Interplay of Gene Expression Noise and Ultrasensitive Dynamics Affects Bacterial Operon Organization

J. Christian J. Ray[1,2], Oleg A. Igoshin[1,2]*

1 Department of Bioengineering, Rice University, Houston, Texas, United States of America, 2 Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America

Abstract

Bacterial chromosomes are organized into polycistronic cotranscribed operons, but the evolutionary pressures maintaining them are unclear. We hypothesized that operons alter gene expression noise characteristics, resulting in selection for or against maintaining operons depending on network architecture. Mathematical models for 6 functional classes of network modules showed that three classes exhibited decreased noise and 3 exhibited increased noise with same-operon cotranscription of interacting proteins. Noise reduction was often associated with a decreased chance of reaching an ultrasensitive threshold. Stochastic simulations of the *lac* operon demonstrated that the predicted effects of transcriptional coupling hold for a complex network module. We employed bioinformatic analysis to find overrepresentation of noise-minimizing operon organization compared with randomized controls. Among constitutively expressed physically interacting protein pairs, higher coupling frequencies appeared at lower expression levels, where noise effects are expected to be dominant. Our results thereby suggest an important role for gene expression noise, in many cases interacting with an ultrasensitive switch, in maintaining or selecting for operons in bacterial chromosomes.

Introduction

The organization of genes into operons is a prominent feature of bacterial chromosomes [1] that appear in some eukaryotes as well [2]. An operon is typically characterized as a promoter followed by multiple genes that are cotranscribed so that each transcription initiation event produces a polycistronic messenger RNA (mRNA) encoding multiple gene products [3]. Hypotheses explaining the emergence and maintenance of operons include proportional coregulation [4,5,6,7,8,9], horizontal transfer of intact "selfish" operons [10], emergence via gene duplication [11], coproduction of physically interacting proteins to speed their association [12,13], evolvability of co-regulation for interacting protein products [14], and reduction of intrinsic noise [15]. Current evidence favors some hypotheses more than others, but fails to indicate a definitive explanation for how operons are maintained in bacterial chromosomes [4,11,12,13].

Arising from stochasticity of individual biochemical reactions and low copy numbers of reactants per cell, intrinsic noise plays a central role in network dynamics [16,17]. In bacteria, intrinsic noise is most evident in gene expression, caused by translational bursting arising from small numbers of mRNA producing many proteins [18,19] and transcriptional bursting arising from slow activation-deactivation cycles of transcriptional activity by unknown mechanism [20,21]. Extrinsic noise, caused by uncertainties in global parameters and states including those characterizing

transcriptional and translational machinery, also contributes to overall biochemical noise [22]. Noise in protein levels is commonly characterized by coefficient of variation (CV), the normalized root-mean square deviation of the protein levels from their mean value ($CV = \sigma/\mu$, where σ is the standard deviation and μ is the mean) but other measures such as autocorrelation and covariance between concentrations of different proteins can give additional insights.

The effects of intrinsic noise on operon maintenance are not well characterized, but covariance between protein levels arising from intrinsic noise depends on transcriptional coupling (co-expression from an operon) of the corresponding genes [15]. The order of genes within an operon may also affect noise [23]. Therefore, we hypothesize that noise-related effects contribute to the evolutionary maintenance of operons. Studies of several specific systems corroborate that correlative effects of transcriptional coupling alter posttranslational dynamics [6,24,25]. However, it is still not clear how different classes of protein interactions and co-expression from an operon may interact to alter biochemical noise.

In this study we assessed these effects for different types of posttranslational interaction between gene products. In several types of interactions, the noise difference between cotranscribed and uncoupled configurations was amplified by the existence of a zero-order ultrasensitive switch [26]. We related the results to an intact naturally occurring system with simulations of cotranscribed

Author Summary

In some species, most notably bacteria, chromosomal genes are arranged into clusters called operons. In operons, the process of transcription is physically coupled: a single pass of the RNA polymerase enzyme reading that region of the chromosome simultaneously produces messenger RNA encoding multiple proteins. So far, we do not have a satisfying explanation for what evolutionary forces have maintained operons on bacterial chromosomes. We hypothesized that different types of interactions between operon-coded proteins affect how strongly operons are selected for between two genes. The proposed mechanism for this effect is that operons correlate gene expression noise, changing how it manifests in the post-translational network depending on the type of protein interaction. Mathematical models demonstrate that operons reduce noise for some types of interactions but not others. We found that operon-dependent noise reduction has an underlying dependence on surprisingly high sensitivity of the network to the ratio of proteins from each gene. Databases of genetic information show that *E. coli* has operons more frequently than random if operons reduce noise for the type of interaction various gene pairs have, but not otherwise. Our study thus provides an example of how the architecture of post-translational networks affects bacterial evolution.
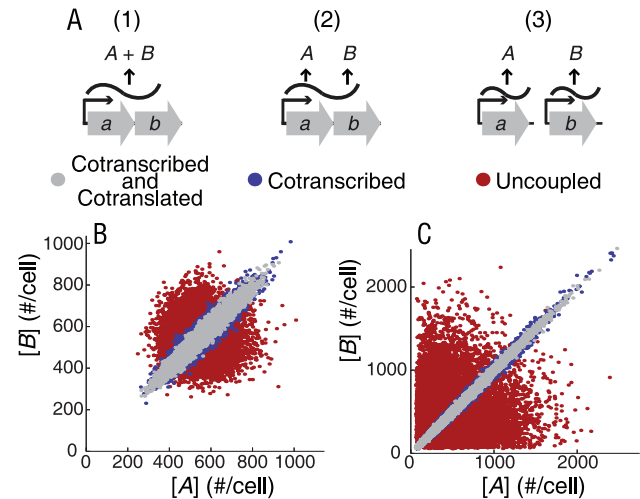
Figure 1. Types of coupling in protein production affect correlations in their fluctuations. A. Two proteins (*A* and *B*) can have coupled translation (1), coupled transcription with independent translation events (2), or uncoupled transcription (3). B, C. Scatter plots showing predicted single-cell distributions of copy numbers for proteins *A* and *B* with different coupling as indicated in models without (panel B) or with (panel C) transcriptional bursting. For simplicity, these models do not include extrinsic noise.
doi:10.1371/journal.pcbi.1002672.g001

and uncoupled configurations of the *lac* operon. To test our predictions bioinformatically, we classified naturally occurring interacting pairs of *E. coli* proteins by their type of interaction and analyzed the effect of chromosomal distance between pairs of genes with interacting products. Finally, we used single-cell protein copy number data to determine differences in operon frequencies at high and low expression levels in *E. coli* for physically interacting protein pairs.

Results

Operons correlate intrinsic protein fluctuations

Intrinsic gene expression noise is correlated in a cotranscribed two-gene configuration, but this correlation was not seen in an uncoupled configuration. Relationships between the fluctuations of two proteins can be quantitatively characterized by the covariance of concentrations for proteins A and B (σ_{AB}). Using the linear noise approximation (LNA; see Materials and Methods) [18,27,28], we calculated a normalized covariance

$$v_{AB} \equiv \frac{\sigma_{AB}}{\langle A \rangle \langle B \rangle} = \frac{\langle AB \rangle - \langle A \rangle \langle B \rangle}{\langle A \rangle \langle B \rangle} =$$

$$\begin{cases} \frac{\tau_{mRNA}}{\langle mRNA \rangle (\tau_{mRNA} + \tau_{protein})}, & \text{if co-transcribed} \\ 0, & \text{if transcriptionally uncoupled} \end{cases} \quad (1)$$

where angle brackets represent average copy-number of each molecular species, and τ_{mRNA} and $\tau_{protein}$ are the characteristic timescales of mRNA and protein decay. Increased covariance of cotranscribed genes is preserved regardless of whether the translation processes are coupled (with a single ribosome binding site for multiple genes; Figure 1A–C) and regardless of the source of intrinsic noise (from translational bursting only or from both transcriptional and translational bursting; Figure 1B–C). We hypothesized that positive covariance can increase or decrease

CV in relevant network outputs depending on the nature of the interactions between two proteins.

In addition to covariance, the effect of operons on correlation between protein copy number fluctuations can be quantified by another measure, the *degree of decorrelation* (Text S1). This measure is useful to characterize the effect of gene expression level on noise differences between cotranscribed and uncoupled proteins (Table S1), and to assess the effects of expression level on frequencies of operon occurrence in bacterial genomes.

