Chapter 1

Introduction

The vision of synthetic biology was first laid out as early as 1974 by the Polish geneticist Wacław Szybalski [38]: "the real challenge will start when we enter the synthetic phase of research in our field. We will then devise new control elements and add these new modules to the existing genomes or build up wholly new genomes. This would be a field with an unlimited expansion potential and hardly any limitations to building 'new better control circuits' and 'new better lambdas', or finally other organisms, like a 'new better mouse'". These early words have been shown to be surprisingly close the the mark, as evidenced by the great strides made in synthetic biology in the last eighteen years.

We may define synthetic biology as the attempt at using principles from engineering disciplines to design and build novel functionalities out of biological components. One of the major paradigms within synthetic biology is the bottom up design of novel genetic circuits, defined as networks of gene regulatory elements and other molecular machinery, to control cellular behavior. Examples of genetic circuits include the seminal work on oscillators [15] and bistable modules [23] and the subsequent work on logic gate circuits [45], feedforward circuits [24], spatial control of expression [64], and circuits leveraging non-coding RNAs for regulation [65], among others.

One of they key paradigms in engineering is the use of simplified prototyping environments for testing system designs. Examples include wind tunnels [3] in aeronautics and breadboards [52] for circuit prototyping in electrical engineering. In synthetic biology, this role is increasingly being played by cell-free protein expression systems [60]. Constituted of cell extracts or purified cellular machinery mixed with buffers containing energy

molecules and other resources, these test tube based platforms provide numerous advantages for genetic circuit prototyping. Firstly, since the DNA encoding the genetic circuit is not constrained by the need for DNA replication, restrictions due to plasmid selection markers and antibiotic compatibility are lifted. This allows for the rapid exploration of genetic circuit variants by adding circuit component containing DNA species in different combinations. Furthermore, time-consuming cloning and transformation steps may be bypassed by using linearized DNA in the form of polymerase chain reaction (PCR) products or *de novo* synthesized fragments, which speeds up the design-build-test cycle time considerably. These advantages of combinatorial multiplexing and quicker cycle times are further compounded by using liquid handling robots [55] and microfluidic devices [47], which cell-free gene expression is particularly well suited for.

We consider the cell-free transcription-translation system developed by Noireaux [60], or TX-TL for short, as our prototyping platform. TX-TL uses an *E. coli* cell extract as a source of the gene expression machinery and may be implemented in batch or continuous flow modes [47]. In either mode, the extract used must be harvested by the lysis of *E. coli* cells, and different batches of these extracts can show significant variability in gene expression capacity. In Chapter 2 we introduce the calibration-correction method, a model-based methodology for calibrating extract batches, and correcting circuit behavior based on these calibrations. We find that naive implementations of this methodology have certain performance limitations, and that the method may be improved by considering parameter non-identifiability of the models used, and designing modifications to the method based on these considerations. In Chapter 3 we observe that the analogy to prototyping platforms in traditional engineering disciplines may be extended by using computer aided design (CAD) software to model genetic circuits before any physical prototyping is attempted. To this end, we describe `txtlsim`, a MATLAB® Simbiology® based toolbox for prototyping genetic circuits in the TX-TL cell-free expression system. This chapter also discussed the capabilities of a sub-toolbox for performing Bayesian parameter inference using ensembles of experimental data. In Chapter 4, we discuss the theoretical justifications for certain modeling choices made for modeling transcription and translation reactions in `txtlsim`.

Specifically, we discuss the model order reduction of detailed multi-step elongation models of transcription and translation, while maintaining the ability of the reduced models to track nucleotide and amino acid usage. In the process of performing the model reduction, we show that reaction extent coordinates form a more natural coordinate system compared to the traditionally used species concentration coordinates to bring our model into the two timescale form of singular perturbation theory [39]. Finally, in Chapter 5, we end with some concluding remarks.