Cell

Tradeoffs and Optimality in the Evolution of Gene Regulation

Frank J. Poelwijk,[1,2] Marjon G.J. de Vos,[1] and Sander J. Tans[1,*]
[1]AMOLF Institute, Science Park 104, 1098 XG Amsterdam, The Netherlands
[2]Present address: Green Center for Systems Biology and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX 75390-9050, USA
*Correspondence: tans@amolf.nl
DOI 10.1016/j.cell.2011.06.035

SUMMARY

Cellular regulation is believed to evolve in response to environmental variability. However, this has been difficult to test directly. Here, we show that a gene regulation system evolves to the optimal regulatory response when challenged with variable environments. We engineered a genetic module subject to regulation by the *lac* repressor (LacI) in *E. coli*, whose expression is beneficial in one environmental condition and detrimental in another. Measured tradeoffs in fitness between environments predict the competition between regulatory phenotypes. We show that regulatory evolution in adverse environments is delayed at specific boundaries in the phenotype space of the regulatory LacI protein. Once this constraint is relieved by mutation, adaptation proceeds toward the optimum, yielding LacI with an altered allosteric mechanism that enables an opposite response to its regulatory ligand IPTG. Our results indicate that regulatory evolution can be understood in terms of tradeoff optimization theory.

INTRODUCTION

The capability to regulate behavior and physiological state in response to the environment is a fundamental property of all living systems. How novel regulatory phenotypes emerge and adapt in populations challenged by the conflicting demands of variable environments has long fascinated biologists (Agrawal, 2001; Pigliucci, 2001; DeWitt and Scheiner, 2004). The general prediction from theory is that regulatory responses are favored to evolve when the selective environment fluctuates more slowly than the typical timescale of responses (Scheiner, 1993; Pigliucci, 2001). However, the outcome of evolution in variable environments may depend on various unknown factors, such as constraints of physical, biochemical, and genetic origin (Maynard Smith et al., 1985); the competition between different regulatory phenotypes (van Tienderen, 1997); and the precise strength and direction of selection on regulation (Scheiner, 1993; Via et al., 1995; Pigliucci, 2001). Therefore, even though experiments have shown that regulation can be beneficial (Pigliucci, 2001) and can be altered by phenotypic screening (Scheiner, 1993), most experimental work on the evolution of regulation by mutation and natural selection has remained indecisive and difficult to explain in terms of the causal selective forces and constraint (Scheiner, 2002).

Conceptually, addressing this issue is straightforward. Selection at the phenotypic level can be revealed by quantifying the dependence of fitness on the relevant phenotypic parameters (Lunzer et al., 2005; Weinreich et al., 2006; Poelwijk et al., 2007). Constraints may be identified by experimental evolution: an evolutionary response according to selection indicates adaptation, whereas conversely, a lack of such a response points to a constraint (Miller et al., 2006). In practice, however, it is nontrivial to determine the relation between regulation and fitness. The overall growth rates of two known *Escherichia coli* regulatory mutants have been measured in a variable environment (Suiter et al., 2003), but our limited insight into the genetic basis of regulatory changes hampers extending this approach to systematically assay the full range of possible regulatory responses.

To overcome these obstacles, we have developed a synthetic approach. Synthetic systems (Benner and Sismour, 2005) can be engineered to contain the two core elements of regulatory evolution: a cellular phenotype that confers a benefit in one environment and a burden in another and a regulation system that senses the environment and modulates the phenotype. By designing a sensed environmental cue that can be varied separately from the environmental factor that confers the burden or benefit, one can quantify the relation between expression regulation and fitness, prior to adaptation and without the need for a comprehensive library of regulatory mutants. Using this approach, we present a case study of optimality in evolution in variable environments throughout the various levels of biological organization, from the environment down to molecular mechanisms and genotype.

RESULTS

Experimental System and Fitness in Constant Environments

To quantify the selective forces and evolutionary change of regulation in variable environments, we constructed an experimental system in *Escherichia coli* consisting of two genetic modules (Figures 1A and 1B). The first module comprises an operon that harbors the *sacB* and *cmR* genes, which affect the growth rate, as well as the *lacZ*α gene, which is used to quantify the operon expression level *E*. *E* is controlled by the *lac* repressor, LacI, which constitutes the regulatory module. The system as

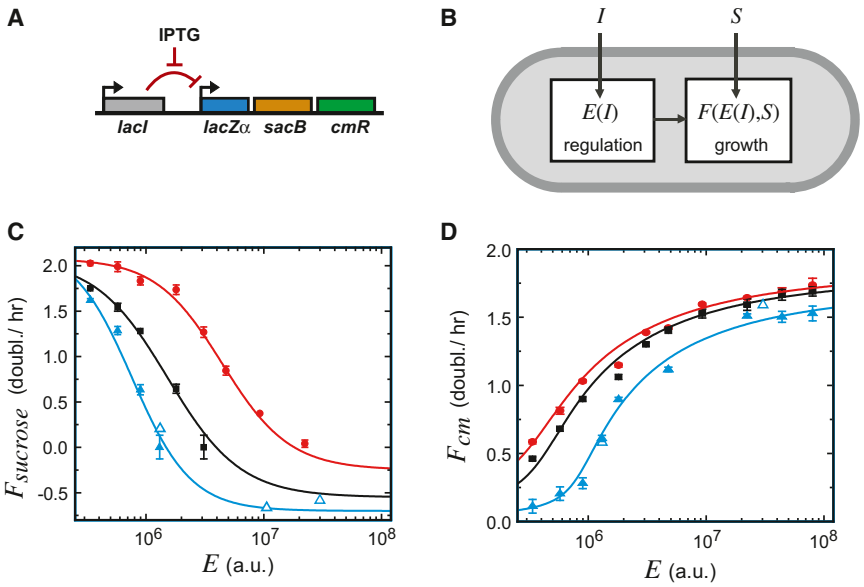

Figure 1. Phenotype and Fitness Characterization

(A) Schematic of the synthetic operon and regulation system.

(B) Functional system representation. *F(E,S)* describes the dependence of fitness (growth rate) on operon expression and concentration of selective agent (sucrose or Cm). *E(I)* is the dependence of expression on the concentration of the environmental cue IPTG.

(C) Measured *F(E,S)* relations for sucrose media. Sucrose concentrations (w/v): 0.15% (red), 0.25% (black), and 0.40% (blue). Error bars are standard errors (n = 3). (Closed symbols) Expression and fitness values as achieved by IPTG induction. (Open symbols) Data obtained by competition assays (Extended Experimental Procedures). Expression standard error is smaller than the symbol size.

(D) Measured *F(E,S)* relations for Cm media. Cm concentrations: 25 μg/ml (red), 40 μg/ml (black), and 80 μg/ml (blue). Curves in (C) and (D) are fits to a growth model (Extended Experimental Procedures).

See also Figure S1.

constructed (termed *WT* hereafter) responds to increasing concentrations of the environmental cue isopropyl-β-D-thiogalactopyranoside (IPTG) by increasing *E*.

In media containing sucrose, expression of the operon by induction with IPTG leads to a reduced growth rate (Figure 1C and Figure S1 available online), as determined by monitoring the optical density of growing populations. Before inoculation in the sucrose media, the cells were grown without sucrose but with the IPTG concentration of interest in order to achieve the steady-state expression levels. The observed negative effect on growth rate is due to the sucrose-polymerizing activity of levansucrase encoded by *sacB*, which leads to the accumulation of large polysaccharide chains that interfere with cell wall formation (Gay et al., 1985). At the highest levels of operon expression, we observed negative growth rates. The corresponding decrease in the number of viable cells was quantified using competition and plating assays (Extended Experimental Procedures). We note that the cells lack the genes to metabolize sucrose and use it as a carbon and energy source. Thus, sucrose-containing media confer a selective pressure to decrease operon expression.

For media that contain the antibiotic chloramphenicol (Cm), the growth rate is suppressed at low operon expression levels (Figure 1D and Figure S1). When operon expression is increased by induction, the growth rate is progressively restored. This beneficial effect of operon expression is due to the inactivation of the antibiotic by the CmR gene product. Cm media thus confer a selective pressure to increase expression.

Increasing the Cm or sucrose concentrations predominantly shifts the point of half-maximum growth along the expression axis and does not significantly affect the maximum growth rate. We find that the growth expression data in sucrose and in Cm media are well-described by the functions $F_{sucrose}(E)$ and $F_{Cm}(E)$, which are based on a simple model of the underlying reaction kinetics (Extended Experimental Procedures).

Tradeoffs for Fixed Expression Phenotypes in Variable Environments

Environments that vary in time between sucrose and Cm result in fluctuating demands on operon expression. The simplest phenotype is one that exhibits a single fixed expression level in both the sucrose and the Cm environment (Figure 2A). Variable environments confront such unregulated phenotypes with a tradeoff: high fitness in sucrose media—resulting from low operon expression levels—will be at the expense of low fitness in Cm media. Conversely, high fitness in Cm media—due to high expression levels—will entail low fitness in sucrose media.

To gain insight into the optimization of the total fitness of fixed expression phenotypes in sucrose-Cm variable environments, we plotted $F_{sucrose}(E)$ versus $F_{Cm}(E)$ for a range of *E* (Figure 2B). We find that, for the lower Cm and sucrose concentrations, this so-called tradeoff curve (Figure 2B, red line) is concave and bulges outward toward optimal growth in both environments (Figure 2B, cross), whereas for the high concentrations, it is convex and bulges inward away from optimal growth (Figure 2B, blue line). Shapes of tradeoff curves and their consequences for selection have been extensively analyzed on a theoretical level. As pointed out by Levins, who originally introduced the idea in the 1960s (Levins, 1968), concave tradeoff curves favor intermediate expression phenotypes that do averagely well in both environments, whereas convex tradeoff curves favor extreme expression phenotypes that perform well just in one environment (Figure 2C). In our study, these contrasting cases depend on the Cm and sucrose concentrations that may be viewed as the amplitude of the environmental variations. The observed change in convexity is a direct consequence of the nonlinear biochemical processes that underlie growth and may well be a general feature of fitness tradeoffs in biological systems.