Effects of operons on biochemical noise depends on the type of protein interaction

We surveyed databases of *E. coli* biochemical networks [29,30,31] to identify simple two-gene modules of larger networks that represent different ways that two proteins can directly or indirectly interact. The modules represent simple models of the following interactions: catalysis of subsequent steps in a linear metabolic pathway (Figure 2A), redundant catalysis of the same metabolic step (Figure 2B), catalysis of diverging reactions following a branch point in a metabolic pathway (Figure 2C), redundant transcriptional regulation of a downstream gene (Figure 2D), physical binding between two proteins (Figure 2E), and covalent modification of one protein by another (Figure 2F). The list may not be fully comprehensive, but represents several classes of interactions between proteins that are building blocks of larger networks. For each module we constructed a mathematical model to calculate CV for interacting proteins transcribed from the same and different operons (hereafter referred to as cotranscribed and uncoupled configurations, respectively). We then determined differences in CV for relevant network outputs between cotranscribed and uncoupled configurations. The simulations were controlled by keeping the same mean and CV for total protein from each gene between configurations. The CV calculations were performed at stationary state, both numerically (stochastic simulation algorithm; [32]) and analytically (LNA [28]
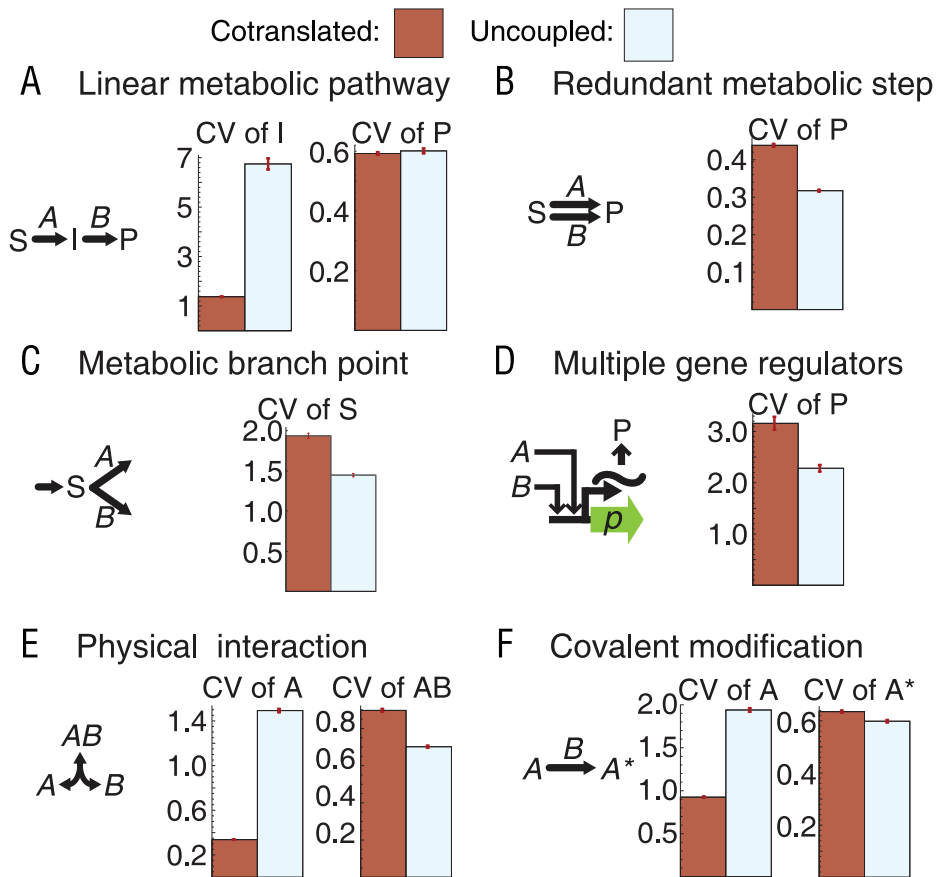
Figure 2. Noise levels in physiologically relevant variables depend on transcriptional coupling. We computed coefficient of variation [CV] = σ/μ, where σ is the standard deviation and μ is the mean, for simplified modules representing modes of interaction between two proteins: A) catalysis of subsequent steps in a linear metabolic pathway, B) redundant catalysis of the same metabolic step, C) catalysis of metabolic steps following a branch point, D) redundant regulation of a downstream gene p encoding protein P, E) physical interactions resulting in heterodimer formation, and F) covalent modification of one protein by another. In metabolic modules, S, I and P represent substrate, intermediate, and product, respectively. Complete reaction diagrams and parameters are given in supplemental tables (S4, S5, S6). Error bars represent one standard deviation from bootstrap resampling. Results correspond to the single ribosome binding site model (translational coupling), but hold qualitatively for multiple ribosome binding sites as well.
doi:10.1371/journal.pcbi.1002672.g002

using Paulsson's [18] normalization). The results of the simulations demonstrate that predicted differences in CV for each metabolic module depend on the type of interaction between proteins (Figure 2).

For the linear metabolic pathway module, cotranscription of two enzymes from the same operon results in lower CV for metabolic intermediate. Without transcriptional coupling, metabolic intermediate concentrations are prone to large spikes (Figures 2A and S1A, B, J, K). Notably, no significant differences between cotranscribed and uncoupled configurations are evident in metabolic product CV (Figure 2A), indicating that metabolic flux is not significantly different between the two groups. Intuitively, a spike occurs when flux from the upstream enzyme exceeds the maximal flux capacity of the downstream enzyme resulting in large increase of metabolic intermediate concentration. This increase exceeds the saturation point for the enzyme converting it to product, making product concentration insulated from these spikes.

In contrast, the metabolic modules with redundant enzymes (Figures 2B and S1C, L) and with a branch point (Figures 2C and S1D, M) show an increase in metabolite CV when the two enzymes are in the same operon. In these cases, lower correlations

between enzyme fluctuations reduce the chance of simultaneous stochastic drops in concentration of both enzymes.

Similarly, cotranscription of multiple (redundant) gene regulators from the same operon results in increased CV in the regulated gene as compared to the uncoupled regulator configuration (Figures 2D and S1E, N). Here, we assumed that the gene regulatory logic was an OR gate (i.e., each regulator by itself or both together would have the same effect). Noise in the output from AND gate logic (i.e., a multi-subunit regulator) is expected to follow the noise pattern of the physical interaction module (below).

Consistent with a previous study [25], the physical protein interaction module under transcriptional coupling shows a strong reduction in fluctuations of monomer concentrations. (Figures 2E and S1F, G, O, P). With strong binding, the concentration of each free monomer changes from nearly zero when its partner is in excess to a finite value when the monomer itself is in excess. These fluctuations are more common when the binding partners are not in the same operon, so the noise is therefore high. Cotranscription slightly increases heterodimer CV compared to the uncoupled configuration (species AB; Figure 2E), but to a much lesser extent than its reduction of CV in monomer concentrations. In the limit

of strong binding, nearly all of one monomer is bound, so the effect on monomer noise is dominant.

For the covalent modification module (Figures 2F and S1H, I, Q, R), different gene configurations cause small changes in the CV of the modified form of protein A (A^*) that can be of either sign depending on parameter values, whereas the unmodified form (A) has consistently lower CV in the cotranscribed configuration (Figures 2F and S1H, Q).

Analytical approach confirms predicted effects of cotranscription on intrinsic noise

The stochastic simulation approach (Figure 2) gives decisive results, but only for the parameter values tested. To determine how generally the simulation results hold in the face of different parameter values, we used the LNA to analytically determine noise differences (here quantified as CV^2) between cotranscribed and uncoupled forms of each network module. For each molecular species denoted by j, we calculated the noise difference between cotranscribed (v_j^C) and uncoupled (v_j^U) configurations as $\Delta v_j = v_j^U - v_j^C$. If the value is positive, the cotranscribed configuration has lower CV^2 (and therefore, lower CV); if it is negative, the uncoupled configuration has lower CV^2. A more complete analysis for each module is presented in Text S2. Here we highlight the main results.

1. Linear metabolic pathway: We find a lower bound on the difference in metabolic intermediate:

$$\Delta v_I >> v_{AB} > 0 \qquad (2)$$

Thus cotranscribed enzymes are generally predicted to have much lower noise in metabolic intermediate than uncoupled enzymes.

2. Redundant metabolic step: We find a difference in the product

$$\Delta v_P \approx -2 v_{AB} H_{AP} H_{BP} \qquad (3)$$

where H_{AP} and H_{BP} are the logarithmic gains (sensitivities) of product P to proteins A and B, respectively. Because $\Delta v_P < 0$, the uncoupled configuration is predicted to have lower noise.

3. Metabolic branch point: For substrate noise, we find

$$\Delta v_s \approx -v_{AB} \frac{2 H_{As} H_{Bs}}{H_{ss}(H_{ss}-1)} \qquad (4)$$

where H_{As}, H_{Bs}, and H_{ss} are the logarithmic gains of substrate flux ratio (V_s^+ / V_s^-, the ratio of the fluxes producing and consuming substrate) in response to changes in A, B and s, respectively. While H_{As} and H_{Bs} are positive, H_{ss} is negative (Text S2). As the enzymes approach saturation, $|H_{ss}|$ becomes small. As a result, Δv_s is large and positive, predicting much lower noise in the uncoupled configuration (Text S2).

