

# Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression

Stekel, Dov; Jenkins, Dafyd

DOI:  
[10.1186/1752-0509-2-6](https://doi.org/10.1186/1752-0509-2-6)

*Document Version*  
Peer reviewed version

*Citation for published version (Harvard):*  
Stekel, D & Jenkins, D 2008, 'Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression', *BMC systems biology*, vol. 2, pp. 6. <https://doi.org/10.1186/1752-0509-2-6>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

## **Strong negative self regulation of Prokaryotic transcription factors increases the intrinsic noise of protein expression**

*BMC Systems Biology* 2008, **2**:6 doi:10.1186/1752-0509-2-6

Dov J Stekel (d.j.stekel@bham.ac.uk)  
Dafyd J Jenkins (djj134@bham.ac.uk)

**ISSN** 1752-0509

**Article type** Research article

**Submission date** 24 May 2007

**Acceptance date** 18 January 2008

**Publication date** 18 January 2008

**Article URL** <http://www.biomedcentral.com/1752-0509/2/6>

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

© 2008 Stekel and Jenkins, licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Strong negative self regulation of Prokaryotic transcription factors increases the intrinsic noise of protein expression

Dov J Stekel\*<sup>1</sup>, Dafyd J Jenkins<sup>1</sup>

<sup>1</sup>Centre for Systems Biology, School of Biosciences, University of Birmingham, Edgbaston B15 2TT, UK

Email: Dov J Stekel\* - d.j.stekel@bham.ac.uk; Dafyd J Jenkins - dj134@bham.ac.uk;

\*Corresponding author

## Abstract

---

**Background:** Many prokaryotic transcription factors repress their own transcription. It is often asserted that such regulation enables a cell to homeostatically maintain protein abundance. We explore the role of negative self regulation of transcription in regulating the variability of protein abundance using a variety of stochastic modeling techniques.

**Results:** We undertake a novel analysis of a classic model for negative self regulation. We demonstrate that, with standard approximations, protein variance relative to its mean should be independent of repressor strength in a physiological range. Consequently, in that range, the coefficient of variation would increase with repressor strength. However, stochastic computer simulations demonstrate that there is a greater increase in noise associated with strong repressors than predicted by theory. The discrepancies between the mathematical analysis and computer simulations arise because with strong repressors the approximation that leads to Michaelis-Menten-like hyperbolic repression terms ceases to be valid. Because we observe that strong negative feedback increases variability and so is unlikely to be a mechanism for noise control, we suggest instead that negative feedback is evolutionarily favoured because it allows the cell to minimize mRNA usage. To test this, we used *in silico* evolution to demonstrate that while negative feedback can achieve only a modest improvement in protein noise reduction compared with the unregulated system, it can achieve good improvement in protein response times and very substantial improvement in reducing mRNA levels.

**Conclusions:** Strong negative self regulation of transcription may not always be a mechanism for homeostatic control of protein abundance, but instead might be evolutionarily favoured as a mechanism to limit the use of mRNA. The use of hyperbolic terms derived from quasi-steady-state approximation should also be avoided in the analysis of stochastic models with strong repressors.

---

## Background

Recent innovations in synthetic biology and real-time imaging have revealed that the abundance of individual proteins in single cells is subject to significant variation, both between cells, and temporal variation within single cells, typically measured in unicellular organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* [1–5]. Such variability is expected and confirmed by mathematical models of protein production, which have demonstrated that protein abundance is subject to random fluctuations resulting from intrinsic and extrinsic noise associated with transcription, translation and mRNA and protein degradation [6–9].

It is important for cells to control the abundance of individual proteins. There are many strategies that cells can employ, including the control of transcription, translation and the degradation of mRNA and protein. One important strategy is the employment of transcription factors — proteins that either positively or negatively regulate transcription.

Many transcription factors have the capacity to control their own transcription, usually in a negative fashion. This is particularly the case in prokaryotes; for example in the best studied organism *E. coli* K12, 79 of the 146 transcription regulators listed on EcoCyc [10] control their own expression — the majority of these are negative feedback. Typically, these regulators are associated with operons containing other proteins, sometimes on the same strand (e.g. *E. coli* NikR), but more commonly divergently transcribed (e.g. the *E. coli* proteins CynR, AraC, SoxR, MelR and many others). Therefore this mechanism is not just controlling the transcription factor itself, but also a gamut of associated proteins, typically working together in the same functional system.

Because so many prokaryotic genes are controlled by negative self-regulating transcription factors, it would appear that such that self-regulation is favoured by evolution. This begs the question of why it is favoured: what is the functional role of negative self-regulation in transcription systems?

The most commonly quoted answer comes from engineering principles. Negative feedback as a mechanism of control is as common in engineering as it is in biology, and therefore it has been natural to conclude that it must be playing similar roles. A simple engineering example is the thermostat, which uses negative feedback to maintain a desired temperature in a room. If the temperature is too cold, heating is switched on; if it is too warm, heating is switched off. In biology, there are many macro-physiological examples of homeostatic control using negative feedback, for example the control of blood sugar level using insulin and glucagon.

This would lead to the view that negative self regulation enables homeostatic control; the cell can use negative self regulation to maintain protein expression at a desired level [11–13]. Other authors have also demonstrated that negative feedback can shift the noise spectrum from low to high frequencies [14–16]. Control of noise would appear to be particularly important in the light of the stochasticity of protein expression observed in real cells. Because protein expression is subject to intrinsic and extrinsic noise, it is even more important to provide homeostatic control of that expression.

A very different explanation is suggested by Rosenfeld *et al.* [17]. They analyze ODE models of negative self regulation and conclude that such systems are able to substantially reduce the response time of protein production in the event of environmental change.

However, there are fundamental differences between transcription systems and the control of temperature or of relatively abundant metabolites (e.g. glucose). The number of molecules in transcription systems is necessarily small. Even if a transcription factor itself is relatively abundant, one of the most important molecules, the DNA molecule, is present in only in a small number of copies, depending on cell cycle and the proximity of the associated gene to the origin of replication. In mathematical terms, models based on ordinary differential equations (ODEs) can often be used to describe the average behaviour of a large population of cells. However, evolution acts on individual cells, and differential equation models of transcription regulation are not valid at the individual cell level. It is impossible to understand the functional role of transcription motifs without evolutionary context, and so it is vital to explore mathematical models that are valid at the single-cell level.

Therefore we have carried out a theoretical investigation of the role of negative transcription regulation on the variability of protein expression using stochastic models. The models that we analyze are similar to those studied by Thattai and van Oudenaarden [7] and Simpson *et al.* [14].

## Models of Negative Self Regulation

Thattai and van Oudenaarden analyzed a stochastic model for a negatively self-regulated gene. In this model, there are six processes: protein binding to DNA, protein-DNA complex dissociation, mRNA production, mRNA degradation, protein production and protein degradation. Of course, each of these processes themselves consists of many sub-processes, for example mRNA production includes the binding of RNA polymerase, initiation, multiple elongation steps and termination. Some authors have built more complex mathematical models that explicitly include these sub-processes [18]. The model also does not include a number of important cellular processes, notably DNA and cell replication.

The model is constructed by considering each of the possible transitions that can take place. This defines a continuous time Markov Chain with the following transitions:

- $D \mapsto D - 1; P \mapsto P - 1$  at rate  $k_{on}DP$
- $D \mapsto D + 1; P \mapsto P + 1$  at rate  $k_{off}(1 - D)$
- $M \mapsto M + 1$  at rate  $k_m D$
- $M \mapsto M - 1$  at rate  $\gamma_m M$
- $P \mapsto P + 1$  at rate  $k_p M$
- $P \mapsto P - 1$  at rate  $\gamma_p P$

$D$ ,  $M$  and  $P$  represent the numbers of DNA, mRNA and protein molecules respectively.  $k_{on}$  represents the rate of protein binding to DNA,  $k_{off}$  is the dissociation rate of the DNA-protein complex,  $k_m$  is the rate of transcription,  $k_p$  the rate of translation,  $\gamma_m$  the rate of degradation of mRNA and  $\gamma_p$  the rate of degradation of protein. (Here, we have slightly changed the notation of Thattai and Van-Oudenaarden). Thattai and Van Oudenaarden derive an elaborate term for the fano factor, also known as noise strength, which is defined as the ratio of protein variance to protein mean, in two steps. First, they make the commonly-used quasi-steady-state (QSS) approximation that the rate of binding and dissociation of the transcription factor to the DNA is faster than the dynamics of the mRNA and protein production. This leads to a simpler system in which the rate of transcription is given by a Michaelis-Menten-like hyperbolic term:

- $M \mapsto M + 1$  at rate  $\frac{k_m}{1+(P/k_d)}$

- $M \mapsto M - 1$  at rate  $\gamma_m M$
- $P \mapsto P + 1$  at rate  $k_p M$
- $P \mapsto P - 1$  at rate  $\gamma_p P$

$k_d$  is the strength of the transcription factor binding site and is defined as  $k_{off}/k_{on}$ . (We have written these equations with no Hill coefficient, although it is possible to include such a coefficient into the model at this stage).

