

# Abstracted Stochastic Analysis of Type 1 Pili Expression in *E. coli*

Hiroyuki Kuwahara\*, Chris Myers†, and Michael Samoilov‡

\* School of Computing, University of Utah, Salt Lake City, UT 84112, U.S.A.

kuwahara@cs.utah.edu

† Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112, U.S.A.

myers@ece.utah.edu

‡ Howard Hughes Medical Institute and Department of Bioengineering, University of California, Berkeley  
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.

MSSamoilov@lbl.gov

**Abstract**—With the aid of model abstractions, biochemical networks can be analyzed at different levels of resolution: from low-level quantitative models to high-level qualitative ones. Furthermore, an ability to change the level of abstraction can be very useful when dealing with many biological systems, including gene regulatory networks. These systems typically have too many components and states to be practically studied using all-inclusive low-level models, yet they often manifest enough dynamical and functional complexity, making an entirely high-level qualitative representation similarly inadequate — thus necessitating a search for some intermediate level of abstraction. Finally, while most abstractions used in modeling of biochemical networks have traditionally been performed manually, doing so accurately in a large system is a tedious and time-consuming process that is highly susceptible to errors during model transformation. To address these issues, we have developed a methodology and implemented an automated modeling and analysis tool with variable abstraction level capabilities. In this paper, we use it for the analysis of switching in Type 1 pili expression dynamics and, in particular, for the problem of estimating the effect of H-NS and Lrp regulatory protein levels on phase variation rates in *E. coli*. Such behavior is notoriously difficult to study due to the size of the associated gene regulatory network and the characteristically stochastic dynamics involved, which result in very high analytical and computational demands. Here, we show how, by using our system, we are able to automatically abstract the switch network and accurately predict *E. coli* afimbriation rates, while, at the same time, accelerating the required computations by up to two orders of magnitude.

## I. INTRODUCTION

Type 1 pili are the foremost virulence factor in Uropathic *E. coli*, which is believed to be responsible for 70-90 percent of urinary tract infections [1]. The pili are 1-2 $\mu$ m long and 7nm-thick helical rods with a 3nm-wide tip, which contains two adapter proteins and adhesins capable of mediating *E. coli* attachment to the mannose-containing receptors found on the surface of many host tissue cells [1], [2]. Type 1 pili are thus thought to aid the infection and colonization process by enhancing the ability of *E. coli* to stick to host cells and by thus enabling them to colonize the bladder. However, while pili provide a means for infection, there are some disadvantages

for *E. coli* in being piliated. For example, a highly piliated population leads to preferential activation of the host immune system, which can rapidly clear the infection. In addition, the expression of pili also makes it harder for *E. coli* to divide.

The expression of Type 1 pili in *E. coli* is phase variable [3], with cells randomly switching between fimbriate (ON) and afimbriate (OFF) phases. Phase variation of Type 1 fimbriae is mediated by the inversion of a 314bp chromosomal region located upstream of *fimA*, the gene encoding the structural pili subunit [4]. This DNA inversion requires either FimB or FimE site-specific recombinases that differ in both specificity and activity. Whereas FimB promotes recombination with little orientational bias, FimE promotes recombination largely in the ON-to-OFF direction [2], [5]. Experimental results also indicate that the total ON-to-OFF switching rate is much faster than fimB-promoted switching rate alone [1].

Figure 1 shows the genetic circuit controlling Type 1 pili expression. Genes *fimA* to *fimH* are transcribed only when the *fim* switch is in the ON position. The *fim* switch can be inverted by recombinases FimB and FimE when they are bound to the four adjacent half-sites with probability rates regulated by the levels of other proteins in the network. Thus, the dynamics of *E. coli* phase variation is manifestly discrete and stochastic, and cannot be accurately modeled by any of the classical continuous and deterministic methods.

Global regulator proteins Lrp, H-NS, and IHF also play important roles in phase variation by, among other things, acting as sensors of the environmental conditions. For example, Lrp binds to the 3 Lrp-sites and changes *fim* switching rates. As the concentration of Lrp increases with temperature, it correspondingly communicates this parameter to the piliation decision switch. Analogously, the concentration or the activity of H-NS is temperature and medium dependent. It has been shown that H-NS binds to the regions that contain the *fimB* and *fimE* promoters [7] and represses the transcriptions of *fimB* and *fimE* [8]. As the temperature increases, the effect of H-NS on the repression of *fimB* transcription increases, which leads to a decrease in the concentration of FimB and a corresponding decrease in the fimbriate phase of the population. This is not unexpected as the ambient temperature elevation may

