Forward engineering of synthetic biologically AND gates

Kavita Iyer Ramalingam 1, Jonathan R. Tomshine 1, Jennifer A. Maynard 2, Yiannis N. Kaznessis ∗

Department of Chemical Engineering & Materials Science, University of Minnesota, 421 Washington Ave SE, Minneapolis, MN 55455, United States

Abstract

The field of synthetic biology has produced genetic circuits capable of emulating functional paradigms seen in digital electronic circuits. Examples are bistable switches, oscillators, and logic gates. The present work combines detailed mechanistic-kinetic models and stochastic simulation techniques as well as the techniques of in vivo molecular biology to study the potential of a synthetic, single promoter AND gate. This device is composed of elements of the tet, lac, and λ-phage promoters and is responsive to the commonly used inducers IPTG and aTc, producing GFP as an output signal. The quantitative behavior of the AND gate phenotype is studied both in vivo and in vitro as a function of promoter topology. The model is constructed from kinetic data obtained from the literature and yields clearly defined ON/OFF logical behavior at realistic inducer concentrations. These behaviors are matched with observed in vivo data obtained through fluorescence-activated cell sorting. The effect of incomplete repression by weaker LacI repressor is also investigated and quantified. The simulation results, coupled with in vivo data, not only identify important design degrees of freedom, but also provide parameters that can be used to guide future synthetic designs using these common regulatory elements.

1. Introduction

A plethora of synthetic gene regulatory networks has been created by arranging naturally occurring regulatory proteins to produce novel biological phenotypic functions [1–4]. To date, the network paradigms that have been most thoroughly studied are the bistable switch [5–8], the oscillator [7,9], and various logic gates [10–17].

In this work we combine multiscale models with synthetic bioengineering experiments to design, build, and characterize a high-fidelity logical AND gate in bacteria. A reductionist modeling approach is pursued. We adopt the molecular biology dogma that there are universal molecular mechanisms underlying the emergence of phenotypes. Consequently, we represent all gene expression molecular level events with reactions, including all the biomolecular interactions involved in transcription, translation, regulation and induction. This way, we relate each model reaction to a real in vivo reaction event and any biological system can be modeled by a network of reactions.

Because biological interactions regularly occur away from the thermodynamic limit, the simulations of reaction networks we conduct are stochastic, i.e., they generate probability distributions of molecular concentration. These distributions are directly comparable to experimentally observed variation. Detailed models then provide the opportunity to gain molecular level insight [18–22].

Experimentally, we constructed an in vivo synthetic-hybrid system consisting of multiple operators within a single promoter. The operator sequences employed are derived from three unrelated natural regulatory elements: the tetracycline (tet), lactose (lac) and λ-phage operators arranged logically within a single transcriptional unit. Specifically, we built six single promoter regulatory motifs by shuffling tet and lac operator sites (T and L, respectively) in and around the P λ (λ-phage): LLT, LTL, TLL, TLT, TTL and LTT (Fig. 1; detailed sequence information is available in supplement). The promoters drive expression of green fluorescent protein (GFP). The regulatory architecture is designed such that each operator’s position efficiently interferes with RNA polymerase (RNP) promoter binding while causing the least perturbation to promoter function [23]. To study the logical gate fidelity among the designed variants, the synthetic promoter sequences were incorporated in the reporter vector pGLOW to direct the transcription of GFP. We engineered the system in an Escherichia Coli strain that constitutively expresses lactose repressor (LacI) and tetracycline repressor (TetR) proteins.

The output fluorescence is then dependent on the input of two small molecule inducers, anhydrotetracycline (aTc) and isopropyl-β-thiogalactoside (IPTG). IPTG and aTc interact with LacI and TetR, respectively, which then free the operator sites for RNP to bind and initiate transcription. The simplicity of these designs is

© 2009 Elsevier B.V. All rights reserved.
recently, Cox et al. [25] presented a powerful, combinatorial method for quickly generating single promoter motifs with a wide variety of logical gate phenotypic behavior. The methods presented herein served to validate the approach. It is in principle applicable to any gene regulatory network, and this reductionist approach lays with the detailed incorporation of all biomolecular interactions describing all of the known interaction events in physiological experimental conditions for comparison of promoter activity with a non-functional promoter sequence. Each of the six synthetic-hybrid promoter variants was characterized under identical experimental conditions for comparison of promoter activity and AND logic gate behavior.

2. Methods and models

2.1. Promoter and plasmid construction, strains

Functional synthetic modules of six designed promoters were constructed using standard molecular biology techniques. They were designed using naturally existing, well-characterized genetic elements from Tet, Lac and λ-phage operons differing in relative positions within the transcriptional unit (Fig. 1; details in Supplementary material). The architecture was based on the modular system of Lutz and Bujard [24,26] however differing in the individual elements used (sequences of lac and tet operators and λ-phage promoter). All promoter/operator sequences, transcriptional start site and the ribosome binding sites are obtained from published sequences [24]. Two overlapping synthetic oligonucleotides (∼110 bp each) corresponding to the sequence of each synthetic-hybrid promoter were assembled using polymerase chain reaction (PCR) amplification with outside primers corresponding to the terminal 20 bp of each larger oligonucleotide. After gel purification, each promoter variant was introduced in the pGLOW-TOPO plasmid (Invitrogen, Carlsbad, CA) upstream of Cycle 3 GFP for use in in vivo promoter activity assays. Integrity of the promoter sequences was confirmed by DNA sequencing and visual verification of constitutive GFP expression in Top 10 cells (lacI−, tetR−). Subsequently, plasmids were transformed into DH5α Pro (lacI+, tetR+) to assess promoter function in the presence of repressors.