The second step is to use a Taylor series to linearize the model about the protein steady state. It is then possible to derive the variance for the linearized model using moment equations. Since the model is now linear, no moment closure techniques are required and an exact solution can be found.

However, somewhat surprisingly, Thattai and van Oudenaarden have presented the results of their model only for weak repressors with range of  $k_d$  between  $100nM$  and weaker (their Figure 3b). In contrast, real self-regulating repressors typically have much stronger values of  $k_d$ , in the range  $0.01nM$  to  $100nM$ .

Examples of strong negative self-regulators include *E. coli* NikR, with  $k_d$  of  $0.015nM$  [19], and *E. coli* PurR, with  $k_d$  of  $0.1nM$  [20]. Other self-regulating transcription factors in the  $1nM$  to  $100nM$  range include *E. coli* ChbR with  $k_d$  of  $1nM$  [21]; KorB from the RK2 plasmid with  $k_d$  of  $9.3nM$  [22]; and *E. coli* Lrp with  $k_d$  of  $35nM$  [23]. Interactions weaker than  $100nM$  are typically regarded as non-specific.

Therefore it is difficult to draw conclusions about realistic systems from their presented results.

Simpson *et al.* use a Langevin approach to derive a frequency-dependent analysis of the same system.

They make the same QSS assumption as Thattai and Van Oudenaarden to introduce a hyperbolic term for the repression and also linearize the system about its steady state. They derive a simpler expression for the fano factor and demonstrate that the fano factor for the negatively regulated system is equal to the fano factor of the unregulated system divided by  $1 + T(0)$ , where  $T$  is the loop transmission which measures the level of resistance of the system to changes in protein level. Importantly, they also demonstrate that the frequency of noise is shifted from lower to higher frequencies in the presence of negative regulation.

### Approaches Taken in This Work

We analyze this system using three different approaches that complement and extend these important contributions. First, we analyze the model using mathematical approaches similar to those above. However, in contrast with Thattai and Van Oudenaarden, we examine the model for physiologically realistic values of  $k_d$ . By demonstrating that the system has two distinct behaviours in different parameter regions, and

observing that only one of these parameter ranges is relevant for physically stable proteins with realistic values of  $k_d$ , we are able to employ stronger approximations and derive a very simple expression for protein variance that we discuss in the light of the results of Thattai and Van Oudenaarden and Simpson *et al.*. Second, we run stochastic simulations with realistic parameters and demonstrate that the results of mathematical analyzes such as ours or those of other authors are only good when the dynamics of DNA binding and dissociation are fast relative to changes in protein abundance. This allows a QSS approximation to be made that leads to the appearance of a hyperbolic, Michaelis-Menten or Monod type term for transcription of gene expression seen in the majority of models [7, 14, 17]. However, when the promoter dynamics are slower, as must be the case for stronger negative repression, as the DNA-protein complex is more stable, we demonstrate that these results cease to be valid, and we use computer simulations to establish the behaviour of the system — which is qualitatively and quantitatively different from systems with weaker negative regulation.

Third, we make use of *in silico* evolution as an approach to understanding the behaviours of the models. Real biological systems are the result of millions of years of evolution — an experiment that is impossible to repeat during a human life-span. Using *in silico* evolution, we can apply the principles of variation and natural selection to a mathematical model, and evolve “organisms” with parameters that are particularly good at solving a set task (e.g. to minimize the noise of protein expression). This allows us to study systems in an evolutionary context, and to perform experiments in evolution that are not feasible with wet biology. We use this approach to compare the effectiveness of unregulated and negatively self regulated systems at controlling protein noise, minimizing protein response time and minimizing mRNA usage. We conclude that the role of negative self regulation in transcription networks may not always be the homeostatic control of protein, especially if the regulation is strong. We put forward a new hypothesis that it may be a strategy for energy minimization that allows the production of protein with minimum access to DNA and use of ribonucleotides. We propose a set of experiments that could be carried out to verify or falsify our results.

## Results and Discussion

### Mathematical Models: No Regulation

The means and variances for the model in which there is no regulation provide a useful baseline to compare with the models with negative self regulation. In this model,  $k_{on}$  is set to zero so that only the transitions for mRNA and protein production and degradation appear. All parameters in these models are expressed



in terms of molecules per cell. To convert parameters to molar units, it is necessary to multiply parameters by  $10^{-9}$ , which corresponds to a spherical cell of diameter  $1.5\mu m$ .

Using Equation 11 (see methods), it is possible to derive differential equations for both the mean and variances of the mRNA and protein molecules (see additional file 1), denoted by  $\hat{M}$ ,  $\hat{P}$  (means) and  $\text{var}(M)$  and  $\text{var}(P)$  (variances), given by:

$$\hat{M} = \frac{k_m}{\gamma_m} \quad (1)$$

$$\text{var}(M) = \frac{k_m}{\gamma_m} \quad (2)$$

$$\hat{P} = \frac{k_m k_p}{\gamma_m \gamma_p} \quad (3)$$

$$\text{var}(P) = \hat{P} \left( 1 + \frac{k_p}{\gamma_m + \gamma_p} \right) \quad (4)$$

There are four important consequences of this model: (i) The number of mRNA molecules follows a Poisson distribution — this result is already established in classical mathematics as the model for the number of mRNA molecules is equivalent to the standard birth-death process. (ii) The variance of the number of protein molecules is proportional to the mean, so that the ratio of variance to mean (the fano-factor) is a natural measure of intrinsic noise alongside the ratio of standard deviation to mean (coefficient of variation). (iii) The variance of the number of protein molecules is at best the variance associated with Poisson noise (when  $k_p \ll \gamma_m + \gamma_p$ ); for larger values of  $k_p$ , the variability is greater than Poisson variability. (iv) For biologically realistic parameters for transcription, translation and mRNA and protein degradation, variance divided by mean would range between 8 and 500 molecules per cell.

### Mathematical Models: Negative Self Regulation

The first part of the analysis proceeds in the same way as Thattai and Van Oudenaarden by making the QSS approximation and introducing a hyperbolic term for the production of mRNA. We now observe that the reduced model has two distinct behaviours depending on whether the dynamics are occurring in a saturated or non-saturated state (for full derivation, see additional file 1). First, if  $\hat{P}_u$ , the steady state mean level of protein in the system without regulation, equal to  $k_m k_p / \gamma_m \gamma_p$ , is less than  $k_d/4$ , then the model is approximately the same as the unregulated model. This is the non-saturated state. Insufficient protein is produced for the effect of the repression to be relevant. The more interesting case is when  $\hat{P}_u > k_d/4$ . This is the saturated case; it is also the more physiological case for any stable protein because

the protein will continue to be made until it switches itself off — which can only be when the concentration exceeds the  $k_d$ . In this case, the rate of transcription is approximately  $k_m k_d / P$ . The steady state mean values for  $\hat{M}_n$  and  $\hat{P}_n$  are given by:

$$\hat{M}_n = \sqrt{\frac{k_m \gamma_p k_d}{\gamma_m k_p}} \quad (5)$$

$$\hat{P}_n = \sqrt{\frac{k_m k_p k_d}{\gamma_m \gamma_p}} \quad (6)$$

In line with other authors, we derive an expression for the variance of the model that has been linearized about its steady state. The linearization is achieved by making the substitution

$$P = \hat{P}_n \left( 1 + \left( \frac{P}{\hat{P}_n} - 1 \right) \right) \quad (7)$$

By using the Taylor expansion for  $(1 + x)^{-1}$  it then is possible to derive a linear model that approximates the nonlinear model and for which it is possible to derive analytic terms for moments (see additional file 1 for details of the mathematics). The protein variance of the linearized model is given by:

$$\text{var}(P) = \hat{P}_n^2 \left( 1 + \frac{k_p - \gamma_m}{2(\gamma_m + \gamma_p)} \right) \quad (8)$$

Although the Poisson-like nature of the noise makes the fano coefficient a natural description of variability, a more standard measure is the dimensionless coefficient of variation equal to standard deviation divided by mean. By combining Equations 6 and 8 it can be seen that the coefficient of variation is proportional to  $1/\sqrt[4]{k_d}$ :

$$\text{cv}(P) = \sqrt[4]{\frac{\gamma_m \gamma_p}{k_m k_p k_d}} \sqrt{1 + \frac{k_p - \gamma_m}{2(\gamma_m + \gamma_p)}} \quad (9)$$

This means that the coefficient of variation actually increases as the strength of the repressor increases.

