Queueing up for enzymatic processing: correlated signaling through coupled degradation

Natalie A Cookson[1,5]**, William H Mather**[2,5]**, Tal Danino**[2]**, Octavio Mondragón-Palomino**[2]**, Ruth J Williams**[3,4]**, Lev S Tsimring**[4,6] **and Jeff Hasty**[1,2,4,6,]*

[1] Molecular Biology Section, Division of Biological Science, University of California, San Diego, CA, USA; [2] Department of Bioengineering, University of California, San Diego, CA, USA; [3] Department of Mathematics, University of California, San Diego, CA, USA and [4] BioCircuits Institute, University of California, San Diego, CA, USA
[5] These authors contributed equally to this work
[6] These authors are co-senior authors
* Corresponding author. Molecular Biology Section, Division of Biological Science, University of California, 9500 Gilman Drive, San Diego, La Jolla, CA 92093-0412, USA. Tel.: +1 858 822 3442; Fax: +1 858 534 7664; E-mail: hasty@ucsd.edu

High-throughput technologies have led to the generation of complex wiring diagrams as a post-sequencing paradigm for depicting the interactions between vast and diverse cellular species. While these diagrams are useful for analyzing biological systems on a large scale, a detailed understanding of the molecular mechanisms that underlie the observed network connections is critical for the further development of systems and synthetic biology. Here, we use queueing theory to investigate how 'waiting lines' can lead to correlations between protein 'customers' that are coupled solely through a downstream set of enzymatic 'servers'. Using the *E. coli* ClpXP degradation machine as a model processing system, we observe significant cross-talk between two networks that are indirectly coupled through a common set of processors. We further illustrate the implications of enzymatic queueing using a synthetic biology application, in which two independent synthetic networks demonstrate synchronized behavior when common ClpXP machinery is overburdened. Our results demonstrate that such post-translational processes can lead to *dynamic* connections in cellular networks and may provide a mechanistic understanding of existing but currently inexplicable links.
Molecular Systems Biology **7**: 561; published online 20 December 2011; doi:10.1038/msb.2011.94
Subject Categories: metabolic & regulatory networks; synthetic biology
Keywords: ClpXP; protein degradation; queueing; synthetic biology

Introduction

Evolutionary pressure has driven organisms to develop into energy efficient machines, which conserve resources by minimizing biosynthetic costs and producing only the required amount of certain costly proteins (Warner *et al*, 2001; Vilaprinyo *et al*, 2010; Barton *et al*, 2010). An implication of maintaining only the minimal amount of critical machinery is the potential to overload important pathways during times of stress and to place unanticipated burden on important cellular workhorses. This can lead to the development of 'waiting lines' for biochemical processing which may cause apparent correlations between seemingly disconnected components that share the same processing pathways. In that sense, an analogy can be drawn to multiclass queueing theory (Kelly, 1979; Bramson, 1998; Williams, 1998; Bramson and Dai, 2001), which we employ here to provide a unifying model for describing how 'waiting lines' for processing by a common enzyme (the 'servers') can lead to correlations between two otherwise uncoupled proteins (the 'customers').

Generally, when strong correlated behavior is observed between two proteins in response to some perturbation, it is assumed that there is a direct coupling mechanism in place, such as correlated transcription. However, recent studies have revealed a lower degree of correlation between mRNA and protein levels than expected, indicating the need to search for other coupling mechanisms that may lead two protein species to follow similar trends in concentration (Gygi *et al*, 1999; Futcher *et al*, 1999; Greenbaum *et al*, 2003; Guo *et al*, 2008). Here, we use several experimental approaches along with a new application of queueing theory to reveal that a seemingly minor form of indirect coupling between cellular species can lead to surprisingly strong correlated behavior. As a model system, we use the native *E. coli* protease, ClpXP, as the 'server' and impose various static and dynamic conditions of underloading and overloading the cells with protein 'customers' targeted for destruction by this complex. We demonstrate that the transition from an underloaded to an overloaded regime can manifest itself in significant cross-talk between two independent networks, where the induction level of one protein substantially affects the mean and variability of the

other protein. We also apply queueing theory to a native stress response system and demonstrate the use of this coupling property as a signaling mechanism to alert the cell to a dangerous environment.

The correlated behavior that arises from this type of indirect coupling can have many implications, both in the analysis of native cellular networks and in the design and construction of synthetic networks. For example, systems biology employs high-throughput technologies to reconstruct cellular networks and generate high level wiring diagrams (Alon *et al*, 1999; Golub *et al*, 1999; Ideker *et al*, 2001; Sauer, 2004; Li *et al*, 2004). While useful as tools for analyzing and understanding biological networks on a large scale, determining *how* these components are connected is the next critical step in understanding what the underlying interactions are and how they lead to the observed cellular behaviors. Similarly, the field of synthetic biology relies on a fundamental understanding of the relationship between cellular networks and the behaviors that emerge from their complex interactions (Elowitz and Leibler, 2000; Gardner *et al*, 2000; Hasty *et al*, 2002; You *et al*, 2004; Del Vecchio *et al*, 2008; Tabor *et al*, 2009; Tan *et al*, 2009; Danino *et al*, 2010; Kim and Sauro, 2010). As many of the emerging studies in synthetic biology aim to develop circuits that exhibit specific, dynamic behaviors, understanding and utilizing both direct and indirect coupling mechanisms will become essential to designing successful synthetic systems.

Results

To illustrate how queueing theory can be used to analyze indirect coupling, suppose that two proteins X_1 and X_2 are involved in signaling pathways that do not directly interact, but they are processed downstream by the same enzyme. If the enzymatic 'servers' (S) are in abundance relative to the the number of target 'customer' molecules (x_1 and x_2), then there are no waiting lines and this corresponds to an *underloaded* system in queueing theory (Figure 1A, left). On the other hand, if the number of servers is small compared with the number of customers, then the system becomes *overloaded* as the customers must wait in line to be processed (Figure 1A, right). Such an overloaded system introduces coupling between the different types of customers. For example, consider an increase in the number of X_1 molecules on the right side of Figure 1A. Conceptually, a rise in the number of X_1 molecules will increase the mean waiting time for the processing of X_2, leading to a decrease in the effective rate at which X_2 leaves the system. In other words, for fixed arrival rate of X_2, the mean number of X_2 will increase as the number of X_1 is increased, even though there is no direct coupling between the two protein species.

If the arrival of the two proteins X_1 and X_2 in Figure 1A is governed by Poissonian statistics, with production rates λ_1 and λ_2, respectively, the enzymatic system can be mapped to a queueing system where analytical formulae for the steady-state distributions can be derived (Mather *et al*, 2010; see Mather *et al*, 2011 for discussion of transients). While the resulting formulae are somewhat complex, they can be used to predict several generic properties that should arise in the context of coupled enzymatic processing. For example, for a fixed production rate λ_2 of X_2, the mean level of X_2 increases as the production rate for X_1 is increased. Initially, this increase is slow since the system is in the underloaded regime, but as
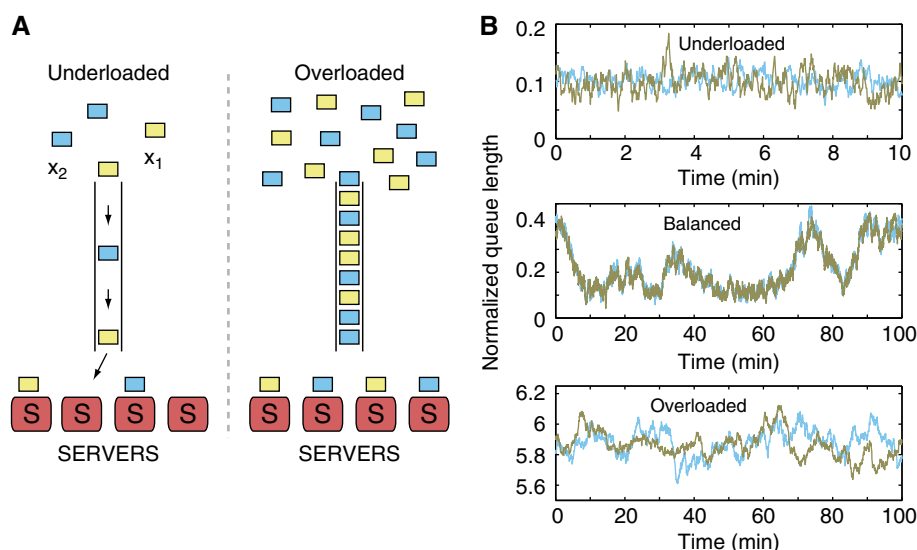

