Modelling *fim* expression in *Escherichia coli* K12

Patrick de Vries[1], Colin G. Johnson[1], and Ian C. Blomfield[2]

School of Computing[1] and School of Biosciences[2], University of Kent, Canterbury, Kent, CT2 7NF
{pd79,c.g.johnson}@kent.ac.uk
http://www.cs.kent.ac.uk

Abstract. Fimbriae are structures in *Escherichia coli*, the expression of which is controlled by the *fim* operon. Understanding this expression is important because the fimbriae are important virulence factors.
This expression can be studied using targeted mutations to the DNA, which can be used to disable binding or transcription of a protein. However, this can be problematic as only the net effect is observed. Turning off expression of a protein may enhance *fim* expression, but deactivating this protein may also repress another protein that functions as an activator of *fim* expression. The net result may be that *fim* expression goes down, so it would seem at first glance that the disabled protein was an activator of *fim* expression and not a repressor.
In order to understand this complex network of interactions, an agent based model of *fim* expression has been created. The subject of this paper is to introduce this model and to use it to disambiguate between a number of hypotheses about this system. Parameters such as binding probability will be optimised using a genetic algorithm. The final model and parameters show a good match to experimental data.

1 Introduction

Fimbriae are hair-like attachments that *Escherichia coli* (*E.coli*) bacteria use to attach themselves to host cells and subsequently enter them. Because of the bacteria's ability to penetrate cells, *E.coli* bacteria infections are very hard to treat and so it is imperative we learn more about the way the fimbriae are regulated.

The main method of investigating the processes within a bacterial cell is by making focused mutations of the DNA. By directed disabling of the production of protein, or by changing binding sites within the DNA new information on protein expression can be gained. However, a mutation can have further effects within the cell then just the process focussed upon. therefore an effect attributed to a DNA fragment or protein can in fact be a different mechanism.

One can try using a computer model to simulate the process, but for this one needs parameters to feed the model, such as binding affinities etc. The aim of this paper is to model this process using the experimental data currently available, i.e. data on replacement mutations.

A key regulator of *fim* expression is the protein FimB. The regulation of FimB expression is not well understood at present. One theory is that H-NS (Histone-like nucleoid-structuring) protein [1] represses *fimB* expression and that SlyA—a protein first discovered in *Salmonella*—antagonises H-NS and reduces *fimB* repression [2]. Experimental data [3] concerning these interactions has been produced via replacement mutations, where the binding sites that control this expression are deleted and the consequent behaviour of the system observed. This information will form the core data input for tuning parameters in our model.

2 Previous Work

Previous attempts have been made at modelling aspects of *E.coli*, for example by means of differential equations, either focused on the individual cell and the processes within [4, 5], or on the entire population [6]. Using differential equations on the whole population can be a good method for predicting global properties such as cell growth, but since biological systems are inherently not continuous, these models will ignore the stochastic nature of the system. For this reason we can use stochastic agent-based models such as those of Karmakar and Bose[7] and Ramsey et al. [8]. Karmakar and Bose describe a stochastic model for transcription factor-regulated gene expression, however this is limited to a broad conceptual model because the detailed parameters are not matched to any experimental data . Ramsey et al. discuss a modelling environment for stochastic and deterministic models and compare results for complex—but well known—regulatory networks using both a deterministic and a stochastic approach. For this example data is available, but it is not clear what could be done if the data were limited or not available. Parameters necessary for this are binding affinity/probability and extent of interaction between different bound proteins. Usually, binding probability of a protein to the DNA is found by gel-shift experiments [9].

A significant difficulty in understanding *fim* expression is that there is no direct way of measuring binding affinity of the protein SlyA. Normally, when doing these gel shifts at different concentrations of the protein clear bands appear for the parts where the protein is bound to the DNA [10]. For unknown reasons gel shifts with SlyA produce irregular banding. The only band with a consistent location is that of the unassociated DNA. If gel shifts would produce consistent results we could have used a similar method as Valeyev et al. [11] used in their model for calcium-calmodulin interaction.

There are many hypotheses for how *fim* expression is regulated in *E.coli* [12, 9]. A main regulatory process in the expression is controlled by a fragment of DNA that can be expelled and reinserted in the opposite direction [13, 14]. It can be seen as a switch turning from OFF to ON and back [15, 16]. It is also known that the switch is regulated by the proteins FimB and FimE, where FimB is expected to turn the switch from OFF to ON and FimE favours the OFF position [12]. The regulation of FimB is the main focus of this paper.