2.2. In vivo promoter activity studies

E. coli strain DH5α Pro was used for all in vivo promoter activity assays in the presence of IPTG and aTc inducers. A promoter-less pGLOW variant (containing a non-functional DNA fragment) served as a negative control. Overnight cultures were inoculated 1:100 into fresh LB medium containing 200 μg/ml ampicillin and 50 μg/ml spectinomycin. Inducer concentrations varied from 0 to 2 mM of IPTG and 0 to 200 ng/ml aTc, resulting in a matrix of 36 different inducer pair combinations, each of which was monitored for GFP output over a 24-h period. Specifically, cultures were maintained at 37°C with shaking (250 rpm). At 3, 6 and 9 h time points samples were taken and the cells diluted 1:10 to restrict growth to logarithmic phase and retain constant cellular parameters (e.g., $\sigma^{70}$ levels). Samples were monitored for growth state by OD$_{600}$.

2.3. GFP quantification using flow cytometry

In vivo GFP fluorescence was measured using a Becton Dickinson FACS Calibur flow cytometer equipped with a 488 nm argon laser and a 515–545 nm emission filter (FL1) at low flow rate. Samples were fixed with paraformaldehyde (PFA) to halt GFP production and degradation after harvesting at each time point. One milliliter of cell culture was centrifuged to collect the cells, fixed for 15–30 min in 4% paraformaldehyde (PFA) and re-suspended in phosphate buffered saline (PBS). For each sample, 100,000 gated events were collected and analyzed using Cellquest software (BD Biosciences). GFP fluorescence detected by the FL1 channel was represented as the mean fluorescence versus the normalized population distribution after subtraction of background fluorescence. Background fluorescence was determined using two sets of controls. First, from non-induced cells harboring functional pGLOW plasmids and second, from induced cells maintaining a pGLOW variant with a non-functional promoter sequence. Each of the six synthetic-hybrid promoter variants was characterized under identical experimental conditions for comparison of promoter activity and AND logic gate behavior.

2.4. Kinetic models and parameters

The set of reactions modeling the LTT logical AND gate synthetic circuit is presented as an example in Table 2. The novelty of the approach lays with the detailed incorporation of all biomolecular interactions describing all of the known interaction events in the transcription, translation, repression, and induction processes. This reductionist approach results in large, complex reaction networks that are difficult to simulate. However, although the resulting
models are challenging to solve, the approach is general enough to be applicable to a wide variety of synthetic gene network constructs, such as oscillators, bistable switches, tetracycline-inducible networks among others [19,21,27,28].

More than sixty reactions comprise the network of components used to simulate the AND gates. Many reactions are reversible, and these are represented as pairs of irreversible reactions. All reactions are modeled as initially occurring in a well mixed volume of 10−15 L which represents a cell, and each cell is assumed to contain one copy of the simulated plasmid. That is, the cell contains one “molecule” of each DNA species. Cell growth is handled by allowing the reaction volume to double over a period of time (60 min) followed by an instantaneous halving of volume to represent cytokinesis, thus matching the log phase growth maintained in the in vivo work. Each of the 60–70 reactions demands at least one kinetic parameter, and in some cases (particularly transcriptional or translational elongation modeled as gamma-distributed random processes), two parameters. These kinetic parameters can be obtained from literature directly or reasonably inferred from published thermodynamic parameters. These kinetic parameters can be obtained from literature directly or reasonably inferred from published thermodynamic data [29–35]. Indeed, one of the reasons the tetracycline and lactose operon systems were chosen is that they are exceptionally well studied, and the kinetic and thermodynamic data necessary to populate Table 2 are available in the literature.

Initially, in the first round of modeling before we conducted any experiments we did not account for the leakiness of the promoters. The simulations revealed a high-fidelity AND gate for all six promoters. Indeed, in the absence of leakiness all the promoters with double-tetO (TTL, TLT, LTL) had the exact same behavior. So did the three promoters with double-lacO. In supplementary text we present results of the simulations with the reaction network without leakiness.

After we conducted the experiments it became clear that leakiness of the promoters as a function of the position of the lactose operator(s) relative to the −35 and −10 sequences of the promoter (promoter topology) is a critical model parameter. It also became quickly clear that it is unavailable in literature sources. Without additional kinetic information, the behavior of the models would depend only on the number of lacO or tetO sites present. That is, all promoters containing two lacO sites and one tetO site would be considered equivalent, regardless of topology, as the initial simulation results suggested. Although the leakiness of promoters with lacO has been discussed in the literature [24,26], there was no existing information available to quantify the differences between designs of equivalent composition (number and type of operator sites) but differing topology.

To address this issue, two reactions were added in each model, reactions 2 and 3 of Table 2, to capture the finite probability of RNAP binding the promoter and initiating transcription by displacing a bound LacI repressor protein from the promoter (“leakiness”). As these two reactions essentially represent the same event, they are constrained to have the same kinetic parameter. This parameter was then used to fit the model results to the experimental measurements, producing values for the kinetic constants of these reactions as explained in supplement and listed in Table 1.