Figure 1 Coupling via common enzymatic machinery: connection to queueing theory. (**A**) Rate-limited processing can couple the numbers of different job types in a queue. Different jobs (yellow and cyan squares) are forming a queue to be processed and removed from the queue by a fixed number of servers with finite processing capacity. If the arrival rate exceeds the maximum processing rate, the servers become overloaded, the queue lengthens dramatically, and the numbers of jobs competing for the attention of the servers become tightly correlated. (**B**) Individual stochastic trajectories for a queueing system in three different conditions demonstrate correlation resonance. We consider the system with $L=100$ servers, each with a processing rate of $\mu=10\,\text{min}^{-1}$. Jobs renege, that is, abandon the system, at the first-order rate $\gamma=\ln2/20.0\,\text{min}^{-1}$. Yellow corresponds to jobs of type 1 and count x_1, while blue corresponds to jobs of type 2 and count x_2. Trajectories are normalized by their arrival rates. (Top) Underloaded condition, with $\lambda_1=0.25\,\mu L$, the arrival rate of job 1, and $\lambda_2=0.5\,\mu L$, the arrival rate of job 2. The total rate of arrival $\Lambda=0.75\,\mu L$ is less than the total processing rate μL. (Middle) Balanced condition, with $\lambda_1=0.5\,\mu L$, $\lambda_2=0.5\,\mu L$, and $\Lambda=\mu L$. (Bottom) Overloaded condition, with $\lambda_1=0.75\,\mu L$, $\lambda_1=0.5\,\mu L$, and $\Lambda=1.25\,\mu L$.
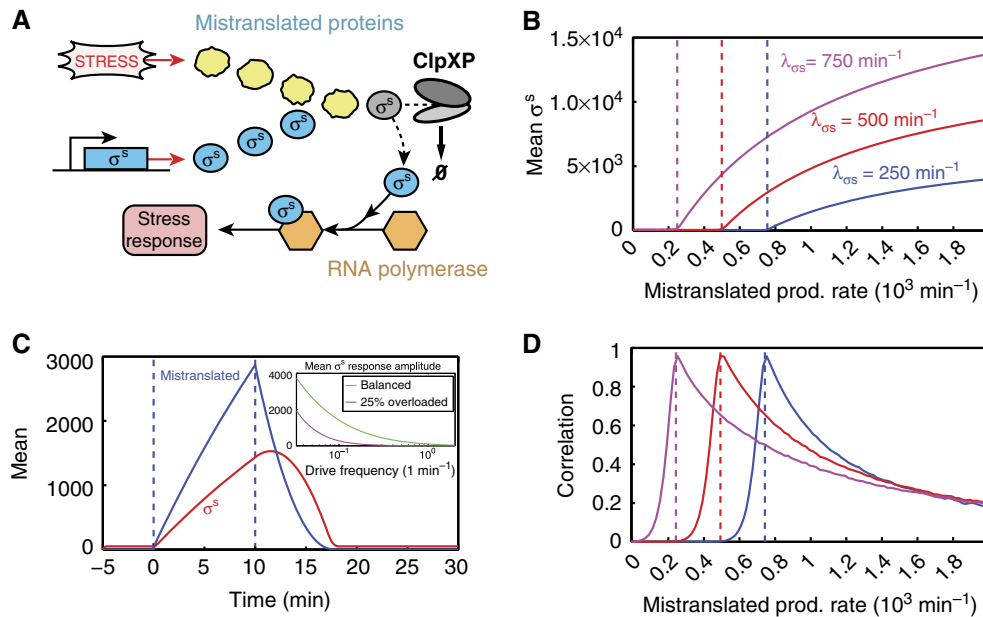
Figure 2 The *E. coli* stress response network employs queueing as a signaling mechanism to ensure the most rapid response possible to adverse conditions. (**A**) Certain types of stress cause the accumulation of a large amount of aberrant proteins, which are targeted for rapid degradation and compete with the master stress regulator, σ^s, for a limited amount of ClpXP machinery. This leads to a decrease in the effective degradation rate of σ^s, allowing it to build up rapidly and initiate an immediate response. See the Supplementary Information for a precise definition of the model and for simulation details. (**B**) For a stochastic queueing model with 100 ClpXP molecules that each have processing rate $\mu=10$ min^{-1}, with cells dividing every 20 min, a scan of the mean steady-state level of σ^s with respect to the stress level (mistranslated protein production rate λ_m) demonstrates a very sensitive response of the system once the system has crossed the balance point (respectively colored dashed lines). Different σ^s basal production rates $\lambda_{\sigma s}$ are indicated in the figure. (**C**) The dynamic response of σ^s to a 10-min pulse of stress demonstrates that queueing provides for a very fast and dynamic response. Importantly, the response is highly specific temporally, as σ^s only remains at a high level during the time that excess aberrant proteins are around. We assume the basal rate $\lambda_{\sigma s}=500$ min^{-1}, while $\lambda_m=1000$ min^{-1} during the pulse of stress but $\lambda_m=0$ min^{-1} otherwise. (Inset) The mean response amplitude of σ^s to a periodic stress is strongest, especially at fast frequencies, when the system is on average near balance or slightly overloaded. The rate λ_m is taken to be a constant plus a sinusoid with amplitude 100 min^{-1} and given frequency. (**D**) The adaptive response leads to positive correlations between σ^s and mistranslated protein levels for a broad range of parameters, peaking near the balance point (respectively colored dashed lines). Parameters are the same as those used in (B).

waiting lines begin to lengthen, the system transitions to the overloaded regime and the mean of X_2 rapidly rises as λ_1 is increased. At the transition point between the underloaded and overloaded regimes (the 'balance point'), the sum of λ_1 and λ_2 is equal to the processing capacity of the queueing network. Interestingly, queueing yields a fast, adaptive response for stress that is dynamic and noisy. Under certain conditions, fluctuations in X_1 and X_2 are highly coupled, with correlations peaking near the balance point (Figure 1B).

The stress response to nutrient starvation in *E. coli* has recently been shown to elicit a rapid and adaptive response through the saturation of downstream enzymatic machinery (Fredriksson *et al*, 2007); that is, in the context of queueing theory, the stressed system operates in the overloaded regime. This network relies on the functionality of the protease ClpXP, a well-studied example of shared enzymatic machinery that targets many types of proteins for degradation (Figure 2A; Levchenko *et al*, 2000; Sauer *et al*, 2004). In low stress conditions, the concentration of the master stress regulator, the sigma factor σ^s, is maintained at a low level through its rapid degradation via ClpXP. However, when the cell is subjected to nutrient starvation, an increased number of mistranslated proteins are targeted for degradation by ClpXP and they compete for a limited number of the proteases. This increases the half-life of σ^s, which builds up, signals starvation, and jump-starts the stress response. Queueing

theory provides several insights into both the dynamic behavior of this response and the benefits that this type of indirect regulation can confer on a network. First, by looking at how the mean level of σ^s depends on the relative stress level (Figure 2B), we observe that queueing yields a very sensitive response that would normally be assumed to imply some type of strong cooperativity (i.e., a very high Hill coefficient). Second, looking at the temporal response of the σ^s level to a transient stress, it is clear that queueing confers a very fast and dynamic response. Importantly, the response is highly specific temporally, as σ^s only remains at a high level during the time that excess aberrant proteins are around (Figure 2C). Scanning the production rate of mistranslated proteins, positive correlations occur for a large range of parameters, peaking near the balance point (Figure 2D; Mather *et al*, 2010).

To systematically investigate the dynamic properties of a biological system in which queueing acts as a regulatory mechanism, we constructed a synthetic system to overexpress two different tagged proteins from separate and uncorrelated promoters (Figure 3A). The $P_{LtetO-1}$ promoter, used to drive expression of YFP (yeast-enhanced yellow fluorescent protein, venus), is tightly repressible by the *Tet* repressor (TetR) and can be regulated over a range of up to 5000-fold by supplying doxycycline to the culture (Lutz and Bujard, 1997). The hybrid promoter, $P_{lac/ara-1}$, used to drive expression of CFP (yeast-enhanced cerulean fluorescent protein), is tightly repressed by
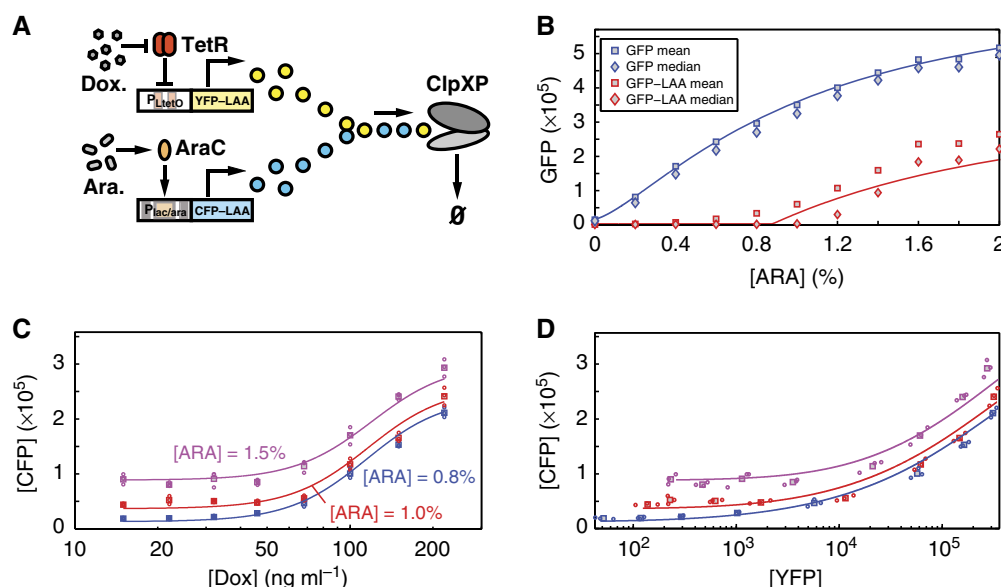
Figure 3 Coupled enzymatic degradation of yellow and cyan LAA-tagged fluorescent proteins by ClpXP machinery in *E. coli*. (**A**) Schematic network diagram of the synthetic circuit. YFP is produced by the $P_{LtetO-1}$ promoter, which is repressed by TetR in the absence of doxycycline. CFP is produced by the $P_{lac/ara-1}$ promoter, which is activated by AraC in the presence of arabinose. Both CFP and YFP molecules are tagged with identical LAA tags and are targeted for degradation by the ClpXP complexes. (**B**) Induction plots for a single fluorescent protein (GFP) produced by the $P_{lac/ara-1}$ promoter. IPTG is held at 1 mM to fully relieve repression by lacI, and production is tuned by addition of arabinose. Blue and red symbols indicate untagged and LAA-tagged protein, respectively. Squares are mean protein counts, while diamonds are median protein counts. Solid lines are steady-state model fits to the data (including those in **C** and **D**). The red line stochastic queueing model prediction for enzymatic protein degradation compares favorably with the data. (**C**) Mean steady-state expression of CFP as a function of doxycycline concentration at three different levels of arabinose in triplicate flow cytometry measurements. Strong coupling is observed between CFP and YFP. Protein counts are reported using a combination of two-color flow cytometry and western blots. (**D**) The means of CFP and YFP increase simultaneously as the doxycycline concentration is increased. The color of the symbols corresponds to (**C**). In both (**C**) and (**D**), results for the stationary state of a fitted stochastic queueing model are included as solid lines. Supposing a doubling time $\tau_d \approx 30$ min, we find an enzymatic degradation rate $\mu = 7.6 \times 10^3$ min^{-1} for the model provides a good fit to the plotted results. Values for the production rates of YFP at given dox concentrations and for the production rates of CFP at given arabinose concentrations were determined from a best fit to the data. The qualitative result of this fit is that CFP only measurably increases when YFP becomes comparable in magnitude, consistent with a slightly overloaded queue. See Supplementary Information (Supplementary Figures S1, S5, and S6) for fitting details and parameters. Source data is available for this figure in the Supplementary Information.

the *Lac* repressor (LacI) and activated by AraC. It can be regulated over a range of up to 1800-fold in the presence of IPTG and arabinose in the culture (Lutz and Bujard, 1997). For high expression from the $P_{lac/ara-1}$ promoter, we used 1 mM of IPTG in all samples to fully relieve repression by LacI, and we used various levels of arabinose to tune the induction level of CFP. Both YFP and CFP were tagged on their C terminus with the well-characterized 11-residue 'LAA' tag (AANDENYALAA), marking them as targets for rapid degradation by ClpXP (Keiler *et al*, 1996). To ensure stable synthesis and maintenance of the regulatory proteins, the synthetic system was transformed into an *E. coli* DH5 α Z1 host that produces constitutive levels of TetR, LacI and AraC off of the chromosome (Lutz and Bujard, 1997), and controls were performed to ensure that there was no direct cross-talk between the two promoters (see Supplementary Figure S4).

