STOCHASTIC MODELING OF THE PAP PILI EPIGENETIC SWITCH

Brian Munsky[1], Aaron Hernday[2], David Low[2] and Mustafa Khammash[1,*]
[1]Department of Mechanical and Environmental Engineering
[2]Biomolecular Sciences and Engineering,
University of California, Santa Barbara
Santa Barbara, CA 93106

Abstract

The following report uses a new stochastic model for the numerical study of the Pap pili epigenetic switch in *Escherichia Coli*. The model focuses on the period immediately following DNA replication during which the cell's fate is decided as a result of a few critical stochastic chemical reactions. Gene methylation and protein-gene binding events are modeled as Markovian state transitions through which the Pap gene can reach any of 16 distinct epigenetic configurations; some of which allow for the production of the feedback regulator PapI. These patterns are composed with an infinitely variable PapI population level for an infinite number of possible system states. The resulting Chemical Master Equation is analytically approximated using the Finite State Projection method and compared with experimental data involving variations in the concentration of DNA adenine methylase. Cells with no PapI are considered to be OFF, and cells with PapI are considered to be ON. The model successfully captures all experimentally observed traits, and suggests further analysis and experimental testing.

Keywords

Pyelonephritis-Associated Pili, Epigenetic Switch, Stochastic Simulation.

Introduction

Each year urinary tract infections affect over 26 million women in the United States (American Medical Association). Over 90% of the *Escherichia Coli* bacteria isolated from these infections are covered with small hair-like structures known as Pyelonephritis-Associated Pili, or Pap, (O'Hanley et al., 1985). From the perspective of E-coli survival within a host organism, pili expression is both beneficial and detrimental—pili enable E-coli to bind to host epithelial cells, establish colonies and feed off host organisms. Without the binding capabilities of pili, E-coli colonies would be more easily flushed from the host (i.e. during urination). Conversely, pili production consumes a significant portion of the cell's energy, thus weakening individual bacteria. Further, pili to host attachment may irritate the host and trigger an immune response. Thus, it is beneficial for any population descending from a single ancestor cell to have different pili expression phenotypes. This variation in expression comes as a result of an epigenetic switch—two cells with the exact same DNA can have vastly different expression: one expresses pili (phase ON) and one does not (phase OFF).

Previous experimental research conducted by David Low's group at UCSB (1985, 1994, 2002, 2003) has produced a vast amount of understanding regarding the Pap system switching mechanism. The key element of the system is the Pap gene (see Figure 1, top) which controls the transcription of the PapBA sequence of messenger rna's necessary for pili expression. There are two areas to which regulators bind and alter the output of the PapBA promoter. These are the *proximal* area (sites 1, 2 and 3) and the *distal* area (4, 5 and 6). The two most influential global regulators are: leucine-responsive regulatory

* Correspondence: khammash@engr.ucsb.edu

protein (lrp), which binds to sites 1-6, and DNA adenine methylase (Dam), which methylates the four GATC sequences found at the top and bottom strands at sites 2 and 5 (Hernday et al., 2002). In addition to the global regulators, the Pap-encoded local regulator protein (PapI) is produced within and is specifically linked to the Pap network (Hernday et al., 2003). Depending upon how the regulators alter the epigenetic structure of the Pap gene, the PapBA promoter may be active or inactive. The key-ingredients for the active cell (see Figure 1, bottom) is Dam methylation of the top and bottom GATC sequences in site 2 and lrp bound to distal sites 4, 5, and 6 (Hernday et al., 2003). PapI is produced only when the gene is in a production state. PapI increases the affinity of lrp for the distal sites, and this acts as a positive feedback regulator (Hernday et al., 2003).