4. Multiple gene regulators: The difference for protein production stimulated by two regulators with OR logic is

$$\Delta v_P = -v_{AB} H_{Am_P} H_{Bm_P} \frac{\tau_{mRNA}}{\tau_{mRNA} + \tau_{protein}} \qquad (5)$$

where H_{Am_P} and H_{Bm_P} are the logarithmic gains of mRNA flux ratio (V_{mRNA}^+ / V_{mRNA}^-, the ratio of mRNA production and degradation fluxes) in response to regulators A and B, respectively. Their absolute values are in the interval [0, 1]

with positive values for activators and negative values for repressors. Thus, when both regulators are activators or repressors, $\Delta v_P < 0$, predicting more noise if regulators are in the same operon.

5. Physically interacting proteins: With strong interactions, we have

$$\Delta v_A = \Delta v_B \approx -v_{AB} \frac{H_{AB}}{1 - H_{AB}^2} \qquad (6)$$

where H_{AB} is the logarithmic gain of the flux ratio of B (V_B^+ / V_B^-, the ratio of fluxes producing unbound B to the fluxes consuming it) to protein A. We show that H_{AB} is negative with its absolute value approaching 1 from below. Therefore, $\Delta v_A \gg v_{AB} > 0$, predicting that monomer noise is significantly decreased by covariance. At the same time, the difference in noise in heterodimer is negative and limited in absolute value to $|\Delta v_{complex}| < \frac{v_{AB}}{3}$, predicting that complex noise slightly increases with cotranscription.

6. Covalent modification module: We have

$$\Delta v_{A^*} \approx -v_{AB} \frac{2}{H_{mA}+1}$$
$$\Delta v_A \approx v_{AB} \frac{2(H_{mA}-1)}{H_{mA}+1} \qquad (7)$$

where H_{mA} represents the logarithmic gain of V_A^+ / V_A^- (the ratio of flux producing unmodified protein A to that consuming it) to concentration of mRNA. For realistic parameter values at which protein modification flux significantly exceeds the degradation flux, $H_{mA} \gg 1$ (Text S1). In this regime, Δv_{A^*} is small and negative whereas Δv_A remains positive and significant, predicting a lower noise in the cotranscribed configuration.

To summarize, we found that differences in noise between cotranscribed and uncoupled configurations in stochastic simulations are qualitatively consistent with the analytical approach. Notably, in all the cases the magnitude of the differences in CV^2 between two configurations is proportional to the value of covariance v_{AB}, but in many cases the coefficient of proportionality is very large. This qualitatively suggests posttranslational interactions in some modules are capable of amplifying noise differences between cotranscribed and uncoupled proteins. However, the LNA method likely underestimates the magnitude of non-linear amplification. We further explore these amplification mechanisms in the next section.

Ultrasensitivity arising in non-redundant protein interactions

Timecourse simulations predict that uncorrelated fluctuations of the two enzymes in a linear metabolic pathway result in large bursts of metabolic intermediate (Figure 3A, B). This suggests that higher noise in the transcriptionally uncoupled linear metabolic pathway arises at least in part from the increased probability of occasionally crossing an ultrasensitive threshold. Indeed, a sharp threshold in the intermediate of the linear metabolic pathway arises when the enzyme-mediated consumption of a product saturates, leading to non-linear degradation [33]. The ultrasensitive threshold is crossed when the downstream enzyme saturates, and the flux from the upstream enzyme exceeds its maximal value ($V^+/V^- > 1$ in Figure 3C). Because V^+ and V^- are proportional to
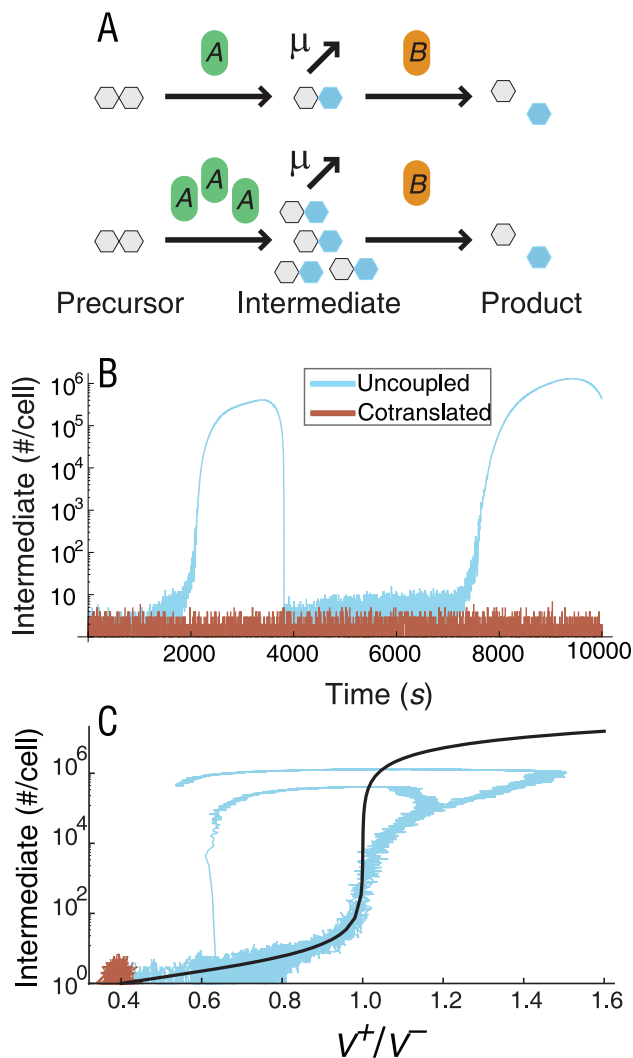
Figure 3. An ultrasensitive response amplifies noise differences between cotranscribed and uncoupled linear metabolic pathway modules. A. When copy numbers of enzymes A and B are matched, transient changes in production and consumption flux are matched, resulting in maintenance of a low concentration of metabolic intermediate. An increase in expression of *A*, unmatched by a change in expression of B, can cause the production flux of metabolic intermediate to exceed the saturation point of flux through enzyme *B*, resulting in accumulation of metabolic intermediate. B. Simulated timecourse of metabolic intermediate in cotranscribed and uncoupled configurations of the linear metabolic pathway model. C. Steady state response of metabolic intermediate to changes in the ratio of production flux to consumption flux (solid line), with stochastic simulation timecourses of intermediate in cotranscribed and uncoupled linear metabolic pathway module configurations plotted with respect to changing flux balance. Results represent the single ribosome binding site model (translational coupling), but are qualitatively the same for multiple ribosome binding sites as well.
doi:10.1371/journal.pcbi.1002672.g003

their enzyme levels, the numerator and denominator of the ratio fluctuate together when both enzymes are in the same operon. Therefore transcriptional coupling lowers noise in the flux ratio and making it unlikely to cross the threshold $V^+/V^- = 1$. When the enzymes are uncoupled, simulations show more variability in the V^+/V^- ratio, allowing the ratio to cross the threshold with consequent large spikes in metabolic intermediate. Thus, the

ultrasensitive switch amplifies noise differences already present between cotranscribed and uncoupled configurations.

Differences in noise between cotranscribed and uncoupled configurations of all of the non-redundant modules can be amplified by ultrasensitive switches in a similar manner (Figure S2). The metabolic branch point module undergoes the same type of non-linear degradation effect as the linear metabolic pathway, but in the branch point transcriptionally coupled enzyme pairs are more likely to fluctuate downward and saturate simultaneously than uncoupled enzymes. This effect leads to a higher likelihood of substrate buildup in the cotranscribed configuration (Figure S2B). The physical interaction and covalent modification modules undergo molecular titration [34], resulting in an ultrasensitive switch for monomers (physical interaction module) or unmodified protein (covalent module) that depends on the ratio of protein production fluxes (Figure S2C, D). Cotranscription of the two genes prevents the switch from amplifying transcriptional noise by reducing fluctuations in this ratio. Sensitivity analysis of mean-field models shows that the existence of ultrasensitive switches does not depend on strict parameter regimes (Text S3).

Detailed *lac* operon model confirms effects of operons on intrinsic noise in an intact system

To explore how conclusions drawn from models of simple network modules apply to a more complicated realistic network, we implemented stochastic simulations of a detailed *lac* operon model that is based on a previous deterministic model [35]. The stochastic model includes enzymatic steps reminiscent of a linear metabolic pathway with permease-mediated lactose import and conversion by β-galactosidase to allolactose and β-D-galactose+β-D-glucose (Figure 4A, Tables S7 and S8). Feedback and gene regulation are present with derepression of *lacY* and *lacZ* expression caused by allolactose binding to LacI (Figure 4A, Tables S7 and S8).