An alternative controlled comparison is to vary  $k_d$  while retaining the same protein expression. One natural way of doing this is to vary the RNA polymerase promoter strength (implicitly included in  $k_m$ ) in concert with  $k_d$  so that their product is constant and consequently protein level remains constant as repression increases. This can be thought of in terms of the cell employing different strategies to control a protein to a set level, ranging from a weak promoter and weak feedback through to a strong promoter with strong feedback. In that case, it can be seen from Equation 9 that the coefficient of variation should also be independent of repressor strength.

The repressor system is able to show some improvement in repression when cooperativity is included in the model. With a Hill coefficient of  $n$ , the protein variance equation becomes:

$$\text{var}(P) = \hat{P}_n \left( 1 + \frac{k_p - n\gamma_m}{(n+1)(\gamma_m + \gamma_p)} \right) \quad (10)$$

However, the result that the fano factor is independent of repressor strength, and the equivalent result for the coefficient of variation, still hold in this system.

This result might appear to be different from that of Thattai and van Oudenaarden [7], but in fact there is no conflict between these results. The mathematical analysis is representative of a realistic parameter range in which it is possible to make stronger approximations than Thattai and van Oudenaarden and thus to derive a formula that is simpler and clearer. Thus we have demonstrated that the formula of Thattai and van Oudenaarden has the asymptotic property that variability is independent of  $k_d$  for strong repressors. When the two formulae are plotted against alongside other in this range, they give almost identical results (Figure 1).

The result is also consistent with, although slightly different from, the result of Simpson *et al.* [14]. Where the protein is fully saturated, their loop transmission term  $T$  would be equal to 0, and so the fano factor would be independent of the  $k_d$  and in fact equal to the fano factor for the unregulated system. Our analysis goes a step further and shows that even for a range where the negative regulation is effective, the noise should be independent of  $k_d$ .

### Computer Simulations

Because of the many approximations needed to obtain a closed form mathematical estimate for the variance of the protein level, we also ran computer simulations for the dependence on variance on the promoter strength and other parameters. Figure 1 shows how the fano factor and the coefficient of variation depend on repressor strength for different realistic values of  $k_d$ , using a set of realistic parameters for mRNA and protein stability (variations in these parameters are explored in Figure 2). The model is controlled both by holding all other parameters constant, or by adjusting the value of  $k_m$  so that the expected protein level is constant.

The figure demonstrates a number of important points. First, for weaker repressors, with  $k_d > 1nM$ , the expressions that we derive is consistent with both the simulated data, in that the fano factor and the coefficient of variation are approximately independent of  $k_d$  when protein level is controlled. Second, for all parameter ranges, there is very close fit between our simple expressions for protein noise and the more

complicated expression of Thattai and Van Oudenaarden, with the exception of very weak repressors in Figure 1(d) where their expression fits the data slightly better. Most importantly, however, is that as repressor strength increases, with  $k_d < 1nM$ , both the fano factor and the coefficient of variation increase to much higher levels than those predicted by linearized QSS. For the parameters used in Figure 1, when  $k_d < 0.1nM$ , both the fano factor and coefficient of variation have risen to levels greater than those of the equivalent unregulated model.

In Figure 2 we demonstrate that the increase in noise for strong repressors is qualitatively (although not quantitatively) independent of choice of parameters, by showing graphs of fano factor against  $k_d$  for less stable protein (Fig 2a), more stable protein (Fig 2b), more stable mRNA (Fig2c) and the inclusion of a cooperativity (Hill) coefficient of 2 (Fig 2d). Although the graphs show the same qualitative behaviour, one important quantitative difference is the value of  $k_d$  for which the repressed system becomes more noisy than the unregulated system. With a very stable protein (13 hours) the negative regulator is always less noisy for realistic values of  $k_d$ ; however, this parameter is difficult to interpret for exponentially growing cells in which protein turn-over would be limited by the dilution rate and so is likely to be of relevance only in stationary phase. Interestingly, when cooperativity of protein binding is included, the protein abundance is noisier for weaker values of  $k_d$  than without cooperativity. Very similar behaviours are observed for other realistic parameter values (see additional file 1).

Figure 3 shows a time course for part of two simulations, one with an average  $k_d$  of  $1nM$  and the other with a strong  $k_d$  of  $0.01nM$ . In order to compare the two behaviours on the same axes, we have controlled the two simulations by adjusting  $k_m$  so that the expected protein level is the same. In both cases, protein production is in bursts, coincident with the synthesis of mRNA when the repressor dissociates from the DNA. But while with average  $k_d$ , protein abundance is adjusted reasonably quickly, when the  $k_d$  is strong, the bursts are slow and irregular. With these particular parameters, the bursts are quite slow relative to the cell cycle time, and therefore it is likely that DNA replication and cell division will interact significantly with protein synthesis and contribute extrinsic noise. These results also appear to contrast with those of Simpson *et al.* in that strong negative repression appears to shift the noise to lower frequencies rather than higher ones.

### **In Silico Evolution**

Four different *in silico* evolution experiments were performed: the first to minimize the standard deviation of protein expression; the second to minimize the rise time to half the steady state of protein expression;

the third to minimize the rise time to the steady state protein expression; and the fourth to minimize mRNA abundance. In each experiment, the model with no regulation was compared with the model with regulation. The results are shown in Figure 4.

In Figure 4a, it can be seen that the best evolved standard deviation of the unregulated system is 10.4 molecules per cell; the minimum standard deviation of the system with negative regulation is 8.2 molecules per cell. This represents a 21% decrease in standard deviation of protein expression. Observe that the Poisson noise level for 100 copies of the protein would be 10 molecules per cell; the unregulated system has evolved to match the Poisson noise, while the regulated system has evolved a noise level slightly below that. In Figure 4b it can be seen that the best evolved rise time of the unregulated system to half its steady state level is 16.9s; the best evolved rise time of the regulated system is 13.9s. This represents an 18% decrease in rise time for the self regulated system. In Figure 4c it can be seen that the best evolved rise time of the unregulated system to its steady state level is 42.6s; the best evolved rise time of the regulated system is 20.0s. This represents an 53% decrease in rise time for the self regulated system. In Figure 4d, it can be seen that the best evolved mRNA level of the unregulated system is a time average of 0.069 molecules per cell; the best evolved mRNA level of the system with negative regulation is 0.020 molecules per cell. This represents a 71% decrease in mRNA usage to 29% of the level of the unregulated system. Therefore there is a clear hierarchy of improvements of these three goals: modest improvements in the reduction of noise and rise-time to 50% steady-state level, good improvements in rise time to steady state level, and very substantial improvements in mRNA usage.

## Discussion

The number of molecules of a protein in a single cell varies in time, resulting from stochasticity in the processes of transcription, translation and degradation. This variability extends also to variability between individual cells in a population. This variation can be seen in experiments that track gene and protein expression in single cells [1, 3] and in experiments that track protein expression in a population of cells [2]. Many transcription systems include genes that negatively regulate their own expression [10]. It has been proposed that the role of such negative regulation is to ensure homeostatic control of the protein products a view deriving from our understanding and use of negative feedback in engineering [11, 12]. In this study, however, we demonstrate the reverse: strong negative control of gene expression results in increased variability.

We have carried out an investigation of how the level of variation depends on the rates of key processes:

transcription, translation, degradation of mRNA and protein, and the binding and dissociation of the transcription factor to the DNA regulatory site. The analysis has been on three levels: mathematical models, computer simulations and *in silico* evolution.

Analytic solution for the linearized version of the negative feedback loop suggests that the fano factor should be independent of repressor strength and that the coefficient of variability should increase with repressor strength, or also be independent of repressor strength if the model is controlled for constant protein level. While these results are themselves surprising, they are in fact implicit in the results of other authors [7, 14], and our derivation provides a simple mathematical formula to capture the behaviour in terms of other parameters.