3 Materials and Methods

Our work consists of two main parts, an agent-based model for the *fim* expression and a parameter optimisation model using genetic algorithms.

3.1 Agent-based Model

In this case the hypothesis tested will be the assumption that the protein H-NS will act as a repressor for *fimB* expression and the protein SlyA will be acting as an antagoniser of H-NS preventing it from binding to the DNA.

The main components of this system are the regulatory region for the *fimB* gene and the proteins that bind to that region, which are SlyA and H-NS. These proteins and the binding sites are represented by entities in the model, which interact according to the description given in the remainder of this section.

Two SlyA binding sites have been identified, called O_{SA1} and O_{SA2}—there is also a possible third site called O_{SA3}. The sites O_{SA1} and O_{SA3} overlap not only with each other (by one base pair) but also with H-NS binding sites. This is shown in Figure 1.

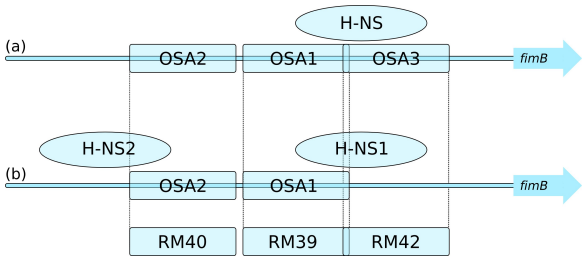

Fig. 1. (a) assumption 1 – Three SlyA binding sites, O_{SA1}, O_{SA2} and O_{SA3} and the overlap with the binding site for H-NS. (b) assumption 2 – Two SlyA binding sites, O_{SA1} and O_{SA2} overlap each with an H-NS binding site.

H-NS represses *fimB* expression and is antagonised by SlyA. FimB in turn will switch the *fim*-switch (*fimS*) ON which will start the translation of the *fim* operon to form the actual fimbriae. While the switch is turned ON FimE will be formed which will stimulate the switch to turn OFF.

It is unknown how high the binding affinity is of SlyA to any of the three possible binding sites or how strong the effect of SlyA is on the binding of H-NS. Using an agent-based model with a genetic algorithm to supply the binding affinities for the different binding sites and the effects of binding on the repression will circumvent this problem.

A number of variants on the model have been hypothesised, with different assumptions about the interaction between the binding sites. These are given in Table 1. Figure 1a shows model 1 from Table 1, whereas the various interactions

in models 2-5 can be understood with reference to Figure 1b. Figure 1 also shows the regions RM40, RM39 and RM42, which are the regions that are replaced by a non-functional DNA fragment in the various replacement mutation experiments.

Model	Summary
1	Assumption 1, where effect of SlyA on H-NS is expected to be the same as the Replacement mutation.
2	Assumption 2. H-NS effect on *fimB* expression is when bound always 100%
3	H-NS has a cumulative effect on *fimB*expression, but H-NS effect can vary.
4	H-NS has an independent variable effect on *fimB* expression, but repression is 100% when both sites are occupied.
5	H-NS only has a variable effect on *fimB* expression, but only when both sites are occupied.

Table 1. summary of differences in the 6 models

In total there are several parameters to optimize, subdivided into 5 sets. Their function is described in table 2

Sets	Explanation
Set 1	Effect of binding of SlyA to O_{SA1}, O_{SA2} and O_{SA3}
Set 2	Binding probability of SlyA and H-NS to the different sites
Set 3	Concentration of either FimB or FimE at which switching probability is 50%
Set 4	Effect of different replacement mutations
Set 5	Repressing effect of H-NS binding on *fimB* expression

Table 2. Explanation of the different parameters

In the simulation the population of *E. coli* bacteria start out as 50 afimbriate cells, growing, dividing and dying for 1000 iterations, where the colony grows to approximately 30,000 cells, consisting of a mixture of the fimbriate and afimbriate types. At each iteration, each cell individually checks the amount of FimB and FimE protein and based on that decides whether to switch the production of fimbriae ON or OFF. FimE promotes the ON-to-OFF switch and is

only produced when the switch is turned ON. FimB production depends on the repression by H-NS and production is independent of the switch, although FimB promotes the switch OFF-to-ON. This reflects the best current knowledge about the functioning of the biological system.