We simulated three inducer concentrations representing minimal *lac* operon induction (1.39 μM extracellular lactose concentration or 835 molecules/femtoliter), intermediate induction (83.0 μM or 50,000 molecules/femtoliter), and excess inducer with maximal *lac* operon induction (~5,000 μM or 3×10^6 molecules/femtoliter). Timecourses suggest that transcriptional coupling between *lacY* and *lacZ* (wild-type situation) eliminates the large fluctuations in allolactose (Figures 4B and S3) and intracellular lactose (not shown) observed in the transcriptionally uncoupled form of the system. This is consistent with a reduction in the correlation between permease and β-galactosidase (*lacY* and *lacZ* gene products, respectively) in time (Figure 4C).

At all inducer concentrations, the uncoupled configuration displays higher CV in allolactose than did the cotranscribed configuration (Figure 4D). This difference is most pronounced in the minimal induction region and gradually reduced with increasing *lacY-lacZ* induction. At the same time, there is little difference in protein CV between cotranscribed and uncoupled configurations of the model at most inducer levels. In both configurations the CV monotonically decreases with higher expression.

The primary consequence of cotranscription of *lac* proteins in the same operon is a reduction in fluctuations of intracellular lactose and allolactose. These fluctuations may prevent disruption of other sugar uptake pathways by, for example, interfering with inducer exclusion mechanisms [36]. Physiological benefits of noise reduction are also consistent with reports that excessive lactose import is associated with significant lowering of growth rate in *E. coli* [37 and references therein,38]. Thus, there may be a selective pressure to maintain high covariance between permease and β-
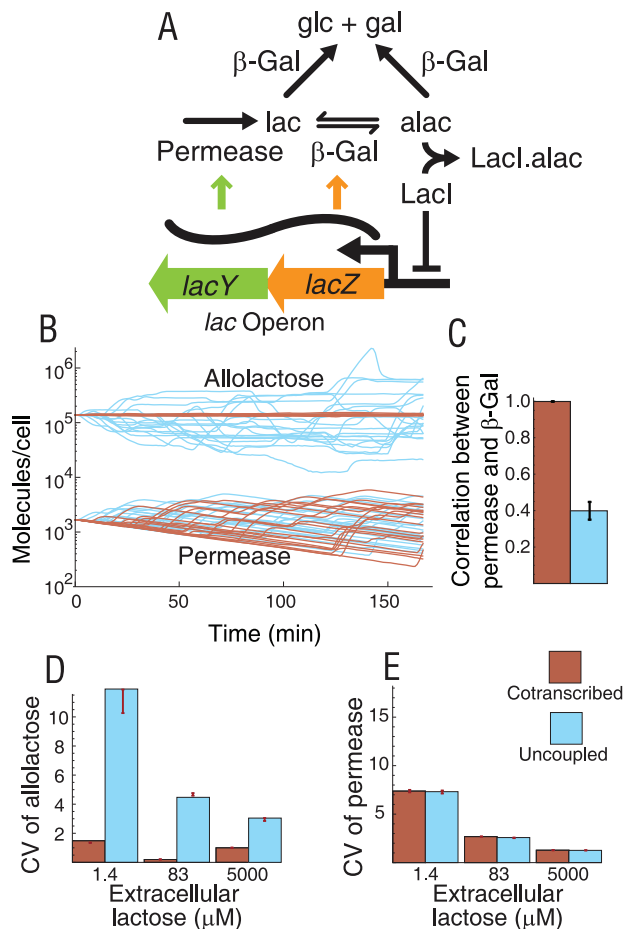
Figure 4. Allolactose noise depends on transcriptional coupling in simulations of the *E. coli lac* operon. A. Simplified reaction diagram of the model used. The complete reaction scheme and parameter values are given in Tables S7 and S8. Arrows depict flux arising from the mechanisms in the model. B. Predicted dynamics of allolactose (Alac) and permease (LacY) in excess inducer (extracellular lactose at 5,000 µM) are represented by 20 individual timecourses. C. Time correlation between permease and β-galactosidase (LacZ) for 100 timecourses. Error bars represent standard error of the mean. D. Coefficients of variation (CV = σ/µ, where σ is the standard deviation and µ is the mean) of metabolite allolactose at three extracellular lactose concentrations. E. CV of permease protein at three extracellular lactose concentrations. Bootstrapped mean and standard deviation of CVs (red bars in panels D and E) diverge from simulated CV when the distribution is highly skewed. Concentration 1.39 µM represents minimal induction of the *lac* operon; 83.0 µM represents mid-range induction; and 5,000 µM represents an excess of inducer with maximal *lac* operon induction. Error bars represent one standard deviation from bootstrap resampling.
doi:10.1371/journal.pcbi.1002672.g004

galactosidase resulting from the wild-type genetic structure of the *lac* operon.

Operon incidence in *E. coli* is correlated with noise reduction

To determine if global operon organization in *E. coli* correlates with predicted noise differences, we characterized frequencies of gene membership in the same operon bioinformatically (Table 1). We first assigned membership of known *E. coli* K12 MG1655 [39] biochemical networks into patterns corresponding to the 2-gene modules (Figure 2) using data on *E. coli* operons [31], metabolic

pathways [29], gene regulation networks [31], covalent modification [30], and physical protein interactions [40,41]. Many natural networks fall into more than one class (e.g., common bacterial signal mediators, two-component systems, have physical interactions between the sensor and the regulator [42] and are also in the covalent modification class). For the metabolic and gene regulation network modules, we eliminated physically interacting pairs to ensure that those included had true functional overlap and were not acting as subunits of a larger enzyme or regulator. Thus, the only systems that are members of more than one class are in members of both the covalent modification and physical interaction modules. In each class we created controls with randomized operon assignment of the genes (see Materials and Methods).

Proteins in the linear metabolic pathway, physical interaction, and covalent modification modules appear in the same operon significantly more frequently than do randomized controls ($p \ll 10^{-6}$; Table 1). On the other hand, redundant metabolic nodes and multiple gene regulators are significantly less likely to be in the same operon than randomized controls ($p \ll 10^{-6}$; Table 1). Metabolic branch points show a bias toward being uncoupled, but falls just short of being statistically significant ($p = 0.071$). These findings hold even after we divide each class into essential and nonessential genes using data from Taniguchi *et al* [43]; Table S2). Thus operon overrepresentation, when it occurs, is present in essential genes, consistent with previous results contradicting the selfish operon hypothesis [12]. Our results establish a correlation between operon organization of protein pairs and their function that is consistent with noise minimization and avoidance of ultrasensitivity.

Higher incidence of cotranscribed interacting proteins at low than high expression levels

To separate the specific effect of noise from that of other factors affecting selection for operons, such as proportional coregulation, we considered whether the tendency toward operon membership of posttranslationally interacting protein pairs is related to gene expression levels [23]. Intrinsic noise is stronger for genes with low expression levels [43], covariance of protein concentrations is more pronounced (Equation 1) and the degree of decorrelation is higher (Table S1). Therefore, if noise is an evolutionary factor driving operon formation, levels of gene expression may be inversely correlated with operon patterns. On the other hand, if coregulation of mean expression levels is the dominant factor in selecting for operons, the frequency of transcriptional coupling may be directly correlated with gene expression levels because the cost of differential regulation would be highest at the highest expression levels. As a result, any trend in coupling frequencies with gene expression levels would favor one hypothesis and disfavor the other.

We used a dataset of average single-cell mRNA and protein copy numbers in *E. coli* [43] to explore this trend for constitutively expressed physically interacting protein pairs (other network modules have insufficient data for such analysis). Because different conditions can shift gene expression levels and the dataset is only available for one condition, we chose to focus on the subset of interacting proteins that are constitutive, i.e., not predicted to undergo any regulation in RegulonDB. Each gene's protein or mRNA copy number was considered once, along with a binary variable indicating whether or not the protein product interacts with a same or non-same operon protein. Further details are given in Materials and Methods.

We divided the set into two subsets of expression level, one below and one above the median copy number (Figure 5). The

Table 1. Operon organization trends in *E. coli* relate to noise-minimizing transcriptional coupling patterns.