More importantly, however, we show that these results are only applicable when the DNA-protein binding dynamics are fast relative to the dynamics of mRNA and protein production and degradation. Instead, computer simulations reveal that as repressor strength increases, so too does the noise. While the noise for weak repressors (with  $k_d$  much bigger than  $1nM$ ) approaches the asymptotic limit derived from the mathematical analysis, the fano factor and coefficient of variation of strong repressors (with  $k_d$  much less than  $1nM$ ) are very much higher than predicted by theory. For very strong repressors (with  $k_d$  less than  $0.1nM$ ), the noise can be greater than in an unregulated system. This is likely to be physiologically important, as many auto-regulating operators have  $k_d$ s in this range, for example *E. coli* NikR, with  $k_d$  of  $0.015nM$  [19], or *E. coli* PurR, with  $k_d$  of  $0.1nM$  [20].

This result would appear to be in contrast with previous theoretical results [7], despite analyzing the same model. However, there are two important differences between the two analyses: first, that work investigated values of  $k_d$  ranging from  $100nM$  and weaker (their Figure 3), which is weaker than many physiological repressors that operate in the range between  $0.01nM$  and  $100nM$ , and also outside the range in which the approximations used in the mathematical derivation cease to be valid. Furthermore, by observing the fact that for stable proteins the protein abundance will be in excess of the  $k_d$ , we are able to derive a novel, clearer and simpler numerical expression for protein variability, and which gives essentially the same results.

The reason for this increase in noise is that when the repression is strong, protein molecules are produced in bursts, which result in highly variable protein levels [24, 25]. These bursts are happening on a timescale related to the repressor strength that is slow relative to the dynamics of protein production and degradation. As a result, the standard QSS approximation, based on a separation of timescales between transcription factor binding events and protein production and degradation events, that leads to a

hyperbolic term is not valid. Therefore the mathematical derivations that depend on it (including our own) cease to describe the behaviour of the model. This result is consistent with other situations in which stochastic models can behave differently from classical chemical kinetics, for example when stochastic switching can occur [26].

This also explains the apparent discrepancy between our results and those of Simpson *et al.*. They have shown, both with theory and experiments [15], that negative repression shifts the noise from low frequency to high frequency; our simulations have suggested that for some realistically strong values of  $k_d$ , the noise increases and appears to be shifted to lower frequencies. However, the result of Simpson *et al.* is also derived using the QSS approximation that we have shown is valid for weaker repressors ( $k_d > 1nM$ ) but not for stronger repressors ( $k_d < 1nM$ ). Moreover, their experimental results were obtained using a TetR system that has a  $k_d$  of  $5.6nM$  [27], which is in the range for which we would expect the QSS approximation to hold. Thus our findings are consistent with these results; we would predict that a similar experiment carried out with a stronger repressor (e.g. NikR or PurR) would give a quite different result. In fact, the behaviour we observe in models of strong repressors is consistent with the behaviour that the same authors derive for the open loop circuit that can be dominated by operator noise [28].

Another interesting point to emerge from this analysis is that for strong repressors, the timescale of fluctuations is slow relative to the rate of DNA or cell replication. The models that we have analyzed do not include terms for DNA or cell replication. Therefore it would appear likely that the interaction between protein production and replication may be quite strong. We would expect two consequences. Firstly, DNA replication may add significant extrinsic variability in these systems. Depending on position in the genome relative to the origin of replication, any given gene may be present in one, two or more copies at different times in the cell cycle [8]. This will influence both the mean and variance of protein expression, and a far more involved analysis would be necessary to evaluate the contributions of this effect to protein variability. Second, there is the capacity for epigenetic inheritance of protein abundance in these systems; this may be of benefit in some biological situations, for example stress responses.

A further interesting comparison is with the work of Kepler and Elston [6]. They apply a slow timescale approximation to a self-activating system and are able to derive an expression for the steady state probability density of protein abundance. This allows them to show that a self-activator can demonstrate bistable dynamics with a bimodal distribution of protein abundance. A similar approach might be successful for deriving an analytic expression for the noise in a negative self regulator that would be valid for physiologically strong repressors where the normal QSS cannot be used. However, the details would

necessarily be different, as Kepler and Elston derive their result by considering the limit as average protein level tends to infinity — a procedure that may not be realistic when considering a repressor system. Therefore it appears that negative self regulation may not lead to better control of protein variability. And yet negative self regulation is favoured by evolution; if it is not playing an important role in regulating protein variability, then it must be performing a different function. One proposal is that negative regulation can speed up the time scale of the response of protein production to an environmental change [17]. We put forward a new proposal that strong negative regulation may provide a mechanism to produce protein for minimum use of resource — in particular mRNA usage and access to DNA. In the final part of our studies, we ran *in silico* evolution experiments to see how well different configurations of transcription control can adapt to different tasks. This allowed us to compare the effectiveness of negative self regulation at regulating noise, at mediating a rapid response time and to minimizing mRNA usage. We compared the negative feedback loop to the unregulated system. The negative feedback system only enjoyed modest improvements of about 20% over the unregulated system when minimizing the level of protein noise. Some improvement is to be expected as the system has five parameters as opposed to four, and this is supported by the analytic results. Better improvements of about 37% were seen in reducing response times, supporting the hypothesis of Rosenfeld *et al.* [17]. However, far more substantial improvements are seen when the systems are adapting to minimize the average mRNA levels, with 73% improvement with negative feedback as compared with the unregulated system.

Our hypothesis about minimizing mRNA usage makes sense in the context of plasmids, where many successful plasmids are of minimal burden to their hosts. It is common for the central regulators of plasmids to negatively regulate their own transcription, for example the KorB regulator in RK2 [29] or the  $\omega$  regulator in pSM19035 [30]. Such mechanisms allow for a burst of gene expression on entry into a new cell to allow the plasmid replication, segregation and conjugation apparatus to become established, and then for the plasmid to switch off their genes and thus have minimal impact on the host once sufficient copy number is achieved. We propose that a similar idea holds for many constitutive bacterial genes too. Our hypothesis is also complementary with the rapid response hypothesis, in that rapid protein production following environmental change (e.g. cell division) is entirely consistent with subsequent shutting down of protein production. It is particularly interesting that using Rosenfeld *et al.*'s definition of the rise time to 50% of the steady state value produces very little improvement, while using a definition of rise time to the steady state value produces much better improvements. The reason for this is likely to derive from the fact that Rosenfeld *et al.* make use of ODE models, while we are using stochastic simulations. The ODE



models cannot achieve the steady state value so it is necessary to use a fixed proportion of the steady state, and 50% is a natural proportion (analogous to  $k_m$ ). With stochastic models, however, the steady state is a mean about which the protein abundance varies, and so is always achieved. Thus it is quite reasonable to define the rise time as first passage time to the mean. It is possible that the improved performance of the negative self-regulator in reducing first passage time to mean value as opposed to half mean value is because the negative regulator can take advantage of the stochasticity of the system and allow an "overshoot" of protein production which can then be attenuated by negative regulation. This overshoot advantage is not seen with the half mean first passage time. It would appear that that the mean first passage time associated with stochastic models is a different property from rise time of the mean associated with ODE models. Moreover, Rosenfeld *et al.*'s study makes use of ODEs using the standard QSS and hyperbolic terms. Thus further investigation of rise times in stochastic, single-cell models is warranted. Furthermore, although strong negative repression might not be particularly effective at noise control, it is possible that cells might have evolved more elaborate mechanisms for noise control. These might include both negative and positive regulation, multiple transcription factors, or regulation at the mRNA or protein levels. An example of elaborate regulation that has been found to minimize noise is in the *E. coli* heat shock response [31]. It is likely that other mechanisms exist; *in silico* evolution techniques could prove valuable in identifying such potential mechanisms.

### *Experimental Validation*

The results of these analyses lend themselves quite naturally to experimental validation or falsification. The experiment would involve constructing a low-copy-number plasmid with a suitable negatively self-regulating transcription factor and its binding site, with very strong  $k_d$ , controlling also the expression of GFP. A series of mutants would be made with mutations to the operator sequence so as to produce a series of weaker repression circuits. Bacteria lacking the aforementioned transcription factor can be transformed with this plasmid and grown. The mean and variance of GFP fluorescence in a population of cells would then be measured using flow cytometry. This procedure would be repeated for each of the mutant plasmid vectors. The DNA-protein  $k_d$  values could also be measured for each operator sequence using biophysical techniques (accepting that such *in vitro* measurements may not always represent an accurate *in vivo*  $k_d$ ). From this data, it would be possible to plot protein variability against  $k_d$  and thus determine whether or not variability really does increase within a physiological range of  $k_d$  values.