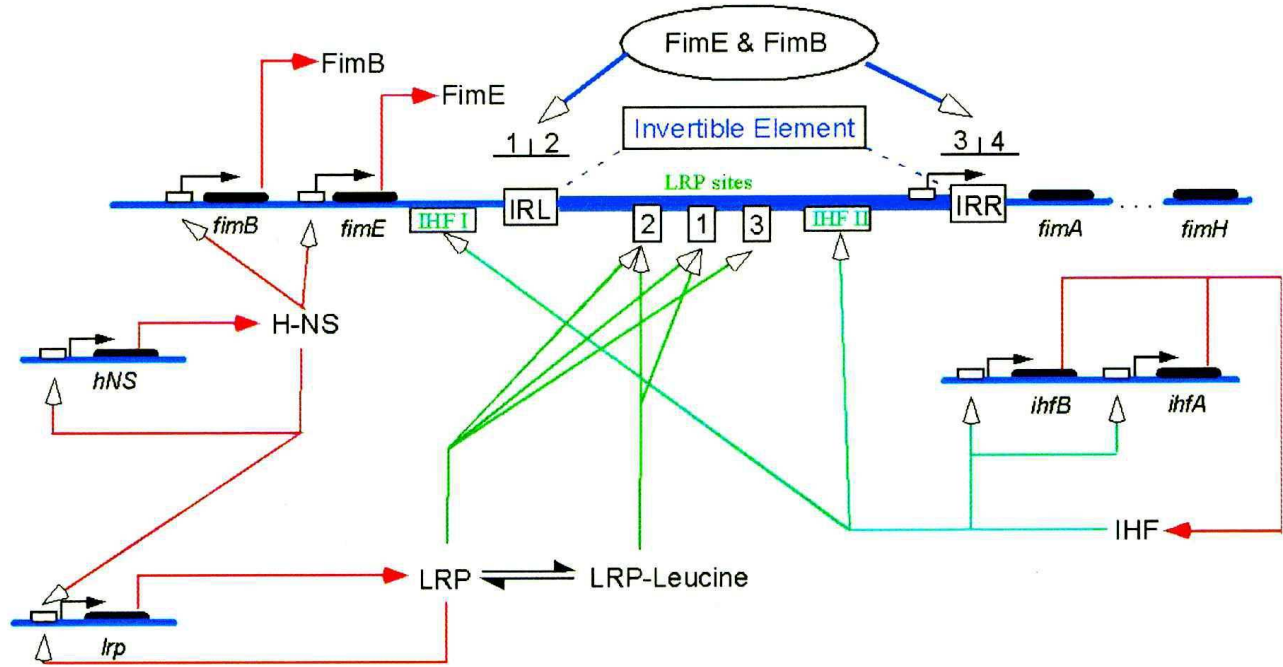


Fig. 1. Type 1 pili genetic regulatory network (courtesy of [1]). Structural pili subunits are encoded by *fimA* and are transcribed only when the *fim* switch is in the ON position. Recombinases FimB and FimE bind the four adjacent half-sites and invert the switch with different rates. FimE is strongly biased in the ON-to-OFF direction, while FimB is close to fair. Proteins H-NS and Lrp regulate the recombinases in a temperature-dependent manner, with H-NS acting as a repressor. It has been further proposed that IHF is needed for any observable phase variation as it plays a structural role during switching via the ability to introduce sharp bends into the DNA [6].

correspond to an increase in immune system activity, which is more sensitive to piliated cells. The main question of interest here is thus: how does *E. coli* sense environmental changes and control its piliation levels?

This paper examines how *E. coli* phase switch is controlled by the global regulators. Such behavior is notoriously difficult to study using an all-inclusive, low-level model due to the size of the associated gene regulatory network and the characteristically stochastic dynamics involved, which result in very high analytical and computational demands. On the other hand, this type of system often manifests too much dynamical and functional complexity so that rendering it in an entirely high-level qualitative representation becomes similarly inadequate.

To address these issues, we have developed a generalized model abstraction methodology and implemented a modeling and analysis tool at variable abstraction level capabilities. Our approach enables one to generate models with multiple levels of abstraction—including intermediate levels of abstraction that may well be suited for an efficient quantitative analysis of a large system—easily and accurately, and to accordingly analyze systems such as Type 1 pili expression in *E. coli* more promptly and efficiently by making the process systematic and automatic [9]. This paper first considers a detailed mathematical model of Type 1 pili phase variation, and then demonstrates how our model abstraction approach can help analyze the effect of Lrp and H-NS levels in the ON-to-OFF switching probability of Type 1 pili expression.

## II. WOLF AND ARKIN'S COMPUTATIONAL MODEL

Numerous methods have been proposed for modeling genetic regulatory networks [10], [11]. While many traditional approaches have relied on some differential equation representation inferred from the set of underlying biochemical reactions, there has been a growing appreciation of their limitations [12], [13], [14], [15]. In particular, differential equation analysis of genetic networks generally assumes that the number of molecules in a cell is high and their concentrations can be viewed as continuous quantities with the underlying reactions occurring deterministically. However, in natural genetic networks, these assumptions frequently do not hold. For example, DNA molecules are typically present in single digit quantities while some promoters can lead to substantial fluctuations in transcription/translation rates and essentially non-deterministic expression characteristics [16], [17]. In these situations, accurate genetic regulatory network modeling requires the use of a discrete and stochastic process description usually encapsulated in the *master equation formalism* [18]. This approach describes well-stirred biochemical systems at the individual reaction level by exactly tracking the quantities of each molecular species and by treating each reaction as a separate random event. It further allows the exact discrete-stochastic simulation of system behavior via Gillespie's *Stochastic Simulation Algorithm* (SSA) [19].