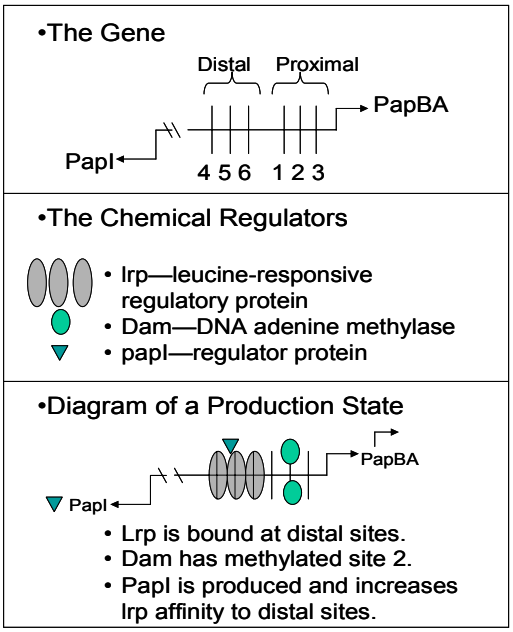

Figure 1. (Top) Schematic of the Pap-Gene. (Middle) Key regulatory components of the Pap network. (Bottom) Diagram of the gene in its pap Production phase.

The pap pili switch is presently the topic of research for many research groups. The first such model was recently published by Jarboe et al. (2004), in which the switch has been modeled using Gillespie's Stochastic Simulation Algorithm (1977) in order to explore the effects of DAM methylation and competition with lrp binding; growth rate; and initial state dependence on the pap switching behavior. This model was later simplified using a Markov chain model to describe the gene regulation prior to simulation of the transcription and translation events (Zhou, et al., 2005). Most recently, a hybrid Boolean/stochastic model of the pap pili switch has recently been developed by Shoemaker and Doyle who have demonstrated how molecular noise may lead to disparities between cell's genotype and its corresponding phenotype (Shoemaker, 2005). The present study uses a new unpublished stochastic model and the Finite State Projection (FSP) method (Munsky, 2005) to quantitatively explain the transitions between cells with no pap activity to those with pap activity and how these transitions are affected by changes in the concentration of DNA adenine methylase.

Model Description

In order for an OFF cell to switch to become an ON cell, the DNA replication process first expels bound lrp and splits the DNA into two daughter strands: one hemimethylated on the top (Hemi-T) and the other on the bottom (Hemi-B) (Hernday et al., 2002, 2003). At this point the Pap gene is open to subsequent lrp-binding and Dam methylation events under the following assumptions (Hernday et al. 2002, 2003): (1) lrp binds and unbinds cooperatively at all three proximal sites or at all three distal sites. (2) Lrp binding at the distal and proximal locations is mutually inhibitive. (3) Lrp blocks Dam methylation at both top and bottom sites. (4) The binding affinities of lrp at the proximal and distal sites are dependant upon the methylation pattern of the Pap gene and the population of PapI. (5) Lrp binding and unbinding is assumed to be in equilibrium. (6) Dam methylation is not reversible. (7) The concentrations of lrp and Dam are assumed to be known constants.

For these six assumptions, Figure 2 shows the sixteen possible methylation patterns which can be attained by any pap gene. Each methylation pattern can have four different lrp binding configurations: no bound lrp, lrp bound at proximal sites, lrp bound at distal sites, or lrp bound at all sites. Thus there are total of 64 (or 2^6) possible gene states. For fast lrp binding and unbinding events (assumption 5), this state dimension can be reduced to 16, where reactions depending upon a certain lrp configuration are scaled by the probability of that lrp configuration. This assumption speeds up computational time, but there is no significant difference between the results from the 16 and 64 gene-state models (results not shown). In addition to the 16 possible gene states, the model allows for the population of PapI to change through stochastic production and degradation events. In theory, the population PapI can reach any integer value, and thus there are an infinite number of possible states. Production of PapI can only occur when the gene has methylation pattern 11, 14 or 15 and lrp is bound only to the distal sites. PapI degradation is allowed to occur when the gene has any other methylation pattern. Any system with at least one molecule of PapI is considered to be ON.