Table 1

<table>
<thead>
<tr>
<th>System</th>
<th>$k$ (Lmol⁻¹ s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTT</td>
<td>$6.23 \times 10^5$</td>
</tr>
<tr>
<td>TTL</td>
<td>$4.54 \times 10^5$</td>
</tr>
<tr>
<td>TTL</td>
<td>$1.09 \times 10^5$</td>
</tr>
</tbody>
</table>

It should also be noted that, while the kinetic constants for reactions 2 and 3 (as well as the levels of LacI and TetR expressed by the DH5αPro cells in use) were fit to experimental data after-the-fact and thus not available a priori, the rest of the model was constructed before experimental work began. While the absence of these parameters led to some quantitative error in this initial round of modeling (see supplementary text), the general viability of the proposed system was verified in advance of any lab work—a significant benefit of “model-driven designs”.

Finally, we should note again that although the model is complex, the approach to build it is not. Indeed we have pursued the same modeling approach in all of our previous work and recently we presented and made publicly available a software tool that codifies this approach so that a user can generate complex reaction networks of arbitrary synthetic gene network constructs by only entering interacting components and defining regulatory relations [36,37]. With the tool, we are also making available all the files for the AND gate reaction network.

2.5. Algorithms and simulations

The numerical simulations are carried out using the multiscale simulation algorithm developed in our group [18,20,28,37,38] rather than by simply solving a system of ordinary differential equations. This is necessitated by the inherently stochastic nature of biomolecular interactions [39], the small size of a bacterium, and the dilute nature of some of the reactants. Indeed, promoter and operator sites may be present in quantities as small as one copy per cell, rendering continuous-deterministic models distinctly false (see Supplementary material for further discussion and a presentation of the discrete nature of reactions used to model the AND gates).

Furthermore, the kinetic constants span twenty orders of magnitude, resulting in a model that is stiff to propagate in time. Thus, while there is the need to simulate parts of the network discretely and stochastically, simulating with a simple kinetic Monte Carlo algorithm would be computationally intractable. The algorithm developed in our group, dynamically determines appropriate stochastic-discrete, stochastic-continuous and deterministic-continuous modeling regimes for each molecular species and each reaction and uses the appropriate modeling formalism to numerically propagate the system forward in time. While computationally efficient, it is also accurate—never generating negative species populations.

A disadvantage of stochastic simulations is that an ensemble of simulations must be conducted for meaningful results to be obtained. In the present case, 1000 trajectories are generated for each aTc/IPTG pair of concentrations (a grid of 36 pairs), resulting in 36,000 simulations for each of the six reaction networks. On the other hand, the time-dependent probability distribution of species-numbers can only be adequately sampled with stochastic simulations, providing phenotypic distributions (as in Figs. 2 and 4) that are directly comparable to experimentally observed variability.

3. Results

We have simulated and experimentally constructed six different promoters in E. coli combining tetracycline (T) and lactose (L) operators in the three positions around the promoter −35 and −10 positions (TTL, TLT, LTL, LTT, TTL, TLL). We induced them, both in silico and in vivo, with 36 different concentrations of IPTG and aTc inducers. We tested their behavior and evaluated whether they behave like an AND gate, turning on the expression of GFP downstream if both IPTG and aTc are present in the culture.

In terms of simulation results, the stochasticity of all biomolecular interactions comprising the synthetic bio-logical gate, as detailed in Section 2, results in probability distributions of GFP.
concentrations. Fig. 2 depicts the results of two simulated systems, the LTT (a) and TTL (b) promoters, in the form of a histogram of GFP levels at different time points for the case of the highest inducer levels investigated (6 × 10^6 molecules/cell or 1 mM IPTG and 259 molecules/cell or 200 ng/mL of aTc). The steady-state GFP expression is at its highest at these inducer concentrations, compared to all 36 different concentration pairs used. Superimposed on the histograms are lines representing the mean and a one standard deviation interval. This figure shows that the distribution of states does not change significantly between 6 and 9 h, indicating that a steady-state has been achieved. This is the case for all six systems simulated, under all inducer concentrations.

Using flow cytometry, probability distributions are measured experimentally for all six promoters in all 36 inducer pair concentrations at 3, 6, and 9 h post-induction. A compilation of both experimental and simulation results at 6 h post-induction are provided in Fig. 3 in the form of mean GFP level. The color-coded, two-dimensional surfaces are constructed from the 36 mean GFP levels at the different IPTG and aTc inducer concentrations. For example, Fig. 4 presents the model-generated distributions calculated for the TTL system at a fixed time point (6 h post-induction) at various inducer concentrations. Using the mean value of each distribution, corresponding to different aTc–IPTG concentrations, we construct the three-dimensional plots of GFP as a function of aTc and IPTG.

More specifically, to compare the simulation with the experimental results we determine the average number of GFP molecules per cell at 6 h, averaging over 1000 simulation trajectories, and the average fluorescence strength at 6 h, averaging over 100,000 cytomtery measurements. The third column of Fig. 3 depicts the square of the difference between the experimental and simulation values of mean GFP per cell as evaluated at each point in the aTc–IPTG plane. Note that only one system containing one tetO site and two lacO sites is displayed, since the behavior of all three (LIT, LTL, TTL) was identical (more in Section 4).

As discussed in detail in Section 2, a first round of simulations, which did not take into account promoter leakiness, showed that we could expect an AND gate behavior from the promoters we constructed. The results from these simulations, however, did not differentiate between the different topologies. In other words, the simulations of TLT, TTL, and LTT systems gave identical results. So did the simulations of LIT, LTL, and TTL systems.