Our queueing analysis of the enzymatic decay of highly expressed proteins predicts that at certain levels of expression, the mean level of one protein will be significantly coupled to the mean of the other. To test this hypothesis, we planned to hold production of one protein constant, while tuning the level of the other. As a first step, we acquired induction data for the $P_{lac/ara-1}$ promoter driving GFP, to determine several arabinose levels to fix for the constant condition. Interestingly, we found a striking difference in the shapes of the induction curves for tagged and untagged variants of the fluorescent

reporter (Figure 3B). While the untagged protein responded in the expected manner (Lutz and Bujard, 1997), the tagged protein data showed clear evidence of queuing as predicted by the theoretical model (Figure 2B), with a sharp bend in the curve indicating the balance point, where ClpXP transitions from the underburdened to overburdened state.

We then used two-color flow cytometry to generate induction curves for the two systems, fixing arabinose at three levels and tuning YFP production with doxycycline, and we were surprised by the extent to which the inducer for one species affected the other species. For example, as doxycycline was varied over a full induction range for YFP, we observed up to a 12-fold increase in the mean level of CFP (Figure 3C). Likewise, we observed unambiguously that the two proteins exhibited highly correlated behavior as we swept doxycycline and plotted the mean levels of CFP as a function of the mean levels of YFP (Figure 3D). This result highlights an important point in the general context of gene-regulatory and signaling networks; were this data collected in the absence of a mechanistic understanding, one would likely conclude that YFP is some form of inducer for CFP.

To place this queueing phenomenon in the context of native cellular behavior, it is important to develop a quantitative understanding of the system's balance point and the processing capacity of ClpXP in healthy *E. coli* cells. To convert mean fluorescence data to a more informative protein count, we
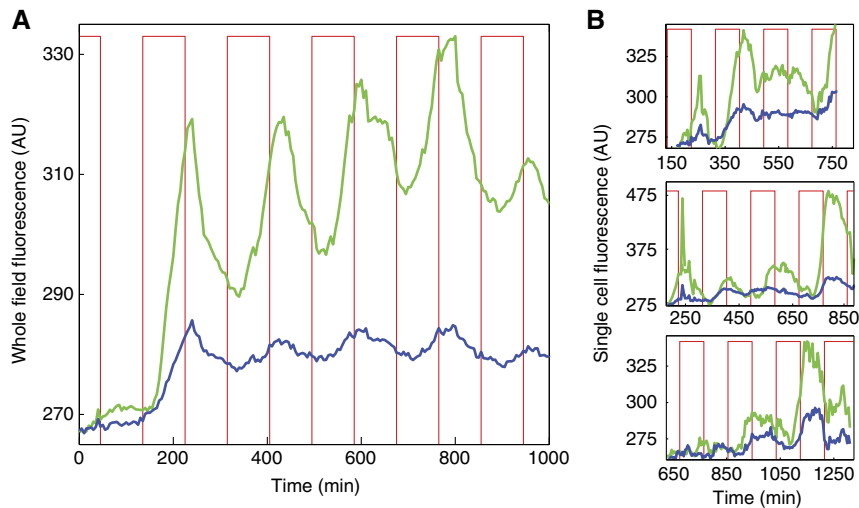
Figure 4 Dynamic behavior of a synthetic signaling network. (**A**) Using a microfluidic platform capable of generating a time-dependent induction signal and multi-color single-cell fluorescence measurements, a large population of *E. coli* expressing the synthetic network were subject to a periodic series of doxycycline pulses (red), such that the system at zero doxycycline relaxed to a dim state, but the system at high doxycycline relaxed to a bright state. The total YFP and CFP fluorescence, integrated over the entire colony, demonstrates the direct response of the $P_{LtetO-1}$ promoter, producing YFP (green) as TetR is periodically deactivated, as well as the indirect response of the CFP signal (blue) due to the time-dependent saturation of the ClpXP machinery. (**B**) Trajectories for several individual cells demonstrate the response at a single-cell level. Source data is available for this figure in the Supplementary Information.

used a combination of flow cytometry and quantitative western blots (see Supplementary Figures S2 and S3) to calculate a scaling factor with which to directly compare YFP and CFP mean fluorescence levels as reported by the flow cytometer. We then employed western blots to determine a combined mean number of tagged YFP and CFP molecules per cell for various induction levels. The combination of these two approaches provides the numerical information necessary to translate the typical 'arbitrary units' to a number of proteins and yields a balance point value of 7.6 K tagged proteins produced (and degraded) per minute for this system. While healthy cells are believed to be in the underloaded regime (Moore and Sauer, 2005), in which they would not possess queues of excess tagged proteins, it is clear that certain stressful conditions, like carbon starvation, will push the degradation system far beyond the balance point (Neher *et al*, 2006; Ferenci, 2007; Fredriksson *et al*, 2007; Bougdour *et al*, 2008; Hengge, 2009). In fact, this appears to be a beneficial design feature of the *E. coli* stress response network, as the degradation delay allows the master regulator, σ^s, to build up rapidly for immediate response but to be removed and recycled quickly after the system is repaired.

The favorable comparison between model and experiment indicates that queueing theory provides a quantitative approach for describing coupled enzymatic processing (Figure 3B–D, solid lines). The model fits were derived through use of a fitting algorithm to determine model parameters μ, γ, K, a Hill function-based parameterization for λ_1 (production rate of YFP for a given doxycycline), and the parameterization for λ_2 (production rate of CFP for a given arabinose). Interestingly, although the model involves many parameters, the shape of the fitted curves depends only on a few parameters, particularly the enzymatic degradation rate μ (see Supplementary information). This indicates that the model describes the system well and that we are not overfitting the data.

To further investigate the implications of enzymatic queueing, we designed a microfluidic platform based on previous designs (Bennett *et al*, 2008; Ferry *et al*, 2011) to drive and monitor the signaling responses of the two networks at the single-cell level. We drove production of YFP with a periodic square-wave signal of doxycycline (see Materials and methods summary) and used two-color microscopy to observe the response of both the YFP and CFP signals. Whole-field fluorescence of a population of *E. coli* cells demonstrates how the coupling of the two proteins through the shared degradation pathway serves to drive one system in response to the behavior of the other, as both the YFP (Figure 4A, green) and CFP (Figure 4A, blue) signals oscillate with the driving signal (Figure 4A, red). Similar trends can be observed in the fluorescence trajectories of individual cells (Figure 4B). Correlated behavior between the two reporters is observed in response to the driving signal as well as in long-term expression trends (Figure 5).

Based on these results, we hypothesized that the effect of queueing through coupled enzymatic processing could have significant implications in the developing field of synthetic biology. As most synthetic biology approaches employ degradation tags to enforce the necessarily quick turnover of key network components, waiting times for processing by ClpXP (or similar machinery in other biological systems) may be exploited to generate synchronized behavior among seemingly independent circuits. To test this hypothesis, we constructed a synthetic two-color system involving two independently produced fluorescent proteins: one (GFP–LAA) produced periodically by a synthetic oscillator and the second (CFP–LAA) produced by a separately tunable promoter (Figure 6A). We built a new single-plasmid version of a previously reported synthetic oscillator (Stricker *et al*, 2008) with the activator (araC–LAA), repressor (LacI–LAA), and reporter (GFP–LAA) all on a p15A plasmid. For the second
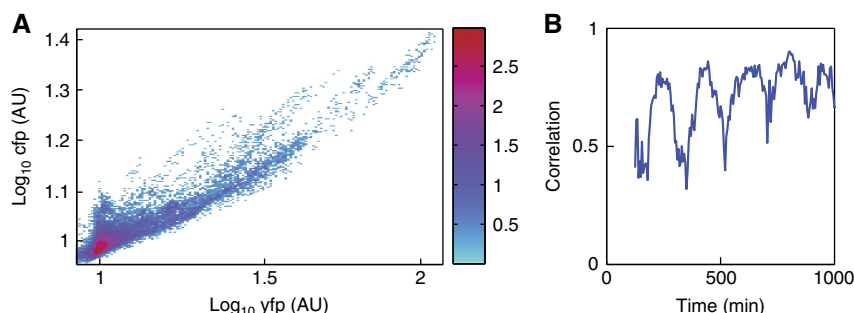
Figure 5 Correlations between two independent reporter proteins. (**A**) A two-dimensional histogram depicts correlations between YFP and CFP levels in individual cells throughout the entire experiment duration from Figure 4. The value in each rectangular bin is log scaled (value of $\log_{10}(1+n)$, with *n* the number of counts in a bin). (**B**) The Pearson's correlation coefficient between the two reporters as a function of time. Correlation is reported for the middle 80% of cells by CFP brightness (eliminating outliers that contribute bias to the correlation coefficient), and the estimated correlation coefficient for typical cells is observed to oscillate with the drive period.
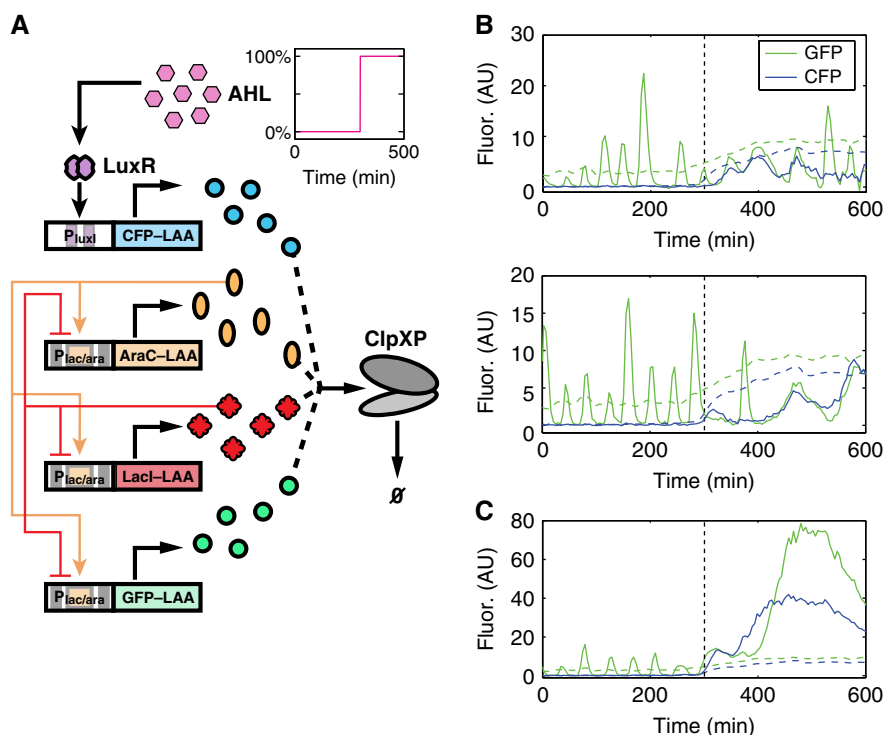

