in Chapter 2, and can be thought of as the need for a lack of *emergent* behavior when system behavior has to be predicted from the combination of part models. We also note that when parameter covariation between parameters from different subsystems exists in the whole system's model, a generalization of the CSP fixing method might be applicable.

In Chapter 3, we discussed two tools to help with modeling of genetic circuits. `txtlsim` is a MATLAB® Simbiology® based toolbox bringing computer aided design capabilities to the TX-TL cell-free prototyping platform. This tool should help reduce the reliance on intuition alone for the design of new circuits, and in doing so allow for more rapid and larger scale design iterations. We also introduce mcmc_simbio, which we have packaged as a sub-toolbox within `txtlsim`, that performs Bayesian inference on parameters from multiple models and experimental data sets. The main utility of this toolbox, after its nominal use for estimating parameters for characterization purposes, is in studying the identifiability of parameters informed by different choices of model-experiment sets. This is a consequence of the fact that the parameter inference capabilities are both *Bayesian* and *concurrent*. The use of Bayesian inference allows us to directly visualize parameter identifiability, and indeed explore it a priori using artificial data, as was done in Section 2.7.1 in Chapter 2. The concurrence feature of the parameter inference allows for a common set

of parameters to be identified by an entire ensemble of different experiments, with each experiment informing some appropriate subset of parameters. This allows for different sets of experiment designs to be proposed, and then checked for parameter identifiability. Combined with the type of results in Chapter 2, which show that non-identifiability need not be completely eliminated from models to meet modeling objectives, these tools form a potentially powerful framework for applying computational design to synthetic biology applications.

Finally, in Chapter 4, we developed a mathematical framework for model reduction of transcription and translation reactions while still accounting for resource consumption. Tracking resource consumption is important in finite resource environments like TX-TL, and might even be important for quantifying the metabolic load that synthetic gene circuit expression places on cells. Detailed elongation models are able to capture resource binding and consumption, but are unnecessarily complicated for the purposes of modeling gene circuits. We showed that the incorporation of a *consumption reaction*, which consumes resources at a rate that is a function of the transcript production rate, allows for the reduction of the detailed elongation models into a single step (or a few steps, if nascent polymers need to be modeled), while still accounting for the consumption of resources.

In the process of performing the model reduction, we showed that species concentration coordinates did not allow for a separation of timescales argument to be made using singular perturbation theory. We noted that the timescale separation occurs in the rates of the reactions, and that the timescale separation in the rates of change of species concentrations is simply a consequence of this fact. Thus, while timescale separation can only be achieved for some network topologies when species concentration coordinates are used (specifically, when there is at least one species that only participates in slow reactions), it should be generally achievable when the system is written in our transformed coordinates. Thus, we show that these coordinates are a more natural way to perform timescale separation in biochemical networks.

In conclusion, modeling and computation hold promise of accelerating the develop-

ment of synthetic biology as a field, and will become indispensable as the scope of the systems we try to engineer surpasses what we can intuit.