Wolf and Arkin have previously demonstrated the importance of stochasticity in the *fim* system. They used a four-state

TABLE I  
STATE TABLE OF *fim* SWITCH DNA BINDING (COURTESY OF [1]).

State	P <sub>IHF</sub>	P <sub>FimE/B</sub>	Lrp-A	Lrp-3	$\Delta G$	$\alpha$	$n$	$j$	$k$	$m$	$l$
1/9	-	-	-	-	0	0	0	0	0	0	0
2/10	IHF	-	-	-	$\Delta G_2$	0	1	0	0	0	0
3/11	IHF	FimE	-	-	$\Delta G_3 / \Delta G_7$	$\alpha_1 / \alpha_4$	1	0	1	0	0
4/12	IHF	FimB	-	-	$\Delta G_4 / \Delta G_8$	$\alpha_2 / \alpha_3$	1	1	0	0	0
5/13	IHF	FimE	Lrp*	-	$\Delta G_{31a} / \Delta G_{71a}$	$\alpha_{11a} / \alpha_{41a}$	1	0	1	2	0
6/14	IHF	FimE	Lrp*	Lrp	$\Delta G_{31b} / \Delta G_{71b}$	$\alpha_{11b} / \alpha_{41b}$	1	0	1	2	1
7/15	IHF	FimB	Lrp*	-	$\Delta G_{41a} / \Delta G_{81a}$	$\alpha_{21a} / \alpha_{31a}$	1	1	0	2	0
8/16	IHF	FimB	Lrp*	Lrp	$\Delta G_{41b} / \Delta G_{81b}$	$\alpha_{21b} / \alpha_{31b}$	1	1	0	2	1
17/27	-	FimE	-	-	$\Delta G_{33} / \Delta G_{77}$	0	0	0	1	0	0
18/28	-	FimB	-	-	$\Delta G_{44} / \Delta G_{88}$	0	0	1	0	0	0
19/29	-	FimE	Lrp*	-	$\Delta G_{331a} / \Delta G_{771a}$	0	0	0	1	2	0
20/30	-	FimE	Lrp*	Lrp	$\Delta G_{331b} / \Delta G_{771b}$	0	0	0	1	2	1
21/31	-	FimB	Lrp*	-	$\Delta G_{441a} / \Delta G_{881a}$	0	0	1	0	2	0
22/32	-	FimB	Lrp*	Lrp	$\Delta G_{441b} / \Delta G_{881b}$	0	0	1	0	2	1
23/33	-	-	Lrp*	-	$\Delta G_{3a} / \Delta G_{7a}$	0	0	0	0	2	0
24/34	-	-	Lrp*	Lrp	$\Delta G_{33a} / \Delta G_{77a}$	0	0	0	0	2	1
25/35	IHF	-	Lrp*	-	$\Delta G_{33b} / \Delta G_{77b}$	0	1	0	0	2	0
26/36	IHF	-	Lrp*	Lrp	$\Delta G_{31a} / \Delta G_{71a}$	0	1	0	0	2	1

Markov model to analyze the expression of Type 1 pili [1]. Their phase variation network model, shown in Figure 1, included five species: FimB, FimE, IHF, Lrp\*, and Lrp. The Wolf and Arkin model assumes that the configuration of the *fim* switch can be described by 36 states based on how the molecules of five species involved are bound to the switch DNA as shown in Table I.

Their model is then described by the master equation for the *fim* switch based on these 36 possible states as:

$$\frac{dP(on, t)}{dt} = f(1 - P(on, t)) - gP(on, t)$$

with

$$f = \frac{\sum_{s \in \text{OFF}} \alpha_s \exp\left(\frac{-\Delta G_s}{RT}\right) [\text{IHF}]^{n_s} [\text{FimE}]^{j_s} [\text{FimB}]^{k_s} [\text{Lrp}^*]^{m_s} [\text{Lrp}]^{l_s}}{1 + \sum_{s \in \text{OFF}} \exp\left(\frac{-\Delta G_s}{RT}\right) [\text{IHF}]^{n_s} [\text{FimE}]^{j_s} [\text{FimB}]^{k_s} [\text{Lrp}^*]^{m_s} [\text{Lrp}]^{l_s}}$$

and

$$g = \frac{\sum_{s \in \text{ON}} \alpha_s \exp\left(\frac{-\Delta G_s}{RT}\right) [\text{IHF}]^{n_s} [\text{FimE}]^{j_s} [\text{FimB}]^{k_s} [\text{Lrp}^*]^{m_s} [\text{Lrp}]^{l_s}}{1 + \sum_{s \in \text{ON}} \exp\left(\frac{-\Delta G_s}{RT}\right) [\text{IHF}]^{n_s} [\text{FimE}]^{j_s} [\text{FimB}]^{k_s} [\text{Lrp}^*]^{m_s} [\text{Lrp}]^{l_s}}$$

where  $P(on, t)$  is the probability that the switch is in the ON position at time  $t$ . This model is used to infer qualitative information about the system. For example, they are able to answer questions such as why the network has two recombinases, and how the *fim* network architecture determines that the environment is at 37 °C and responds by increasing switching rates and the piliation level of the population.