	Network module	Low noise configuration	Number of pairs	f^a	f_{rand} ($\mu \pm \sigma$)	Trend in *E. coli*[b]
A.	Linear metabolic pathway	Coupled	2471	0.038	0.0036 ± 0.0011	Coupled ($<<10^{-6}$)
B.	Redundant enzymes	Uncoupled	114	0	0.01 ± 0.00094	Uncoupled ($<<10^{-6}$)
C.	Metabolic branch point	Uncoupled	2036	0.0025	0.0047 ± 0.0015	Uncoupled (0.071)
D.	Multiple gene regulators	Uncoupled	1368	0.0015	0.043 ± 0.0054	Uncoupled ($<<10^{-6}$)
E.	Physical interaction	Coupled	3938	0.35	0.0030 ± 0.00073	Coupled ($<<10^{-6}$)
F.	Covalent modification	Coupled	201	0.23	0.0047 ± 0.0047	Coupled ($<<10^{-6}$)

[a]Same operon pair fraction.
[b]Numbers represent *p*-value vs rand.
doi:10.1371/journal.pcbi.1002672.t001

fraction of protein pairs sharing the same operon is higher in the low-expression subset for protein (bootstrap test $p<0.01$) and mRNA (bootstrap test $p<0.05$) copy numbers. This suggests that evolutionary selection against decorrelation (Table S1) significantly contributes to maintenance of operons in the chromosome.

Discussion

A longstanding question in evolutionary biology is how non-transcriptional dynamics [44] affect selection of particular genetic architectures. By relating chromosomal patterns to protein network structures in *E. coli*, we see a compelling case for post-translational dynamics altering the probability of operon membership of genes depending on the nature of their interaction.

Covariance and noise reduction in non-redundant interactions

Because enzymes often operate close to saturation [45], resolving metabolic flux imbalances may prevent widespread accumulation of intermediate, which is potentially toxic [46,47,48]. Simulations of a detailed *lac* operon model in our study corroborate the results of the simpler linear metabolic module, suggesting a role for intrinsic noise in selecting for *lac* operon architecture (in addition to the stochastic effects previously examined in this system [49]). Simulations that include extrinsic

Figure 5. Dependence on expression level for frequencies of gene pairs of physically interacting proteins sharing operons. A dataset of single-cell protein and mRNA copy numbers in *E. coli* [43] shows reduced frequency of coupling for abundant proteins (A) and mRNA (B). Error bars represent one standard deviation from bootstrapping the data 1,000 times. Significance levels were determined by a bootstrap test.
doi:10.1371/journal.pcbi.1002672.g005

noise as a correlating factor indicate that it does not reduce metabolite noise as well as the stronger correlations caused by cotranscription (Text S4, Figure S4).
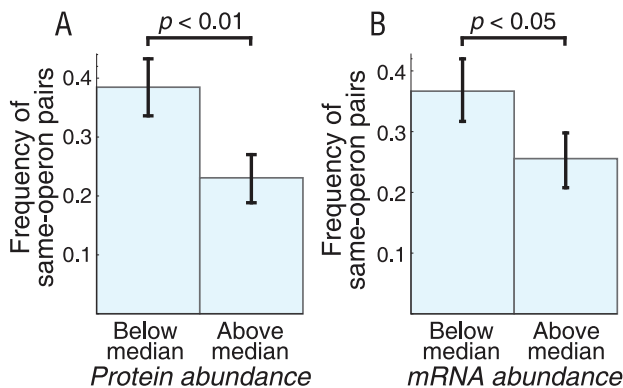
Many metabolic operons are large (and with complex evolutionary histories; [50]), but the length of a metabolic pathway is often longer than that of a typical operon, leading to the question of where optimal operon break points for metabolic pathways may lie. Our results suggest that break points occur predominantly where the intermediate is not toxic or where it is processed by multiple downstream enzymes, such as at branch points and metabolic steps with redundant enzymes. Metabolite spikes could also potentially be buffered by reversibility of catalytic reactions, though the reversible step in the *lac* operon did not prevent intermediate spikes. Furthermore, if portions of metabolic pathways that are divided by intermediates with relatively low toxicity undergo upregulation as needed, there may be a trade-off between reduction of toxic intermediate spikes and just-in-time transcription [51] in the evolution of metabolic networks.

Our analysis suggests that pairs of enzymes after a branch point can have lower noise (CV) if they are not cotranscribed (Figure 2C), but with a less consistent CV difference between cotranscribed and uncoupled configurations than the other modules (Figure S1 D,M). Therefore, the noise hypothesis predicts patterns of transcriptional coupling to be weaker than in other modules, as we observe to be the case in *E. coli* (Table 1).

The simple physical protein interaction module in our study (Figure 2E) may result in one of two different types of physiologically meaningful output variables: an active heterodimer, in which the genes make up subunits of a functional complex, or an active monomer, in which its activity is negatively regulated by the binding partner (as with sigma-antisigma systems [52]). In either case reduction in monomer noise is justified; in the latter case, to reduce noise in the physiologically relevant output. In the former case, lower noise represents a reduction in inefficient protein production that can reduce promiscuous interactions with other parts of the network. Heterodimer noise is smaller for the uncoupled configuration because upward fluctuations in its concentration are limited to being no larger than the minimum of [A] and [B] and those concentrations are less likely to simultaneously fluctuate upward simultaneously.

The covalent modification system (Figure 2F) in its uncoupled configuration has reduced fluctuations in the unmodified protein (A) compared with the uncoupled configuration. Noise effects of transcriptional coupling may therefore be important in covalent modification systems where the unmodified form of the protein is capable of interacting with other systems (Text S2).

Higher-order chromosome structure, such as bacterial chromatin [53,54] and regulatory factors such as bidirectional promoters

and transcriptional terminators [55] affect the spatial proximity of genes. Operons could also play a role in spatial proximity, as suggested by the selfish operon hypothesis [10]. We explored whether chromosomal proximity can explain operon membership in linear metabolic and physically interacting gene pairs. Our bioinformatic analysis suggests that the prevalence of operons cannot be solely explained by a proximity bias of interacting gene pairs in the *E. coli* chromosome (Text S4, Figure S5).

Noise differences in non-redundant modules arise from ultrasensitivity

A striking feature of the non-redundant protein interaction modules is that they all contain a zero-order ultrasensitive switch, which arises as a side-effect of saturation. This effect amplifies differences in CV between cotranscribed and uncoupled forms of the modules (Figures 3 and S2) and may degrade performance when its threshold is crossed. In each two-gene module, the operon architecture that avoids crossing the ultrasensitive threshold is significantly over-represented in *E. coli* (Table 1). Signatures of selection against noise in these modules thus likely represent selection against performance-degrading ultrasensitivity as well.

Gene expression level and operon prevalence

Gene pairs encoding constitutive physically interacting proteins are significantly more likely to be in the same operon if their expression levels are low (Figure 5). This trend could be explained by slow protein diffusion in the crowded intracellular environment, as cotranscribed gene products are more likely to be present at the same subcellular location. However even slow diffusion ($<1~\mu m^2/s$) across a typical bacterial length of $\sim 1~\mu m$ is much faster than the expected time lag between translation of two proteins given typical ribosomal speeds of 12–21 AA/s [56]. Therefore, increased biochemical noise (here, measured as decorrelations between uncoupled proteins) at low expression levels are the most likely explanation of the observed trend. We argue that these noise effects are detrimental to the performance of some protein interaction networks.

The opposite trend would be expected if proportional expression of mean concentrations or other mechanisms are the primary selective pressure on operon maintenance. In general, genes with high expression levels may operate under greater evolutionary pressure than genes with low expression levels [57,58] and therefore their deviation from optimal chromosomal organization is less likely. Arguably, noise minimization is the only selective force that is expected to be more important for genes with low expression levels than for genes with high expression levels [23].

Redundant proteins and selection against operons

Partial functional redundancy of proteins allows one protein to compensate for a downward fluctuation in concentration of the other protein, thereby reducing noise with uncorrelated protein fluctuations (Table 1; Figure 2). Therefore, just as noise minimization may explain operon membership for non-redundant interactions, it may also explain the lack of redundant proteins in operons. Differential regulation of the genes can additionally play an important role in keeping redundant interactions transcriptionally uncoupled. In yeast metabolic pathways, apparently redundant enzymes are differentially expressed in different pathways depending on external conditions [59]. This type of mechanism, if present *in E. coli*, may also explain why no redundant enzymes are in the same operon. Similarly, different growth conditions may result in different regulators affecting downstream expression of the same genes. Further work is necessary to distinguish the noise reduction hypothesis more

decisively from differential gene regulation as a selective force in redundant pairs; differential regulation may be physiologically important in some cases and not in others.