Figure 6 Coupled degradation can serve to indirectly couple gene circuits. (**A**) Circuit diagram for a variant of a synthetic gene oscillator discussed previously (Stricker *et al*, 2008), expressed alongside an AHL-inducible, LAA-tagged CFP (see Supplementary Information for details). Competition for enzymatic degradation among LAA-tagged proteins allows CFP to interact with the oscillator by slowing the degradation of the oscillator components. Since fast enzymatic decay is thought to strongly influence the period and robustness of certain gene oscillators (Wong *et al*, 2007; Mather *et al*, 2009), the character of GFP oscillations should be closely tied to CFP expression. (**B**) Cells containing the circuit in (A) were grown in a microfluidic device to test the influence of CFP production on GFP oscillations (see Supplementary Information for details). Two single-cell trajectories for GFP and CFP fluorescence (solid lines) show regular oscillations in GFP fluorescence in the absence of external AHL, that is, at low CFP fluorescence. However, addition of 15 nm AHL (time of induction start is indicated by a vertical dotted line) introduces CFP into the system, causing the GFP oscillations to slow and the CFP signal to oscillate as a result of indirect coupling due to queueing. Coupling is also observed in the mean fluorescence across a region of cells (mean fluorescence as dashed lines), where increasing mean CFP fluorescence is associated with an increase in mean GFP fluorescence. (**C**) A representative trajectory where both CFP and GFP become and remain bright simultaneously. Positive coupling between trajectories is apparent.

component, we placed the *LuxI* promoter driving CFP–LAA along with a constitutively produced LuxR on a high copy pUC19 vector, such that the expression of CFP is inducible by the addition of AHL to the medium.

We loaded this system into our microfluidic platform and imaged cells for several hours without AHL, and we found

regular oscillations of GFP–LAA with an average period of 28 min, while CFP–LAA levels were negligible. However, following induction with 15 nM AHL, the mean levels of both CFP and GFP increased, and CFP trajectories rapidly became correlated with GFP trajectories. Some cells produced irregular oscillations of both colors with much longer periods, while

others stopped oscillating altogether (Figure 6B and C). These results indicate that the additional production of tagged proteins by an independent circuit had a dramatic effect on the behavior of the synthetic oscillator, due to an increased burden on the degradation machinery.

Discussion

The steady-state induction characteristics and dynamic coupling observed in the two synthetic systems provide unambiguous evidence of how queueing for common enzymatic processing can induce coupling when there is a significant abundance of proteins relative to the number of functional enzymes. If such behavior was observed in a native or uncharacterized system, it would likely be assumed that these two proteins were coupled in a direct way, such as by coordinated gene expression or by a protein–protein interaction. As large-scale wiring diagrams have become a post-sequencing paradigm for depicting the interactions between vast and diverse cellular species, a major challenge is the deduction of the molecular interactions that underlie the observed correlations. Our results demonstrate that strong correlated behavior can be observed between two components that are only indirectly coupled via an overloaded enzymatic process and suggest that indirect coupling sources should be considered when evaluating native systems or designing and constructing functional synthetic networks.

By quantifying the number of tagged proteins observed in our experiments instead of using the typical 'arbitrary units,' we were able to generate a more precise numerical queueing model and to approximate the true 'balance point' in the natural system, above which the number of tagged proteins will likely lead to correlated behavior. While healthy cells are believed to be in the underloaded regime, in which they would not possess significant queues of excess tagged proteins, it is clear that stressful conditions can push the degradation system beyond the balance point, causing correlations between many proteins as a result. In fact, this may be a beneficial design feature of some stress response networks, where the degradation delay allows the required proteins to build up rapidly for immediate response but to be removed and recycled quickly after the system is repaired. There are likely many other examples of enzymes that exist in small numbers by design to confer specific behaviors to biological networks. For example, the yeast scaffold protein Ste5 is maintained in limited amounts to concentrate its targets near the common backbone (Chapman and Asthagiri, 2009).

Similarly, it may be beneficial to exploit these indirect links when designing new synthetic systems. As many of the interesting studies in synthetic biology focus on the creation of circuits that exhibit precise, dynamical behavior, targeted degradation of key network components has become an almost essential characteristic of synthetic systems. Since our study demonstrates how the use of degradation tags can lead to unexpected correlations, the results will have important implications on efforts to establish a 'forward engineering' methodology. That is, it may be both critical to consider coupled degradation in the modeling of genetic circuits and potentially interesting to intentionally incorporate waiting lines into the design criteria of novel synthetic systems to enhance the desired behavior.

Materials and methods

Strains and flow cytometry

The plasmids were constructed using the pZE24-mcs2a cloning plasmid, which has a kanamycin resistance marker and the hybrid $P_{lac/ara-1}$ promoter upstream of a multiple cloning site (mcs) (Lutz and Bujard, 1997). For the dual-color plasmid, pNO-2cLAA, the sequence for CFP was tagged by PCR with a carboxy-terminal ssrA tag (AANDENYALAA) (Keiler *et al*, 1996) and inserted between the *Kpn*I and *Hind*III sites of the mcs, creating pZE24-CFP–LAA. The YFP fragment was similarly tagged and inserted onto the pZS31-mcs2a cloning plasmid, which contains a chloramphenicol resistance marker and the $P_{LtetO-1}$ promoter upstream of an mcs. The fragment of this plasmid containing the marker gene and $P_{LtetO-1}$ driving YFP was copied by PCR and inserted into the *Sac*I site of pZE24-CFP–LAA, creating the final plasmid pNO-2CLAA, containing the independently controlled fluorescent proteins. The plasmids for comparing the tagged and untagged GFP induction curves were constructed similarly on the same plasmid backbone.

For the synthetic oscillator coupled to a constitutive system, the medium copy oscillator (pTDCL8) was constructed by combining oscillator components from pJS167 and pJS169 (Stricker *et al*, 2008) onto a single p15A plasmid. The pJS167 module (from *Sac*I up to *Avr*II) was copied by PCR with Kan resistance and flanking *Avr*II and *Nhe*I sites, and it was inserted via ligation at the *Avr*II site of the pZA14-LacI vector, maintaining uniqueness of the *Avr*II site and creating an undigestable *Avr*II-*Nhe*I hybrid site. The pLuxI-CFP plasmid (pZU25-CFP–LAA) was constructed by building pZE25-CFP–LAA and inserting the promoter + CFP module by PCR onto a pUC19 vector via flanking *Avr*II sites (U=pUC19, 5=luxIp promoter).

Flow cytometry data were taken with a Becton-Dickinson LSR II Cell Analyzer, fitted with 405 and 488 nm lasers. The cells were grown overnight in non-inducing medium: LB plus kanamycin for plasmid selection. The cells were passed in the morning into LB plus kanamycin plus various levels of inducer, doxycycline, arabinose/IPTG, or both. The cells were grown in a 37°C shaker at 300 r.p.m. After 3 h, the cells were harvested by centrifugation, resuspended in sterile PBS, and put on ice until they were ready to be sampled. Using the LSR II, 100 000 cells were assayed using MATLAB (The MathWorks, Inc.) and interfacing software. The SBML file of the model used is archived at BioModels (MODEL1111150000).

Microfluidics and microscopy

Image acquisition was performed on a Nikon Eclipse TI epifluorescent inverted microscope outfitted with fluorescence filter cubes optimized for YFP and CFP imaging and a phase-contrast-based autofocus algorithm. Images were acquired using a Photometrics CoolSNAP HQ2 cooled CCD camera, controlled by Nikon Elements software. For the signaling experiment, images were acquired every 1 min in phase contrast, to provide the optimal temporal coverage to suit the automated tracking program. Fluorescent images in the CFP and YFP channels were acquired every 5 min. The cells were imaged inside a microfluidic device with an upstream switch, with the ability to mix or switch between two different media sources. A custom application written in LabVIEW (National Instruments, Austin, Texas) controlled linear actuators, to which two reservoirs containing inducing and non-inducing medium were attached. Using this algorithm, a square wave of 3 h period was generated, subjecting the cells to alternating 90 min terms of 220 ng ml^{-1} and 0 ng ml^{-1} doxycycline. For the synthetic circuit experiment, CFP–LAA was induced with 15 nM AHL (*N*-3-Oxo-hexanoyl-homoserine lactone) from the *LuxI* promoter as described in the main text. Cells were imaged every 30 s in brightfield and every 4.5 min in both CFP and YFP channels. Controls were performed to ensure that IPTG and Dox did not affect the incorrect promoter and that lacI and tetR do not repress the incorrect promoter, as well as to ensure there was not significant overlap between the CFP and YFP channels, even when imaging CFP and GFP.

The microfluidic experiments were performed as previously described (Danino *et al*, 2010). Briefly, 50 μl of an overnight culture was diluted in 50 ml of LB (Difco) + antibiotics the day of the experiment.