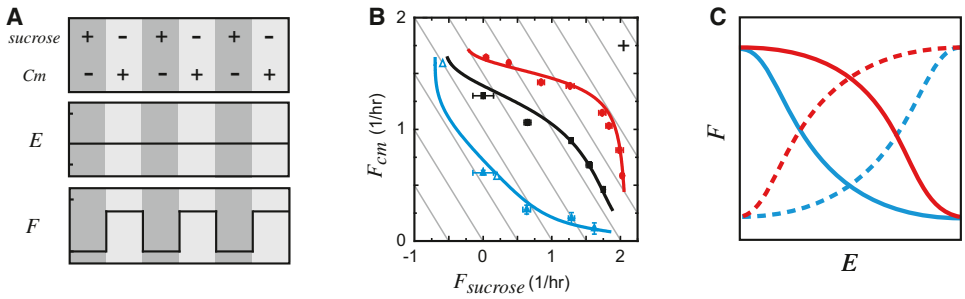

Figure 2. Tradeoffs for Fixed Expression Phenotypes in Variable Environments

(A) Variation of environmental and system parameters. Here, sucrose alternates with Cm, and the operon expression level *E* is constant across both environments, whereas the growth rate *F* may vary.

(B) Tradeoff curves in sucrose-Cm variable environments for different sucrose and Cm concentrations. For any *E*, the resulting growth rate in Cm $F_{Cm}(E)$ (Figure 1D) is plotted against the growth rate in sucrose $F_{sucrose}(E)$ (Figure 1C). (Red line) 0.15% sucrose, 25 μg/ml Cm. (Black line) 0.25% sucrose and 40 μg/ml Cm. (Blue line) 0.40% sucrose and 80 μg/ml Cm. The cross indicates the optimal growth in both environments at low sucrose and Cm concentrations (red line). The diagonal isoclines indicate equal overall growth rate $F_{tot} = (F_{sucrose}(E)+F_{Cm}(E))/2$ when alternating between sucrose and Cm for equal periods of time. Error bars are standard errors (n=3).

(C) Schematic diagram illustrating the origin of the switch from concave to convex tradeoff curves, as observed in (B). The lines represent the expression growth relations *F*(*E*) for sucrose (solid lines) and for Cm (dashed lines). Red indicates lower concentrations of sucrose and Cm, and blue indicates higher concentrations. At fixed medium expression *E*, the red curves show a near optimal growth in both environments resulting in a tradeoff curve that bulges outward (B). For the blue growth conditions, such a condition does not exist. Here, the highest fitness is attained by phenotypes with a fixed high or low expression.

Fitness of Regulated Phenotypes in Variable Environments

Regulated phenotypes, which can sense changes in the environment and respond by adjusting their expression, can escape from tradeoff curves for unregulated phenotypes (Figure 2B) and acquire higher fitness. When Cm is supplemented with the cue IPTG (1 mM; see Figure 3A), the *WT* phenotype in fact exhibits a nearly optimal response: with IPTG, the high expression level E_1 is favorable in Cm, whereas without IPTG, the resulting low expression level E_0 is favorable in sucrose. As observed in the previous section, however, the best unregulated phenotypes perform almost as well for the lower Cm and sucrose concentrations due to the concave tradeoff curve (Figure 2B, red line). The resulting limited advantage of regulation is not caused by weak selection in each medium separately, as evidenced by the large growth rate differences (Figures 1C and 1D, red curves) but, rather, by the precise shape of the growth expression curves $F_{sucrose}(E)$ and $F_{Cm}(E)$. For higher sucrose and Cm concentrations though, the tradeoff curve becomes convex (Figure 2B, blue curve), thus lowering the maximally attainable fitness for unregulated phenotypes and consequently increasing the selective advantage of regulated phenotypes.

To compare the performance of different phenotypes systematically, we consider their overall growth rate in variable environments as a function of E_0, the expression level in the environment without IPTG (and with sucrose), and $P = E_1/E_0$, the fold change in expression when switching to the environment with 1 mM IPTG (and with Cm). The separate growth rates in the sucrose and Cm media are then found by substituting the value for E_0 in the function $F_{sucrose}(E)$ and E_1 in $F_{Cm}(E)$ (Figures 1C and 1D). We determine the overall growth rate, F_{tot}, as the arithmetic mean of the separate growth rates weighted by the fraction of time spent in each environment. We use the arithmetic mean (as opposed to the geometric mean) of the local growth rates because growth rates—unlike the number of offspring—are Malthusian fitness parameters (Roff, 2001). In Figure 3B, we plot F_{tot} as a function of E_0 and *P* for the high sucrose and Cm concentrations. This fitness function contains a single optimum near WT (low E_0 and high *P*). The peak is well separated from $P = 1$, consistent with the strong selection for regulated over nonregulated phenotypes.

Competition in Variable Environments

Having determined the relation between expression phenotype and fitness (Figure 3B), the question is whether it accurately captures the competition between different regulatory mutants in variable environments. We tested this in two ways. First, we competed *lacI* mutants one to one. We employed *lacI* mutants that were generated using error-prone polymerase chain reaction (PCR) and displayed contrasting expression phenotypes (Figure 3B, crosses). The different alleles were cloned into our plasmid, and pairs of them were grown in sucrose and Cm environments, either with or without IPTG, whereas their relative abundance was followed by plating. Again, cells were grown nonselectively before and after selection, and care was taken that growth was not limited by other growth factors such as carbon or oxygen. These competition results were compared to fitness differences predicted by the fitness function $F_{tot}(E_0,P)$, based on the uninduced (E_0) and induced (E_1) expression levels of each mutant, which showed a good agreement (Figure 3C).

Second, we investigated the competition within a large population of cells with randomly mutated regulatory systems. For this purpose, we again mutated the *lacI* coding sequence with error-prone PCR. Sequencing indicated an average mutation rate of $\sim 3 \times 10^{-3}$ per base pair. The mutated *lacI* coding regions were placed back into the original plasmid, resulting in a diverse population of $\sim 10^6$ variants. Note that only the regulatory system was genetically diversified. The rest of the plasmid, including the selection genes, as well as the chromosomal background, remained identical.

To characterize the phenotypic diversity prior to selection, 35 clones were randomly picked from the population and mapped

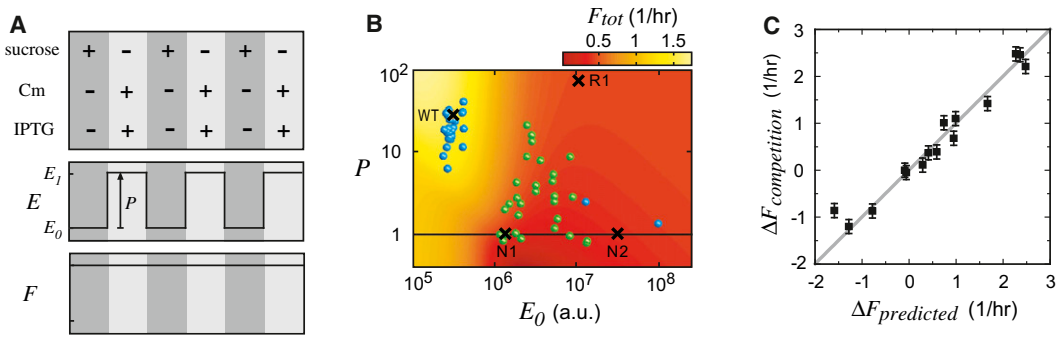

Figure 3. Fitness and Evolution in a Variable Environment with Stabilizing Selection

(A) Variation of environmental and system parameters. Here, sucrose alternates with Cm plus IPTG. In this environment, the operon expression level of the system as constructed (*WT*) will vary between a low E_0 in sucrose and a high E_1 in Cm plus IPTG, resulting in high growth rates in both environments. P is the fold change in operon expression ($P = E_1 / E_0$). Note that the selection experiments (dots, B) involve one sucrose/Cm cycle.

(B) Selective landscape and mapped regulatory variants. Color indicates the overall growth rate in the variable environment defined in (A), determined with the data from Figures 1C and 1D ($F_{tot} = (F_{sucrose}(E_0)+F_{Cm}(E_1))/2$), as a function of system parameters E_0 and P. Sucrose, Cm, and IPTG concentrations are 0.40%, 80 μg/ml Cm, and 1 mM, respectively. Different *lacI* regulatory variants are mapped on the landscape based on their measured E_0 and P values. (Crosses) WT and three previously isolated *lacI* mutants. (Green spheres) 35 randomly chosen isolates from a diverse population obtained by *lacI* mutagenesis. (Blue spheres) After selection in the corresponding variable environment. Growth time is 6 hr in each medium. Prior to inoculation in new medium, the cells are grown nonselectively (neither sucrose nor Cm) to adjust to the new IPTG level.

(C) Fitness differences from competition experiments between genetic *lacI* variants against fitness differences predicted by the fitness landscape (Extended Experimental Procedures).

in phenotype space based on their measured E_0 and P values. The phenotypes did not appear distributed equally throughout the E_0-P plane but were, rather, located toward higher E_0 and lower P down to $P = 1$ (Figure 3B), consistent with the expectation that most random mutations in the repressor will deteriorate the ability to repress. The original *WT* phenotype was not present in this sample of 35 variants.