Eukaryotes, operons, and noise

Improvement of dynamic performance of simple networks arising from cotranscription of interacting genes from the same operon raises the question of why operons are rare in eukaryotes. Eukaryotic cell volumes are much higher than prokaryotes, likely lowering the effect of intrinsic noise relative to the dominant effect of extrinsic noise [60]. Nevertheless, such benefits may still be present in some systems, and there are mechanisms that allow correlating gene expression noise in eukaryotic cells without polycistronic loci. Genes located near each other have correlated transcriptional bursts that likely arise from chromatin decondensation [61,62]. Clusters of co-expressed genes, particularly metabolic genes, appear in eukaryotic chromosomes at a rate higher than would be expected randomly [63,64]. Co-expressed, functionally related genes at distant genomic loci also appear to migrate together for co-transcription from discrete transcription initiation complexes [65,66,67]. These mechanisms, arising from the increased size and structural complexity of eukaryotic chromatin over prokaryotic chromosomes, can correlate gene expression noise with similar dynamic benefits to operons.

Concluding remarks

We have developed a theory predicting that operon membership can increase or decrease noise in different types of protein interactions. Bioinformatic analysis finds that naturally occurring operon patterns in *E. coli* correlate with reduction of biochemical noise. Nevertheless, it would be interesting to explore operon coupling frequencies in bacterial stress response systems known to favor population-level heterogeneity, such as stress responses in *B. subtilis*; the amplification of noise by underlying ultrasensitive switches in non-redundant network modules may be a potential mechanism of population-level heterogeneity.

The existence of implicit ultrasensitive switches also underscores the idea that dramatic non-linearities are likely present in many simple protein interaction networks. Our results suggest that ultrasensitive switches are likely undetectable in the wild-type configurations of well-adapted systems as a result of selection against them, but may be present in conditions with lower selective pressure, or recent evolutionary events. These switches nevertheless have important implications for genome evolution. Their effects, and the mechanisms for avoiding them, may in turn shape larger biochemical networks by changing global noise properties, and will be an important factor in designing synthetic networks.

Materials and Methods

Symbolic manipulations and data analysis were performed in Mathematica 7.0 and 8.0 (Wolfram Research, Champagne, IL). We predicted intrinsic noise characteristics with stochastic simulations at stationary state using the StochKit (http://engineering.ucsb.edu˜/cse/StochKit/) tau-leaping routine for 10,000 runs of each model (except in the *lac* operon model, for which 1,000 runs of each condition were done). Initial model construction and test runs were done with Copasi (www.copasi.org). Simulations with an extrinsic noise representation were done in Copasi as detailed below. All models were represented with elementary reaction steps; in models involving gene regulation, we defined a promoter variable as always present at one copy per cell.

Stochastic simulations

Unless otherwise specified, each network module was tested with promoter-mediated noise, represented by promoters switching between "on" and "off" states of mRNA production. This process has estimated switching rates of $k_{goff} = 0.0028$ s^{-1} for switching to the "off" state and $k_{gon} = 0.00045$ s^{-1} for switching to the "on" state [20]. In models including a gene regulation step, we assumed binding and unbinding of regulators to be independent of promoters switching between on and off states. Without promoter-generated bursting, gene expression noise largely arises from low mRNA copy numbers per cell and the effect of transcriptional coupling is qualitatively similar (Figure 1). Furthermore, analytical results from LNA do not include the effects of bursty transcription, showing that we arrive at qualitatively similar results without transcriptional bursts.

We distinguish between three types of coupling between production of two proteins in stochastic simulation reaction schemes. Transcription may be coupled or uncoupled (i.e., proteins in the same or separate operons) and when transcription is coupled, proteins may be cotranslated (single ribosome binding site for both) or translationally uncoupled (two ribosome binding sites). These three cases represent simplified extremes; intermediate translational linkage (e.g., read-through from multiple ribosome binding sites) is possible but was not further considered here. Figure 1 illustrates the three cases with promoter-mediated gene expression noise. For simplicity of presentation, we compare transcriptionally uncoupled with cotranslated models in the main text.

Tables S4, S5, S6, S7, S8 give reaction schematics and parameters for the gene expression and posttranslational models used in the main text. Parameter values were chosen to be of the correct order of magnitude for realistic expression levels and binding kinetics. To ensure a fair comparison between cotranscribed and uncoupled configurations, production and degradation rates of mRNA species for proteins A and B are identical. The degradation rate k_{deg} corresponds to the value expected from a dilution rate for typical *E. coli* doubling every half hour.

Analytical determination of protein covariance

The basis of noise differences between networks with proteins in the same operon and those with proteins in separate operons is the covariance between the expressed proteins. We used LNA to analytically characterize noise and covariance as follows. For the mean values of copy numbers (denoted by angular brackets):

$$\frac{d}{dt}\begin{bmatrix}\langle m_1 \rangle \\ \langle A \rangle \\ \langle m_2 \rangle \\ \langle B \rangle\end{bmatrix} = \begin{bmatrix}k_m - k_{mdeg}\langle m_1 \rangle \\ k_{tsn}\langle m_1 \rangle - k_{deg}\langle A \rangle \\ k_m - k_{mdeg}\langle m_2 \rangle \\ k_m \langle m_i \rangle - k_{mdeg}\langle B \rangle\end{bmatrix} \quad (8)$$

where $i = 1$ with proteins A and B in the same operon, and $i = 2$ with proteins A and B in separate operons. Then we solved the fluctuation-dissipation matrix equation at steady state ($\mathbf{M}\sigma + \sigma\mathbf{M}^T + \Omega\mathbf{N} = 0$) for σ, where $\mathbf{M}$ is the Jacobian of the (macroscopic) system, Ω is cell volume, and $\mathbf{N}$ is the diffusion matrix [18]. Characterizing intrinsic noise as $v_j = \text{CV}_j^2 = \sigma_{jj}/\mu_j^2$ and $v_{ij} = \sigma_{ij}/(\mu_i\mu_j)$ with indices i and j taking values corresponding to molecular species (A, B, m_1 and m_2), we follow the methods of [18] to obtain:

$$v_A = \frac{\sigma_{AA}}{\langle A \rangle^2} = \frac{1}{\langle A \rangle} + \frac{\tau_{mRNA}}{\langle m_1 \rangle (\tau_{mRNA} + \tau_{protein})}. \quad (9)$$

and v_{AB} as in Equation 1. Note that σ_{AA} is the variance, or the square of the standard deviation.

To analytically approximate noise of physiologically relevant variables in the simple network modules (Figure 2), we made the following simplifications to make the systems tractable. For metabolic steps with a substrate as a dependent variable, we assumed a Michaelis-Menten propensity. For the covalent modification module, we assumed a simple mass-action with no saturation or complexes. For the multiple gene regulator module, we used a Hill equation propensity for regulated mRNA production. Details of the analysis are in Text S2. Mean-field models are given in Table S3.

Trends in *E. coli* K12 MG1655 genes

Bioinformatic analyses used pairs of interacting genes extracted from databases of *E. coli* K12 MG1655 as described below. To determine the randomized control, we needed to account for potential biases resulting from dataset size and other features of chromosome organization that we were not attempting to test. For instance, if we randomly assigned genes to extant operons in *E. coli* across the entire chromosome, the less frequently occurring modules would have much less same-operon membership than the modules with larger numbers of members, and would not be a useful control. We chose to randomize the genes extracted from the pairs within each module to set the random control for each class. Thus, for a list of gene pairs

$$\boldsymbol{g} = \{(g_{11}, g_{12}), (g_{21}, g_{22}), \ldots, (g_{n1}, g_{n2})\}$$

we determined a randomized case by flattening $\boldsymbol{g}$ into

$$\mathbf{g}_f = \{g_{11}, g_{12}, g_{21}, g_{22}, \ldots, g_{n1}, g_{n2}\},$$

randomly permuting the order of the genes and then re-pairing them to determine the frequency. This process was repeated 1000 times to determine the parameters of the randomized distribution.

Metabolic networks. The *E. coli* metabolic network was extracted from Kyoto Encyclopedia of Genes and Genomes (KEGG; [29]). We partitioned the network into pairs of adjacent steps and assigned a value of 1 to each pair in the same operon (using the operon membership dataset from RegulonDB; [31]) and 0 to each pair in separate operons. This gave a frequency f_{met} of subsequent same-operon metabolic steps. Several metabolic steps were found to be catalyzed by enzymes with subunits from multiple genes. Because such interactions have a large bias in favor of operon membership (Table 1), we eliminated them from the analysis to ensure operon membership biases result from metabolic, and not physical, interactions. We then repeatedly randomly assigned the metabolic enzymes to operons to generate a predicted distribution of background operon membership. The resulting distribution is approximately normal, allowing a parametric determination of f_{met} significance ($f_{met} = 91/2417 = 0.038$; mean $\mu_r = 0.0036$; standard deviation $\sigma_r = 0.0011$; $p \ll 10^{-6}$).