When cells reached an OD_{600} of 0.1, cells were spun down and resuspended in 5 ml of fresh media and loaded into the device. Image processing was performed using a custom application tied into the ImageJ image processing suite. Images were segmented by creating a binary mask to identify individual cells, and cells were tracked from frame to frame using a combination of quantitative characteristics. The output of this process is the single-cell trajectory information provided in the main text.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alon U, Barkai N, Notterman D, Gish K, Ybarra S, Mack D, Levine A (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* **96:** 6745

Barton M, Delneri D, Oliver S, Rattray M, Bergman C, B 'ahler J (2010) Evolutionary systems biology of amino acid biosynthetic cost in yeast. *PloS One* **5:** e11935

Bennett MR, Pang WL, Ostroff NA, Baumgartner BL, Nayak S, Tsimring LS, Hasty J (2008) Metabolic gene regulation in a dynamically changing environment. *Nature* **454:** 1119–1122

Bougdour A, Cunning C, Baptiste P, Elliott T, Gottesman S (2008) Multiple pathways for regulation of σ^S (RpoS) stability in Escherichia coli via the action of multiple anti-adaptors. *Mol Microbiol* **68:** 298–313

Bramson M (1998) State space collapse with application to heavy traffic limits for multiclass queueing networks. *Queueing Syst* **30:** 89–148

Bramson M, Dai JG (2001) Heavy traffic limits for some queueing networks. *Ann Appl Probab* **11:** 49–90

Chapman S, Asthagiri A (2009) Quantitative effect of scaffold abundance on signal propagation. *Mol Syst Biol* **5:** 313

Danino D, Mondragón-Palomino O, Tsimring L, Hasty J (2010) A synchronized quorum of genetic clocks. *Nature* **463:** 326–330

Del Vecchio D, Ninfa A, Sontag E (2008) Modular cell biology: retroactivity and insulation. *Mol Syst Biol* **4:** 161

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403:** 335–338

Ferenci T (2007) Sensing nutrient levels in bacteria. *Nat Chem Biol* **3:** 607–608

Ferry M, Razinkov I, Hasty J (2011) Microfluidics for synthetic biology from design to execution. *Methods Enzymol* **497:** 295

Fredriksson A, Ballesteros M, Peterson C, Persson O, Silhavy T, Nystrom T (2007) Decline in ribosomal fidelity contributes to the accumulation and stabilization of the master stress response regulator σS upon carbon starvation. *Genes Dev* **21:** 862

Futcher B, Latter G, Monardo P, McLaughlin C, Garrels J (1999) A sampling of the yeast proteome. *Mol Cell Biol* **19:** 7357

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403:** 339–342

Golub T, Slonim D, Tamayo P, Huard C, Gaasenbeek M, Mesirov J, Coller H, Loh M, Downing J, Caligiuri M, Bloomfield CD, Lander ES (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286:** 531

Greenbaum D, Colangelo C, Williams K, Gerstein M (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* **4:** 117

Guo Y, Xiao P, Lei S, Deng F, Xiao G, Liu Y, Chen X, Li L, Wu S, Chen Y (2008) How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochim Biophys Sin* **40:** 426–436

Gygi S, Rochon Y, Franza B, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19:** 1720

Hasty J, McMillen D, Collins JJ (2002) Engineered gene circuits. *Nature* **420:** 224–230

Hengge R (2009) Proteolysis of σ^S (RpoS) and the general stress response in Escherichia coli. *Res Microbiol* **160:** 667–676

Ideker T, Thorsson V, Ranish J, Christmas R, Buhler J, Eng J, Bumgarner R, Goodlett D, Aebersold R, Hood L (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292:** 929

Keiler K, Waller P, Sauer R (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271:** 990

Kelly FP (1979) *Reversibility and Stochastic Networks*. John Wiley and Sons: Chichester

Kim K, Sauro H (2010) Fan-out in gene regulatory networks. *J Biol Eng* **4:** 1–14

Levchenko I, Seidel M, Sauer R, Baker T (2000) A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289:** 2354

Li S, Armstrong C, Bertin N, Ge H, Milstein S, Boxem M, Vidalain P, Han J, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF *et al* (2004) A map of the interactome network of the metazoan C. elegans. *Science* **303:** 540

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res* **25:** 1203–1210

Mather W, Bennett M, Hasty J, Tsimring L (2009) Delay-induced degrade-and-fire oscillations in small genetic circuits. *Phys Rev Lett* **102:** 068105

Mather W, Cookson N, Hasty J, Tsimring L, Williams R (2010) Correlation resonance generated by coupled enzymatic processing. *Biophys J* **99:** 3172–3181

Mather W, Hasty J, Tsimring L, Williams R (2011) Factorized time-dependent distributions for certain multiclass queueing networks and an application to enzymatic processing networks. *Queueing Syst* **69:** 313–328

Moore SD, Sauer RT (2005) Ribosome rescue: tmRNA tagging activity and capacity in Escherichia coli. *Mol Microbiol* **58:** 456–466

Neher S, Villén J, Oakes E, Bakalarski C, Sauer R, Gygi S, Baker T (2006) Proteomic profiling of ClpXP substrates after DNA damage reveals extensive instability within SOS regulon. *Mol Cell* **22:** 193–204

Sauer R, Bolon D, Burton B, Burton R, Flynn J, Grant R, Hersch G, Joshi S, Kenniston J, Levchenko I, Neher SB, Oakes ES, Siddiqui SM, Wah DA, Baker TA (2004) Sculpting the proteome with AAA + proteases and disassembly machines. *Cell* **119:** 9–18

Sauer U (2004) High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol* **15:** 58–63

Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J (2008) A fast, robust and tunable synthetic gene oscillator. *Nature* **456:** 516–519

Tabor J, Salis H, Simpson Z, Chevalier A, Levskaya A, Marcotte E, Voigt C, Ellington A (2009) A synthetic genetic edge detection program. *Cell* **137:** 1272–1281

Tan C, Marguet P, You L (2009) Emergent bistability by a growth-modulating positive feedback circuit. *Nat Chem Biol* **5:** 842–848

Vilaprinyo E, Alves R, Sorribas A (2010) Minimization of biosynthetic costs in adaptive gene expression responses of yeast to environmental changes. *PLoS Comput Biol* **6:** 184–1037

Warner JR, Vilardell J, Sohn JH (2001) Economics of ribosome biosynthesis. *Cold Spring Har Symp Quant Bio* **66:** 567–574

Williams RJ (1998) Diffusion approximations for open multiclass queueing networks: sufficient conditions involving state space collapse. *Queueing Syst* **30:** 27–88

Wong W, Tsai T, Liao J (2007) Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol Syst Biol* **3:** 130

You L, Cox RSI, Weiss R, Arnold FH (2004) Programmed population control by cell-cell communication and regulated killing. *Nature* **428:** 868–871

Queueing up for enzymatic processing:
Correlated signaling through coupled degradation
(Supplementary Information)

Natalie A. Cookson, William H. Mather, Tal Danino,
Octavio Mondragón-Palomino, Ruth J. Williams, Lev S. Tsimring, Jeff Hasty

SBML Model

The SBML file of the model used for this study has been archived at the BioModels Database (MODEL1111150000).

Flow Cytometry

Flow cytometry data was taken with a Becton-Dickinson LSR II Cell Analyzer, fitted with 405nm and 488nm lasers. Because a wide range of fluorescence intensities arose as inducer levels were scanned, the sensitivity (set by voltage) of photomultiplier tubes in the flow cytometer was varied from sample to sample. We calibrated the photomultiplier tubes by scanning a range of voltages for cells with a constant mean level of fluorescent protein, either YFP or CFP alone. We fit the resulting mean YFP and CFP intensity curves to piecewise-smooth functions of voltage. These functions were used to correct flow cytometry data by scaling all measurements to a common apparent voltage.

In order to be able to directly compare numbers of fluorescent proteins of YFP and CFP, we used a plasmid containing two copies of the $P_{LtetO-1}$ promoter, one driving YFP-LAA and one driving CFP-LAA (pZA11-YC-LAA). This was constructed using similar techniques to those described above. Using this strain and the assumption that the two proteins should be produced in equal mean levels due to their tandem arrangement on the same plasmid, we induced cells at various concentrations of doxycycline and measured mean fluorescence. We scaled CFP fluorescence such that the mean CFP fluorescence values at the selected induction levels were essentially the same as the corresponding YFP mean fluorescence values (difference of 0.5% in the typical mean fluorescence), and we were able to determine a conversion factor to compare YFP "arbitrary units" (AU) to CFP AU. This, combined with the Western blot data (see below) enabled the estimation of the total number of each fluorescent protein in each data set.

As another validation of the queueing theory, we compared this dual-color induction data to an almost identical plasmid, but one in which only YFP-LAA was produced from the $P_{LtetO-1}$ promoter. Interestingly, we found that the overall level of YFP fluorescence was significantly lower in the case of expression of only one of the two colors. This falls in line with queueing theory predictions. That is, in the dual-color case, the total number of tagged proteins is doubled, so ClpXP would be more overloaded than when just a single fluorescent protein type is produced. This is further evidence that over-burdening the protease can lead to coupling between the levels of two different proteins.

As described in the main text, we also compared induction curves of tagged and untagged fluorescent proteins, in an effort to determine if the queueing effect could be directly observed in induction data. This also served to demonstrate that the effects observed throughout the experiments were in fact due specifically to the abundance of tags, and not simply side effects of general over-expression. To take this data, we created two plasmids very similar to those used for the two-color study. We used the pZE24-mcs2a cloning plasmid [1], which has a kanamycin resistance marker and the hybrid $P_{lac/ara-1}$ promoter upstream of a multiple cloning site (mcs). We simply replaced the mcs with either GFP or GFP-LAA, and used this to take induction data as described in the main text.