In order to ultimately obtain the observed match between simulations and experiment for each promoter design, the operator position dependent leakiness had to be taken into account in the models. Two more reactions were added in the models and the lacO leakiness reactions (reactions 2 and 3 of Table 2) were fit to experimental data. The values of the kinetic constants for the leakiness reactions are shown in Table 1.

Fig. 5 depicts detailed histograms of experimental flow cytometry data for selected systems (TTL, LTT, and LTL) and induction conditions, also 6 h post-induction. The simulation data are also analyzed in terms of the fraction of cells in an ON or OFF state (where "ON" is defined as greater than 50 molecules of GFP—that is, 5% of the maximum GFP expression level observed under any conditions: 1015 molecules of GFP observed for an LTT promoter design with 2 mM IPTG and 100 ng/mL of aTc) in Fig. 6.

Since the model is detailed, with each simulated reaction step corresponding to a real biochemical reaction, a sensitivity analysis can be undertaken. While there has been no attempt made to exhaustively analyze the response of the model to all 60–70 parameters, selected analyses are presented in Fig. 7 (GFP expression versus repressor generation rates in a fully induced state) and 8 (GFP expression versus leaky initiation rate in the presence of aTc, both with and without IPTG).

4. Discussion

Let us first focus on the experimental results (first column of Fig. 3). A high-fidelity bio-logical AND gate will have high GFP expression levels only at high concentrations of both aTc and IPTG. It is clear that none of the designed bio-logical gates is of perfect, digital fidelity. In the double-tetO systems (promoters containing two tetO sites and one lacO site: LIT, LTL, and TTL) there is always a GFP signal for non-zero aTc concentrations, even without IPTG present. This is clearly the result of leakiness of promoters containing the lactose operator.

Nonetheless, despite the imperfect AND gate phenotype, the double-tetO systems exhibit varying degrees of AND gate functionality with no GFP expressed in the absence of inducers and high GFP levels in response to high inducer concentrations. There is a monotonic increase of output signal with inducer concentrations at low aTc and IPTG levels. The output signal strength reaches a plateau past low IPTG and aTc levels. Increasing the concentration beyond 0.1–1 mM IPTG does not result in significantly stronger fluorescence signals. In prior reports, the lacO_1 location within a promoter sequence has been found to significantly impact promoter repressibility, with a centrally placed lacO_1 in the core promoter region.
Fig. 3. The x and y axes form a grid of 36 pairs of inducer concentrations. The concentrations tested were 0, 1, 10, 50, 100, and 200 ng/mL for aTc and 0, 0.001, 0.002, 0.005, 0.010, 0.100, and 1.000 mM for IPTG (all experimental histograms in Supplementary material). The third column of heat maps shows the square of the difference between the experiments and the simulations. These four pairs of profiles depict the modeled (left) and observed (center) behavior of four promoter designs: (a) LTT, (b) TLT, (c) TTL, and (d) LTL. Only one representative promoter with 2-Lac operator elements is depicted, as none of the 2-Lac designs induces to a significant degree. In all cases, with both experiment and simulation, behavior is depicted 6 h after induction. The plotted model values are the means of 1000 independent stochastic kinetic simulations, whereas experimental values are the means of 100,000 FACS observations.
Table 2