We extracted pairs of enzymes catalyzing the same metabolic step from the database to assess the trend in redundant metabolic steps. Here, $f_{redundant} = 0/114$. The resulting randomized distribution had mean $\mu_r = 0.0047$; standard deviation $\sigma_r = 0.0015$ ($p \ll 10^{-6}$ against this distribution). Enzymes catalyzing steps after a metabolic branch point were determined on the basis of common substrates in KEGG ($f_{branch} = 5/2036$; $p = 0.071$ against a randomized operon membership set of the same genes).

Covalent modifications. Extracting examples of this inter-action type from databases is difficult because of the complex nature of the relationship, which requires interaction with subsequent modification. We used the MultiFun Gene Ontology from EcoCyc [30] to identify protein-related information transfer systems, each of which we manually checked to eliminate irrelevant cases and identify operon membership of the pair. Forty-six of the 201 pairs (23%) had interactions within the same operon. Because some of the identified members have multiple interaction pairs that could not be identified, the 46/201 fraction is somewhat uncertain. Unlike any of the other interaction modules, this module has pairs that overlap with another module (specifically, the physical interaction module). We used the same randomization procedure as for the metabolic pairs above to test the fraction relative to a randomized case of the covalent modification genes. With $\mu_c = 0.0047$, $\sigma_c = 0.0047$, and $p \ll 10^{-6}$, the true fraction is likely greater than the randomized control.

Multiple gene regulators. We used the *E. coli* gene regulation network from RegulonDB [31] to construct a graph of genes regulated by two or more regulators, excluding pairs that form subunits of a larger regulator. Following the method for metabolic networks, we assigned to each pair a value of 1 if the regulators are in the same operon and 0 if the regulators are in separate operons, giving frequency $f_g = 2/350 = 0.0057$. Here, the random distribution has $\mu_{rg} = 0.099$, $\sigma_{rg} = 0.016$, and $p \ll 10^{-6}$.

Physical protein interactions. For physical protein inter-actions we used a functional interaction dataset [40,41] to characterize the frequency of same-operon pairs, $f_p = 0.35$. In the randomized distribution, $\mu_p = 0.0030$, $\sigma_p = 0.00073$, and $p \ll 10^{-6}$.

Operon relationship to protein copy number per cell

We extracted single-cell mRNA expression data (RNAseq) and protein copy number data from Taniguchi *et al* [43]. To ensure a meaningful comparison of expression levels, we considered only genes predicted to be unregulated in RegulonDB. Only the physical interaction module left enough data for analysis. For instance, the number of unregulated pairs in the same operon for the linear metabolic pathway dataset was 5, insufficient to distinguish the established operon membership pattern from noise when parti-tioned between high and low expression. Each average single-cell mRNA or protein copy number was used, along with physical interaction status (1 = same operon; 0 = non-same operon). Proteins with multiple interaction partners within and between operons were represented twice, once for same-operon and once for non-same-operon interaction. We then divided the set into above- and below-median subsets and compared the fraction of same-operon interactions in the subsets using a standard bootstrap resampling test. We resampled 10,000 times with replacement and computed the difference in coupling frequencies between low and high expression as the test statistic. To compute error bars, we used bootstrapping of each bin by sampling each bin with replacement up to the bin size, repeated 1,000 times.

Supporting Information

Figure S1 Conservation of noise relationships in network modules across expression level and parameter variations. Percentage noise difference is given by $\dfrac{CV_L - CV_U}{CV_L}$; blue shades indicate lower CV for the coupled architecture; red, lower CV for the uncoupled architecture. A, J. Intermediate in linear metabolic pathway. Parameter: k_{cat2}. The distributions at low expression

level are highly skewed, giving misleading CV values that do not accurately reflect the higher variability in the uncoupled architecture. B, K. Product of metabolic pathway. Parameter: k_{cat2}. C, L. Product of redundant metabolic step. Parameter: k_{cat1} and k_{cat2} varied simultaneously. D, M. Substrate of a metabolic branch point. Some cases have lower CV in the co-transcribed model because the distribution is bimodal, but the uncoupled model predicts lower actual variability than the co-transcribed model in all cases. Parameter: k_{cat1} and k_{cat2} varied simultaneously. E, N. Protein product of multiple gene regulator network. Parameter: k_d. F, O. Monomer of physical protein interaction module. Parameter: k_b. G, P. Heterodimer of physical protein interaction module. Parameter: k_b. H, Q. Unmodified protein of covalent modification network. Parameter: k_p. I, R. Modified protein of covalent modification network. Parameter: k_p The scale for each variable was set by the largest absolute value. Expression levels (protein copy number/gene/cell) for multiple gene regulator module: L: 1 M: 53 H: 529. For all others: L: 53 M: 529 H: 5285.
(PNG)

Figure S2 Non-redundant two-gene modules undergo an ultrasensitive switch dependent on production and degradation fluxes. Black lines are the mean-field steady state response while orange and blue lines trace timecourses from individual stochastic simulation trajectories. A. Intermediate in the linear metabolic pathway repeated from Figure 3C in the main text. B. Substrate levels at a metabolic branch point in response to changes in the balance between production and total consumption by two enzymes. Spikes are more likely when enzyme-mediated con-sumption fluxes at the branch point covary. C. Quantities of monomer subunit *A* of a heterodimer in response to different relative levels of *A* and *B* monomers. Physically interacting proteins produced asynchronously cross an ultrasensitive thresh-old, which is avoided by cotranscription from the same operon. D. Response of unmodified protein *A* in the covalent modification module to changes in the ratio of *A* to *B*. Unmodified protein *A* undergoes large spikes corresponding to crossing an ultrasensitive threshold when uncoupled.
(PNG)

Figure S3 Predicted dynamics of the *lac* operon system at three inducer concentrations. A. Simulated time courses of permease, allolactose and product. At all three concentrations, the transcrip-tionally uncoupled form of the system induces higher noise in allolactose (metabolic intermediate) concentration, but not product (glucose+galactose) or protein (permease). B. Correlations between permease and β-galactosidase in cotranscribed and uncoupled configurations. Throughout the range of induction the system demonstrates a consistent, significant reduction of correlation between permease and β-galactosidase in the uncoupled form of the system.
(PNG)

Figure S4 Effects of extrinsic noise on transcriptional coupling dynamics for metabolic modules. A. Linear metabolic pathway. B. Redundant metabolic step. Translational and transcriptional rate constants were randomly selected from uniform distributions to mimic global extrinsic noise. The resulting transcriptionally uncoupled protein distributions show a slight correlation between the proteins A and B (r = 0.293 top panel simulations and 0.388 in the bottom simulations). Metabolite noise differences between co-translated and transcriptionally uncoupled architectures are qualitatively unchanged from simulations that do not simulate extrinsic noise, with lower intermediate CV in the linear metabolic

pathway (A) and higher product CV in the redundant metabolic step (B) in the cotranscribed configuration.
(PNG)

Figure S5 Distributions of linear metabolic and physically interacting protein pair interaction chromosomal locus distances in *E. coli* K12 MG1655. A. Gene pair locus distances in linear metabolic interactions are not distinguishable from randomized distances. B. A subset of gene pair locus distances in physical protein interactions have a distinct bias toward close chromosomal proximity. A distance randomization procedure (Text S4) does not indicate that proximity explains operon frequencies in either case.
(PNG)

Table S1 Predicted level of decorrelation between proteins that are uncoupled, cotranscribed, or cotranslated, with or without transcriptional bursting.
(PDF)

Table S2 Essential and nonessential subsets of gene pairs both have significantly high fractions of same-operon pairs *f*.
(PDF)

Table S3 Mathematical models used for linear noise approximation of five simple network motifs in the form $\frac{d\vec{y}}{dt} = \boldsymbol{R}$.
(PDF)

Table S4 Reactions for two genes expressed from the same and separate operons.
(PDF)

Table S5 Post-translational interactions in the linear and redundant metabolic step models.
(PDF)

Table S6 Post-translational interactions in the covalent modification and physical interaction models.
(PDF)

Table S7 Detailed *lac* operon model.
(PDF)

Table S8 Parameter values for detailed *lac* operon simulations.
(PDF)

Text S1 Decorrelation, extrinsic noise and expression levels.
(PDF)

Text S2 Analytical approach confirms predicted effects of cotranscription on intrinsic noise.
(PDF)

Text S3 Analytical determination of ultrasensitive thresholds.
(PDF)

Text S4 Chromosomal proximity of genes does not explain frequencies of metabolic or physical interaction operons.
(PDF)

Author Contributions

Conceived and designed the experiments: JCJR OAI. Performed the experiments: JCJR. Analyzed the data: JCJR OAI. Wrote the paper: JCJR OAI.