The calculation of the binding probabilities of SlyA and H-NS depends on a number of parameters, that are described in Table 2. When H-NS binds it has a maximum effect on the repression of FimB. The effect is reduced when SlyA binds and the repression is reduced by as much as stated in parameter set 1. The binding probability of H-NS can be reduced or enhanced in the case of replacement mutations by values stated in parameter set 5. When SlyA binds to one of the sites it reduces the effect of H-NS inhibiting *fimB* expression by as much as the the respective gene from parameter set 1.

3.2 Parameter Optimisation

The structure for the parameter optimization model is shown in Figure 2. It starts out with generating a population of solutions for the parameters. In each generation the solutions are tested by running the agent-based model described above on each set of parameters in the population and comparing with experimental data, from which a fitness measure is calculated, which is stored in an output-file.

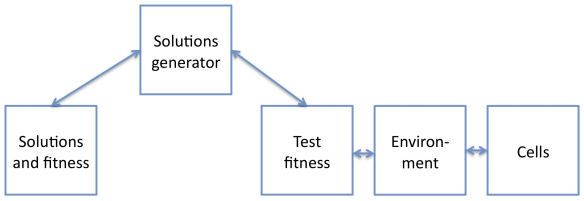

Fig. 2. Structure of the computer model.

The experimental data used is concerned with replacement mutations obtained from *switching experiments* as done by Gally et al. [17] and calculated from β-galactosidase experiments [18] (the relation between β-galactosidase and *switching frequency* as shown by El-Labany et al. [19]; data as used in the model can be found in Figure 3). These mutations include RM39, RM40 and RM42, where RM39 replaces O_{SA1}, RM40 replaces O_{SA2} and RM42 replaces O_{SA3} (as illustrated in Figure 1). Notice that RM39 and RM42 have a direct effect on the binding of H-NS. Two other mutations include ΔSlyA and ΔH-NS[1].

The initial population of parameters is randomly created by sampling from a uniform distribution within sensible ranges. The algorithm then iterates for

[1] In wild type background the absence of H-NS is tested and as with the other experiments either with or without SlyA present.

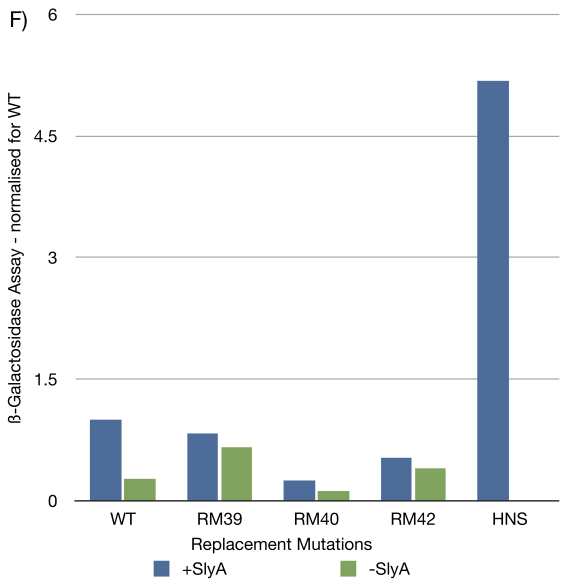

Fig. 3. Experimental results used to test the model's fitness. (WT = wild type bacterium).

50 generations. In each generation, after they have been tested for their fitness, the best solution is kept for the next generation and the rest of the solutions are generated by means of crossover. The candidates for crossover are selected by tournament selection, where from a random selection of four solutions the two strongest are mixed [20, 21]. The new parameter sets are then subjected to random mutation, where one of the parameters is altered, to prevent ending in a local optimum.

Fitness is measured by taking the least square error (LSE) for each model compared with experimental results.

$$\text{LSE} = \sum (F_{i,exp} - F_{i,Model})^2 \tag{1}$$

Where i is the *fimB* expression for each of the different mutants.

For every test of a parameter list a new run of the model is created as described in the previous section.

Reaching an optimum in the parameter optimisation is a good sign the hypothesis may be correct, however, the parameters produced should be scrutinised by comparison with what is known from biology in order to check that the hypothesis produced by the optimisation process is biologically realistic.