We then exposed the population of ~10^6 random variants to a variable-purifying selection. In short, the cells were first grown in the sucrose medium (0.4% w/v) for 6 hr. From the end population, 1/500th was taken and then grown nonselectively at 1 mM IPTG for 3 hr to induce the cells. Next, Cm was added (at 80 μg/ml), and the population continued to grow for another 6 hr. Experiments indicated that the 3 hr of nonselective growth was sufficient to reach steady-state expression and did not significantly influence the overall fitness (Figure S1C). From the final population, we again took isolates randomly, assayed their E_0 and E_1 values, and mapped them onto the E_0-P plane. The group of isolates was clustered at the *WT* phenotype (Figure 3B). These results indicate that the simple fitness function $F_{tot}(E_0,P)$ captures the competition between diverse regulatory variants, which involves the integration of the different fitness gains and losses experienced by the competing variants as they experience the environmental changes, into an overall total fitness. They also confirm that that the expression E is the key phenotypic parameter in determining fitness in our system.

Evolution under Directional Selection in Variable Environments

In order to study adaptation to a novel regulatory function, we defined a variable environment in which the *WT* phenotype is maladapted and can improve by changing the regulatory response. We exploit the decoupling in the system, which allows one to impose a controlled mismatch between cue and selective agent. Here, a medium with sucrose plus IPTG (1 mM) alternates with a medium with Cm (Figure 4A). In the presence of sucrose and IPTG, the high induced operon expression is burdensome, whereas in Cm media without IPTG, a high operon expression would be beneficial but is repressed. Mirroring the case of the unregulated phenotypes, changing expression overall (in both E_0 and E_1) gives rise to a tradeoff: an increase in expression leads to gains in Cm but losses in sucrose, whereas conversely, a decrease in expression yields gains in sucrose but losses in Cm. Improvements in both environments may be realized by changes in the regulatory response, namely when IPTG would lead to repression, and an absence of IPTG to expression. Such an inverse regulatory response is, in fact, observed for the LacI homolog PurR, which represses in the presence of the corepressor guanine (Choi and Zalkin, 1992), suggesting that such inversions have occurred in evolutionary history.

The fitness function corresponding to this variable environment is displayed in Figures 4B and 4C. E_1 and E_0 here, respectively, denote the expression level in the medium with sucrose plus IPTG and the medium with Cm. The *WT* phenotype here maps onto a fitness valley at $P > 1$, indicating its poor performance in both environments. The peak is located at $P < 1$, which reflects the inverse nature of the corresponding phenotype ($E_1 < E_0$).

Next, we monitored evolution of the *WT* system using the same directed evolution approach. In brief, we prepared randomly mutated *lacI* variants and then grew a starting population of ~10^6 cells in a sucrose medium (0.4% w/v) with IPTG (1 mM) for 6 hr. From the end population, 1/500th was taken and grown nonselectively without IPTG for 3 hr. Then Cm was added (to reach 80 μg/ml), and the population was grown for an additional 6 hr.

After selection, a random sample of 35 isolates was mapped within the E_0-P plane, which showed that the population mean

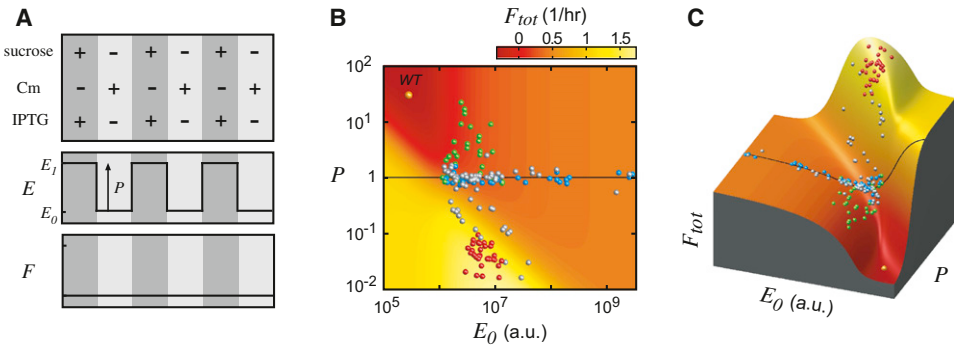

Figure 4. Fitness and Evolution in a Variable Environment with Directional Selection

(A) Variation of environmental and system parameters. Here, sucrose plus IPTG alternates with Cm. In this environment, the operon expression level of the system as constructed (*WT*) will vary between a high E_1 in sucrose plus IPTG and a low E_0 in Cm, resulting in low growth rates in both environments. P is the fold change in operon expression ($P = E_1 / E_0$). Note that the selection experiments (dots, B and C) involve one sucrose/Cm cycle.

(B) Selective landscape and mapped regulatory variants. Color indicates the overall growth rate in the variable environment defined in (A), determined with the data from Figures 1C and 1D ($F_{tot} = (F_{sucrose}(E_0)+F_{Cm}(E_1))/2$), as a function of system parameters E_0 and P. Sucrose, Cm, and IPTG concentrations are 0.40%, 80 μg/ml Cm, and 1 mM, respectively. Different *lacI* regulatory variants are mapped on the landscape based on their measured E_0 and P values. (Green spheres) 35 randomly chosen isolates from a diverse population obtained by *lacI* mutagenesis. (Blue spheres) After selection in the corresponding variable environment. (Gray spheres) After a second cycle of *lacI* mutagenesis and selection in both media. (Red spheres) After a third cycle. Growth time is 6 hr in each medium. Prior to inoculation in new medium, the cells are grown nonselectively (neither sucrose nor Cm) to adjust to the new IPTG level.

(C) Idem, in three-dimensional representation, rotated for visibility.

had shifted toward higher fitness as expected (Figures 4B and 4C, blue spheres). However, no isolates were found below $P = 1$, suggesting that there exists an adaptive constraint that renders crossing of the $P = 1$ boundary a rare or impossible event. Note that $P < 1$ phenotypes might be present in the population at low numbers and absent in the isolates due to insufficient sampling. However, the fitness function does indicate strong selection for $P < 1$ phenotypes: the growth rate of the optimal phenotypes is more than 1.4 dbl/hr higher than the best isolates at $P = 1$, making their initial presence very unlikely ($p < 0.02$, given the relative enrichment factor of $2^{(1.4 \cdot 12)} \sim 1.1 \times 10^5$).

Selection also affected the shape of the phenotypic distribution, which was now spread out broadly along the E_0 axis ($P = 1$) and had become narrow along the P axis (Figures 4B and 4C). The expansion along E_0 can be understood from the constraint: with paths to $P < 1$ poorly accessible, the paths offering higher fitness lead along the E_0 axis toward high or low expression values (Figure 4B, black line). These fixed expression phenotypes may represent evolutionary dead ends, as they can be achieved by simply abolishing repressor-operator binding or IPTG binding capabilities.

To test whether $P < 1$ phenotypes could still emerge before the population was taken over by evolutionary dead ends, we performed a new cycle: 1/500th of the previous culture was taken, and the plasmid DNA was extracted; *lacI* was randomly mutated as before and placed back into fresh plasmids and hosts, which were subsequently grown in the same sucrose-Cm variable environment. The resulting isolates indicated that, whereas about half of the population remained at $P = 1$, the other half emerged below (Figures 4B and 4C, gray spheres), showing that $P = 1$ constituted not a global but a local (breakable) constraint. These results indicate a capacity of the system to evolve a fundamentally altered response to inducer before getting trapped in specialization. The evolved inverse phenotypes were distributed along a downward diagonal, indicating that they all had similar E_1 and differed mainly in E_0.

After a third cycle of the same *lacI* mutagenesis and variable selection, the mean fitness of the population had further increased, first of all through decreased frequencies of fixed expression ($P = 1$) phenotypes, which were no longer observed among the isolates (Figures 4B and 4C, red spheres). In addition, the subpopulation of inverse phenotypes showed further fine-tuned improvements of the response toward smaller E_0 and smaller P. Figures 5A and 5B show induction profiles of evolved inverse phenotypes. The isolates were now distributed near the effective fitness optimum, which is taken as the average of the maximum observed growth rates in the two environments (1.85 db/hr). Some evolved phenotypes (4 out of 35) had a fitness within 10% of the optimum, which is comparable to the variability that we observed between the predicted fitness and results from competition experiments (Figure 3C). Also note that small fitness increases near an optimum entail long fixation times, given the limited number of isolates that can be assayed. For example, a single mutant with a 10% fitness improvement over the rest of the population (of 10^6 cells) would require more than 80 hr of growth to be detected in the sampled isolates, as the required enrichment factor is about 10^4 (which is equal to $2^{(0.1 \cdot 1.85 \cdot 80)}$). In conclusion, we find no evidence for global constraint in E_0 or P, nor a rigid correlation between the two, that prevents access to the landscape optimum.