## Conclusions

There are two important conclusions of these analyses. The first is that the standard quasi-steady-state approximation, in which it is assumed that the dynamics of protein-DNA binding are faster than the dynamics of mRNA and protein synthesis and degradation, and which gives rise to the hyperbolic Michaelis-Menten-like terms typically used in differential equation models of gene expression, is not realistic for a range of biologically important parameter values. New mathematical approximations will have to be derived in order to seek accurate closed-form equations for protein noise in such systems. Great care must be taken not to use hyperbolic terms in mathematical models of single cells, but instead to fully implement the protein-DNA interaction dynamics.

Second, strong negative self regulation can actually increase the noise of protein expression and is unlikely to be a mechanism for control of protein noise. Instead, we postulate that it is a mechanism to minimize the amount of mRNA needed to produce protein at a given level. There are two reasons why it might be advantageous for a cell to do this. First, it reduces the need to access the DNA molecule. A bacterium such as *E. coli* packs 4.5MBp of DNA into a  $1.5\mu\text{m}$  cell: stretched out, the DNA molecule is about 1000 times the length of the cell, and it may be advantageous to minimize the extent to which the DNA needs to unfold and refold. The second explanation is that the energy required to manufacture mRNA. The synthesis of an mRNA molecule costs a cell approximately 1.5 times the energy of the correspondent protein molecule [32], and that does not take into account the energy invested in the nucleotides themselves. A variant of a gene that is able to produce the same amount of protein but using less mRNA may have significant evolutionary advantage over a version using more mRNA, both by saving on energy of mRNA synthesis, and by saving on ribonucleotide use. The negative feedback loop is a highly effective way of achieving this, and so is favoured by evolution.

Thus we see that negative self repression is not a single mechanism for homeostasis but is a mechanism that might be performing quite different functions depending upon the strength of the repressor. Weaker negative self-repressors can reduce noise and shift noise from low to high frequencies; stronger negative self-repressors increase noise, particularly at low frequencies. Negative self-repressors can speed up response times. And negative self-repressors, particularly strong self-repressors, allow protein production for minimal energy cost.

## Methods

### Derivation of Means and Variances

In order to derive equations for the means and variances of mRNA and protein expression, we make use of the equation:

$$\frac{d\langle f(\mathbf{x}) \rangle}{dt} = \left\langle \sum_{\text{events}} \text{rate of event} \times \text{change to } f(\mathbf{x}) \text{ due to event} \right\rangle \quad (11)$$

In this equation,  $\mathbf{x}$  represents a vector of variables and  $f(\mathbf{x})$  represents a function of  $\mathbf{x}$ . It is straightforward to derive this equation from the Master Equation [33].

### Stochastic Simulations

Stochastic simulations of the systems studied have been carried out using the Gibson-Bruck algorithm [34], which has been implemented into our own Java-based simulator (source code is available from the authors on request). The simulations for Figure 1 were run with 1,000,000 protein production or degradation events. Parameters for these simulations were chosen for no particular gene but with a wide range of realistic parameters to explore the general behaviour of the model.

### In Silico Evolution

*In silico* evolution has been carried out as a genetic algorithm in a relatively standard way. An initial population of 25 individuals is generated; each individual has randomly chosen parameters. The no regulation model has four parameters and the negative regulation model has a fifth parameter,  $k_{off}$ :  $k_{on}$  is fixed at the physiological diffusion-limited level of 0.01 molecules per cell per second.

Initial parameters are chosen at random from the log (to base 10) of a normal distribution with mean of  $-2$  and standard deviation 1 for all parameters (with appropriate units). At each round of the simulation, the best 25 individuals from the previous generation are selected (elitism); a further 25 individuals are generated by mutations, taking each individual and adding the log (to base 10) of normal noise with mean 0 and standard deviation 0.2 to each parameter (with appropriate units). The use of log normal mutations allows the model evolution to explore parameters at all orders of magnitude. A further 25 individuals are generated by recombination: two parents (elite or mutated) are chosen at random, and each parameter of the offspring is selected at random from one of the parents. The elite strategy was found to be superior to a Boltzmann random selection strategy (unpublished results) with faster convergence to similar limiting fitnesses.

In all individuals, we have constrained the rate of production of mRNA and the rate of production of protein to be within physiological realistic range. Thus mRNA and protein synthesis are each constrained to have maximum values of  $1.0s^{-1}$ . We have also constrained the rate of protein degradation to be no slower than a realistic cell cycle time; thus  $\gamma_p$  can be no smaller than  $0.0001995s^{-1}$  which corresponds to 83.5 minutes.

The fitness of each of the 75 individuals is determined on the basis of a simulations of the model with 100,000 protein production and degradation events, and initial mRNA and protein levels of 0. The models were required to have a mean protein abundance of 100 molecules. The rise time is given as the first passage time either to half the required mean number of molecules, i.e. 50 molecules (to be consistent with the methodology of Rosenfeld et al.) or to the required mean number of molecules.

The fitness functions used are:

1. Deviance from protein mean + protein standard deviation
2. Deviance from protein mean + protein rise time
3. Deviance from protein mean + 10 times mRNA level

The scaling factor on the mRNA level is to ensure that the fitness functions weight their components comparably. In Figure 4 it can be seen that protein standard deviation is approximately 10 molecules per cell; rise times are approximately 20s; mRNA abundances are approximately 0.05 molecules per cell.

Therefore in fact the fitness function is quite conservative for mRNA abundance and greater improvements could be seen with a higher weighting.

At each generation, the best 25 models are selected for the next generation. The genetic algorithms were run for 30 generations and each repeated 20 times. The best evolved models for each scenario were then simulated for 1,000,000 events to determine means and standard deviations. The rise times for the best models were evaluated as the average of 20 repeats.

### **Authors' contributions**

Dov Stekel devised the work, carried out the mathematical research and wrote the paper. Dafyd Jenkins implemented the stochastic simulation and evolutionary computation environments. Both authors have read and approved the final version of the manuscript.

## Acknowledgments

The authors thank Alun Lloyd, Elsi Coates, Eva Hyde, Warren Smith and Michael Yudkin for helpful discussion and Lewis Bingle and Pete Lund for comments on the manuscript. DJJ is funded by BBSRC Strategic Research Studentship BBS/S/H/2005/12006. Computer simulations were carried out using the Birmingham Biosciences High Performance Compute Cluster, funded by BBSRC REI grant BB/D524624/1.

## References

1. Elowitz M, Levine A, Siggia E, Swain P: **Stochastic gene expression in a single cell.** *Science* 2002, **297**:1129 – 1131.
2. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O’Shea E, Pilpel Y, Barkai N: **Noise in protein expression scales with natural protein abundance.** *Nature Genetics* 2006, **38**:536 – 643.
3. Cai L, Friedman N, Xie X: **Stochastic protein expression in individual cells at the single molecule level.** *Nature* 2006, **440**:358 – 362.
4. Dublanche Y, Michalodimitrakis K, Kümmerer N, Foglierini M, Serrano L: **Noise in transcription negative feedback loops: simulation and experimental analysis.** *Mol. Sys. Biol.* 2006, **msb4100081**:E1.
5. Hooshangi S, Weiss R: **The effect of negative feedback on noise propagation in transcriptional gene networks.** *CHAOS* 2006, **16**:026108.
6. Kepler T, Elston T: **Stochasticity in transcriptional regulation: origins, consequences and mathematical representations.** *Biophysical Journal* 2001, **81**:3116 – 3136.
7. Thattai M, van Oudenaarden A: **Intrinsic noise in gene regulatory networks.** *Proc. Natl. Acad. Sci. USA* 2001, **98**:8614 – 8619.
8. Swain P, Elowitz M, Siggia E: **Intrinsic and extrinsic contributions to stochasticity in gene expression.** *Proc. Natl. Acad. Sci. USA* 2002, **99**:12795 – 12800.
9. Paulsson J: **Summing up the noise in gene networks.** *Nature* 2004, **427**:415 – 418.
10. Keseler I, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, Paulsen I, Peralta-Gil M, Karp P: **EcoCyc: a comprehensive database resource for Escherichia coli.** *Nucleic Acids Res.* 2005, **33**:D334 – 337.
11. Becskei A, Serrano L: **Engineering stability in gene networks by autoregulation.** *Nature* 2000, **405**:590 – 593.
12. Wall M, Hlavacek W, Savageau M: **Design principles for regulator gene expression in a repressible gene circuit.** *J. Mol. Biol.* 2003, **332**:861 – 876.
13. Shinar G, Dekel E, Thustly T, Alon U: **Rules for biological regulation based on error minimization.** *Proc. Natl. Acad. Sci. USA* 2006, **103**:3999 – 4004.
14. Simpson M, Cox C, Sayler G: **Frequency domain analysis of noise in autoregulated gene circuits.** *Proc. Natl. Acad. Sci. USA* 2003, **100**:4551 – 4556.
15. Austin D, Allen M, McCollum J, Dar R, Wilgus J, Sayler G, Samatova N, Cox C, Simpson M: **Gene network shaping of inherent noise spectra.** *Nature* 2006, **439**:608 – 611.
16. Cox C, McCollum J, Austin D, Allen M, Dar R, Simpson M: **Frequency domain analysis of noise in simple gene circuits.** *CHAOS* 2006, **16**:026102.
17. Rosenfeld N, Elowitz M, Alon U: **Negative autoregulation speeds the response times of transcription networks.** *J. Mol. Biol.* 2002, **323**:785 – 793.
18. Kierzek A, Zaim L, Zielenkiewicz P: **The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression.** *J. Biol. Chem.* 2001, **276**:8165 – 8172.