### III. OUR FULL PHASE VARIATION MODEL

Our model uses the same master equation as Wolf and Arkin's model to specify the configuration of the DNA switch region with most parameter values derived from [1]. Our model, however, uses a lower-level quantitative reaction-based abstraction which models reaction-level molecular processes that largely satisfy the Markovian requirement of the SSA.

The model is constructed by first reverse-engineering the underlying reactions from the equilibrium statistical thermodynamics model that describes the probability of each switch DNA configuration shown in Table I. This is accomplished using the hypothesis that the binding and unbinding reactions are much more rapid as compared to the associated switching or gene expression rates [20]. Thus, from the standard free energy relationship,  $\Delta G = -RT \ln(k_f/k_r)$ , the corresponding binding and unbinding reactions are estimated using a rapid unbinding rate constant of  $1.0\text{s}^{-1}$ .

For example, the reaction-based model for switch state 6 is reverse-engineered as shown in Figure 2. In this graphical representation, a reaction that is connected to a species with a double arrow is a shorthand to show a reversible reaction, and species connected to a reaction with letters,  $r$ ,  $p$ , and  $m$  are a reactant, product, and modifier<sup>1</sup> for that reaction, respectively. In this reaction scheme, the new species P<sub>finm</sub> represents the open binding sites, while S<sub>6</sub> corresponds to the bound configuration where one molecule of IHF binds to P<sub>IHF</sub>, one molecule of FimE binds to P<sub>FimE</sub>, and one molecule of Lrp binds to each of the three Lrp binding sites (notice that Lrp\* and Lrp represent the same species in our model). In other words, S<sub>6</sub> could be thought of as the switching dynamics analog of the *closed complex* configuration in transcription modeling. The association and dissociation rate constants for binding are deduced from  $k_{s6}/k_{-s6} = \exp(-\Delta G_{31b}/RT)$ . The switching rate  $\alpha_{11b}$  is adjusted to appropriate units (from (/cell/min) to (s<sup>-1</sup>)) and used as the rate constant in the reaction for the switch to move to the OFF position when the DNA-protein complex is in state 6.

<sup>1</sup>In this paper, a modifier is a species which is neither produced nor consumed by a reaction.

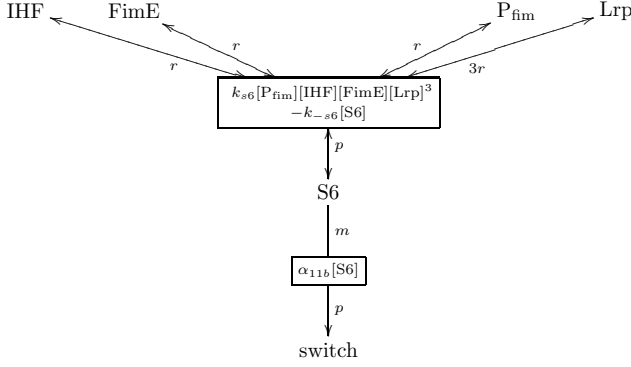


Fig. 2. Reverse-engineered full model reaction of the switch inversion through state 6.

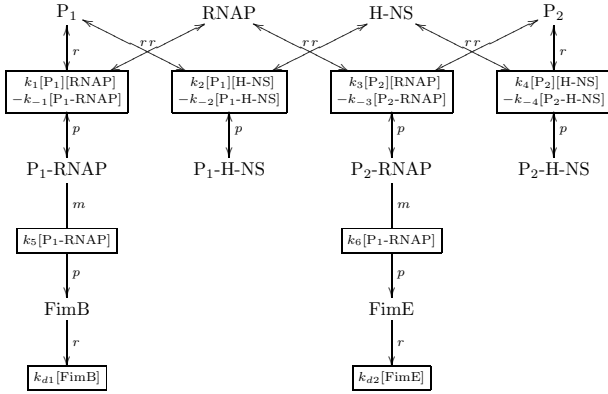


Fig. 3. Reaction-level model sub-network of FimB and FimE regulation.