<table>
<thead>
<tr>
<th>Number</th>
<th>General transcription and translation reactions</th>
<th>k</th>
<th>Source</th>
<th>TetR repressions, 2nd Tet operator</th>
<th>k</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RNAp:lacP + lacO + tetO + tetR2 → RNAP:lacP</td>
<td>1.00E+07</td>
<td>3</td>
<td>33 tetR2 + tetO2 → tetR2:tetO2</td>
<td>100000000</td>
<td>2*</td>
</tr>
<tr>
<td>2</td>
<td>(See below)</td>
<td></td>
<td>34 tetR2:tetO2 → tetR2 + tetO2</td>
<td>400.000000</td>
<td>2*</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(See below)</td>
<td></td>
<td>35 tetR2:tetO + tetR2 → tetR2:tetO2+aTc</td>
<td>100000000</td>
<td>6*</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>RNAp:lacP + RNAP + lacO + tetO + tetR2</td>
<td>0.01</td>
<td>4</td>
<td>36 tetR2:tetO2 + aTc + tetR2::tetO2</td>
<td>100000000</td>
<td>6*</td>
</tr>
<tr>
<td>5</td>
<td>RNAp::DNA+RNAp + gfp::mRNA + RNAP + RNAP:DNA</td>
<td>30</td>
<td>5</td>
<td>37 tetR2:tetO2 + aTc + tetR2::tetO2</td>
<td>100000000</td>
<td>6*</td>
</tr>
<tr>
<td>6</td>
<td>gfp::mRNA + rib + rib::gfp::mRNA</td>
<td>100000</td>
<td>6</td>
<td>38 tetR2::tetO2 + aTc + tetR2::tetO2</td>
<td>100000000</td>
<td>6*</td>
</tr>
<tr>
<td>7</td>
<td>rib::gfp::mRNA + rib::gfp::mRNA</td>
<td>33</td>
<td>5</td>
<td>39 tetR2::tetO2 + aTc + tetR2::tetO2</td>
<td>100000000</td>
<td>6*</td>
</tr>
<tr>
<td>8</td>
<td>rib::gfp::mRNA + rib + gfp</td>
<td>33</td>
<td>5*</td>
<td>40 tetR2::tetO2 + aTc + tetR2::tetO2</td>
<td>100000000</td>
<td>6*</td>
</tr>
<tr>
<td>9</td>
<td>Lad repression at lac operator</td>
<td></td>
<td>41</td>
<td>42 tetR2::tetO2 + aTc + tetR2::tetO2</td>
<td>100000000</td>
<td>6*</td>
</tr>
<tr>
<td>10</td>
<td>Non-specific DNA interactions</td>
<td></td>
<td>43</td>
<td>44 lacP + nsDNA → lacP + nsDNA</td>
<td>1000</td>
<td>7*</td>
</tr>
<tr>
<td>11</td>
<td>Lac[4]:lacO1 → lac[4]:lacO1</td>
<td>2E + 09</td>
<td>1</td>
<td>45 lacP + nsDNA → lacP + nsDNA</td>
<td>1000</td>
<td>7*</td>
</tr>
<tr>
<td>12</td>
<td>Lac[4]:lacP1 + lac[4]:lacP1</td>
<td>4.00E-04</td>
<td>1</td>
<td>46 lacP + nsDNA → lacP + nsDNA</td>
<td>0.0041667</td>
<td>7*</td>
</tr>
<tr>
<td>13</td>
<td>Lac[4]:IPTG + lac[4]:IPTG</td>
<td>4.60E-06</td>
<td>1</td>
<td>47 lacP + nsDNA → lacP + nsDNA</td>
<td>1000</td>
<td>7*</td>
</tr>
<tr>
<td>14</td>
<td>Lac[4]:IPTG + lac[4]:IPTG</td>
<td>0.2</td>
<td>1</td>
<td>48 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>15</td>
<td>Lac[4]:lacP1 + IPTG + lacP1 + lac[4]:lacP1</td>
<td>1.00E-06</td>
<td>1</td>
<td>49 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>16</td>
<td>Lac[4]:lacP1 + IPTG + lacP1 + IPTG</td>
<td>0.8</td>
<td>1</td>
<td>50 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>17</td>
<td>Lac[4]:IPTG + lac[4]:IPTG + lac[4]:lacP1</td>
<td>2E + 09</td>
<td>1</td>
<td>51 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>18</td>
<td>Lac[4]:IPTG + lac[4]:IPTG + lacP1 + lacP1</td>
<td>0.4</td>
<td>1</td>
<td>52 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td></td>
<td>TetR repression, 1st Tet operator</td>
<td></td>
<td>53</td>
<td>54 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>19</td>
<td>tetR2::tetO2 + aTc → tetR2::aTc</td>
<td>100000000</td>
<td>6*</td>
<td>55 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>20</td>
<td>tetR2::tetO2 + aTc + tetR2::aTc</td>
<td>0.001</td>
<td>6*</td>
<td>56 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>21</td>
<td>tetR2::tetO2 + aTc + tetR2::aTc</td>
<td>100000000</td>
<td>6*</td>
<td>57 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>22</td>
<td>tetR2::tetO2 + aTc + tetR2::aTc</td>
<td>0.001</td>
<td>6*</td>
<td>58 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>23</td>
<td>tetR2::tetO2 + aTc + tetR2::aTc</td>
<td>100000000</td>
<td>6*</td>
<td>59 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>24</td>
<td>tetR2::tetO2 + aTc + tetR2::aTc</td>
<td>0.001</td>
<td>6*</td>
<td>60 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>25</td>
<td>tetR2::tetO2 + tetR2 + tetR2::tetO2 + aTc</td>
<td>100000000</td>
<td>6*</td>
<td>61 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>26</td>
<td>tetR2::tetO2 + tetR2 + tetR2::tetO2 + aTc</td>
<td>100000000</td>
<td>6*</td>
<td>62 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>27</td>
<td>tetR2::tetO2 + tetR2 + tetR2::tetO2 + aTc</td>
<td>100000000</td>
<td>6*</td>
<td>63 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>28</td>
<td>tetR2::tetO2 + tetR2 + tetR2::tetO2 + aTc</td>
<td>100000000</td>
<td>6*</td>
<td>64 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>29</td>
<td>tetR2::tetO2 + tetR2 + tetR2::tetO2 + aTc</td>
<td>100000000</td>
<td>6*</td>
<td>65 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>30</td>
<td>tetR2::tetO2 + tetR2 + tetR2::tetO2 + aTc</td>
<td>100000000</td>
<td>6*</td>
<td>66 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>31</td>
<td>tetR2::tetO2 + tetR2 + tetR2::tetO2 + aTc</td>
<td>100000000</td>
<td>6*</td>
<td>67 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
<tr>
<td>32</td>
<td>tetR2::tetO2 + tetR2 + tetR2::tetO2 + aTc</td>
<td>100000000</td>
<td>6*</td>
<td>68 lacP + nsDNA → lacP + nsDNA</td>
<td>2.89E-04</td>
<td>7*</td>
</tr>
</tbody>
</table>