4 Results

The LSE for the different models is shown in Table 5 and the change in LSE over the generations is shown in Figure 4. The parameter optimisation shows that most of the different assumptions lead to an optimum where the least square error (LSE) is reduced 10-400 fold.

The plots of *fimB* expression for each mutant is plotted in Figure 5, where Model 1 has a problem with modelling the SlyA mutant, but the other models give a close resemblance with experimental data, also shown by having relatively low error values.

A further test is to look at the different parameters collected from these different models. These are shown in Tables 3-4. No experimental data is available to compare these. However we do know that H-NS binds much stronger to the DNA then SlyA and also is known that replacement mutation RM40 makes site H-NS2 closer to the consensus of an H-NS binding site, so may actually enhance binding of H-NS. Replacement mutation RM39 and RM42 tend to make binding of H-NS to site H-NS2 less likely.

	Parameter Set 1					Parameter Set 2				
Model	O_{SA1}	O_{SA2}	O_{SA3}	$O_{SA1\&2}$	$O_{SA2\&3}$	O_{SA1}	O_{SA2}	O_{SA3}	H-NS1	H-NS2
1	29	40	23	53	9	9	58	9	99	-
2	5	15	-	-	-	31	22	-	100	94
3	27	1	-	-	-	9	43	-	100	94
4	13	3	-	-	-	15	33	-	100	94
5	2	12	-	-	-	4	21	-	92	98

Table 3. Parameter sets 1 and 2. Parameter set 1 shows the effect of SlyA binding on H-NS repression, and parameter set 2 shows binding probabilities.

	Parameter Set 3		Parameter Set 4			Parameter Set 5	
Model	[FimB]	[FimE]	RM39	RM40	RM42	H-NS1	H-NS2
1	100	3	0	8	1	100	-
2	25	92	3	-64	-	100	100
3	100	39	3	-71	-	100	100
4	75	39	-88	-49	-	83	82
5	55	19	-8	-47	-	93	100

Table 4. Parameter sets 3,4 and 5. Parameter set 3 shows the FimB and FimE concentrations in the cell at which the probability of switching ON or OFF the production of fimbriae is 50%, parameter set 4 shows the effect of replacement mutation on the ability of H-NS to repress *fimB* expression (where negative numbers mean that H-NS binding is enhanced), and parameter set 5 shows the extent of H-NS repression of *fimB* expression (note that Model 1 only has one H-NS binding site).