The effect of H-NS is included by introducing production reactions for FimB and FimE. Since the quantitative knowledge of the production reactions of FimB and FimE is insufficient and the parameter values are not known, our model uses the mechanism as well as the parameter values which are consistent with our knowledge of the network. For example, it was shown that a differential modulation from a decrease in H-NS concentration and/or activity leads to a decrease in  $[FimE]/[FimB]$  ratio [1]. The FimB and FimE production mechanism as shown in Figure 3 captures this qualitative relationship.  $P_1$  is the promoter site for *fimB* and  $P_2$  is the promoter site for *fimE*. H-NS can also bind to the promoter sites. When the promoter is occupied by H-NS, RNAP cannot attach to the promoter site, and transcription is repressed. The effect of H-NS is much stronger in transcription of *fimB* than in that of *fimE*. Hence, an increase in concentration of H-NS leads to an increase in  $[FimE]/[FimB]$  ratio. The parameter values in the production mechanism are set as shown in Table II.

Our model also includes degradation reactions for FimB and FimE. The degradation rate constants are chosen so that the changes of the concentrations of the two recombinases are

TABLE II  
PARAMETER VALUES OF THE PRODUCTION REACTION SCHEME IN FIG 3.

Parameter	Value
$k_1$	$1.0 \text{ (nMs)}^{-1}$
$k_{-1}$	$1.0 \text{ (s)}^{-1}$
$k_2$	$10.0 \text{ (nMs)}^{-1}$
$k_{-2}$	$1.0 \text{ (s)}^{-1}$
$k_3$	$10.0 \text{ (nMs)}^{-1}$
$k_{-3}$	$1.0 \text{ (s)}^{-1}$
$k_4$	$0.01 \text{ (nMs)}^{-1}$
$k_{-4}$	$1.0 \text{ (s)}^{-1}$
$k_5$	$0.4367 \text{ (s)}^{-1}$
$k_6$	$0.10033 \text{ (s)}^{-1}$ if the switch is on $0.0 \text{ (s)}^{-1}$ otherwise

small over time in an ODE simulation given that the initial concentrations of the recombinases are 100nM and the initial concentration of H-NS is 10nM. Hence, in a Monte Carlo simulation, the concentration of FimB tends to significantly decrease over time if  $[H-NS] \gg 10\text{nM}$  while it tends to significantly increase over time if  $[H-NS] \ll 10\text{nM}$ , provided that the initial concentration of FimB is 100nM. Therefore, the characteristics of the ratio  $[FimE]/[FimB]$  can be controlled by adjusting the initial concentration of H-NS.

The goal of our analysis is to determine the effect of H-NS and Lrp on the ON-to-OFF switching probability. Our model is simulated for 20,000 runs using the optimized SSA for each combination of  $[H-NS]_0$  and  $[Lrp]_0$  where the range of  $[Lrp]_0$  is chosen to be from 0nM to 20nM with increment of 2nM and  $[H-NS]_0$  is allowed to be chosen from 0nM and 100nM. Each simulation starts with the switch in the ON position and is run for up to one cell generation. If the switch moves to the OFF position at least once, then the simulation is counted as an ON-to-OFF switching event. The ON-to-OFF switching probability is calculated by the number of the ON-to-OFF switching events divided by the total number of simulations with the same initial condition. We assume that *E. coli* is a cylinder  $2\mu\text{m}$  long and  $1\mu\text{m}$  in cross sectional diameter, and that the cell volume is fixed for each simulation.

Our full phase variation model contains 31 species and 52 *irreversible* reactions, and is encoded in SBML format [21]. The simulations of this model take 11.3 hours on a 3GHz Pentium4 with 1GB of memory. Figure 4 shows the changes of the computed ON-to-OFF switching probability as a function of Lrp concentration at different levels of H-NS.

Our results are consistent with those obtained by Wolf and Arkin [1] and, in turn, with the empirical observations. In particular, their analysis showed that the ratio of  $[FimE]/[FimB]$  is crucial to the switching probability in that, when the ratio is low (i.e. when  $[H-NS]$  is low in our model), the switching probability gets lower, and when the ratio is higher (i.e. when  $[H-NS]$  is high), the switching probability gets higher. This H-NS mechanism can be seen in our result as well. Furthermore, our results capture the phenotype tuning motif [1], whereby at intermediate concentrations Lrp activates, while at high levels it inhibits switching.

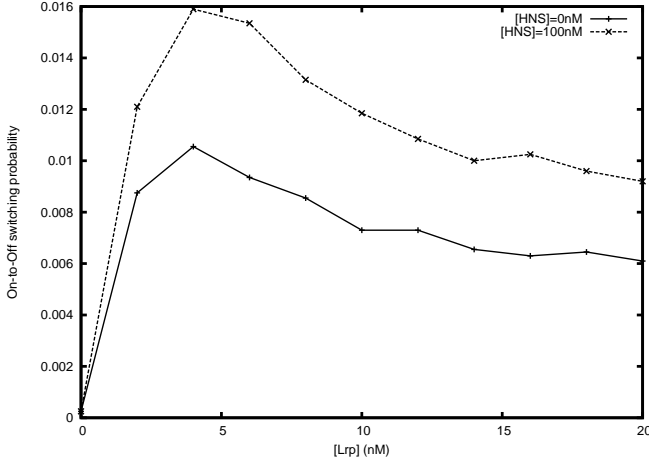


Fig. 4. Regulation of the ON-to-OFF *fim* switching probability by H-NS and Lrp.