* a The indicated reference provides affinity data, so the forward rate is assumed and the reverse rate is calculated to match the data.
* b The affinities at the two distinct levels of aTc induction (a single aTc molecule or two aTc molecules bound to TetR) are estimated, based on a single literature value.
* c This value is based on a 6-h half-life. Effectively, this species does not degrade chemically—rather, the rate of dilution through cell growth is much greater than the degradation rate.
* d This reaction time is gamma-distributed, based on the indicated 1st order constant occurring at each step over a span of 600 nucleotides (200 amino acids).
* e This value has been adjusted to give ~20 polypeptides per mRNA transcript.
* f These parameters are fit to experimental data obtained in this study—they are not obtained from previously published sources.
* g This “degradation” rate matches the rate of dilution due to cell growth. It is intended to represent dilution of DNA-bound species during DNA replication and cell growth.
under identical conditions. To better understand this effect, the TTL and LTT promoters were investigated, with the latter expressing tighter repression than either the TTL or LTT systems, an interesting deviation from the observations of Lanzer and Bujard.[24]

The promoters remained tightly repressed under the following conditions: 0–1 mM IPTG and no aTc and 0–10 ng/mL aTc and no IPTG, defining the off limit of the switch. Induction was observed only above 10 ng/mL aTc concentrations, indicating the system's transition from an AND gate to single input switch. Interestingly, the fidelity of the AND gate appears to improve by moving lacO downstream in the promoter. Although the maximum amount of GFP expressed under full induction (the plateau region) is lower as lacO is moved downstream in the promoter, the leakiness effect on the fidelity of the AND gate appears to be minimal under conditions of strong induction (6 h after the addition of 1 mM IPTG), a condition corresponding to free tetO or increased probability for recruitment of RNAP to initiate transcription. Consequently TTL displayed an enhanced aTc induction threshold concentration compared to LTT or TTL.

The fuzziness of the AND gate can also be viewed in terms of phenotypic diversity in response to inducers. Phenotypic variation as a result of point mutation or altered connectivity in synthetic circuits has been observed[10–12,16,40]. This study indicates that variability in logic behavior can be attained through permutations of operator position within a transcriptional unit. The double-lacO systems, on the other hand, express almost no appreciable GFP levels even at the highest inducer concentrations. This can be explained by relatively high concentrations of LacI constitutively expressed in the E. coli strain DH5αPro. Given the high levels of LacI present, even under conditions of strong induction, most lacO sites remain bound by repressor. For instance, the simulations of the LTT system—the most active promoter—show that under conditions of strong induction (6 h after the addition of 1 mM IPTG), the single lacO site is only free in 6.8% of the cells in the 1000-cell ensemble. In the rest of the cells it is occupied (0.1% of the cells offering the best repression [24]. This trend agrees with the LTT and TTL systems under investigation, with the latter expressing tighter control. However with TTL, lacO positioned downstream of −10 hexamer provided greater repression than either the TTL or LTT systems, an interesting deviation from the observations of Lanzer and Bujard [24].

The promoters remained tightly repressed under the following conditions: 0–1 mM IPTG and no aTc and 0–10 ng/mL aTc and no IPTG, defining the off limit of the switch. Induction was observed only above 10 ng/mL aTc concentrations, indicating the system's transition from an AND gate to single input switch. Interestingly, the fidelity of the AND gate appears to improve by moving lacO downstream in the promoter. Although the maximum amount of GFP expressed under full induction (the plateau region) is lower as lacO is moved downstream in the promoter, the leakiness effect on the fidelity of the AND gate appears to be minimal—-a trend that is also observed in the model (Fig. 8). To better understand this effect, the TTL and LTT promoters were investigated under identical in vivo experimental conditions, producing the qualitatively different histograms in panels (a) and (b) of Fig. 5. Under conditions of increasing aTc concentration and IPTG absent, downstream positioning of lacO in TTL conferred an order of magnitude tighter control, yielding the best AND gate switch among the systems investigated. Thus lacO–repressor complex seems to be performing well at maintaining the regulatable gene circuit in a repressed state under high aTc concentrations (up to 100 ng/mL), a condition corresponding to free tetO or increased probability for recruitment of RNAP to initiate transcription. Consequently TTL displayed an enhanced aTc induction threshold concentration compared to LTT or TTL.

The fuzziness of the AND gate can also be viewed in terms of phenotypic diversity in response to inducers. Phenotypic variation as a result of point mutation or altered connectivity in synthetic circuits has been observed [10–12,16,40]. This study indicates that variability in logic behavior can be attained through permutations of operator position within a transcriptional unit. The double-lacO systems, on the other hand, express almost no appreciable GFP levels even at the highest inducer concentrations. This can be explained by relatively high concentrations of LacI constitutively expressed in the E. coli strain DH5αPro. Given the high levels of LacI present, even under conditions of strong induction, most lacO sites remain bound by repressor. For instance, the simulations of the LTT system—the most active promoter—show that under conditions of strong induction (6 h after the addition of 1 mM IPTG), the single lacO site is only free in 6.8% of the cells in the 1000-cell ensemble. In the rest of the cells it is occupied (0.1% of the cells

Fig. 5. Experimental histograms: Conditions are ± 100 ng/mL aTc and ± 1 mM IPTG, e.g., ± refers to 100 ng/mL aTc and 1 mM IPTG. Panel (a) depicts a TTL promoter design which behaves as a true AND gate. With no inducer, or with aTc or IPTG alone (even at fairly high levels), a single un-induced population is observed with minimal leakiness. Only when aTc and IPTG are both present are the cells induced. Panel (b) depicts an LTT promoter design, where the single lac operator site has been moved upstream of the −35 sequence. In this case, LacI is no longer able to entirely suppress transcription and an intermediate level of induction (leakiness) occurs at conditions of high aTc but no IPTG. A double lacO model system behaves as a true AND gate. With no inducer, or with aTc or IPTG alone (even at fairly high levels), a single un-induced population is observed with minimal leakiness. Only when aTc and IPTG are both present are the cells induced. Panel (c) depicts an LTT promoter design, where the single lac operator site has been moved upstream of the −35 sequence and downstream of the −10. In this case, even high levels of IPTG are unable to induce the system, regardless of aTc concentration. Models suggest that this design may become feasible if the cells' intrinsic rate of constitutive LacI synthesis (and hence steady-state LacI concentration) were lowered. All data is obtained 6 h after induction.