For all flow results, a background subtraction procedure was performed on the raw data (after the voltage correction described above) to arrive at reported YFP and CFP fluorescence statistics. Using data from the experiments discussed in the main text, we defined the background mean of YFP and CFP fluorescence as that derived from cells induced with 1 mM IPTG alone (in the absence of arabinose or doxycycline). The background mean for each color was subtracted from the mean of raw data.
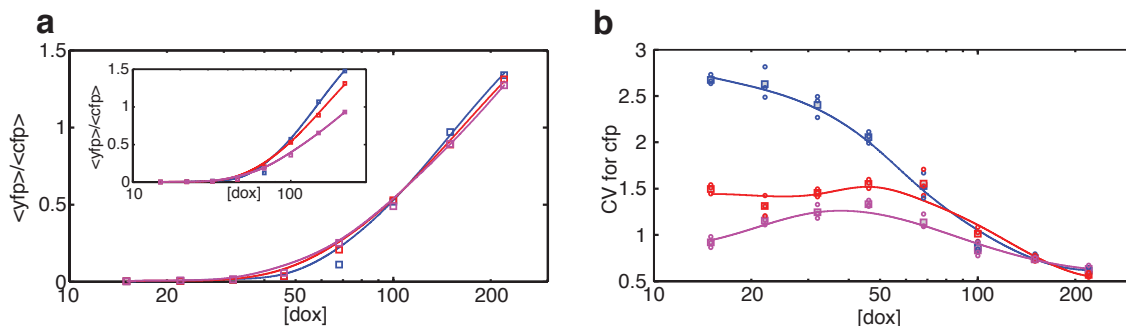

Figure S1: Coupled enzymatic degradation of yellow and cerulean LAA-tagged fluorescent proteins by ClpXP machinery in *E. coli*. (a) According to the stochastic queueing model, the ratio of the two mean concentrations $\langle x_1 \rangle / \langle x_2 \rangle$ is equal to the ratio of the corresponding production rates λ_1/λ_2. In accordance with the model, the ratio $\langle yfp \rangle / \langle cfp \rangle$ exhibits the same dependence on dox concentration for three different levels of arabinose, which allowed us to collapse all the data to a common curve by normalizing them by the mean value over the whole range of employed dox concentrations. Inset shows the same data without collapsing. (b) Coefficient of variation of CFP concentration decreases with increasing dox in the overloaded regime in qualitative agreement with the queueing theory predictions (different symbols correspond to three levels of arabinose concentration similar to panels c-d of Fig. 3 in the main text). Solid lines represent trend lines through the data.

As one further test of the stochastic queueing model, we used the theoretical results to deduce the scaling relationship between the data sets in Fig. 3c of the main text at different levels of arabinose. In other words, if correct, the theory can be used to predict how one can plot the data such that it will collapse onto the same curve. The resulting verification of this prediction further confirmed the general validity of the queueing theory approach (Fig. S1a). Lastly, we calculated the noise (as measured by the coefficient of variation) of the CFP signal as a function of increasing doxycycline (Fig. S1b). The general trend of these curves is also in agreement with the theoretical predictions.

We did not investigate the effect of removing SspB [2], a protein associated with increased affinity of tagged proteins to ClpXP. We anticipate that a moderate decrease in this Michaelis-Menten affinity would not qualitatively change our conclusions.
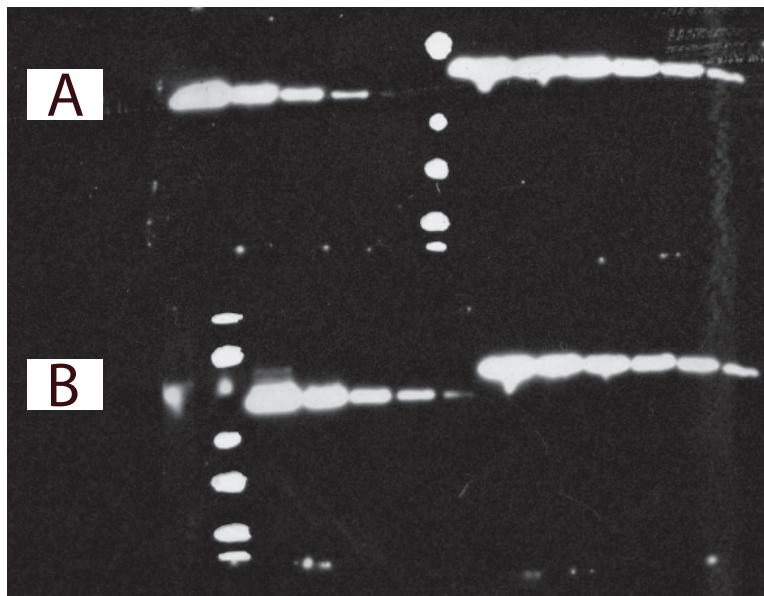
Protein Counts

Figure S2: Inverted image of Western blot film taken for cellular lysate data from two induction levels, where IPTG and arabinose were held at 1 mM and .8%, respectively, and doxycycline was 32 ng/ml (A) and 68 ng/ml (B). An antibody for GFP variants was used to detect the total amount of CFP and YFP inside these samples, when compared to a purified GFP standard (left five lanes of both A and B).

Western blots were performed using standard techniques in order to quantify the number of tagged fluorescent proteins being measured in our flow cytometry data. As a standard, we used purified Enhanced GFP (BioVision 4999-100), supplied in a 1 mg/ml 100 μl aliquot. We chose to measure protein levels in a sample of cells expressing the pNO-2CLAA plasmid, induced to various levels of dual-color expression. For all samples, we used 1 mM of IPTG and 0.8% arabinose. Samples were induced exactly as done for the flow cytometry experiments. Cells were grown overnight without inducers, and then passed 1:1000 into inducers for 3 hours. In order to obtain enough protein for quantitative detection, 50 ml of each sample was harvested by centrifugation after 3 hours. ODs at 600 nm were taken just before centrifugation, in order to quantify cell number (see below).

After centrifuging the samples and aspirating the inducing medium, the cells were resuspended in 100 μl of SDS sample buffer to aid in cell lysis by boiling. The total volume after resuspension was measured in order to obtain an accurate measurement of cell concentration. Cells were then lysed and proteins denatured by subjecting the samples (both lysates and standards) to boiling water for 5 minutes. A 12 lane 12% Tris-Glycine gel was used in order to have enough lanes for a sufficient dilution series of both the cell lysate sample as well as the standard. The standard was diluted to a concentration of 100 ng/μl, and fives samples were loaded on the gel in subsequent 2-fold dilutions, starting with 500ng. Similarly, the cellular lysate was loaded in subsequent 2-fold dilutions, starting with a volume of 15 μl. The gel was run at 125 V for about 100 minutes, followed by a membrane transfer run at 25 V for 90 minutes.

Standard blocking and probing reactions were set up using a GFP polyclonal rabbit antibody (Cell Signaling 2555S) and an Anti-Rabbit IgG (whole molecule)-Peroxidase antibody. After exposing the membrane to the Chemiluminescent Peroxidase Substrate (Sigma, CPS-1), Kodak
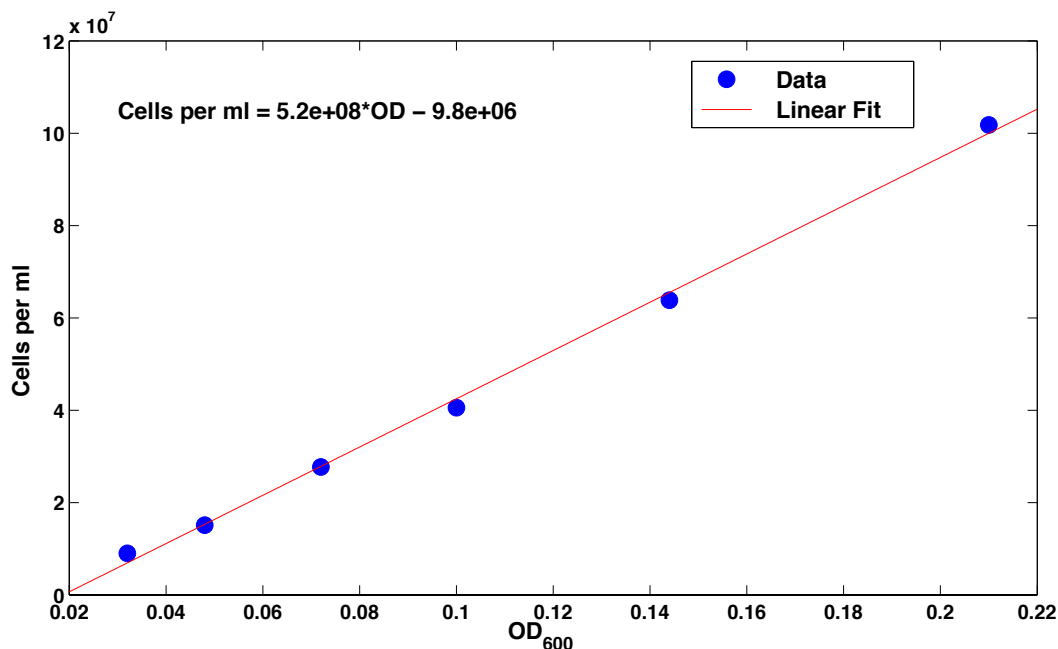
Figure S3: Cell count was correlated with OD by taking several measurements throughout logarithmic growth of each. A linear fit of this data was then used to convert the OD of Western blot samples to a particular concentration of cells.

BioMax Light Film was exposed to the membranes in a dark room for 60 seconds. Once a satisfactory image was taken, processing was performed using ImageJ. Background correction was performed to remove some of the background coloration from the image. The image was then inverted so that bands showed as white on a black background (Fig. S2), and the freehand selection tool was used to quantify the total intensity of each band. Comparing the total intensity of each standard band to the known protein mass loaded on the gel, we were able to obtain a function to convert band intensity to protein mass, and this was used to quantify several lanes of the cell lysate samples that fell within a linear range (where the known two-fold dilution matched a two-fold drop in band intensity). Finally the protein weight measurement was converted to a total protein count per cell, using the weight of a single protein and the number of cells loaded onto the gel. Control experiments were performed to ensure that this antibody binds with equal affinity to GFP, CFP, and YFP. Cells expressing each protein from the same promoter were grown in identical conditions and induced simultaneously. Western blotting was performed as described above to ensure that the same signal was detected from each of the three samples, after normalizing for cell count.