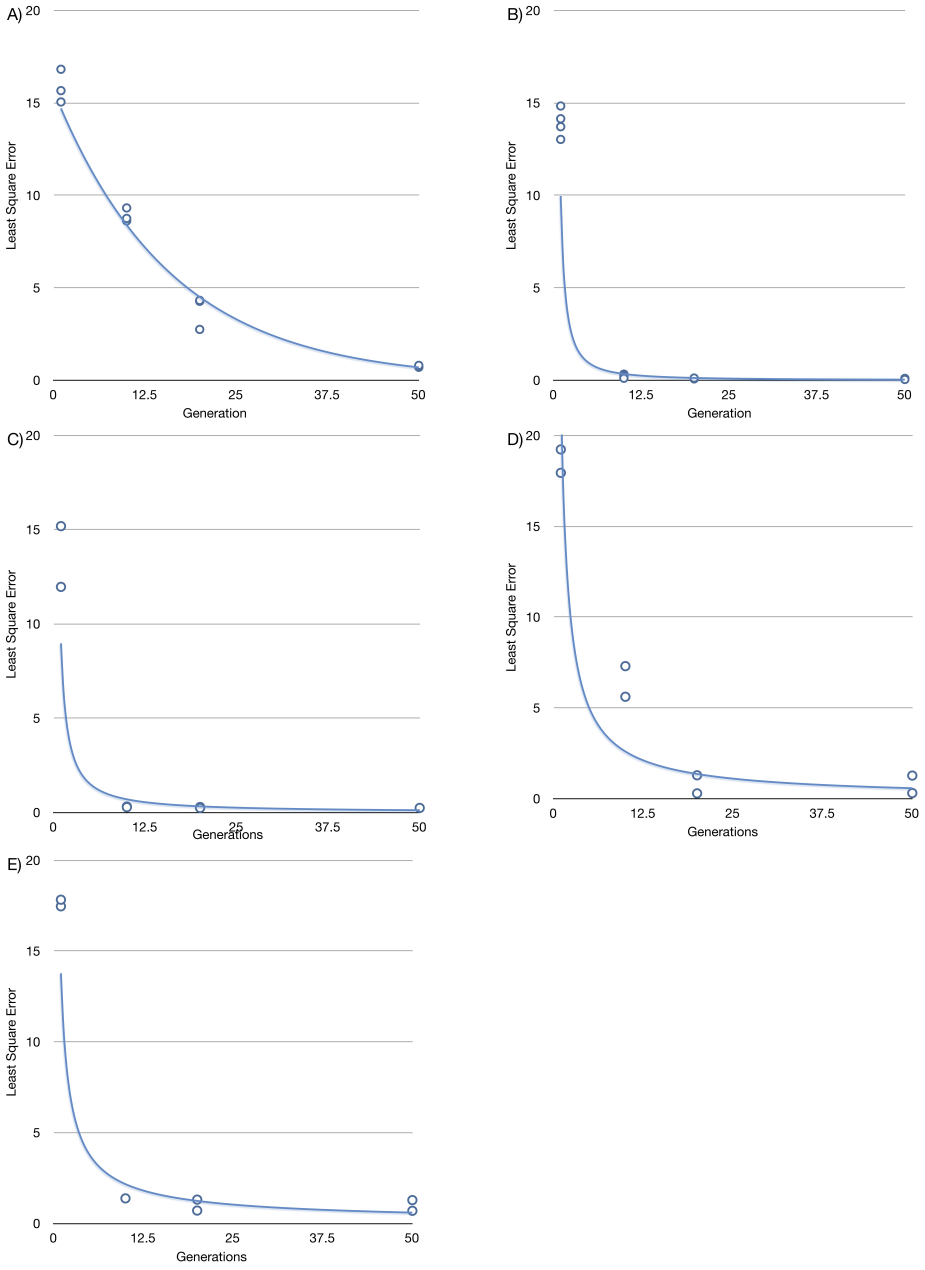

Fig. 4. Error values as a function of generation, samples at initial population, after 10 generations, 20 generations and finally after 50 generations. See A, B, C, D or E for respectively Model 1, 2, 3, 4 or 5.