Fig. 6. The surfaces represent the fraction of simulated cells that are in an OFF state at 6 h post-induction. Panel (a) is an LTT system, while panel (b) is the TTL system. Here, a cell is described as “ON” if it is expressing greater than 50 molecules of GFP per cell. This threshold is arbitrarily set at 5% of the absolute maximum GFP expression observed under any conditions. The “OFF” state is defined as 1 − fraction of Gates OFF. While these panels bear a resemblance to their mean value counterparts in Fig. 3, it should be noted that even with the more active LTT promoter, not all cells are predicted to induce. Indeed, the lower activity of the TTL design is manifested in a very significant fraction of cells that are OFF, expressing very low levels of GFP, even in the fully induced plateau region.
by LacI alone, 88.3% by LacI:IPTG complex, 4.7% by RNAS closed complex, and 0.1% by RNAS open complex). Nevertheless, these 6.8% of cells are sufficient to produce the observed GFP output.

The addition of a second lacO site (in place of a tetO) might be expected to yield a further 93.2% reduction in the number of promoters with both lacO sites free – a total of 0.5% of cells (0.0682 – 0.0046) – and this is precisely what is observed. The simulations for the LTL system show that, under the same conditions mentioned above, both lacO sites are simultaneously free in only 5

cells of the 1000-cell ensemble, or 0.5% of cells (one lacO site was free in 10.9% of cells, and neither site was free in 88.6% of cells).

Furthermore, the leakiness reactions employed in the model (reactions 2 and 3 of Table 2) allow RNAP to dislodge only one lacO-bound LacI molecule. Therefore, in the case of the double-tetO systems, all cells are subject to some form of transcription initiation at all times—either normal initiation at a LacI-free promoter or leaky initiation at a promoter containing its maximum of one bound LacI. The double-lacO systems, on the other hand, possess states at which initiation is strictly forbidden—namely the state in which neither operator is free (which accounts for 88.6% of cells in the LTL system, as discussed above). While these double-lacO designs may still be induced by adding sufficient IPTG, the levels required are beyond the range that can be employed in the in vivo work.

Fig. 3 demonstrates that the model captures the experimentally observed synthetic phenotypes. The fit of the simulated dynamic behavior of a complex network with more than 60 reactions modeled stochastically to experimental flow cytometry measurements is remarkable. The agreement between experimental and simulation results is quantified in the third column of Fig. 3, where the square of the difference between mean observed and simulated GFP expression is presented. The fit between experimental and simulation results is worst at conditions of low but non-zero aTc and non-zero IPTG. In this region, the simulations induce more slowly with increasing aTc than do the experimental results, most likely due to an inaccuracy in the parameters governing aTc induction. This is not entirely unexpected, as the kinetic parameters governing aTc induction are not as well documented as those governing IPTG induction.

It should also be emphasized that aside from the variation in the kinetic parameters of reactions 2 and 3 of Table 2, the models depicted in Fig. 3 are identical in every other respect. This indicates that, given a model of more than 60 individual reactions, the model parameter that should depend on promoter configuration is indeed sufficient to account for the observed variability among the promoter designs. Since these rates were unavailable in the literature, they were fit by minimizing the sum-of-squares of relative expression levels at 3 conditions: no induction, high aTc/no IPTG, and high aTc/high IPTG. Shifting lacO location within the hybrid tet–lac promoter changes the values of the kinetic constant for reactions 2 and 3, decreasing these k values with more favorable lacO placement and consequently improved repression under no IPTG and high aTc conditions.

The data generated in this study, both simulated and experimentally observed, ultimately consist of quantifications of the behavior of individual cells. Although the most straightforward analysis involves the computation of mean values of cell fluorescence or GFP expression level, there is also considerable information contained in the distributions of cell behavior—information that cannot be obtained through an ordinary differential equation simulation.

For instance, of the 6 systems simulated, the highest GFP expression level of any cell at 6 h post-induction was 1015 molecules of GFP (for an LTT promoter design with 2 mM IPTG and 100 ng/mL of aTc). At the other extreme, zero induction at 6 h, greater than 95% of cells had exactly zero GFP expression for all 6 systems tested. If one arbitrarily defines the minimum “ON” state to be ~5% of the absolute maximum GFP expression level – 50 molecules per cell, with “OFF” = 1 – “ON” – then one can analyze the percentage of cells falling on either side of that threshold. This can help determine the noise–induced states of the AND gates.

Even with the most active promoter design, LTT, the model predicts that only 85% of cells are “ON” (>50 molecules of GFP) at conditions of 1 mM IPTG and 100 ng/mL of aTc at 6 h post-induction, with 95% of cells expressing between 20 and 296 molecules of GFP. On the other hand, the TTL design – the least active of the functional
AND gates – gave only 43% of cells in an “OFF” state (at identical conditions) with 95% of cells expressing between 6 and 217 molecules of GFP. Full plots for the simulated fraction of cells in an OFF state at all induction conditions are provided in Fig. 6. One can further understand the nature of the distributions with histograms, as is done in Figs. 2 and 4 (simulation) or Fig. 5 (experiment).