#### IV. MODEL REDUCTION

Though our full model allows for a more detailed description of the network, such a low-level abstraction leads to substantial computational costs for analysis. Thus, going to a higher-level representation and abstracting away dynamically insignificant reactions or species to reduce the complexity of the system can help make our analysis more efficient. For example, the *quasi-steady-state approximation* (QSSA) has long been carried out to reduce the complexity of biochemical networks. Although this type of reduction has traditionally been done manually, this practice gets increasingly more tedious as the size of the network increases, eventually rendering it intractable and potentially leading to significant errors in large model transformations. Our abstraction tool alleviates these problems by automatically and systematically testing network patterns and characteristics to determine which abstraction methods are applicable.

Using this approach, our phase variation model can be transformed into a simplified (“reduced”) model using our automatic abstraction tool to allow for a more efficient analysis of the switching probability. The abstraction tool (outlined in Figure 5) begins with a model composed of a set of chemical reactions which could be simulated using SSA or one of its variants though at a substantial computational cost. To reduce the cost of analysis, the abstraction tool simplifies the original model by applying several abstraction methods that, among others, leverage QSSA and the rapid equilibrium assumption [22]. The result is a new, higher-level abstracted model with less reactions and species which substantially lowers the cost of stochastic analysis. Even though this transformation may cause our stochastic model to be non-Markovian, the simplified model still captures the large scale system behavior, such as the ON-to-OFF switching probability in our phase variation system, so that the trade-off of accuracy and speed is found to be desirable.

With our tool, all the reverse-engineered reactions are first abstracted (reduced) to more simplified reactions. For example,

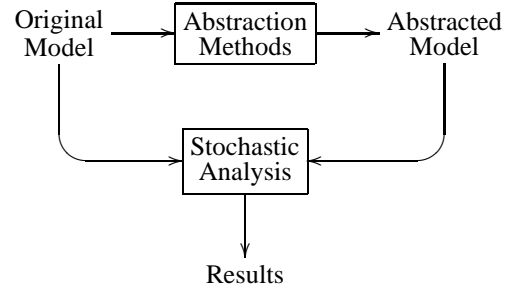


Fig. 5. Automated model abstraction tool flow.

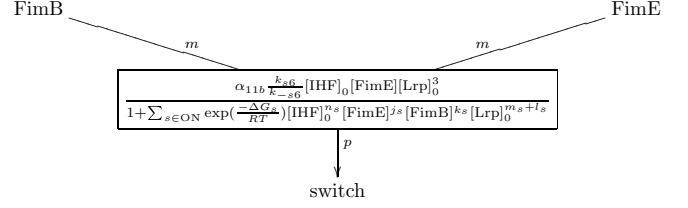


Fig. 6. Reduced model reaction of the switch inversion through state 6.

the reaction scheme for the switch state 6 as shown in Figure 2 is reduced to the reaction scheme shown in Figure 6 by applying the rapid equilibrium approximation. After applying these abstractions, the tool finds that the values of IHS and Lrp do not change over time. Thus, in this transformed reaction, IHS and Lrp are replaced with constants whose values are set to the corresponding initial concentrations. The production reaction scheme of FimB and FimE is also reduced as shown in Figure 7 by assuming that bindings and unbindings of promoter sites are rapid. After applying model reduction, the original model is transformed to a model with 3 species, (FimB, FimE, and switch) and 5 reactions (production and degradation reactions for FimB and FimE shown in Figure 7 and the combined ON-to-OFF switching reaction).

In order to compare the reduced abstracted model with the original model, we have performed the same number of simulations using the same simulator. Figure 8 shows the results from the simplified model. The results are in close agreement with the results from our original model. However, the computational gains from the model abstraction

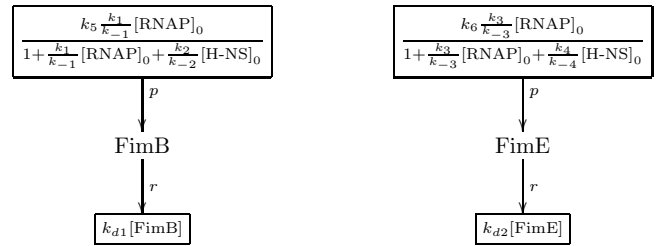


Fig. 7. Reduced reaction-level sub-network of FimB and FimE regulation.