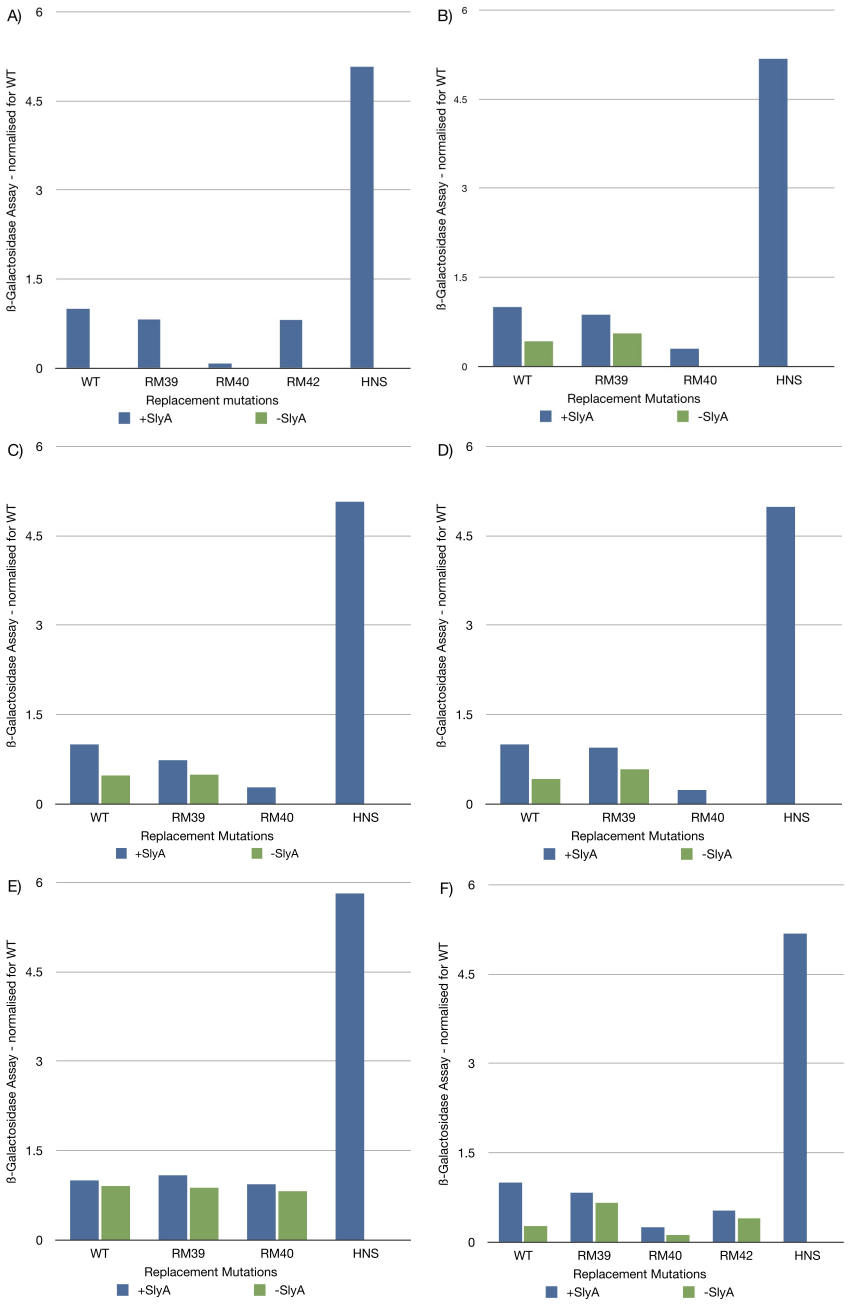

Fig. 5. *fimB* expression for each model. See A, B, C, D or E for respectively Model 1, 2, 3, 4 or 5 and F for experimental values

Model	Error
1	0.7920
2	0.0566
3	0.0611
4	0.1013
5	1.1901

Table 5. Results for the 5 models, Error values

5 Discussion

In general the model corresponds well with the experimental data as the graphs for *fimB* expression are nearly identical to that of the biological experiments. This is also shown in the small error value in Table 5. The real measure however is in the parameters.

Knowing the biological mechanism on which this model is based, we can see that only the parameters of Model 3 seem to make any sense. This is evident from the effect of SlyA binding to O_{SA1} or O_{SA2} that have a little, but still significant effect on H-NS repression, as does the replacement mutation RM39. RM40 however is quite capable of increasing binding of H-NS to the site and increases its binding probability to a near 100%.

At the time of writing it is unclear how the two H-NS sites interact when repressing *fimB* expression. The results of this study would suggest the two H-NS sites act independently, although their effect on the repression is cumulative.

The third SlyA site has only been included in model 1, and its omission from the other models is supported by an examination of the DNA sequence, which shows that the O_{SA3} has less resemblance to the SlyA binding site consensus compared with the other two sites (See Table 6).

Binding site	Genetic code
SlyA	TTAGCAAGCTAA
O_{SA1}	TTAGCATGATAA
O_{SA2}	CTAGGGACCTAA
O_{SA3}	ATAGCCACTAA

Table 6. Genetic code of the different binding sites

Further work is needed to investigate the remaining H-NS sites. There are two more sites in the same region on the DNA, although it is accepted that these sites are not under control by any of the SlyA sites mentioned here. A further SlyA site (O_{SA4}) has been identified, but under normal circumstances is not found to be occupied by SlyA, but the site does overlap with the H-NS2 site and under some circumstances its effect on antagonising H-NS is still significant.

6 Conclusions

An agent-based model has been presented for the regulation of expression of *fimB* in *E. coli* with regard to the regulatory proteins SlyA and H-NS. A number of hypotheses have been presented for the effect of these proteins on the expression, and optimisation of parameters against experimental data from replacement mutation experiments has been used to disambiguate between these hypotheses. One hypothesis has clearly been identified as the most likely candidate for explaining the experimental data.

Future work will focus on the influence of other H-NS and SlyA binding sites on the system.