Genetic Basis of Local Constraint

Tracking evolution within phenotype fitness landscapes can identify local adaptive constraints but does not reveal the genetic architecture from which they originate. We sequenced several inverse *lacI* alleles to provide some insight into the genetic causes of the observed constraint at $P = 1$ (Table S1). S97P appeared to be a key substitution: it occurred in the majority of

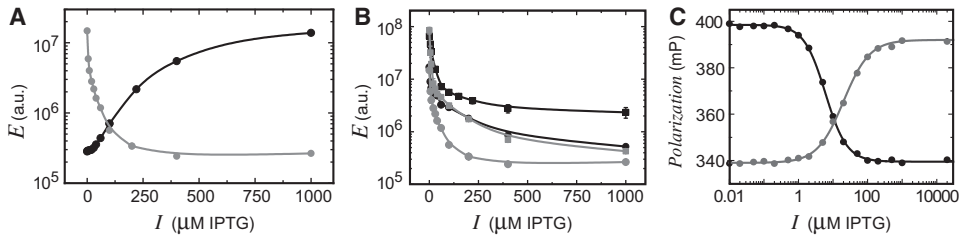

Figure 5. Induction Curves

(A) Measured induction curves for wild-type LacI (black) and an evolved inverse LacI phenotype M-inv-1 (gray).

(B) Induction profiles of a number of evolved inverse regulation mutants (at 1 mM from low to high expression: M-inv-1, M-inv-2, M-inv-3, M-inv-4; sequences given in Table S1). Standard errors (n = 3) are smaller than the symbol size.

(C) Fluorescence polarization profiles of wild-type (gray) and the inverse regulation mutant M-inv-1 of Figure 5A (black). Fluorescence polarization (see Extended Experimental Procedures) here is a measure for the binding of purified repressor protein to TAMRA-labeled 18 base pair symmetric *lac* operator DNA. Polarization is recorded as a function of IPTG concentration, and higher polarization indicates a higher fraction of operator DNA bound by repressor. The data is fitted with a Hill curve with K_D = 6.0 μM and Hill coefficient 1.4 for wild-type and K_D = 20 μM and Hill coefficient 1.15 for M-inv-1.

Error bars are standard errors over three or four measurements and smaller than the symbol size. See also Table S1.

genotypes and was essential for achieving the optimal inverse response. In a *WT* background, however, S97P yielded a *P* = 1 phenotype, as was determined by constructing this mutant using site-directed mutagenesis, and as was also corroborated by previous studies (Suckow et al., 1996; Flynn et al., 2003). It has been suggested that this serine residue, which is located at the dimer interface of the repressor, is central to the IPTG-induced allosteric transition in LacI (Flynn et al., 2003; Zhan et al., 2010). This may explain the observed interference with induction by the S97P mutation.

Unlike the recurring S97P substitution, other genetic changes found in the inverse *lacI* alleles were diverse. For instance, one inverse phenotype harbored the mutations S69Y, Q131P, M242I, and a stop at L346 in addition to S97P, whereas another contained K108N, E235D, and Q352E together with S97P (Table S1). Though we could not identify a specific pattern or known functional effect for these additional mutations, their diversity indicated the presence of different possible mutational routes from S97P and *P* = 1 to the optimum inverse phenotype.

In order to study the functional relevance of the observed mutations and their interactions, we engineered *lacI* variants containing subsets of the mutations found in one evolved inverse phenotype (M-inv-5; see Table S1). We found that, out of the six nonsynonymous mutations, three (S97P, R207L, and T258A) were sufficient in a *WT* background to confer an inverse phenotype (*P* ~0.02). Next, we engineered the double mutants S97P-R207L and S97P-T258A, which both gave a value for *P* of order 1, as did the single mutant S97P. Thus, in the S97P background, R207L and T258A individually were nearly neutral but together conferred inversion, which indicates epistasis between R207L and T258A. R207L also exhibited epistasis with S97P; in the T258A background that displayed a *WT* phenotype (*P* of order 10), R207L was again neutral, but in the S97P-T258A background, R207L transformed a *P* = 1 phenotype into the inverted phenotype. Overall, the results suggest that S97P is a rare and unique gateway at *P* = 1 from *WT* at *P* > 1 to the optimum at *P* < 1 and support the notion that the phenotypic clustering at *P* = 1 was due to genetic constraint.

Molecular Mechanism of Evolved Phenotypes

Next, we questioned what molecular mechanism could underlie the regulatory change. One might imagine that the transcription factor now activates rather than represses by binding elsewhere in the regulatory region. However, no mutations were found in the DNA recognition domain of LacI, which would be required for this scenario. The inversion could also be explained by a mechanism in which the nonspecific affinity to DNA is increased, as has been shown previously for LacI inversion (Pfahl, 1976; Miller and Schmeissner, 1979). An increased overall affinity then leads to binding to other cellular DNA, titrating transcription factor away from the operator and thus producing expression without IPTG. In this scenario, the allosteric effect of IPTG on LacI actually remains the same as for *WT*, namely a decreased affinity for DNA. With IPTG, the mutant LacI would thus be liberated to start the operator search, which can result in repression if binding affinity to the operator is significant. However, given the recurrence of mutation S97P and its supposed role in the allosteric mechanism, we surmised that the inversion might be based on a modification of the allosteric transition. In this scenario, the evolved LacI would bind the operator with IPTG and have a low affinity for the operator in the absence of IPTG.

To discriminate these two scenarios, we first measured the expression response to the compound ONPF. ONPF is an anti-inducer for *WT* LacI that leads to increased DNA affinity, as seen by a lower expression level than achieved by *WT* repression. For inverted phenotypes with unchanged allostery but increased nonspecific DNA affinity, addition of ONPF can be expected to further increase nonspecific affinity and hence should *not* reduce expression, as has indeed been shown previously (Chamness and Willson, 1970). In contrast, we find that our evolved inverted phenotypes *do* show reduced expression in response to ONPF, with values for P ranging between 0.25 and 0.59 for the five isolates indicated in Table S1. This result is in agreement with the allosteric scenario. In order to provide a direct test of changes in operator affinity, we performed operator binding assays with purified *WT* and an evolved inverse LacI protein (as assayed in Figure 5) in the absence of nonspecific DNA (Figure 5C). These data support the allosteric model: the

evolved LacI is able to bind the operator but only in the presence of IPTG.

DISCUSSION

Recent decades have brought important advances to our understanding of how cellular regulation evolves, for instance by rewiring of regulatory networks (Isalan et al., 2008), cooption of existing transcription factors to new regulatory regions (True and Carroll, 2002) and ligands (Bridgham et al., 2006), evolving transcription factor protein functions (Hoekstra and Coyne, 2007), recombination of regulatory protein domains (Chothia et al., 2003), and by the convergent evolution toward generic network motifs (Milo et al., 2002). These efforts at the genetic, structural, and functional levels are complemented by studies of fitness consequences of regulatory changes in constant environments (Cooper et al., 2003, 2008; Woods et al., 2006) and some in variable environments (Suiter et al., 2003; Mitchell et al., 2009). Fitness in variable environments has been extensively studied within evolutionary ecology, wherein the ability to alter the expressed phenotype in response to environmental changes is typically referred to as phenotypic plasticity (Pigliucci, 2001; DeWitt and Scheiner, 2004). However, direct experimental study of the adaptive evolution of phenotypic plasticity by mutation and natural selection (Vasi et al., 1994) is a challenge because of limited insight into the genetic basis of the studied regulatory phenotypes and the selective forces in variable environments (Scheiner, 2002).

Here, we find that that the evolutionary trajectory in variable environments is altered by internal constraint of the regulatory system, but the evolutionary outcome is the optimal solution of a fitness tradeoff problem and hence determined by selection. The measured expression fitness relations of a founding genotype in constant environments predicted the optimal regulatory response to a challenging variable environment, whereby expression changes that are beneficial in one environment are detrimental in another. Cells were subsequently shown to evolve to the optimum response, in which the sensed cue (originally an inducer) acts as a corepressor. These findings indicate that regulatory evolution by selection in variable environments may be understood within the framework of multiobjective optimization theory (Sawaragi et al., 1985), which addresses the maximization of overall performance in the presence of tradeoffs between conflicting objectives. It will be intriguing to test the limits of prediction by optimization principles, for instance by studying more elaborate regulatory phenotypes and environments, and by determining the fundamental conflicts between objectives that are imposed by physico-chemical constraints.

The observed absence of global constraints in the synthetic system studied here raises the question of whether transcriptional regulation phenotypes found in microorganisms similarly are near-optimal biological functions. Insight into how transcriptional regulation systems achieve optimality would help to explain the adaptive origins of the highly specialized repertoire of signal detection and transmission capabilities found in transcription factor families (Madan Babu and Teichmann, 2003), as well as the observed convergence toward generic regulatory network motifs (Milo et al., 2002).

Local constraints within the *lac* regulatory system that could be broken by selection affected the evolutionary trajectory. Whereas the population predominantly followed the direction of strongest selection within phenotype space, genetic constraint caused delay and deviation at a distinct boundary defined by a single parameter (at $P = 1$). Upon the fixation of mutations that provide the functional innovation—in which IPTG acts as a corepressor rather than an inducer—a diverse set of genetic changes offered a fine-grained variation in both P and E_0 that is essential to developing optimality in a novel function. The picture emerging from these observations is a genotype space consisting of multiple regions that confer different regulatory functions, which are distinct and connected by just a few mutations. The data highlight the potential of regulatory evolution by changes in protein coding regions, which complement other mechanisms such as the co-option of existing regulatory proteins to new regulatory regions.

Central to our strategy was the engineering of a model system that allowed us to independently measure the cross-environmental tradeoffs, which in turn informed us of the selective forces and constraint in regulation. This genetic engineering approach may offer a starting point for quantitative models describing the adaptive evolution of biological complexity, relating environmental conditions, regulatory architecture, adaptive constraint, and competition dynamics.