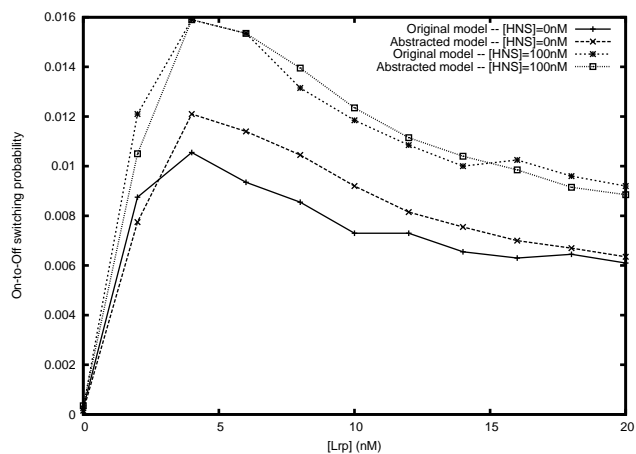


Fig. 8. Regulation of the ON-to-OFF *fim* switching probability by H-NS and Lrp. Comparison of the reduced and full model results.

are significant. The abstract model simulations only take 1.0 hour which is a speed-up of about 11 times compared with the runtime of the original model.

## V. FINITE STATE SPACE GENERATION

While the simplified model contains significantly less species and reactions, the state space of the model is still infinite. In order to make the state space finite, the upper limit of each molecular count could be specified. To enforce an upper molecular count limit, the abstraction tool can insert an artificial reaction. For example, to limit the upper molecular count of FimB to 300, the tool may insert a reaction:  $300 \text{ FimB} \rightarrow 299 \text{ FimB}$ , with a very high reaction rate. With this reaction inserted in our model, whenever the number of FimB molecules reaches 300 in a Monte Carlo simulation, the above reaction fires at a very high probability before the production reaction of FimB fires again. Consequently, the FimB count decreases to 299, and it gives an artificial upper limit of FimB molecules of 300. One could also use other techniques to specify the upper limit of each molecular count. For example, mass conservation conditions may be explicitly inserted, thus also eliminating some of the species. Alternatively, to limit the upper molecular count of FimB to 300, our abstraction tool could make the reaction rate function of FimB production piecewise with the expression: zero reaction rate when the FimB count is at least 300.

Suppose the upper limits of the molecular counts of both FimB and FimE are chosen to be 300. The state space of the model is still very large with a state count of 181,202 (90,601 for the states with switch ON and another 90,601 for switch OFF). To reduce the state space further without lowering the molecular count upper limits, the abstraction tool can increase the stoichiometry of FimB and FimE in their production and degradation reactions. For example, if the stoichiometry of FimB in its production and degradation reactions are amplified from 1 to 10 as shown in Figure 9, and if the artificial reaction to specify the upper limit of the FimB count is  $300 \text{ FimB} \rightarrow 290 \text{ FimB}$ , then the number of

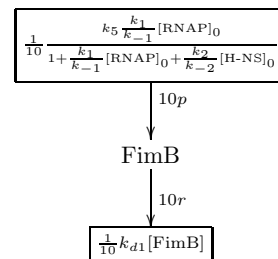


Fig. 9. Reduced FimB reactions with amplified stoichiometry.

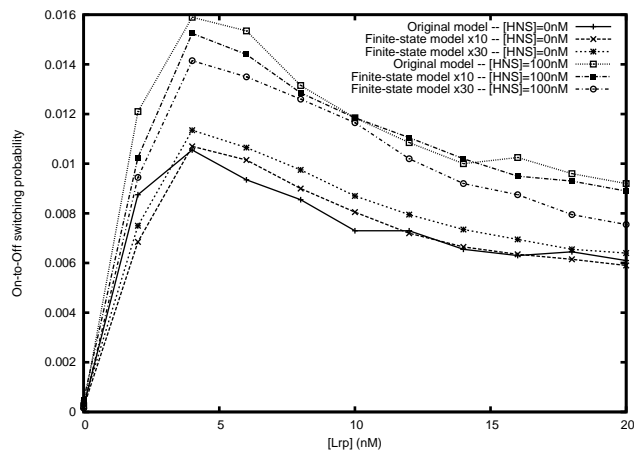


Fig. 10. Regulation of the ON-to-OFF *fim* switching probability by H-NS and Lrp. Comparison of the full model results with those from two different levels of finite state space abstraction.

distinct values that the FimB counts can take decreases from 301 to 31 given the initial count of FimB is a multiple of 10 and at most 300.

If the possible number of states for the FimB count is reduced to 31 with this method, the state space of the model is reduced to 1,922. The computational cost of the stochastic simulation is significantly reduced by this stoichiometry amplification as well since the kinetic laws of the production and degradation reactions are reduced by the factor of the stoichiometry amplified. The entire simulation of this finite-state model with stoichiometry of 10 takes only 14 minutes.

Notably, this method allows the degree of abstraction to be controlled at various levels by using the stoichiometry amplifier as a smoothing parameter. Thus, to reduce the state space as well as the simulation time even further, a more aggressive stoichiometry amplification can be used. For example, if the stoichiometry of FimB and FimE are amplified to 30, then the state space can decrease to only 242. Furthermore, the entire simulation of this reduced state-space model only takes 5 minutes which is a speed-up of more than 135 times compared with the runtime of the original model. Figure 10 shows the comparison of the results from these two finite-state models at different levels of abstraction compared with the results from the full model.