References

1. Remus T. Dame, Martijn S. Luijsterburg, Evelyne Krin, Philippe N. Bertin, Rolf Wagner, and Gijs J. L. Wuite. DNA bridging: a property shared among H-NS-like proteins. *Journal of Bacteriology*, 187(5):1845–1848, 2005.
2. David Corbett, Hayley J. Bennet, Hamdia Askar, Jeffrey Green, and Ian S. Roberts. SlyA and H-NS regulate trascription of the *Escherichia coli* K5 capsule gene cluster, and expression of *slyA* in *Escherichia coli* is temperature dependent, positively autoregulated, and independent of H-NS. *Journal of Biological chemistry*, 282(46):33326–33335, 2007.
3. Baljinder K. Sohanpal, Simon Friar, Joanne Roobol, Jacqueline A. Plumbridge, and Ian C. Blomfield. Multiple co-regulatory elements and IHF are necessary for the control of *fimB* expressionin response to sialic acid and n-acetylglucosamine in *Escherichia coli* K-12. *Molecular Microbiology*, 63:1223–1236, 2007.
4. Dominique Chu and Ian C. Blomfield. Orientational control is an efficient control mechanism for phase switching in the *E. coli fim* system. *Journal of Theoretical Biology*, 244(3):541–551, 2007.
5. Dominique Chu, Jo Roobol, and Ian Blomfield. A theoretical interpretation of the transient sialic acid toxicity of a *nanR* mutant of *Escherichia coli*. *Journal of Molecular Biology*, 375:875–889, 2008.
6. Ting Chen, Hongyu L. He, and George M. Church. Modeling gene expression with differential equations. *Pac Symp Biocomput*, pages 29–40, 1999.
7. Rajesh Karmakar and Indrani Bose. Stochastic model of transcription factor-regulated gene expression. *Physical Biology*, 3:200–208, 2006.
8. Stephen Ramsey, David Orrell, and Hamid Bolouri. Dizzy: Stochastic simulation of large-scale genetic regulatory networks. *Journal of Bioinformatics and Computational Biology*, 3:415–436, 2005.
9. Paula L. Roesch and Ian C. Blomfield. Leucine alters the interaction of the leucine-responsive regulatory protein (Lrp) with the *fim* switch to stimulate site-specific recombination in *Escherichia coli*. *Molecular Microbiology*, 27(4):751–761, 1998.
10. James K. Lithgow, Fouzia Haider, Ian S. Roberts, and Jeffrey Green. Alternate SlyA and H-NS nucleoprotein complexes control *hlyE* expression in *Escherichia coli* K-12. *Molecular Microbiology*, 66:685–698, 2007.
11. Najl V Valeyev, Declan G Bates, Pat Heslop-Harrison, Ian Postlethwaite, and Nikolay V Kotov. Elucidating the mechanism of cooperative calcium-calmodulin interactions: a structural systems biology approach. *BMC Systems Biology*, 48(2), 2008.

12. Per Klemm. Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. *The EMBO Journal*, 5(6):1389–1393, 1986.
13. Ian C. Blomfield, Don H. Kulasekara, and Barry 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*. *Molecular Microbiology*, 23(4):707–717, 1997.
14. Aileen M. Adiciptaningrum, Ian C. Blomfield, and Sander J. Tans. Direct observation of Type 1 fimbrial switching. *EMBO reports*, 10(5):527–532, March 2009.
15. Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, January 2000.
16. Denise M. Wolf and Adam P. Arkin. 15 minutes of *fim*: Control of phase variation in *E.coli*. *OMICS: A Journal of Integrative Biology*, 6(1):91–114, January 2002.
17. David L. Gally, Jacqueline A. Bogan, Barry I. Eisenstein, and Ian C. Blomfield. Environmental regulation of the *fim* switch controlling Type 1 fimbrial phase variation in *Escherichia coli* K-12: Effects of temperature and media. *Journal of Bacteriology*, 175(19):6186–6193, 1993.
18. Jeffrey H Miller. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, 1972.
19. Sammia El-Labany, Baljinder K. Sohanpal, Maryam Lahooti, Robert Akerman, and Ian C. Blomfield. Distant *cis*-active sequences and sialic acid control the expression of *fimB* in *Escherichia coli* K-12. *Molecular Microbiology*, 4:1109–1118, 2003.
20. Melanie Mitchell. *An Introduction to Genetic Algorithms*. The MIT press, 1996.
21. A.E. Eiben and J.E. Smith. *Introduction to Evolutionary Computing*. Springer-Verlag, 1998.