EXPERIMENTAL PROCEDURES

Strains

In all selection experiments, *Escherichia coli* K12 strain MC1061 (Casadaban and Cohen, 1980) was used, which carries a deletion of the complete *lac* operon. Genotype of MC1061: *F⁻ΔlacX74 mcrB1 e14⁻ (mcrA0) rpsL150 (StrR) galE15 galK16 Δ(ara,leu)7697 araΔ139 λ⁻ hsdR2(rk⁻,mk⁺) spoT1*. This strain was obtained from Avidity LLC, Denver CO, USA, as electrocompetent strain EVB100 (containing an additional chromosomal *birA*). All growth and expression measurements, as well as the selection and competition experiments, were performed in Defined Rich medium (Teknova, Hollister, CA, USA), with 0.2% glucose as carbon source, and supplemented with 1 mM thiamine HCl. For protein expression, we used the BLIM/pTara system (Wycuff and Matthews, 2000) with an arabinose-inducible T7 polymerase and lacking a native *lac* repressor. After transformation of the pRSET-B (Invitrogen) plasmid expressing *lacI* into BLIM/pTara cells, all growth was performed in M9T medium (Wycuff and Matthews, 2000) containing 0.5% glucose and the appropriate antibiotics.

Plasmids

Two plasmids were constructed based on the pZ vector system (Lutz and Bujard, 1997) in which the expression of the selection module is regulated by *lacI* (pRD007). The selection module consists of the coexpressed genes *lacZα*, *cmR*, and *sacB* under control of the Ptrc promoter from pTrc99A (Amann et al., 1988) (which is amplified until base pair −300 before start). Reporter gene *lacZα* consists of the first 364 base pairs of *lacZ*, amplified from the chromosome of strain MG1655 (CGSC stock center). Chloramphenicol resistance gene *cmR* originates from the pZ vector system. The levan sucrase coding sequence *sacB* was amplified from plasmid pKNG101, obtained from the BCCM/LMBP Plasmid and DNA Library Collection (Belgium), accession number LMBP 5246. Two reporter plasmids (pRepLacZω and pReplacZ) were created for measuring expression either in *cis* or in *trans*, respectively, by deleting pTrc99A for *lacI* and Ptrc and inserting a constitutive P_{lacI^q}-*lacZω* fragment or by deleting pTrc99A for *lacI* and *Ptrc* and inserting the MG1655 *Plac-lacZ* fragment. Between *cis* and *trans* expression levels, an empirical relation was observed $E_{cis} = 2.3 \times 10^4 \cdot E_{trans}^{0.32}$. For the production of wild-type and mutant *lac* repressor protein, the *lacI* coding sequence was inserted

directly downstream of the enterokinase cleavage site of expression plasmid pRSET-B (Invitrogen). Plasmids and sequences are available upon request.

Mutagenesis

Mutants were created using the Stratagene Genemorph II Random Mutagenesis kit. Mutagenized product was restricted and ligated into the (nonmutated) selection vector and subsequently transformed into *E. coli* strain MC1061 by electroporation. Pool sizes were routinely between 5×10^5 and 1×10^7. Throughout this experiment, mutagenesis conditions were constant. In order to determine the mutation rate, a random sample of mutants was sequenced after one mutagenesis round, yielding an average mutation rate of 0.003/bp (n = 9).

In Vitro Binding Assay

Fluorescence polarization was measured using purified repressor and 3′-carboxytetramethylrhodamine (TAMRA)-labeled 18 base pair symmetric operator DNA. Polarization values were recorded in a 384-well plate in a Victor 3V plate reader (Perkin Elmer) at 531 nm excitation and 595 nm emission. Further details are given in Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, one figure and one table and can be found with this article online at doi:10.1016/j.cell.2011.06.035.

ACKNOWLEDGMENTS

We thank Kathleen S. Matthews for kindly providing the BLIM cells and pTara plasmid for our expression of the *lac* repressors. We thank Kim Renders and Roland Dries for technical assistance and Bertus Beaumont, Tom Shimizu, Ron Milo, and members of the Tans and ten Wolde groups for critical reading of the manuscript. We thank Walraj Gosal and other members of the Ranganathan lab for help with lac repressor purification. This work is part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijke Onderzoek (NWO).

Received: December 23, 2010
Revised: May 5, 2011
Accepted: June 16, 2011
Published online: July 28, 2011

REFERENCES

Agrawal, A.A. (2001). Phenotypic plasticity in the interactions and evolution of species. Science *294*, 321–326.

Amann, E., Ochs, B., and Abel, K.J. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. Gene *69*, 301–315.

Benner, S.A., and Sismour, A.M. (2005). Synthetic biology. Nat. Rev. Genet. *6*, 533–543.

Bridgham, J.T., Carroll, S.M., and Thornton, J.W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. Science *312*, 97–101.

Casadaban, M.J., and Cohen, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. J. Mol. Biol. *138*, 179–207.

Chamness, G.C., and Willson, C.D. (1970). An unusual lac repressor mutant. J. Mol. Biol. *53*, 561–565.

Choi, K.Y., and Zalkin, H. (1992). Structural characterization and corepressor binding of the Escherichia coli purine repressor. J. Bacteriol. *174*, 6207–6214.

Chothia, C., Gough, J., Vogel, C., and Teichmann, S.A. (2003). Evolution of the protein repertoire. Science *300*, 1701–1703.

Cooper, T.F., Rozen, D.E., and Lenski, R.E. (2003). Parallel changes in gene expression after 20,000 generations of evolution in Escherichiacoli. Proc. Natl. Acad. Sci. USA *100*, 1072–1077.

Cooper, T.F., Remold, S.K., Lenski, R.E., and Schneider, D. (2008). Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in Escherichia coli. PLoS Genet. *4*, e35.

DeWitt, T.J., and Scheiner, S.M. (2004). Phenotypic Plasticity: Functional and Conceptual Approaches (Oxford: Oxford University Press).

Flynn, T.C., Swint-Kruse, L., Kong, Y., Booth, C., Matthews, K.S., and Ma, J. (2003). Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. Protein Sci. *12*, 2523–2541.

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C.I. (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J. Bacteriol. *164*, 918–921.

Hoekstra, H.E., and Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. Evolution *61*, 995–1016.

Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., Garriga-Canut, M., and Serrano, L. (2008). Evolvability and hierarchy in rewired bacterial gene networks. Nature *452*, 840–845.

Levins, R. (1968). Evolution in Changing Environments (Princeton: Princeton University Press).

Lunzer, M., Miller, S.P., Felsheim, R., and Dean, A.M. (2005). The biochemical architecture of an ancient adaptive landscape. Science *310*, 499–501.

Lutz, R., and Bujard, H. (1997). 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*, 1203–1210.

Madan Babu, M., and Teichmann, S.A. (2003). Evolution of transcription factors and the gene regulatory network in Escherichia coli. Nucleic Acids Res. *31*, 1234–1244.

Maynard Smith, J., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., and Wolpert, L. (1985). Developmental constraints and evolution: A perspective from the Mountain Lake conference on development and evolution. Q. Rev. Biol. *60*, 265–287.

Miller, J.H., and Schmeissner, U. (1979). Genetic studies of the lac repressor. X. Analysis of missense mutations in the lacI gene. J. Mol. Biol. *131*, 223–248.

Miller, S.P., Lunzer, M., and Dean, A.M. (2006). Direct demonstration of an adaptive constraint. Science *314*, 458–461.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824–827.

Mitchell, A., Romano, G.H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., Dahan, O., and Pilpel, Y. (2009). Adaptive prediction of environmental changes by microorganisms. Nature *460*, 220–224.

Pfahl, M. (1976). lac Repressor-operator interaction. Analysis of the X86 repressor mutant. J. Mol. Biol. *106*, 857–869.

Pigliucci, M. (2001). Phenotypic Plasticity: Beyond Nature and Nurture (Baltimore: Johns Hopkins University Press).

Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., and Tans, S.J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. Nature *445*, 383–386.

Roff, D.A. (2001). Life History Evolution (Sunderland, MA: Sinauer Associates).

Sawaragi, Y., Nakayama, H., and Tanino, T. (1985). Theory of multiobjective optimization (Orlando: Academic Press).

Scheiner, S.M. (1993). Genetics and Evolution of Phenotypic Plasticity. Annu. Rev. Ecol. Syst. *24*, 35–68.

Scheiner, S.M. (2002). Selection experiments and the study of phenotypic plasticity. J. Evol. Biol. *15*, 889–898.

Suckow, J., Markiewicz, P., Kleina, L.G., Miller, J., Kisters-Woike, B., and Müller-Hill, B. (1996). Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. J. Mol. Biol. *261*, 509–523.

Suiter, A.M., Bänziger, O., and Dean, A.M. (2003). Fitness consequences of a regulatory polymorphism in a seasonal environment. Proc. Natl. Acad. Sci. USA *100*, 12782–12786.

True, J.R., and Carroll, S.B. (2002). Gene co-option in physiological and morphological evolution. Annu. Rev. Cell Dev. Biol. *18*, 53–80.

van Tienderen, P.H. (1997). Generalists, Specialists, and the Evolution of Phenotypic Plasticity in Sympatric Populations of Distinct Species. Evolution *51*, 1372–1380.

Vasi, F., Travisano, M., and Lenski, R.E. (1994). Long-Term Experimental Evolution in Escherichia-Coli. 2. Changes in Life-History Traits during Adaptation to a Seasonal Environment. Am. Nat. *144*, 432–456.

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S.M., Schlichting, C.D., and Van Tienderen, P.H. (1995). Adaptive phenotypic plasticity: consensus and controversy. Trends Ecol. Evol. (Amst.) *10*, 212–217.

Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. Science *312*, 111–114.

Woods, R., Schneider, D., Winkworth, C.L., Riley, M.A., and Lenski, R.E. (2006). Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. Proc. Natl. Acad. Sci. USA *103*, 9107–9112.