Cell counts were done using a hemacytometer. Cells were grown in inducing conditions and sampled every 20 minutes over a 3 hour period around the OD sampled for the Western blot data. Cell count was plotted vs OD over this range and a good linear fit was achieved (Fig. S3). Using this linear fit, we were able to calculate cell count for the ODs at which our cultures were sampled for the Western blots.
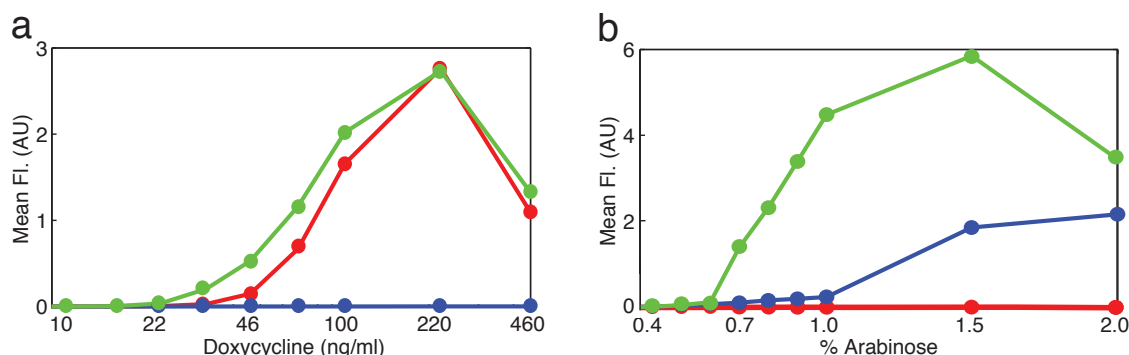
Figure S4: Control experiments were performed to ensure that there was no cross-talk between the inducers. (a) Doxycycline is shown to not induce the expression of CFP (blue), but induces YFP (red). Addition of arabinose increases YFP expression (green) due to slower overall ClpXP processing. (b) Arabinose does not induces expression of YFP (red), but induces CFP (blue). Addition of doxycycline increase CFP expression (green).
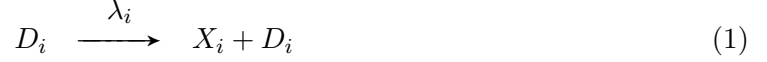
Experimental Controls

Several control experiments were performed to ensure that the coordinated behavior in our two experimental systems was due to degradation-based coupling and not just an artifact due to some other phenomenon. First, the comparison of the tagged and untagged induction curves in the main text served to demonstrate that the effects we are seeing are due to the tag and not just side effects of over-expression. In the case of the dual-tunable signaling network, another primary concern was ensuring that the two inducers did not interfere with the other promoters (i.e. that there is no cross-talk between the two promoters). To test this, we induced cells with each of the two inducers independently, and ensured that each color was only induced by the appropriate inducer (Fig. S4). In panel a, it is clear that doxycycline strongly induces YFP (red) and not CFP (blue). When arabinose is introduced in addition to doxycycline, the YFP levels increase (green), as CFP is now being produced as well, causing an increased burden on ClpXP. The reverse is true as well; in panel b, arabinose alone is seen to induce CFP (blue) and not YFP (red), however the addition of doxycycline causes increased levels of CFP. This is further evidence of queueing theory, in addition to a good control for crosstalk between the two promoters.

A similar control was performed for the dual-color synthetic circuit experiment. That is, the oscillator strain was induced with AHL, and we saw no effect on period of the oscillator, indicating that AHL and LuxR do not interfere with the ara/lac promoter. As another general control that the addition of tagged proteins causes the observed effect, as opposed to it being some other artifact of over expression, we tested the synthetic oscillatory system in conjunction with a high level of a general, untagged protein (for this purpose, we used Pn25 driving TetR, untagged on a p15A plasmid). When producing a large amount of untagged TetR (approximately 100,000 copies per cell [1]) alongside the oscillator, we saw no difference between this and the normal behavior of the oscillator. This control provides evidence that there is no apparent effect on period when expressing untagged proteins. In addition, the experimental acquisition of the two induction curves (described above), with the tagged and untagged versions of GFP, served to demonstrate that the effects observed throughout the experiments were in fact due specifically to the abundance of tags, and not simply side effects of general over-expression.

Stochastic Theory - Model

The model considered in this paper involves production and degradation of protein types X_i, where indices $i = 1, 2, ..., m$ identify different protein types. Degradation occurs by the protein binding to a protease P and subsequently being annihilated. Specifically, the model reactions are (rates are rate constants, not including mass action terms)

$$D_i \xrightarrow{\lambda_i} X_i + D_i \tag{1}$$

$$X_i \xrightarrow{\gamma} \emptyset \tag{2}$$

$$X_i + P \underset{\eta_-}{\overset{\eta_+}{\rightleftharpoons}} X_i P \tag{3}$$

$$X_i P \xrightarrow{\mu} P \tag{4}$$

$$X_i P \xrightarrow{\gamma} P \tag{5}$$

where DNA D_i produces protein X_i with rate constant λ_i, X_i is diluted (due to cell growth and division) with rate constant γ, X_i binds to the protease P with rate constant η_+, X_i unbinds from P with rate constant η_-, and P degrades X_i with rate constant μ. Reactions occur with exponentially distributed times. For simplicity of results, we assume that dilution can act on X_i bound to P, though results can be generalized to when dilution does not act on X_i bound to P. We typically assume results for a single effective protease (with an exception at the end of this section), though single protease results can be generalized to many proteases [3]. We assume the count of each DNA D_i is 1 for simplicity.

Using reasonable approximations, we can further simplify Eqs. 1–5. The simplest approximation is to suppose that $\eta_- \approx 0$ and that η_+ is large, such that the reactions in Eqs. 2 and 5 collapse to Eq. 2 and Eqs. 3–4 combine into a single degradation reaction, where the protease chooses one particular protein and degrades it at rate μ; the latter has the same steady-state behavior as when Eqs. 3-4 are replaced by the reaction

$$X_i \xrightarrow{\mu/n} \emptyset \tag{6}$$

where $n = \sum_{j=1}^m x_j$, and x_i is the count of protein type X_i, as in the main text. Similar results can be derived if both η_+ and η_- are sufficiently large. This leads instead to the approximate degradation reaction

$$X_i \xrightarrow{\mu/(K+n)} \emptyset \tag{7}$$

where $K = \eta_-/\eta_+$ is a Michaelis-Menten parameter [3]. In the limit $K \to 0$, the $\eta_- \approx 0$ system is recovered. More details concerning the motivation and derivation of the reduced rates in Eqs. 6–7 appear in Refs. [3, 4].

The reduced system, using either Eq. 6 or Eq. 7 for the enzymatic degradation reactions Eqs. 3–4, can be mapped (preserving the statistical distributions of protein counts) onto a stochastic queueing

model. One such queueing model places each new X_i at a random position in a single queue, while P processes (degrades) the protein at the head of the queue. Dilution can be added by allowing "reneging," whereby any member of the queue (including a member being processed by the server) leaves at an average rate of γ.

In Fig. 2, a version of this stochastic queueing model is used to describe the σ^s stress response system. We consider two substrates, mistranslated proteins, X_m and σ^s proteins, We suppose there are 100 ClpXP "servers" that each have processing rate $\mu = 10$ min^{-1}. Binding of both substrates to ClpXP is assumed fast ($\eta^+ = 10^8$ min^{-1}), and unbinding negligible ($\eta^- = 0$ min^{-1}). Cells are assumed to divide every 20 min, such that $\gamma = \ln 2$ min^{-1}. Production of mistranslated proteins occurs with rate λ_m, and the production of σ^s occurs with rate λ_{σ^s}. The system is overloaded when $\lambda_m + \lambda_{\sigma^s} > 100\,\mu$, balanced when $\lambda_m + \lambda_{\sigma^s} = 100\,\mu$, and underloaded otherwise. The statistics presented in Fig. 2 are derived from a large 32,000 ensemble of simulations. Single trajectories also use an ensemble size of 1. Numerical simulations take advantage of custom code using the CUDA framework for GPU acceleration.

Stochastic Theory - Results

We have carefully derived several relevant results for the above model in another study [3], which applies the theory of multiclass queueing in the context of gene regulation. One key result is that the steady state probability distribution $P(\{x_i\})$ for the set of counts $\{x_i\}$ can be factored into $R(n)$, the probability distribution for the sum, times a multinomial distribution:

$$P(\{x_i\}) = R(n)\,n! \prod_{j=1}^{m} \frac{p_j^{x_j}}{x_j!} \tag{8}$$

where $p_i \equiv \lambda_i / \sum_{j=1}^{m} \lambda_j$. From this, it can be shown that moments of x_i are given in terms of moments of n. In particular,

$$\langle x_i \rangle = p_i \langle n \rangle \tag{9}$$

$$\sigma_i^2 \equiv \langle x_i^2 \rangle - \langle x_i \rangle^2 = p_i(1 - p_i) \langle n \rangle + p_i^2 (\langle n^2 \rangle - \langle n \rangle^2) \tag{10}$$

The moments of n are less general and will depend on the particular model. With the reaction scheme Eqs. 1–2, 7, we find

$$\langle n \rangle = \frac{\alpha \delta M(\alpha + 1, \beta + 1, \delta)}{\beta M(\alpha, \beta, \delta)} \tag{11}$$

$$\langle n^2 \rangle = \langle n \rangle + \frac{\alpha(\alpha + 1)\delta^2}{\beta(\beta + 1)} \frac{M(\alpha + 2, \beta + 2, \delta)}{M(\alpha, \beta, \delta)} \tag{12}$$

with $\alpha \equiv K + 1$, $\beta \equiv (\mu/\gamma) + \alpha$, $\delta \equiv \Lambda/\gamma$, $\Lambda \equiv \sum_{i=1}^{m} \lambda_i$, and $M(\cdot, \cdot, \cdot)$ the confluent hypergeometric function of the first kind.