The solution of the current model is conducted in the stochastic regime using the Finite State Projection Method (Munsky, 2005). Each reaction has a propensity function, a_i, that is dependent upon the concentrations of the reactants and the rate constants (Gillespie, 1977). Suppose

that Reaction k models the transition between methylation patterns i and j. This reaction has the propensity function:

$$a_k = c_k [x_i][DAM]dt \quad (1)$$

With knowledge of these propensities, there are a number of simulations that can be conducted to explore the behavior of the system. Many of these methods rely upon Monte-Carlo simulations such as Gillespie's Stochastic Simulation Algorithm (1977) or similar methods. However, concurrent studies by Munsky, have shown that in certain cases, the CME can be solved analytically (2004) or approximated via the FSP to within a specified error tolerance (2005). In general, the evolution of the probability density vector of every possible state, P, is governed by the infinite dimensional matrix Chemical Master Equation:

$$\dot{P}(t) = A \cdot P(t) \quad (2)$$

where A is a *state reaction matrix* comprised of all the possible reactions from one state to another. For example, Reaction k (from above) adds a negative term to the A_{ii} element and a positive term of the same magnitude to the A_{ji} element. The magnitude of this term is given by $a_k([DAM],[x_i])$, where $[x_i]$ is equal to one (because the reaction starts at state i), and $[DAM]$ is an assumed constant. In general A is an infinite dimensional square matrix, and Eqn.2 cannot be solved exactly. However, using the FSP, we can approximate the full system with a finite system:

$$P(t) = \exp(A_{FSP} \cdot t) \cdot P(0), \quad (3)$$

where A_{FSP} is an appropriately chosen principle submatrix of A (Munsky, 2005). This expression provides full knowledge (to any desired level of accuracy) of the system's probability density vector at any time provided that the initial probability density, $P(0)$, is known.

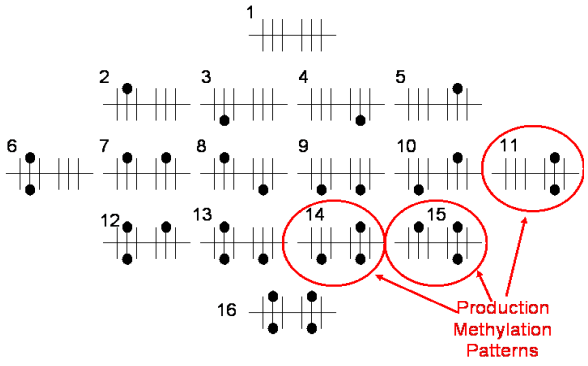

Figure 2. Schematic of the 16 possible methylation patterns. Only the circled "Production Patterns" may result in production of PapI). Transitions between methylation patterns require Dam methylation events, which are strongly dependent upon the interactions of LRP at various sites along the gene.

Results and Discussion

The above described 16 gene configuration model has been used to analyze the Pap system. Relative rates for Lrp binding versus unbinding events depend upon the population of PapI and were chosen based upon the data presented by Hernday et al. (2002—Figures 4a and 6.) In each analysis, the baseline concentrations of free Lrp and Dam were assumed to be known constants. Rates for every Dam methylation event were set corresponding to a remethylation half-life of 4 seconds (in the absence of lrp), based upon the results given by Urig et al., (2002). The generation life was chosen to be 30 minutes, and the initial condition was chosen such that the gene starts with methylation pattern number 1 (no previous methylation). Additional reports are in preparation, which will provide a detailed description of the current model as well detail the choices of all the chosen parameters. The first analysis (Figure 3) was aimed at predicting the OFF to ON switching behavior over a single generation as a function of the concentration of Dam.

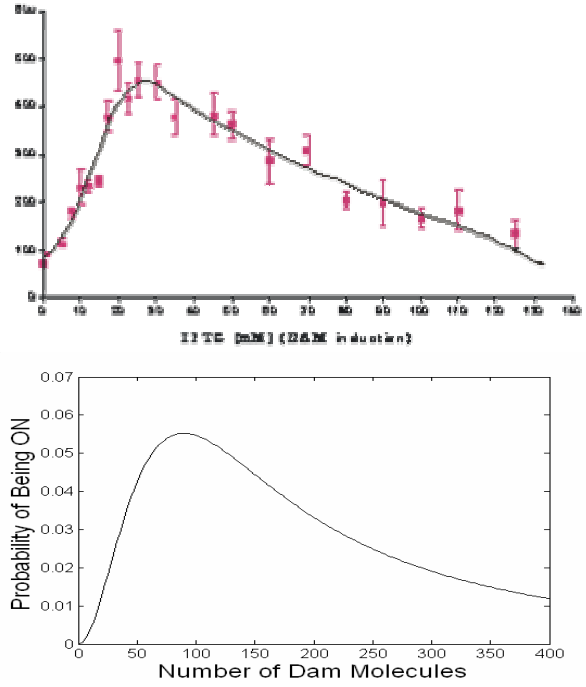

Figure 3. (Top) Experimentally measured Pap transcript levels under different Dam expression levels (the ordinate represents PapBA transcription levels and the abscissa represents relative Dam concentration). These data were obtained with E. coli Pap-lac containing dam under control of plac (Warren et al., 2000). (Bottom) Model prediction of the Pap pili OFF to ON switching behavior in response to varying levels of Dam expression for wild-type Pap.