Wycuff, D.R., and Matthews, K.S. (2000). Generation of an AraC-araBAD promoter-regulated T7 expression system. Anal. Biochem. *277*, 67–73.

Zhan, H., Camargo, M., and Matthews, K.S. (2010). Positions 94-98 of the lactose repressor N-subdomain monomer-monomer interface are critical for allosteric communication. Biochemistry *49*, 8636–8645.

Cell

EXTENDED EXPERIMENTAL PROCEDURES

Growth Conditions

Growth Conditions during Selection

Growth was performed at 37°C in 100 ml erlenmeyer flasks, under vigorous shaking. Culture medium was 20 or 40 ml EZ Rich Defined medium + glucose (Teknova, Hollister, CA, USA, cat. nr. M2105), supplemented with 1 mM thiamine HCl, the appropriate antibiotic, and IPTG when needed. Selective compounds (chloramphenicol, sucrose) were added after 3 hr of pre-selection, after which the cultures were grown for 6 hr. The duration of selective growth was chosen to obtain significant enrichment factors (of up to 10^4), while still maintaining diversity in the population (which starts off at about 10^6). The pre-selection time of 3 hr ($\sim$ 6 generations for unselective growth) was chosen in order to be long enough to reach steady state enzyme expression, either due to protein production (in the order of 3 generations (1)), or due to dilution (in which case 6 generations amounts to a dilution factor of 64). Optical density was monitored at 550 nm and whenever an OD_{550} of 0.1 is reached, a dilution was made into fresh prewarmed selective medium. After selection, cultures were washed, and flash frozen. When transferred to the next environment (without mutagenesis), a threshold dilution of 5 10^2 is applied, which determines the minimum growth rate for mutants to effectively increase in number in the previous environment.

Growth Conditions during Measurement of the Fitness Landscapes

In order to measure growth rates for determination of the fitness landscape, wells of a 96-well plate containing 200 μl of Defined Rich glucose medium with the appropriate amount of IPTG were inoculated with a $2 \cdot 10^4$ x dilution of an O/N (LB) culture, and grown for three hours (pre-selection) at 37°C in a Perkin & Elmer Victor^3 plate reader until an optical density at 550 nm of around 0.0005 was reached (in the plate reader, which corresponds to an OD_{550} of 0.002). As this OD is too low to be determined, the same plate contains 6 wells that were inoculated with a mere $5 \cdot 10^2$ x dilution, which reached a measurable OD of around 0.02 at the same time. At that moment sucrose or chloramphenicol was added.

Optical density at 550 nm was recorded every 4 min, and every 29 min 9 μl sterile water was added to each well to counteract evaporation. When not measuring, the plate reader was shaking the plate at double orbit with a diameter of 2 mm.

From the measured growth curves (see examples in Figure S1) the growth rate was obtained by determining for each well the increase in cell density at t = 6 hr. From this the effective exponential growth rate (or Malthusian parameter, see e.g., (2)) was obtained according to $F = (\log(OD_{t=6}/OD_{t=0})/\log 2)/6$, in doublings per hour. In case the growth rate was high and stationary phase was reached within 6 hr, the initial slope of the growth curve was taken, since in the selection experiments the cultures were always diluted long before reaching stationary phase.

In the main text the fitness F is identified with the Malthusian parameter (2). This is the appropriate measure of the fitness for clonal organisms with variable generation times and overlapping generations, such as bacteria. The average growth rate over multiple environmental conditions is $\overline{F} = \sum_i p_i F_i$, where the p_i's denote the time fraction spent in environmental condition i, and F_i is the fitness for that condition. For the two conditions used in the main text at equal dwelling times, the fitness therefore can be expressed as $F_{tot} = (F_{sucrose} + F_{Cm})/2$ (see caption Figure 2 in the main text).

Growth Conditions during Competition Assays

Two mutants were mixed in a known ratio and subjected to selective environments (see above). After 6 hr of growth for a certain environment (in which the initial inoculation was such that an OD_{550} of just under 0.1 was reached), cultures were washed, and allowed to grow to stationary phase in LB medium. A DNA extraction was performed on the whole population for each culture, of which subsequently around 0.1 ng was electroporated into BioRad EP-Max10B Electro-Competent Cells (cat. no. 170-3330), and directly plated on agar containing Xgal with or without IPTG. As our selection module contains a *lacZ*α gene, complementation with the chromosomally expressed *lacZ*ω allowed for discrimination between the mutants and determination of their ratio, on the basis of their differential expression of *lacZ*α. For example, WT and mutant N1 (Figure 3B) can be distinguished on plates containing Xgal and IPTG, where WT forms blue colonies and N1 remains white. We can then calculate their fitness difference $\Delta F = \log(A_2^1)/(6 \log 2)$, where A_2^1 is the factor by which mutant 1 out-grows mutant 2 during 6 hr of growth. The graph in Figure 3C contains 15 data points for 3 competitions: WT versus N1, N1 versus R1, WT versus N2, each in 5 different environments: 0.4% sucrose, 0.25% sucrose, 0.15% sucrose, 80 μg/ml cm + 1 mM IPTG, 25 μg /ml cm + 1 mM IPTG.

Growth Conditions during the Induction/De-induction Assays

To measure the rates of induction and de-induction after a switch in IPTG concentration, exponentially growing cultures expressing WT *lacI* were grown for four hours in the presence or absence of 1 mM IPTG. Subsequently a time series for 3.5 hr was performed. At each time point 1 mM IPTG was added to a culture that had grown previously without IPTG. For cultures that had previously grown with IPTG, the medium was exchanged for medium without IPTG at each time point (after washing the cells). Growth was monitored such that the OD at 600 nm never exceeded 0.05. After 3.5 hr the expression level for each culture experiencing induction or de-induction for a different period of time was measured by an FDG assay (Figure S1C, left).

Assay of Growth Recovery after Selective Periods

Growth during the periods of non-selective growth was assayed by plating cultures over time and counting the number of colony forming units (CFU). Cultures grown selectively for 6 hr (see above) were washed and grown in Defined Rich medium + glucose

for 3.5 hr. At specific time points a part of the growing culture was taken, appropriately diluted, and spread on an LB plates. Plates were grown overnight at 37°C, after which colonies were counted (Figure S1C center and S1C right).

β-Galactosidase Assay

Assay Conditions

A reporter plasmid expressing either *lacZ* or *lacZω* was cotransformed into the mutant population or clone that is to be assayed. Cell cultures were grown at 37°C in a Perkin & Elmer Victor^3 plate reader, at 200 μl per well in a black clear-bottom 96 well microtiter plate (NUNC 165305). Medium was EZ Rich Defined medium + glucose (Teknova, Hollister, CA, USA, cat. nr. M2105), supplemented with 1 mM thiamine HCl and the appropriate antibiotics. Optical density at 600 nm was recorded every 4 min, and every 29 min 9 sterile water is added to each well to counteract evaporation. When not measuring, the plate reader was shaking the plate at double orbit with a diameter of 2 mm. Cells were fixed after the cultures had reached an optical density of at least 0.015 and at most 0.07 (in the plate reader, which corresponds to an OD_{600} of 0.05 to 0.23), by adding 20 μl FDG-fixation solution (109 μM fluorescein di-β-D-galactopyranoside (FDG, MarkerGene Technologies Inc, Eugene, OR, USA, cat. nr M0250), 0.15% formaldehyde, and 0.04% DMSO in ddH_2O). Fluorescence development was measured every 8 min (exc. 480 nm, em. 535 nm), as well as the optical density at 600 nm. Shaking and dispensing conditions as above. When cells are induced with IPTG, directly before or after fixation, an appropriate amount of inhibitive IPTG was added. Analysis of fluorescence trace is described below.

Analysis of the FDG β-Galactosidase Assay

Using the fluorogenic substrate fluorescein di-β-D-galactopyranoside (FDG) allows for an accurate determination of the β-galactosidase activity over at least 4.5 orders of magnitude. FDG contains two galactose groups that both have to be cleaved in order to release fluorescein.

$$\mathrm{FDG} \xrightarrow{\mathrm{LacZ}} \mathrm{FMG} + \mathrm{galactose}$$

$$\mathrm{FMG} \xrightarrow{\mathrm{LacZ}} \mathrm{fluorescein} + \mathrm{galactose}$$

In ref. (3) a model for the FDG-FMG hydrolysis is proposed. In the concentration range of LacZ and FDG used in our experiments, the increase in fluorescence F is given by (Equation 7 in ref. (3)):

$$\frac{d}{dt}F = k_2 E \frac{S_0}{K_m + S_0}\left(\alpha_p + (\alpha_M - \alpha_p)e^{-Rt}\right) \tag{S.1}$$

where R is the relaxation constant (timescale to reach maximum fluorescence rate), E is the (total) concentration of enzyme, k_2 is the catalysis rate constant of FDG to FMG, and the α's are proportionality factors between the products and their fluorescence, $F = \alpha_P P + \alpha_M M$ (P is product (fluorescein) and M is FMG). K_m is the Michaelis-Menten constant for FDG and S_0 is the initial FDG concentration. We can see that at time $t = 0$ as well as at large t's the rate is proportional to E, though with different proportionality constants (first α_M, then α_P).

Ref. (3) gives measured values for $\alpha_M = 5.3\ \mu M^{-1}$ and $\alpha_P = 150\ \mu M^{-1}$. Although assigning arbitrary units to the fluorescence counts, they are relevant as relative quantities between FMG and fluorescein. Thus at $t = 0$, equation (S.1) reduces to

$$\frac{d}{dt}F = \alpha_M k_2 E \frac{S_0}{K_m + S_0} \tag{S.2}$$

In order to determine the enzyme concentration per cell, fitted slopes at $t = 0$ are divided by the cell density. We use here $\varepsilon \propto E/OD_{600}$, where ε is the LacZ concentration per cell.