Deterministic Approximation for Mean Protein Counts

Though deterministic models do not address the many issues tied to noisy dynamics, e.g. correlations between the counts of the protein species, certain aspects of queueing coupling can be understood using a deterministic analog of the stochastic queueing model. Deriving approximately deterministic processes from chemical reaction networks has a long history, and so we will be brief
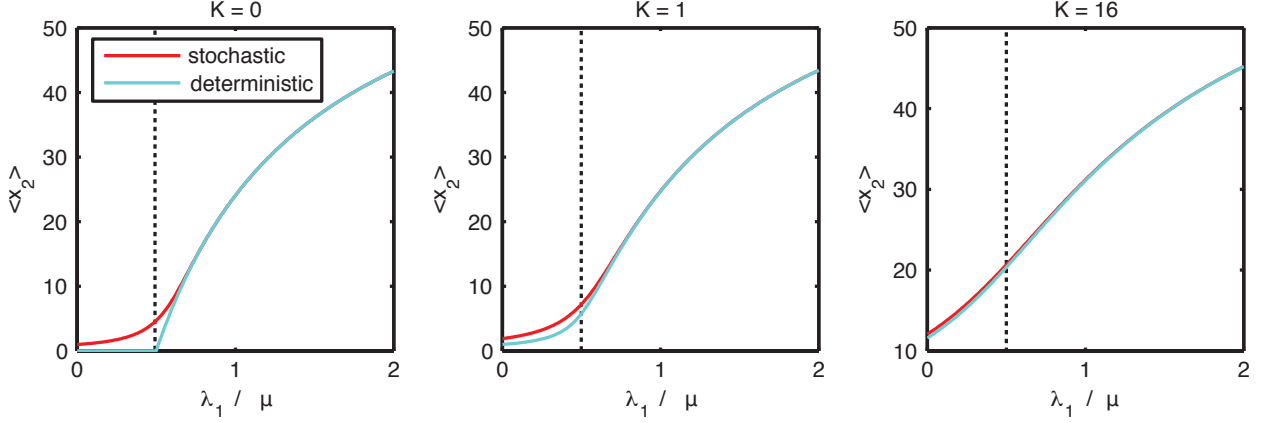
Figure S5: The deterministic approximation at steady state (see Eq. 13) can reasonably approximate the mean concentration of proteins for the stochastic model at steady state, especially when the system is above balance or when the Michaelis-Menten constant K is nonzero. Shown are results for a stochastic model and its deterministic approximation for a system containing one enzyme with processing rate $\mu = 100$ (AU). Proteins of type 2 are produced at constant rate $\lambda_2 = 0.5\,\mu$, and dilution occurs with rate $g = \ln 2$. The indirect response of mean protein level $\langle x_2 \rangle$ is plotted as a function of the normalized production rate λ_1/μ of protein 1. It is seen that deviation of the deterministic approximation from the stochastic mean value is largest in the underloaded regime (left of the dashed line) and smallest in the overloaded regime (right of the dashed line). Furthermore, increasing K reduces the error in the underloaded regime. These results support the general trend that we find in our analysis.

in the details of this section. We use this deterministic approximation primarily to accelerate some of our model fitting routines, but final results in the main text are always derived from stochastic models.

A deterministic approximation for mean protein levels x_i in a stochastic model with multiple servers is given by

$$\frac{dx_i}{dt} = \lambda_i - \gamma x_i - \frac{\mu x_i}{K + n} \tag{13}$$

where λ_i is a production rate constant, γ is the dilution rate constant, K is a Michaelis-Menten molar constant, μ is the total enzymatic processing rate constant, and $n = \sum_{j=1}^{m} x_j$ is the mean total protein count over all types. See [5] for mathematical details of derivation and validity for similar systems. By comparison to numerical simulations, the deterministic model is a reasonable approximation for the *mean* protein levels of the stochastic model (see Fig. S5). Deviation between the deterministic model and stochastic mean levels is largest in the balanced and underloaded regimes, but this deviation tends to be small when compared to the scale μ/γ, especially when the constant K is nonzero.

The steady state solutions $x_i^{(ss)}$ to Eqs. 13 satisfy a relation similar to Eq. 9

$$x_i^{(ss)} = p_i n^{(ss)} \tag{14}$$

where $n^{(ss)}$ is the total protein at steady state, and $p_i \equiv \lambda_i/\Lambda$. For finite K

$$n^{(ss)} = \frac{\Lambda - \mu - K\gamma + \sqrt{(\Lambda - \mu)^2 + K\gamma(2\Lambda + 2\mu + K\gamma)}}{2\gamma} \tag{15}$$

8

which simplifies in the limit $K \to 0$ to

$$n^{(ss)} = \frac{\Theta(\zeta)}{\gamma} \tag{16}$$

where $\zeta = \Lambda - \mu$, and $\Theta(\cdot)$ is the integrated Heaviside step function: $\Theta(\zeta) = \zeta$ if $\zeta \geq 0$ and $\Theta(\zeta) = 0$ if $\zeta < 0$. The solutions in Eqs. 14–16 reveal that components strongly interact when at least two λ_i's are simultaneously nonzero and $\Lambda > \mu$.

Fitting of Steady State Model

Figs. 3b–d in the main text include fits of the stochastic model to the data. Below, we outline the procedure to obtain fitted model parameters.

The fitted model results presented in Figs. 3c–d of the main text were derived through use of a fitting algorithm to determine model parameters μ, γ, K, a Hill function parameterization for λ_1 (production rate of YFP for a given dox level), and a set of 3 values for λ_2 (production rate of CFP for a given arabinose level). At the end of this section, we revisit these best fit values of λ_2 to find they are in reasonable agreement with single fluorescent protein expression data.

Curve fitting was implemented by a Metropolis algorithm. The energetic penalty used for the algorithm was a weighted sum of the square distances between stationary state model mean values (see Eqs. 9 and 11) and mean fluorescence data points. Due to the wide range of YFP fluorescence magnitudes, we used linear distance when comparing CFP fluorescence and logarithmic distance when comparing YFP fluorescence.

Parameters γ and K were not especially important for our fitting. We scaled time by the doubling time τ_d (approximately 30 min.), such that the value of the dilution rate was fixed at $\gamma = \ln 2$ in natural units. Furthermore, we set $K = 0$ with little reduction in the goodness of fit, and setting $K \leq 1000$ or so did not drastically change the results.

We found that the deterministic queueing model's stationary state approximates the overloaded stochastic queueing model's mean values well, and so we used the deterministic model's results for rapid fitting of the data, even though final results are generated from the stochastic model. Arbitrary precision calculations in the Maple 11 software package (Waterloo Maple Inc.) confirmed the stochastic model's mean values were reasonably approximated by the deterministic model with the assumption of overloading.

Using the data from the dox induction curves in Figs. 3c–d of the main text, λ_1 was fit to a shifted Hill function of the form

$$\lambda_1 = B_1 + D_1 \frac{(([\text{dox}]/C_1)^{H_1})}{(1 + ([\text{dox}]/C_1)^{H_1})}. \tag{17}$$

with $H_1 = 3.0782$, $B_1/\mu = 0.0023$, $D_1/\mu = 2.3429$, and $C_1 = 168.2114$ ng/mL. We did not fit λ_2 to a smooth curve, due to a small number of points being available, but we found best fit values $\lambda_2/\mu = 1.0373, 1.1093, 1.2683$.

Other parameters used for Figs. 3c–d in the main text are as follows: Using doubling time $\tau_d \approx 30$ min (for *E. coli*), $\mu = 7.589 \times 10^3$ min^{-1}, $\gamma = \ln 2$, $K = 0$.

We tested consistency of the model fit in Figs. 3c–d by comparing results to the data in Fig. 3b. Here, we found a continuous curve fit for the mean untagged (slow degrading) GFP fluorescence multiplied by γ, providing a continuous parameterization of the apparent production rate of GFP. Before fitting, to account for the difference between single molecule GFP and CFP fluorescence, we scaled GFP fluorescence such that the three values of mean CFP fluorescence at low dox (in
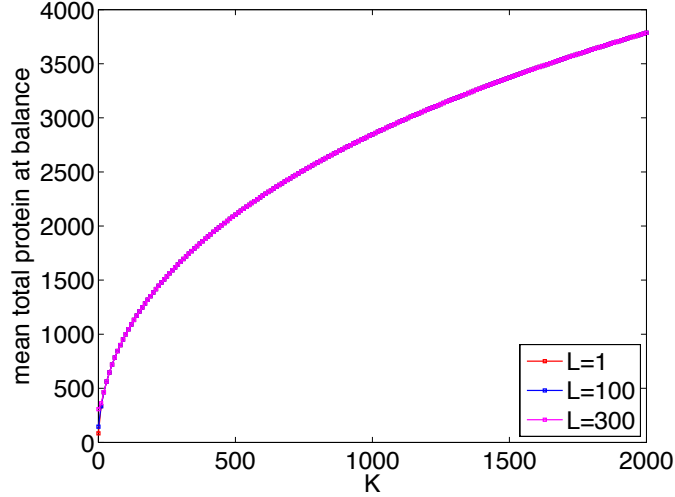
Figure S6: Pictured is the mean total protein at balance for the stochastic queueing model over a range of the Michaelis-Menten constant K. This is approximated here by each protease (copy number L) processing at a reduced rate μ/L and with Michaelis-Menten constant K. The balanced condition is defined here as when the total protein production rate Λ equals the total processing rate μ, i.e. $\Lambda = \mu$. Other parameters are determined from the model fit in this section. We find the mean total protein at balance depends weakly on L for $K > 0$. The theoretical protein level at $K = 1000$ can be considered a moderate perturbation compared to the typical protein level for the experimental data in Fig. 3 of the main text. Note that even for zero K, queue lengths are nonzero.

Figs. 3c–d) were closest in a least squares sense to the corresponding mean GFP fluorescences. We fit the apparent production rate $\lambda_2^* = \gamma < \mathrm{gfp} >$ to the continuous function

$$\lambda_2^* = B_2 + D_2 \frac{(([\mathrm{ARA}]/C_2)^{H_2})}{(1 + ([\mathrm{ARA}]/C_2)^{H_2})}. \tag{18}$$

with $H_2 = 1.3660$, $B_2/\mu = 0.0665$, $D_2/\mu = 3.1039$, and $C_2 = 1.0323$ %. The difference $\Delta\lambda_2$ between λ_2 from panels c,d and λ_2^* from panel b are relatively minor, being $\Delta\lambda_2/\lambda_2 = 0.1093, 0.0144, -0.0850$, respectively, suggesting that the fits are consistent.

In Fig. 3b, using the parameterization λ_2^* and the model parameters determined by Fig. 3c–d, we present the prediction for mean protein count as a solid red curve. This prediction is in agreement with the data in Fig. 3b, suggesting that the data in Fig. 3b and in Figs. 3c–d are in agreement.

Though we found a small value of K (e.g. about a thousand or less) was consistent with our model fit, the effect of larger K on the stochastic queueing model at the balance point was considered (see Fig. S6).

References

[1] R. Lutz and H. Bujard. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I₁-I₂ regulatory elements. *Nucleic Acids Res.*, 25(6):1203–10, Mar 1997.

[2] I. Levchenko, M. Seidel, R. T. Sauer, and T. A. Baker. A specificity-enhancing factor for the ClpXP degradation machine. *Science*, 289(5488):2354–6, Sep 2000.

[3] W.H. Mather, N.A. Cookson, J. Hasty, L.S. Tsimring, and R.J. Williams. Correlation resonance generated by coupled enzymatic processing. *Biophysical Journal*, 99(10):3172–3181, 2010.

[4] K.R. Sanft, D.T. Gillespie, and L.R. Petzold. Legitimacy of the stochastic Michaelis-Menten approximation. *IET Syst. Biol.*, 5:58–69, 2011.

[5] N.G. van Kampen. Power series expansion of master equation. *Canadian Journal Of Physics*, 39(4):551–567, 1961.