Figure 3, top, illustrates qualitative experimental results where the transcription level has been measured as a function of the relative amount of Dam. Using the current model, Figure 3 (bottom) plots the calculated proportion of cells in the ON state after a single generation (30 minutes) for a wild type gene as a function of the Dam concentration. The qualitative similarity in the two plots illustrates that the current model successfully captures the effect of Dam on the system.

Conclusions

This report has utilized a new stochastic model in order to produce preliminary predictions regarding the transient switching behavior of the Pyelonephritis-associated pili (Pap) switch. Specifically, studies have modeled the behavior of the Pap system beginning with an unmethylated daughter cell immediately following DNA-replication. The model considers the stochastic interactions of the Pap-gene with the global regulators DNA adenine methylase (Dam) and leucine-responsive regulatory protein (lrp). The *pap* gene is capable of achieving sixteen different methylation patterns, and the full system can reach an infinite number of distinct possible states: sixteen states for each possible population of PapI. The evolution of the state probabilities over time is governed by an infinite set of linear, time-invariant, ordinary differential equations, which can be analytically approximated to any desired degree of accuracy using the Finite State Projection method.

Preliminary model results have been compared to qualitative experimental data in which the amount of PapBA transcription has been explored as a function of the concentration of Dam. The model successfully captures all experimentally observed trends. Related studies are in preparation and will document testable predictions regarding the sensitivity of the model to genetic mutations, uncertainty in reaction parameters, and variations in other key chemical concentrations such as Leucine Responsive Protein.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant ITR CCF-032657 and by the Institute for Collaborative Biotechnologies through grant DAAD19-03-0004 from the U.S. Army.

References

Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, *81*, 2340.

Hernday, A., Krabbe, M., et al. (2002). Self-perpetuating Epigenetic Pili Switches in Bacteria. *Proceedings of the National Academy of Science,* USA, 99**,** 16470.

Hernday, A., Braaten, B., Low, D. (2003). The Mechanism by which DNA Adenine Methylase and PapI Activate the Pap Epigenetic Switch. *Molecular Cell, 12,* 947.

Jarboe, L., Beckwith, D., Liao, J. (2004). Stochastic Modeling of the Phase-Variable pap Operon Regulation in Uropathogenic Escherichia coli. *Biotechnology and Engineering*, *88,* 189.

O'Hanley, P., Low, D. (1985) Gal-Gal Binding and Hemolysin Phenotypes and Genotyes Associated with Uropathogenic Escherichia Coli. *Embo J*, *14(23)*, 5785.

Munsky, B. E., (2004) Stochastic Simulation of the Pap (pili) Epigenetic Switch. *UCSB Chem. Eng. 154 Class Report,* June 7.

Munsky, B., Khammash, M., (2005) The Finite State Projection Algorithm for the Solution of the Chemical Master Equation. *Submitted to J. Chemical Physics*.

Shoemaker, J., Doyle, F., (2005) *FOSBE Proceedings.*

Urig, S., Gowher, H., et al. (2002). The Escherichia Coli Dam DNA Methyltransferase Modifies DNA in a Highly Processive Reaction. *J. Mol. Bio., 319*, 1085.

Warren, J., Walker, J., et al. (2000). Construction and Characterization of a Highly Regulable Expression Vector, pLAC11, and its Multipurpose Derivatives, pLAC22 and pLAC33. *Plasmid*, *44(2)*, 138.

Zhou, B., Beckwith, D., Jarboe, L., Liao, J. (2005) Markov Chain Modeling of Pyeleonephritis-Associated Pili Expression in Uropathogenic Escherichia Coli. *Biophys J BioFAST, Jan 28, 2005*