19. Chivers P, Sauer R: **Regulation of high affinity nickel uptake in bacteria.** *Journal of Biological Chemistry* 2000, **275**:19735 – 19741.
20. Rolfes R, Zalkin H: **Autoregulation of Escherichia coli purR requires two control sites downstream of the promoter.** *J. Bacteriol.* 1990, **172**:5758 – 5766.
21. Plumbridge J, Pellegrini O: **Expression of chitobiose operon of Escherichia coli is regulated by three transcription factors: NagC, ChbR and CAP.** *Mol. Microbiol.* 2004, **52**:437 – 449.
22. Kostelidou K, Thomas C: **The hierarchy of KorB binding at its 12 binding sites on the broad-host-range plasmid RK2 and modulation of this binding by IncC1 protein.** *J. Mol. Biol.* 2000, **295**:411 – 422.
23. Wang Q, Wu J, Friedberg D, Platko J, Calvo J: **Regulation of the Escherichia coli lrp gene.** *J. Bacteriol.* 1994, **176**:1831 – 1839.
24. Ozdubak E, Thattai M, Kurster I, Grossman A n, van Oudenaarden A: **Regulation of noise in the expression of a single gene.** *Nature Genetics* 2002, **31**:69 – 73.
25. Koern M, Elston T, Blake W, Collins J: **Stochasticity in gene expression: from theories to phenotypes.** *Nature Reviews Genetics* 2005, **6**:451 – 464.
26. Samoilov M, Arkin A: **Deviant effects in molecular reaction pathways.** *Nature Biotech.* 2006, **24**:1235 – 1240.
27. Reichheld S, Davidson A: **Two-way interdomain signal transduction in tetracycline repressor.** *J. Mol. Biol.* 2006, **361**:382 – 389.
28. Simpson M, Cox C, Sayler G: **Frequency domain chemical Langevin analysis of stochasticity in gene transcription regulation.** *J. Theor. Biol.* 2004, **229**:383 – 394.
29. Bingle L, Thomas C: **Regulatory circuits for plasmid survival.** *Current Opinion in Microbiology* 2001, **4**:194 – 200.
30. de la Hoz A, Ayora S, Sitkiewicz I, Fernandez S, Pankiewicz R, Alonso J, Ceglowski P: **Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator.** *Proc. Natl. Acad. Sci. USA* 2000, **97**:728 – 733.
31. El-Samad H, Kurata H, Doyle J, Gross C, Khammash M: **Surviving heat-shock: control strategies for for robustness and performance.** *Proc. Natl. Acad. Sci. USA* 2005, **102**:2736 – 2741.
32. Neidhardt F, Ingraham J, Schaechter M: *Physiology of the Bacterial Cell: a Molecular Approach.* Sunderland, Massachusetts: Sinauer 1990.
33. Keeling M: **Multiplicative moments and measures of persistence in ecology.** *J. Theor. Biol.* 2000, **205**:269 – 281.
34. Gibson M, Bruck J: **Efficient exact stochastic simulation of chemical systems with many species and many channels.** *J. Phys. Chem. A* 2000, **104**:1876 – 1889.

## Figures

### Figure 1 – Dependence of Protein Noise on $k_d$

Dependence of fano factor (variance of number of protein molecules per cell divided by mean number of protein molecules per cell) and coefficient of variation (standard deviation divided by mean) on  $k_d$  of the DNA binding site, for physiological values of  $k_d$  ranging between  $0.01nM$  and  $100nM$ . In all panels,  $k_p = 0.1s^{-1}$ ,  $\gamma_m = 5 \times 10^{-3}s^{-1}$  and  $\gamma_p = 2 \times 10^{-4}s^{-1}$ . In the top two panels (a) and (b),  $k_d$  is varied, and the model is controlled by holding all other parameters fixed. In the bottom two panels (c) and (d),  $k_d$  is varied, and the model is controlled to keep a constant protein abundance of 100 molecules per cell by also

varying  $k_m$ . In the left-hand panels (a) and (c), the fano factor is plotted as a noise measure; in the right-hand panels (b) and (d), the coefficient of variation is plotted. In (a) and (b)  $k_m = 0.1s^{-1}$ . In (c) and (d),  $k_m$  is varied along with  $k_d$  so that mean protein abundance is held constant. Each of the data points is the measure of noise from a stochastic simulation of the model. The black lines show the respective noise measure as derived by our mathematical analysis; the blue lines, where distinguishable from the black lines, show the noise measure as derived by Thattai and van Oudenaarden; the red lines show the noise measure for the equivalent unregulated model. In all panels it can be seen that (i) our noise measures are very close to the expression derived by Thattai and Van Oudenaarden; (ii) the simulations match the noise level for weak values of  $k_d$  greater than  $1nM$ ; (iii) for strong values of  $k_d$  less than  $1nM$ , the level of noise increases with repressor strength, and is very much greater than predicted by the linearized QSS model. (a), (c) and (d) all show that the noise level is predicted to be lower in the regulated system than the equivalent unregulated system. However, the stochastic simulations demonstrate that for strong values of  $k_d$ , the noise of the regulated system can be greater than that of the unregulated system. In (b), the red line would appear to indicate that the unregulated system is consistently less noisy than the regulated system. However, in this panel, because all parameters are held constant, the protein abundance is much higher than in the unregulated system than the regulated system, and because of the Poisson-like nature of the noise (variance proportional to mean), the coefficient of variation is necessarily lower. In (d) it can be seen that when the unregulated system is adjusted so that protein levels are the same, a consistent pattern of behaviour is observed.

### Figure 2 – Dependence of $k_d$ Noise Response on Other Parameters

The qualitative nature of our results are independent of choice of parameter, although the quantitative measures do change. All panels show the fano factor (variance divided by mean) for varying values of  $k_d$  holding all other parameters constant (a very similar pattern would be seen using coefficient of variation and/or controlling for constant protein expression). Importantly, the mathematical formulae only fit the simulations for weak  $k_d$ , and the noise increases for strong repressors. (a) A less stable protein with  $\gamma_p = 2 \times 10^{-3}s^{-1}$  exhibits similar behaviour; all other parameters are as in Figure 1(a). (b) A very stable protein with  $\gamma_p = 2 \times 10^{-5}s^{-1}$  exhibits the same behaviour, except that with these parameters, the increase in fano factor never matches the noise of the unregulated system. (c) A more stable mRNA with  $\gamma_m = 10^{-3}s^{-1}$ . (d) A cooperatively binding protein with Hill coefficient of 2 also shows the same pattern, but this time the noise is greater than the unregulated system for weaker repressors than in the

non-cooperative case.