## VI. CONCLUSIONS

In this work, we have constructed a model of Type 1 pili phase variation in *E. coli* and have analyzed the response of the ON-to-OFF switching probability to the changes in the concentrations of the two global regulators, H-NS and Lrp. Our results quantify how the switching probability is influenced by those two regulators and confirm that the pattern is consistent with those expected empirically and analytically from [1]. Furthermore, we have demonstrated how the models with various levels of abstraction (as automatically generated by our method) can be used to obtain the results that agree with those of the original model, while also producing performance increases of up to two orders of magnitude without significant losses in accuracy.

## REFERENCES

- [1] D. M. Wolf and A. P. Arkin, "Fifteen minutes of *fim*: Control of type 1 pili expression in *e. coli*," *OMICS: A Journal of Integrative Biology*, vol. 6, no. 1, pp. 91–114, 2002.
- [2] H. Kulasekara and I. Blomfield, "The molecular basis for the specificity of *fimE* in the phase variation of type 1 fimbriae of *escherichia coli* k-12," *Molecular Microbiology*, vol. 31, pp. 1171–1181, 1999.
- [3] P. Klemm, "Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *escherichia coli*," *EMBO*, vol. 5, pp. 1389–1393, 1986.
- [4] I. Henderson, P. Owen, and J. Nataro, "Molecular switches - the ON and OFF of bacterial phase variation," *Molecular Microbiology*, vol. 33, pp. 919–932, 1999.
- [5] D. L. Gally, J. A. Bogan, B. I. Eisenstein, and I. C. Blomfield, "Environmental regulation of the *fim* switch controlling type 1 fimbrial phase variation in *escherichia coli* k-12: effects of temperature and media," *J Bacteriol.*, vol. 175, pp. 6186–6193, 1993.
- [6] I. C. Blomfield, D. H. Kulasekara, and B. I. Eisenstein, "Integration host factor stimulates both *FimB*- and *FimE*- mediated site-specific DNA inversion that controls phase variation of type 1 fimbriae expression in *escherichia coli*," *Mol Microbiol*, vol. 23, pp. 705–717, 1997.
- [7] P. B. Olsen, M. A. Schembri, D. L. Gally, and P. Klemm, "Differential temperature modulation by H-NS of the *fimB* and *fimE* recombinase genes which control the orientation of the type 1 fimbrial phase switch," *FEMS Microbiology Letters*, vol. 162, pp. 17–23, 1998.
- [8] T. Atlung and H. Ingmer, "H-NS: a modulator of environmentally regulated gene expression," *Molecular Microbiology*, vol. 24, pp. 7–17, 1997.
- [9] H. Kuwahara, C. Myers, N. Barker, M. Samoilov, and A. Arkin, "Asynchronous abstraction methodology for genetic regulatory networks," in *The Third International Workshop on Computational Methods in Systems Biology*, 2005.
- [10] H. D. Jong, "Modeling and simulation of genetic regulatory systems: A literature review," *J. Comp. Biol.*, vol. 9, no. 1, pp. 67–103, 2002.
- [11] P. Baldi and G. W. Hatfield, *DNA Microarrays and Gene Expression*. Cambridge University Press, 2002.
- [12] A. Arkin, J. Ross, and H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *escherichia coli* cells," *Genetics*, vol. 149, pp. 1633–1648, 1998. [Online]. Available: <http://www.genetics.org/cgi/reprint/149/4/1633.pdf>
- [13] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.
- [14] C. V. Rao, D. M. Wolf, and A. P. Arkin, "Control, exploitation and tolerance of intracellular noise," *Nature*, vol. 420, pp. 231–238, 2002.
- [15] M. Samoilov, S. Plyasunov, and A. P. Arkin, "Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations," *Proceedings of the National Academy of Sciences USA*, vol. 102, no. 7, pp. 2310–5, 2005.
- [16] J. M. Raser and E. K. O'Shea, "Control of stochasticity in eukaryotic gene expression," *Science*, vol. 304, pp. 1811–1814, 2004.
- [17] A. M. Kierzek, J. Zaim, and P. Zielenkiewicz, "The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression," *J. Biol. Chem.*, vol. 276, p. 8165, 2001.
- [18] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A*, vol. 188, pp. 404–425, 1992.
- [19] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, pp. 403–434, 1976.
- [20] G. K. Ackers, A. D. Johnson, and M. A. Shea, "Quantitative model for gene regulation by  $\lambda$  phage repressor," *Proc. Natl. Acad. Sci. USA*, vol. 79, pp. 1129–1133, 1982.
- [21] "Systems Biology Workbench Development Group," <http://www.sbw-sbml.org/>. [Online]. Available: <http://www.sbw-sbml.org/>
- [22] J. Keener and J. Sneyd, *Mathematical Physiology*. Springer, 1998.