FDG expression measurements were compared to the standard Miller assay for β-galactosidase activity (4). We measured an induction curve of wild-type LacI (as expressed from plasmid pRD007), both by using the Miller assay and the FDG assay described above (Figure S1B).

Growth Models and Interpolation of Expression-Growth Curves

In order to interpolate the measured points on the expression-growth relations in Figure 1C and 1D, we use models for the selective action of chloramphenicol and sucrose.

Chloramphenicol Growth

In the presence of a certain concentration of chloramphenicol acetyl transferase (CmR), the internal concentration of chloramphenicol (cm) is reduced and determined by the equilibrium between influx through the cell membrane and acetylation ('inactivation') by CmR. As such, we model the action of cm by comparing the situation with growth under sublethal concentrations of cm. The most basic equation relating growth to the concentration of an inhibitive substance is derived from the Monod form for nutrient limited

growth (5), $\mu \sim KX/X+K$, where μ is the growth rate, X is the concentration of nutrient and K is a constant determining the nutrient concentration that allows half-maximum growth rate. Interestingly, this is the same functional form as the fraction of substrate bound enzyme under Michaelis-Menten kinetics. Now, for sublethal concentrations of cm, whose action is to block protein synthesis upon binding to the ribosomes, it would not be unreasonable to expect the the growth of the cell (as a first-order approximation) to be proportional to the *unbound* fraction of ribosomes, which is given by $K/X+K$. Therefore we adopt the following simple functional form for growth in the presence of chloramphenicol

$$\mu\left([\mathrm{cm}]_{\mathrm{ext}}\right) = \frac{\mu_0}{c_1 [cm]_{\mathrm{int}} + 1} \tag{S.4}$$

where c_1 is a constant, μ_0 the growth rate in absence of cm, and $[\mathrm{cm}]_{\mathrm{ext}}$ and $[\mathrm{cm}]_{\mathrm{int}}$ respectively the cm concentrations outside and inside the cell.

To obtain a relation between the internal and external cm concentration, we express the equilibrium between influx and acetylation of cm by

$$C_{\mathrm{bar,cm}}\left([\mathrm{cm}]_{\mathrm{ext}} - [\mathrm{cm}]_{\mathrm{int}}\right) = r_{\mathrm{acet,cm}} \tag{S.5}$$

Here the influx of cm is either diffusion limited or limited by the permeability of the membrane, which does not matter for the functional form of the equation, and can be expressed as a constant $C_{\mathrm{bar,cm}}$ times the concentration difference between inside and outside. The acetylation rate $r_{\mathrm{acet,cm}}$ is given by

$$r_{\mathrm{acet,cm}} = k_{\mathrm{cat,cm}}[E * \mathrm{cm}] = k_{\mathrm{cat,cm}} \frac{E_{\mathrm{tot}}}{1 + \frac{K_{mEcm}}{[cm]_{\mathrm{int}}}} \tag{S.6}$$

where $k_{\mathrm{cat,cm}}$ is the catalysis rate constant for the acetylation reaction, and $K_{\mathrm{m}Ecm}$ is the Michaelis-Menten constant for CmR. Solving for $[\mathrm{cm}]_{\mathrm{int}}$ in

$$C_{\mathrm{bar,cm}}\left([\mathrm{cm}]_{\mathrm{ext}} - [\mathrm{cm}]_{\mathrm{int}}\right) = k_{\mathrm{cat,cm}} \frac{E_{\mathrm{tot}}}{1 + \frac{K_{mEcm}}{[cm]_{\mathrm{int}}}} \tag{S.7}$$

now yields the expression for the growth rate as a function of the external chloramphenicol concentration (here abbreviated as [cm]), being

$$\mu([cm]) = \frac{\mu_0}{1 + \frac{c_1}{2}\left([cm] - K_{mEcm} - E_{tot}k_{cat,cm} + \sqrt{4[cm]K_{mEcm} + ([cm] - K_{mEcm} - E_{tot}k_{cat,cm})^2}\right)} \tag{S.8}$$

Expression-growth data for media containing cm in Figure 1C were fitted with this equation, using the following parameter set:

[cm]	80 μg/ml	40 μg/ml	25 μg/ml
μ_0	1.7	1.8	1.8
c_1	0.17	0.17	0.17
n	1	1	1
$10^5 \cdot k_{\mathrm{cat}}$	10.8	6.8	4.5
$K_{\mathrm{m}Ecm}$	0.44	0.44	0.44

Sucrose Growth

Sucrose selection is based on the formation of sugar chains (levan) in the periplasmic domain of gram-negative bacteria (6). The enzyme catalyzing this polymerization reaction is levansucrase (SacB) from *Bacillus subtilis*. In the gram-positive *B. subtilis* the enzyme is exported through the inner membrane, where it constitutes a protective poly-sugar layer outside the cell wall. In gram-negative bacteria, which have a second cellular membrane, the enzyme is not exported through the second membrane and therefore accumulates levans in between the cellular membranes, which decreases the cellular growth rate. High expression of the protein in the presence of sucrose is lethal and leads to lysis of the cells. Thus, the rate of levan formation is the factor influencing cell growth. In contrast to chloramphenicol, which is a bacteriostatic, high levan production leads to lysis of cells, and in a population average this

can give rise to a negative growth rate. Therefore the growth as a function of levan formation rate cannot directly be described by the Monod form. However, expecting that the relevant parameter for the toxic effect is the levan formation rate (r_{levan}) *relative* to the instantaneous growth rate, we can write a modified Monod form

$$\mu([\text{sucrose}]) = \frac{\mu_0}{c_0 \frac{r_{levan}}{\mu([sucrose])} + 1} \tag{S.9}$$

where μ_0 is the growth rate in absence of sucrose and c_0 is a constant. (A stronger effect of sucrose was indeed observed when the basal growth rate is lowered [e.g., growth with glycerol as a carbon source instead of glucose]). This can be solved to yield

$$\mu([\text{sucrose}]) = \mu_0 - c_0\, r_{\text{levan}} \tag{S.10}$$

Analogous to the chloramphenicol selection, we write for the rate of levan formation and the equilibrium governing the transport of sucrose through the outer membrane

$$r_{\text{levan}} = k_{\text{cat,sucr}}[E * \text{sucrose}] = k_{\text{cat,sucr}} \frac{E_{tot}}{1 + \frac{K_{mEsucr}}{[sucrose]_{\text{int}}}} \tag{S.11}$$

and

$$C_{\text{bar,sucr}}\left([\text{sucrose}]_{\text{ext}} - [\text{sucrose}]_{\text{int}}\right) = k_{\text{cat,sucr}} \frac{E_{tot}}{1 + \frac{K_{mEsucr}}{[sucrose]_{\text{int}}}} \tag{S.12}$$

We can solve equation (S.12) for $[\text{sucrose}]_{\text{int}}$, substitute this into equation (S.11), which in its turn can be substituted into equation (S.10) to obtain the growth rate as a function of the external sucrose concentration and the expression of *sacB*.

However, all measured expression-growth characteristics for sucrose show a steeper dependency on enzyme concentration than can be obtained by this form. Indeed, SacB mediated formation of levan is a process that needs a levan seed in order to proceed (7,8). Most probably seed formation is also dependent on the enzyme and sucrose concentration. Therefore we phenomenologically alter the equation for the growth rate as a function of levan formation rate into

$$\mu([\text{sucrose}]) = \mu_0 - c_0\, r_{\text{levan}}^{n} \tag{S.13}$$

The obtained function provides good fits for the low-enzyme regime of the expression-growth data, with the following parameter set:

[sucrose]	0.4%	0.25%	0.15%
μ_0	2.18	2.14	2.18
c_0	86	86	86
n	4	4	4
$10^7 \cdot k_{cat}$	4.8	6.4	4.9
K_{mEsucr}	0.35	0.35	0.23

However, at the high [*E*] end, we observe a saturation at higher growth rates than equation (S.13) can account for. There are at least three saturation effects (see also (7,8)) coming into play at high rates of levan synthesis (apart from potential feedback on protein production in 'struggling' cells):

(1) Since the levans are (possibly branching) chains, the autocatalytic seed-effect (see above) of the reaction decreases: attaching a fructosyl-group to an existing long chain does not increase the number of fructosyl-acceptors.
(2) At high levan production rates, there is a high concomitant production of glucose, that has an inhibitory effect on levan formation in two ways:
  (2a) The fructosylation reaction by the *E* * *S* (levansucrase-sucrose) complex branches between levan elongation and fructosylation of glucose (which re-forms sucrose).
  (2b) The competitive inhibition of *E* to *S* binding by glucose. Due to levan formation, the internal sucrose concentration decreases, and the glucose concentration increases.
(3) The formed levans themselves act as an inhibitor at higher concentrations.

Since it is at this stage impossible and will not yield further insight to adapt the model to account for the saturation at high enzyme concentration, we opt for a more phenomenological description. For fits over the complete concentration range of SacB enzyme (Figure 1D), we used

$$\mu([\text{sucrose}]) = \frac{\mu_0 + \mu_{sat}}{c_0\, r_{levan}^{\,n} + 1} - \mu_{\text{sat}} \tag{S.14}$$

which approaches growth rate μ_{sat} for high expression levels.