### Figure 3 – Bursty Protein Production

Two simulations contrasting the behaviour of protein abundance for a moderate repressor with  $k_d$  of  $1nM$  (solid line) and a strong repressor with  $k_d$  of  $0.01nM$  (dashed line). For the  $1nM$  repressor, all other parameters are as in Figure 1(a). For the strong repressor,  $k_m$  has been adjusted so that the both models have an average protein abundance of 100 molecules per cell — this allows both simulations to be plotted on the same axes so that the noise can be easily compared. In both cases, protein is produced in bursts. With the moderate repressor, the bursts are short and protein level is being continuously adjusted about the mean. With a strong repressor (low  $k_d$ ), the bursts are large and coincident with small number of times in this simulation that mRNA is synthesized. This is the source of the additional variability over and above the linearized system. It is also important to observe that the variability in protein abundance — at least for a stable protein — is slow relative to the cell cycle time. This means that extrinsic noise due to DNA and cell replication are likely to contribute very significantly to strongly auto-repressing transcription factors.

### Figure 4 — In Silico Evolution

Simulations of the best models derived from *in silico* evolution of the system without regulation and the system with negative feedback. (a) Evolution to minimize protein standard deviation. The regulated system can achieve about 21% improvement over the unregulated system, with the standard deviation reduced from 10.4 to 8.2 for a protein abundance of 100 molecules per cell. The evolved parameters for the unregulated system are:  $k_m = 1.0s^{-1}$ ;  $k_p = 0.00387s^{-1}$ ;  $\gamma_m = 0.0664s^{-1}$ ;  $\gamma_p = 0.000572s^{-1}$ . The evolved parameters for the negatively regulated system are:  $k_{off} = 0.256s^{-1}$ ;  $k_m = 0.835s^{-1}$ ;  $k_p = 0.294s^{-1}$ ;  $\gamma_m = 1.783s^{-1}$ ;  $\gamma_p = 0.00282s^{-1}$ . (b) Evolution to minimize first passage time to 50% of mean protein abundance. The regulated system only achieves an improvement of 18% over the unregulated system with the rise time (of 20 repeats of the best model) reduced from 16.9s to 13.9s. The evolved parameters for the unregulated system are:  $k_m = 1.0s^{-1}$ ;  $k_p = 1.0s^{-1}$ ;  $\gamma_m = 0.0591s^{-1}$ ;  $\gamma_p = 0.163s^{-1}$ . The evolved parameters for the negatively regulated system are:  $k_{off} = 0.0664s^{-1}$ ;  $k_m = 1.0s^{-1}$ ;  $k_p = 1.0s^{-1}$ ;  $\gamma_m = 0.0494s^{-1}$ ;  $\gamma_p = 0.0146s^{-1}$ . (c) Evolution to minimize first passage time to mean protein abundance. The regulated system achieves an improvement of 53% over the unregulated system with the rise time (of 20 repeats of the best model) reduced from 42.6s to 20.0s. The evolved parameters for the unregulated



system are:  $k_m = 0.903s^{-1}$ ;  $k_p = 1.0s^{-1}$ ;  $\gamma_m = 0.101s^{-1}$ ;  $\gamma_p = 0.0916s^{-1}$ . The evolved parameters for the negatively regulated system are:  $k_{off} = 0.000648s^{-1}$ ;  $k_m = 1.0s^{-1}$ ;  $k_p = 1.0s^{-1}$ ;  $\gamma_m = 0.00139s^{-1}$ ;  $\gamma_p = 0.0183s^{-1}$ . (d) Evolution to minimize mRNA usage. The regulated system is able to achieve a 71% improvement over the unregulated system, reducing mean mRNA levels from 0.069 to 0.020 molecules per cell to achieve a protein abundance of 100 molecules per cell. The evolved parameters for the unregulated system are:  $k_m = 0.000641s^{-1}$ ;  $k_p = 0.468s^{-1}$ ;  $\gamma_m = 0.0148s^{-1}$ ;  $\gamma_p = 0.000200s^{-1}$ . The evolved parameters for the negatively regulated system are:  $k_{off} = 0.0942s^{-1}$ ;  $k_m = 0.270s^{-1}$ ;  $k_p = 1.0s^{-1}$ ;  $\gamma_m = 1.168s^{-1}$ ;  $\gamma_p = 0.000200s^{-1}$ . Note that the systems that minimize noise and mRNA usage evolve stable proteins while the systems that minimize response times evolve more rapidly turned-over proteins.

## **Additional Files**

Additional file 1

File Format: pdf

Title: Supplementary Information

Description: Contains details of mathematical derivations and evolutionary simulations.

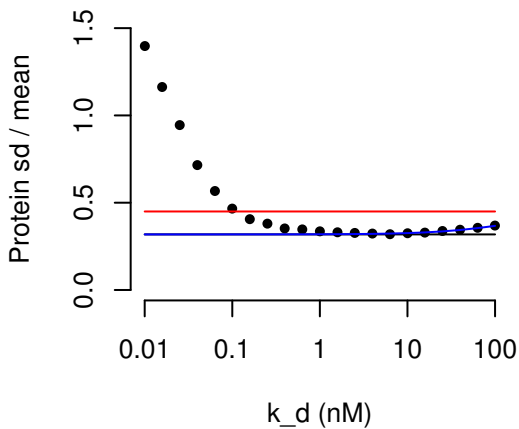
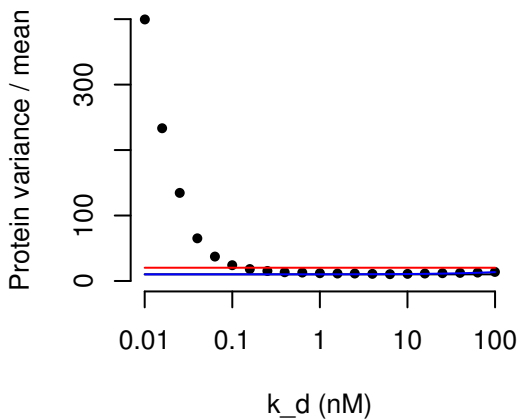
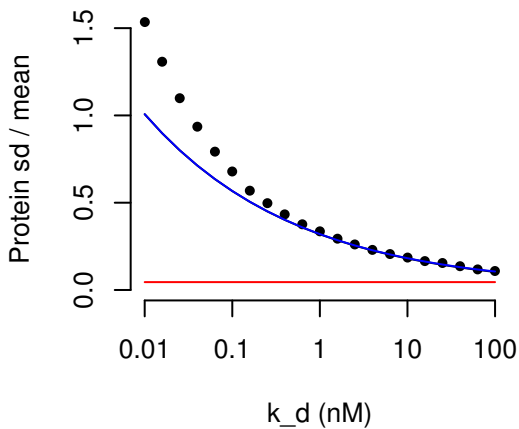
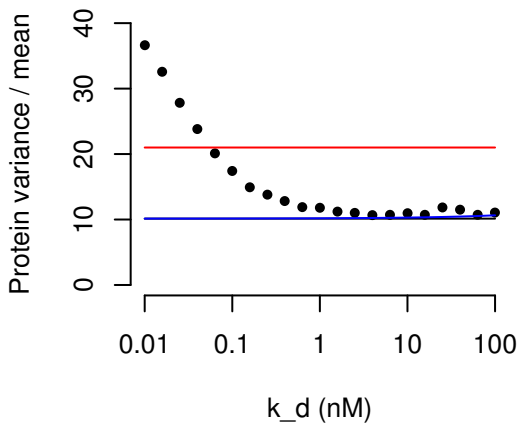