As with the simulation data, the experimental data can also be viewed as individual cellular measurements. When one applies the same criterion for a cell being “ON” – i.e., reaching 5% of the absolute maximum observed fluorescence level at 6 h post-induction – the results are somewhat different. The real cells, unlike the simulated ones, tend to induce more uniformly, with 97% of LTI systems “ON” at conditions of 1 mM IPTG and 100 ng/mL of aTc. As with the simulation, TTL is less active and fewer cells meet the “ON” criterion – 91% – though this is still a much higher value than the simulations predict. Therefore, while the model may be able to capture many of the relevant trends between promoter designs and provide a quantitative prediction of mean values, there remains work to be done in predicting the higher moments of the distributions.

The mechanistically detailed nature of the model employed in this work also provides the opportunity to investigate the effect of changes in kinetic parameters that correspond to those of real biochemical reactions. The generation rates of LacI and TetR will be strain dependent so it may useful to understand how these rates would affect the behavior of the system. We vary the LacI and TetR concentrations and determine GFP expression levels. Fig. 7 shows one of the resulting sets of simulations: a fully induced double-tetO design—in this case with k2 and k3 set to zero. As expected, GFP expression is low at higher repressor generation rates (upper-right) but high at lower rates (lower-left). The gradient is also much steeper with increasing TetR than with increasing LacI, reflecting the tighter binding of TetR and the presence of two tetO operator sites rather than one lacO site. The red circle at the top of the figure indicates the rates employed for all other simulations in this study. These parameters were chosen with the aid of similar plots evaluated for double-lacO systems and zero induction conditions (data not shown).

The level of GFP expression is depicted in Fig. 8 as a function of the leaky transcription initiation rate (at high aTc values), with lower k2 and k3 (i.e., k0 + k1) values producing gates that express less GFP overall, but where the amount of leaky expression is proportionally low compared to the level of fully induced expression. This trend correlates with in vivo observation as discussed previously. As the rate of leaky initiation increases, the amount of GFP expressed ceases to have a strong dependence on the presence or absence of IPTG, as one might expect. At suitably low and commonly employed aTc concentrations (example 10 ng/mL), however, all three systems display AND gate functionality. Interestingly, similar leakiness reactions were not necessary for the tetracycline operator sites. That is, the probability of RNAP binding and transcribing DNA that is already bound by TetR is negligible, though this probability is further reduced by the presence of two TetR sites in each of the three functional gates.

While all three double-tetO synthetic circuits can be considered AND gates over some range of inducer concentrations, we observe that knowledge of operator placement in the promoter is critical for optimal AND gate behavior. Since the three promoters represent all possible placements of a single lacO within a three-operator hybrid promoter, and rates of kinetic leakiness have been obtained in each case, these results foster meaningful predictions for subsequent work.

In summary, we show how the integration of computational and experimental molecular biology can rationalize the synthesis of novel biological functionalities. We have constructed a biological AND gate from a simple, synthetic-hybrid promoter module. From a biological perspective, experimental results identify an effective lacO position where improved repression results due to weak interaction between RNAP and operator-bound repressor. We propose that this interaction plays a vital role in transcriptional regulation as supported by the matching between models and experiments over a wide range of induction conditions, having taken into account the inherently stochastic behavior of the systems. Reactions 2 and 3 and their kinetic constants quantify this biological behavior for the first time and should prove useful in constructing further stochastic kinetic models that involve repression by LacI.

The models created to study this system predict the system’s feasibility, quantify the effect of operator placement and provide an understanding of the improved AND gate function demonstrated by TTL. Certainly, there is not one single, direct method of conducting simulations that precisely outlines future experimental designs. Instead, there is an intimate interplay between simulations and experiments, where information flows back and forth between both computational attempts and experimental ones. In this work, the first models constructed did not include promoter topology dependent leakiness. Nevertheless, they were useful in demonstrating that the designs were promising and would likely work as planned. After the first set of experiments, it became clear that reactions 2 and 3 were required, and that different values for the kinetic parameters would lead to correlation with the designed promoters. What was gained was insight into the mechanism of leakiness and a model that quantitatively captures it. It is not clear how this biological insight would be captured with an a posteriori modeling effort, where minimal models are built to fit observed experimental data.

In conclusion, we believe that this approach of designing and tailoring logical regulation of gene expression can be potentially extended beyond transcriptional regulation to include other modular genetic elements, search for more complex network behaviors and assist in the forward engineering of other biological circuits.

**Acknowledgements**

This work was supported by a grant from the National Science Foundation (BES-0425882, CBET-0425882 and CBET-0644792) and the University of Minnesota Biotechnology Institute. Computational support from the Minnesota Supercomputing Institute (MSI) is gratefully acknowledged. This work was also supported by the National Computational Science Alliance under TG-MCA04N033. In addition, we would like to thank Jin Cui Tomshine for her preparation of the schematic diagrams in Fig. 1.

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bej.2009.06.014.

**References**

[29] A. Revzin, P.H. von Hippel, Direct measurement of association constants for the binding of Escherichia coli lac repressor to non-operator DNA, Biochemistry 16 (1977) 4769–4776.