Repressor Protein Expression Conditions and Fluorescence Polarization

Repressor Protein Expression Conditions

For the production of wild-type and mutant *lac* repressor protein, the *lacI* coding sequence was inserted into expression plasmid pRSET-B (Invitrogen), in which a T7 promoter drives expression of the His-tagged repressor. The BLIM/pTara system (9) with an arabinose inducible T7 polymerase, and lacking a native *lac* repressor, was used for all protein expression. After transformation of the pRSET-B plasmid into BLIM cells containing pTara, all growth was performed in M9T medium (9) containing 0.5% glucose. 100 mg/l ampicillin and 40 mg/l chloramphenicol was used to retain pRSET-B and pTara respectively, except when 1l cultures were grown for expression, in which case the chloramphenicol concentration was lowered to 15 mg/l. Protein expression was induced at absorbance between 0.9 and 1.1 at 600 nm, by addition of 0.25% arabinose, after lowering the temperature to 17°C. Cells were harvested by centrifugation after 16 hr of induction.

Subsequently the cell pellet was resuspended in a 50 mM sodium phosphate buffer at pH 8.0 containing 500 mM NaCl, 20 mM imidazole, 2.5% glycerol, 1 mM DTT, 10 mM $MgCl_2$, 0.1% tween 20, 20 mg lysozyme, and one tablet of protease inhibitor cocktail (Roche). Cells were lysed by sonication, ~2000U of DNase was added and allowed to incubate for 10 min at 4°C. The suspension was cleared by centrifugation at 4°C for one hour at $4.8 \cdot 10^4$g in a Sorvall SS-34 rotor. 0.5 ml of Ni-NTA agarose (QIAGEN) was added and incubated for 30 min at 4°C. The agarose was batch washed 5x in a 50 mM sodium phosphate buffer at pH 8.0 containing 500 mM NaCl, 20 mM imidazole, 2.5% glycerol, 1 mM DTT, and the protein was eluted with 250 mM imidazole. The protein solution was dialyzed overnight into a 50 mM HEPES buffer at pH 8.0 containing 200 mM NaCl, 20 mM imidazole, 1% glycerol and 1 mM DTT.

Fluorescence Polarization

Fluorescence polarization measurements were performed in the dialysis buffer with addition of BSA to 0.05%. Oligonucleotides containing the 18 base pair symmetric *lac* operator (ATTGTGAGCGCTCACAAT) and containing a 3′-carboxytetramethylrhodamine (TAMRA) fluorophore (Integrated DNA Technologies) were hybridized in a 10 mM Tris-Cl buffer at pH 8.5, by cooling down overnight in a water bath from 95°C to room temperature. The polarization assay was performed in a 384 well plate in a Victor 3V plate reader (Perkin Elmer) at 531 nm excitation and 595 nm emission. Each well contained 50 μl dialysis buffer and 10 nM of *lac* operator. IPTG was added at appropriate concentrations. The amount of repressor protein was such that saturating binding could be observed (without IPTG for wild-type and with IPTG for M-inv-1). Each measurement was performed with 3 or 4 replicates.

SUPPLEMENTAL REFERENCES

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C.I. (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J. Bacteriol. *164*, 918–921.

Huang, Z.J. (1991). Kinetic fluorescence measurement of fluorescein di-β-D-galactoside hydrolysis by β-galactosidase: intermediate channeling in stepwise catalysis by a free single enzyme. Biochemistry *30*, 8535–8540.

Miller, J.H. (1972). Experiments in Molecular Genetics (New York: Cold Spring Harbor Laboratory Press).

Monod, J. (1949). The growth of bacterial cultures. Annu. Rev. Microbiol. *3*, 371–394.

Novick, A., and Weiner, M. (1957). Enzyme induction as an all-or-none phenomenon. Proc. Natl. Acad. Sci. USA *43*, 553–566.

Roff, D.A. (2002). Life History Evolution (Sunderland: Sinauer Associates).

Chambert, R., and Gonzy-Tréboul, G. (1976). Levansucrase of *Bacillus subtilis*: kinetic and thermodynamic aspects of transfructosylation processes. Eur. J. Biochem. *62*, 55–64.

Chambert, R., Treboul, G., and Dedonder, R. (1974). Kinetic studies of levansucrase of *Bacillus subtilis.* Eur. J. Biochem. *41*, 285–300.

Wycuff, D.R., and Matthews, K.S. (2000). Generation of an AraC-araBAD promoter-regulated T7 expression system. Anal. Biochem. *277*, 67–73.

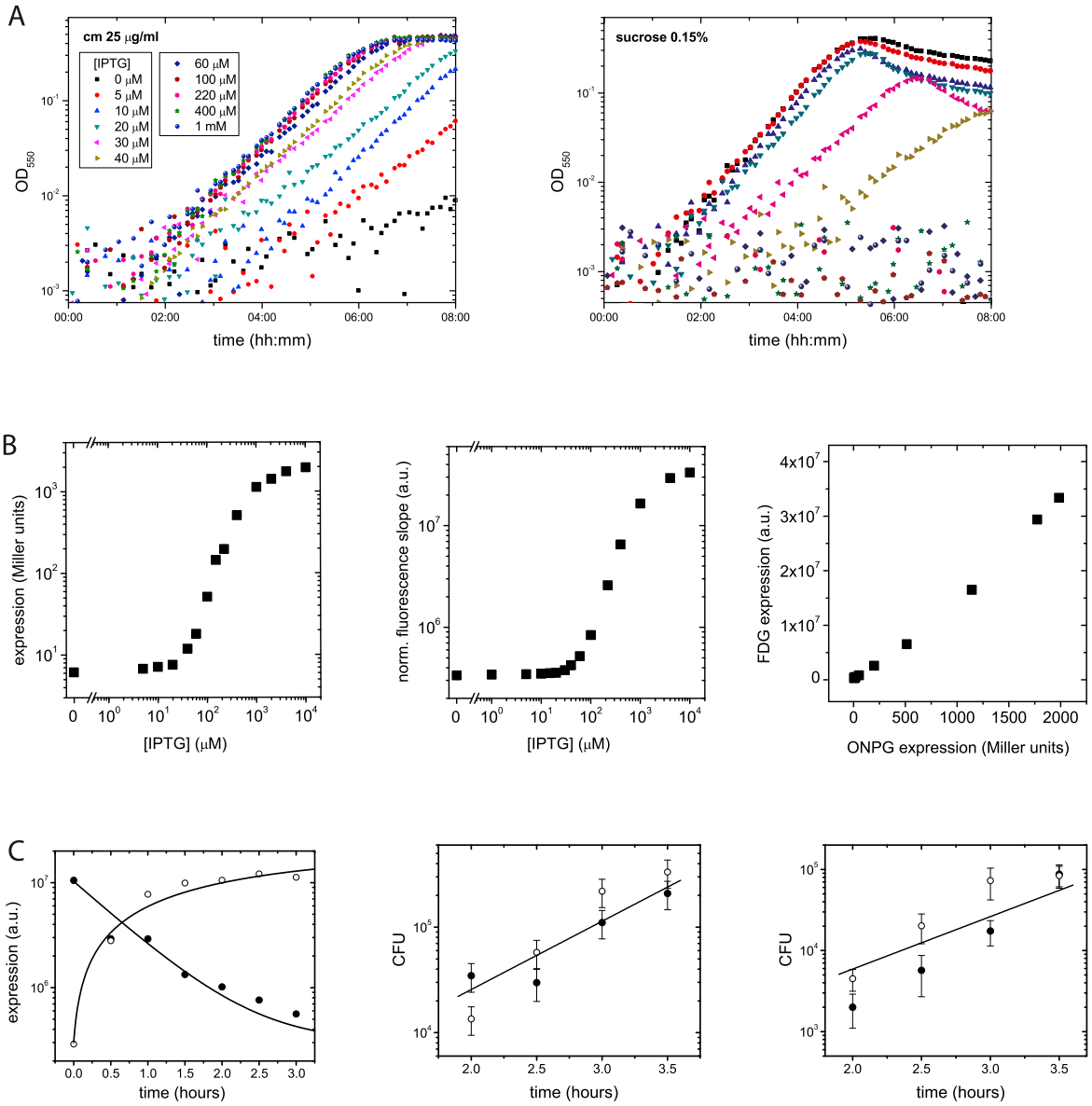

Figure S1. Growth Rate and Expression Measurements, Related to Figure 1

(A) Examples of measured growth curves of cells harboring the selection module, with 25 μg/ml chloramphenicol (left) or 0.15% sucrose (right). Optical density of the culture is recorded as a function of time in a 96-well plate reader. Different concentrations of the inducer IPTG are indicated.

(B) Comparison of Miller assay and fluorescein di-β-D-galactopyranoside (FDG) assay. Shown is the expression in response to the inducer IPTG as determined by the Miller assay (left) and the FDG assay (center), and the FDG expression values against the Miller expression values (right). For details see Supplemental Experimental Procedures B.

(C) left: induction and de-induction of WT *lacI* as a function of time. At t = 0 either 1 mM IPTG is added to a culture growing without IPTG (open symbols) or a culture previously growing at 1 mM IPTG is grown further in the absence of IPTG (solid symbols). See Supplemental Experimental Procedures A4 for details. The curves are theoretical expectations based on a growth rate of 2.15 doubl./hour. Center and right: Recovery after selective periods, as a function of time after transfer to non-selective medium (see Supplemental Experimental Procedures A5). CFU is shown after 80 μg/ml chloramphenicol (center) with 1 mM (solid symbols) and 25 μM IPTG (open symbols), and 0.4% sucrose (right) with 1 mM IPTG (open symbols) and without IPTG (solid symbols). Straight lines indicate 2.15 doubl./hour.