References

1. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3: 318–356.
2. Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, et al. (2002) A global analysis of Caenorhabditis elegans operons. Nature 417: 851–854.
3. Rocha EP (2008) The organization of the bacterial genome. Annu Rev Genet 42: 211–233.
4. Price MN, Huang KH, Arkin AP, Alm EJ (2005) Operon formation is driven by co-regulation and not by horizontal gene transfer. Genome Res 15: 809–819.
5. Price MN, Arkin AP, Alm EJ (2006) The life-cycle of operons. PLoS Genet 2: e96.
6. Iber D (2006) A quantitative study of the benefits of co-regulation using the spoIIA operon as an example. Mol Syst Biol 2: 43.
7. Ray JCJ, Igoshin OA (2010) Adaptable functionality of transcriptional feedback in bacterial two-component systems. PLoS Comput Biol 6: e1000676.
8. Lim HN, Lee Y, Hussein R (2011) Fundamental relationship between operon organization and gene expression. Proc Natl Acad Sci U S A 108: 10626–10631.
9. Laing E, Mersinias V, Smith CP, Hubbard SJ (2006) Analysis of gene expression in operons of Streptomyces coelicolor. Genome Biol 7: R46.
10. Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143: 1843–1860.
11. Lawrence JG (1997) Selfish operons and speciation by gene transfer. Trends Microbiol 5: 355–359.
12. Pál C, Hurst LD (2004) Evidence against the selfish operon theory. Trends Genet 20: 232–234.
13. Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 23: 324–328.
14. Voigt CA, Wolf DM, Arkin AP (2005) The Bacillus subtilis SIN Operon: An Evolvable Network Motif. Genetics 169: 1187–1202.
15. Swain PS (2004) Efficient attenuation of stochasticity in gene expression through post-transcriptional control. J Mol Biol 344: 965–976.
16. Silva-Rocha R, de Lorenzo V (2010) Noise and robustness in prokaryotic regulatory networks. Annu Rev Microbiol 64: 257–275.
17. McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA 94: 814–819.
18. Paulsson J (2004) Summing up the noise in gene networks. Nature 427: 415–418.
19. Berg OG (1978) A model for the statistical fluctuations of protein numbers in a microbial population. J Theor Biol 71: 587–603.
20. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123: 1025–1036.
21. Mitarai N, Dodd IB, Crooks MT, Sneppen K (2008) The generation of promoter-mediated transcriptional noise in bacteria. PLoS Comput Biol 4: e1000109.
22. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A 99: 12795–12800.
23. Kovács K, Hurst LD, Papp B (2009) Stochasticity in protein levels drives colinearity of gene order in metabolic operons of Escherichia coli. PLoS Biol 7: e1000115.
24. Løvdok L, Bentele K, Vladimirov N, Müller A, Pop FS, et al. (2009) Role of translational coupling in robustness of bacterial chemotaxis pathway. PLoS Biol 7: e1000171.
25. Sneppen K, Pedersen S, Krishna S, Dodd I, Semsey S (2010) Economy of operon formation: cotranscription minimizes shortfall in protein complexes. MBio 1: e00177–00110.
26. Goldbeter A, Koshland DE (1981) An amplified sensitivity arising from covalent modification in biological systems. Proc Natl Acad Sci U S A 78: 6840–6844.
27. Alloing G, Martin B, Granadel C, Claverys J-P (1998) Development of competence in Streptococcus pneumoniae: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. Mol Microbiol 29: 75–83.
28. van Kampen NG (2007) Stochastic Processes in Physics and Chemistry. 3rd edition. Amsterdam: North Holland Publishing Company.
29. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36: D480–D484.
30. Keseler IM, Bonavides-Martinez C, Collado-Vides J, Gama-Castro S, Gunsalus RP, et al. (2009) EcoCyc: A comprehensive view of Escherichia coli biology. Nucleic Acids Res 37: D464–D470.
31. Gama-Castro S, Jiménez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Peñaloza-Spinola MI, et al. (2008) RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. Nucleic Acids Res 36: D120–D124.
32. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81: 2340–2361.
33. Buchler NE, Gerland U, Hwa T (2005) Nonlinear protein degradation and the function of genetic circuits. Proc Natl Acad Sci U S A 102: 9559–9564.
34. Buchler NE, Louis M (2008) Molecular titration and ultrasensitivity in regulatory networks. J Mol Biol 384: 1106–1119.
35. Dreisigmeyer DW, Stajic J, Nemenman I, Hlavacek WS, Wall ME (2008) Determinants of bistability in induction of the Escherichia coli lac operon. IET Syst Biol 2: 293–303.

36. Saier MH, Crasnier M (1996) Inducer exclusion and the regulation of sugar transport. Res Microbiol 147: 482–489.
37. Perfeito L, Ghozzi S, Berg J, Schnetz K, Lässig M (2011) Nonlinear fitness landscape of a molecular pathway. PLoS Genet 7: e1002160.
38. Eames M, Kortemme T (2012) Cost-benefit tradeoffs in engineered *lac* operons. Science 336: 911–915.
39. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, et al. (1997) The complete genome sequence of Escherichia coli K-12. Science 277: 1453–1462.
40. Butland G, Peregrín-Alvarez JM, Li J, Yang W, Yang X, et al. (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature 433: 531–537.
41. Peregrín-Alvarez JM, Xiong X, Su C, Parkinson J (2009) The modular organization of protein interactions in *Escherichia coli*. PLoS Comput Biol 5: e1000523.
42. Marles-Wright J, Lewis RJ (2007) Stress responses of bacteria. Curr Opinion Struct Biol 17: 755–760.
43. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, et al. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329: 533–538.
44. Ray JCJ, Tabor JJ, Igoshin OA (2011) Non-transcriptional regulatory processes shape transcriptional network dynamics. Nat Rev Microbiol 9: 817–828.
45. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, et al. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat Chem Biol 5: 593–599.
46. Fell DA (1996) Understanding the Control of Metabolism. London, UK: Portland Press.
47. Levine E, Hwa T (2007) Stochastic fluctuations in metabolic pathways. Proc Natl Acad Sci U S A 104: 9224–9229.
48. Lee SJ, Trostel A, Le P, Harinarayanan R, FitzGerald PC, et al. (2009) Cellular stress created by intermediary metabolite imbalances. Proc Natl Acad Sci U S A 106: 19515–19520.
49. van Hoek M, Hogeweg P (2007) The effect of stochasticity on the *lac* operon: an evolutionary perspective. PLoS Comput Biol 3: e111.
50. Fani R, Brilli M, Liò P (2005) The origin and evolution of operons: the piecewise building of the proteobacterial histidine operon. J Mol Evol 60: 378–390.
51. Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, et al. (2004) Just-in-time transcription program in metabolic pathways. Nat Genet 36: 486–491.
52. Igoshin OA, Brody MS, Price CW, Savageau MA (2007) Distinctive topologies of partner-switching signaling networks correlate with their physiological roles. J Mol Biol 369: 1333–1352.
53. Noom MC, Navarre WW, Oshima T, Wuite GJ, Dame RT (2007) H-NS promotes looped domain formation in the bacterial chromosome. Curr Biol 17: R913-R914.
54. Wang W, Li G-W, Chen C, Xie XS, Zhuang X (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. Science 333: 1445–1449.
55. Hermsen R, Rein ten Wolde P, Teichmann S (2008) Chance and necessity in chromosomal gene distributions. Trends Genet 24: 216–219.
56. Bremer H, Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhart FC, III RC, Ingraham JL, Lin ECC, Low KB et al., editors. Escherichia coli and Salmonella: Cellular and Molecular Biology. 2nd edition. Washington: ASM Press.
57. Gouy M, Gautier C (1982) Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res 10: 7055–7074.
58. Akashi H (2001) Gene expression and molecular evolution. Curr Opin Genet Dev 11: 660–666.
59. Ihmels J, Levy R, Barkai N (2004) Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae. Nat Biotechnol 22: 86–92.
60. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A (2011) Information transduction capacity of noisy biochemical signaling networks. Science 334: 354–358.
61. Becskei A, Kaufmann BB, van Oudenaarden A (2005) Contributions of low molecule number and chromosomal positioning to stochastic gene expression. Nat Genet 37: 937–944.
62. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biol 4: e309.
63. Wang G-Z, Lercher MJ, Hurst LD (2011) Transcriptional coupling of neighboring genes and gene expression noise: evidence that gene orientation and noncoding transcripts are modulators of noise. Genome Biol Evol 3: 320–331.
64. Yi G, Sze S-H, Thon MR (2007) Identifying clusters of functionally related genes in genomes. Bioinformatics 23: 1053–1060.
65. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, et al. (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 36: 1065–1071.
66. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, et al. (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet 42: 53–61.
67. Dong X, Li C, Chen Y, Ding G, Li Y (2010) Human transcriptional interactome of chromatin contribute to gene co-expression. BMC Genomics 11: 704.