Figure 1

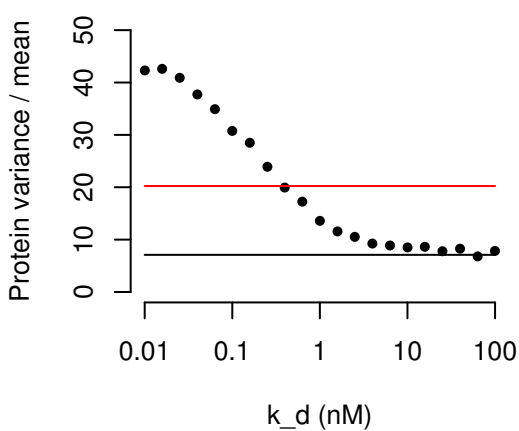
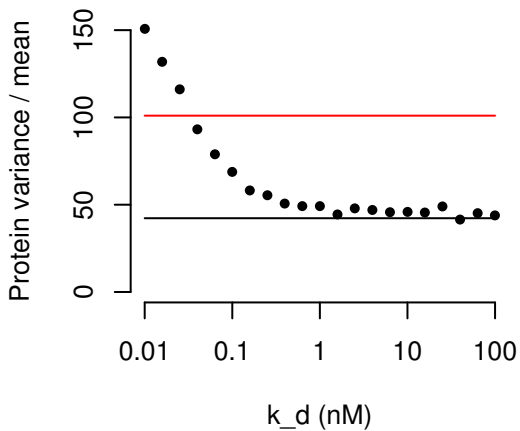
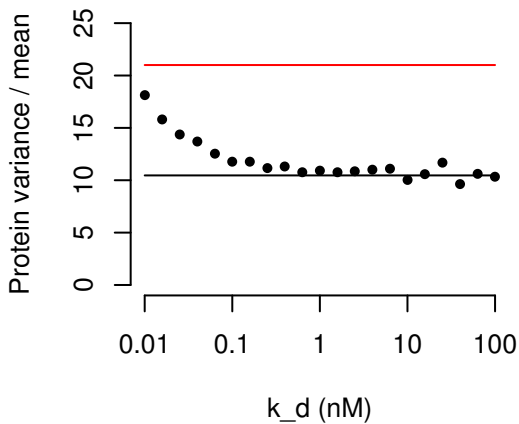
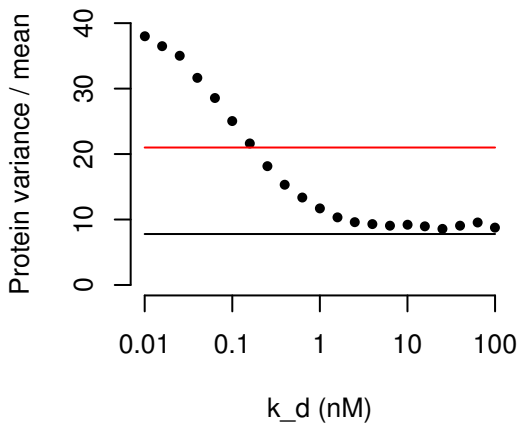


Figure 2

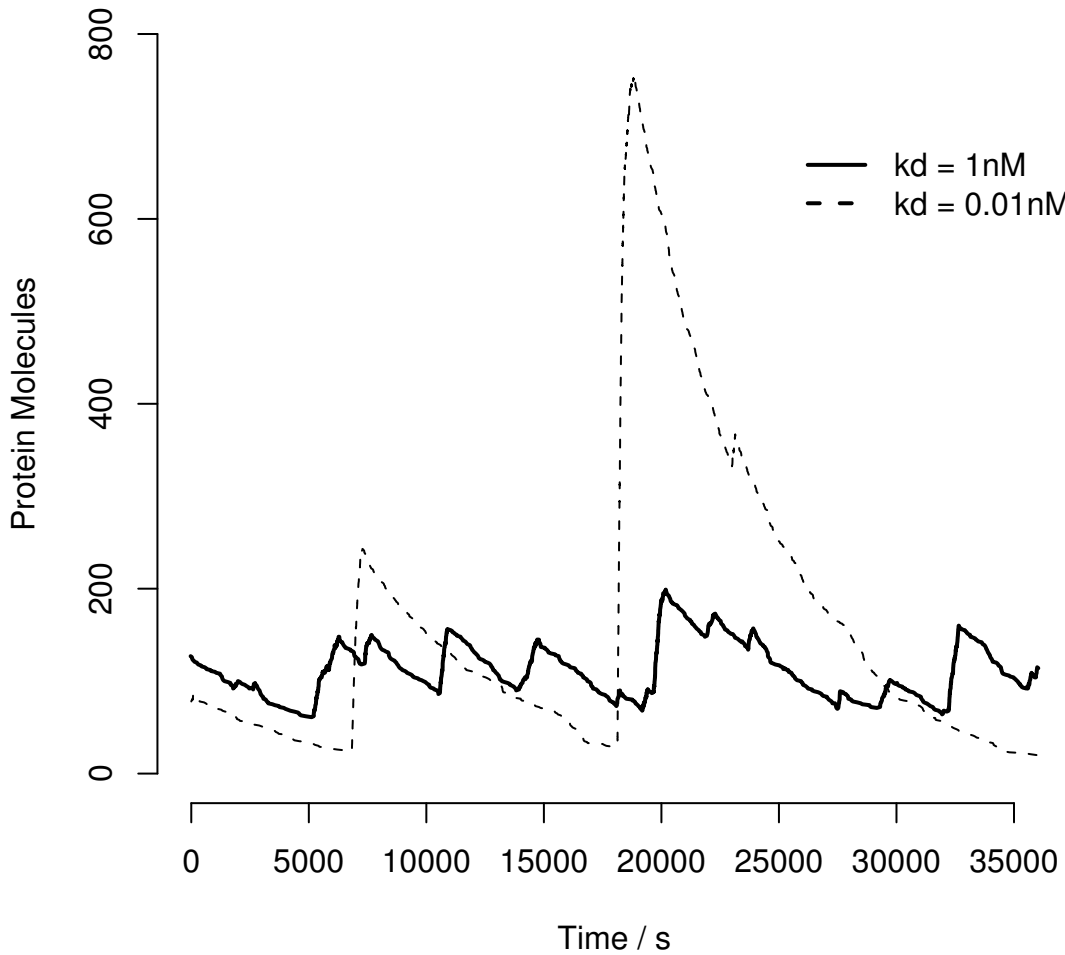


Figure 3

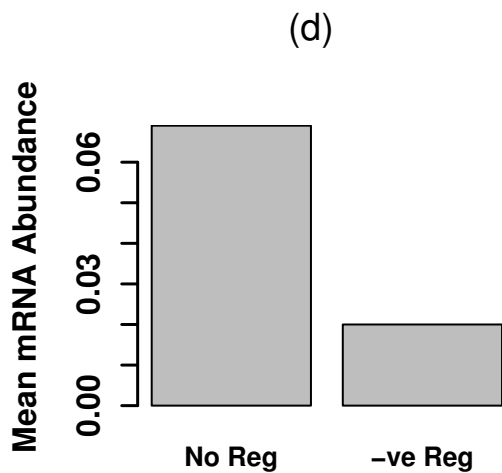
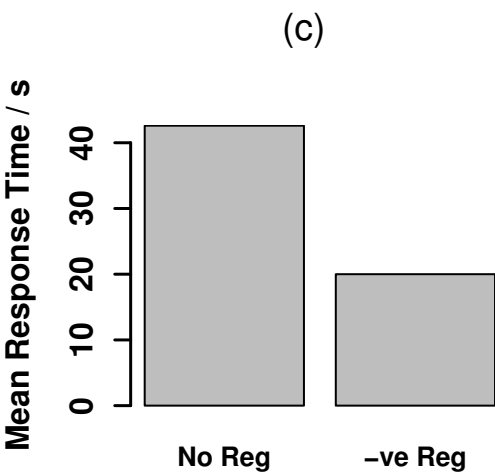
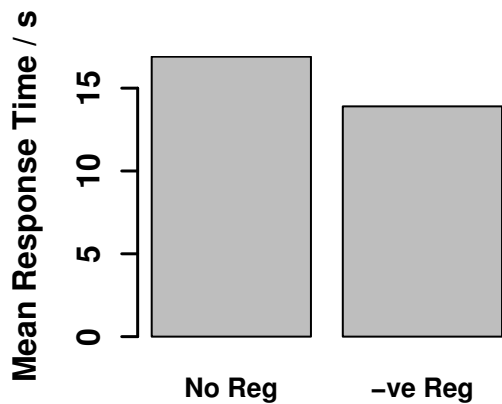
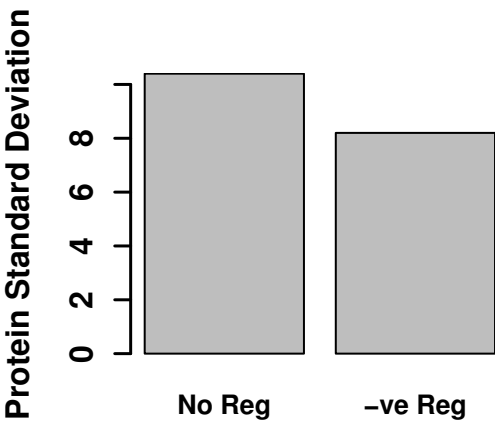


Figure 4

**Additional files provided with this submission:**

Additional file 1: additional\_file\_1.pdf, 831K

<http://www.biomedcentral.com/imedia/1319531674177564/supp1.pdf>