OPEN ACCESS Freely available online

PLoS GENETICS

Nonlinear Fitness Landscape of a Molecular Pathway

Lilia Perfeito[1,2]*, Stéphane Ghozzi[2], Johannes Berg[2], Karin Schnetz[1], Michael Lässig[2]*

1 Institut für Genetik, Universität zu Köln, Cologne, Germany, **2** Institut für Theoretische Physik, Universität zu Köln, Cologne, Germany

Abstract

Genes are regulated because their expression involves a fitness cost to the organism. The production of proteins by transcription and translation is a well-known cost factor, but the enzymatic activity of the proteins produced can also reduce fitness, depending on the internal state and the environment of the cell. Here, we map the fitness costs of a key metabolic network, the lactose utilization pathway in *Escherichia coli*. We measure the growth of several regulatory *lac* operon mutants in different environments inducing expression of the *lac* genes. We find a strikingly nonlinear fitness landscape, which depends on the production rate and on the activity rate of the *lac* proteins. A simple fitness model of the *lac* pathway, based on elementary biophysical processes, predicts the growth rate of all observed strains. The nonlinearity of fitness is explained by a feedback loop: production and activity of the *lac* proteins reduce growth, but growth also affects the density of these molecules. This nonlinearity has important consequences for molecular function and evolution. It generates a cliff in the fitness landscape, beyond which populations cannot maintain growth. In viable populations, there is an expression barrier of the *lac* genes, which cannot be exceeded in any stationary growth process. Furthermore, the nonlinearity determines how the fitness of operon mutants depends on the inducer environment. We argue that fitness nonlinearities, expression barriers, and gene–environment interactions are generic features of fitness landscapes for metabolic pathways, and we discuss their implications for the evolution of regulation.

Citation: Perfeito L, Ghozzi S, Berg J, Schnetz K, Lässig M (2011) Nonlinear Fitness Landscape of a Molecular Pathway. PLoS Genet 7(7): e1002160. doi:10.1371/journal.pgen.1002160

Editor: Ivan Matic, Université Paris Descartes, INSERM U1001, France

Received November 11, 2010; **Accepted** May 14, 2011; **Published** July 21, 2011

Funding: This work was funded by the Deutsche Forschungsgemeinschaft grant SFB 680 and Fundação para a Ciência e Tecnologia grant SFRH\BPD\43224\2008 (LP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lassig@thp.uni-koeln.de (ML); lperfeit@uni-koeln.de (LP)

Introduction

Gene regulation is a major factor of molecular evolution, and changes in gene expression contribute to phenotypic differences between species [1]. Expression levels are under natural selection, which results from a balance between costs and benefits for the organism. For single-cell organisms, fitness benefits include the ability to digest nutrients in different environments. The cost of gene expression, on the other hand, depends on the biophysics of protein production and of protein activity. The cost of protein production has been studied extensively [2–6]. However, enzymatic activities of proteins can also reduce fitness due to energy consumption or toxic effects of the reaction products. What are the relative contributions of these two effects? How do they interact? To address these questions, we have to understand the fitness effects of an entire metabolic pathway, in which protein production is coupled to function and growth. This is the subject of the present paper.

For our analysis, we use the lactose utilization pathway in *Escherichia coli*, which is one of the best characterized molecular pathways [7]. It is coded in a set of genes referred to as the *lac* operon. Several studies have addressed fitness effects associated with expression of the *lac* genes. In particular, production of the *lac* proteins in the absence of lactose has been shown to involve a fitness cost, that is, to reduce the growth rate of a cell population [2–4,8,9]. This cost has been ascribed to transcription and translation of the *lac* genes [10], because toxic effects of the gene products have not been observed. Growth is also reduced by the presence of inducers in the medium, even after the maximum of expression is reached [8]. This fitness cost is likely to arise from inducer transport through the cell membrane [11,12]. Furthermore, the *lac* operon has been used to study the interplay of cost and benefit in the evolution of gene expression [9]. Taken together, these observations make the *lac* operon an ideal system to study the coupled fitness costs of protein production and activity.

Here we determine a fitness landscape of the *lac* pathway by a combined experimental and theoretical approach. We measure the fitness of different regulatory mutant strains in the presence and absence of the *lac* inducer IPTG and of the natural sugar lactose. LacY proteins act as transporters (so-called permeases) for IPTG and lactose (i.e., these molecules are substrates of LacY). We develop a quantitative biophysical growth model to disentangle the fitness contributions of protein production (i.e., transcription and translation) and of protein activity (i.e., intra-cellular transport). The model explains the growth rate of all observed mutants in different inducer environments. Its key element is a *feedback loop* between the *lac* pathway and fitness: at constant rate of protein production, faster cell growth leads to stronger dilution of proteins and lowers the cost of protein activity. In addition, the rate of *lac* gene expression itself can depend on growth [13,14]. Similar growth feedback mechanisms have been argued to play an important role in bacterial drug resistance [15,16], and to generate diversity in an isogenic population [13,15,17].

Our analysis suggests that growth feedback is a pervasive feature of the activity-dependent fitness of metabolic pathways. This feature has important evolutionary consequences. In particular, our model predicts a *fitness cliff*, beyond which populations cannot maintain viable growth, and an *expression barrier*, that is, an upper

Author Summary

The levels of protein produced by an organism are likely to change its fitness, potentially driving the evolution of genetic regulation. Importantly, protein expression generates costs as well as benefits. Here, we use a model genetic system, the *lac* operon of *Escherichia coli*, to investigate different sources of fitness costs. We find that fitness depends not only on the production rate of proteins but also on their enzymatic activity. A simple quantitative model, which is based on the biophysics of protein production and activity, accurately reproduces the experimental results and provides testable predictions. The model describes a feedback cycle between a molecular pathway and the growth rate of cells: pathway activity impedes growth, but growth itself affects the pathway. This feedback can generate dramatic effects, such as gene expression barriers, fitness cliffs, and population extinctions, which can be triggered by small environmental or genetic changes. Our results disentangle the complex interplay of protein production and activity, and they show how these processes shape the evolution of simple organisms.

bound for protein production and activity in viable populations. As a consequence, gene regulation in metabolic pathways is likely to be under stronger selection than the mere cost of protein production would suggest.

Results

Fitness effects of *lac* protein production and activity

There are two generic sources of fitness cost for a molecular pathway: the cost of protein production and the cost of enzymatic activity [2–6]. In the case of the *lac* pathway, fitness depends strongly on the presence of substrates of the *lac* proteins, even when these substrates cannot be used as a carbon source [8,11,12]. One such substrate is IPTG (isopropyl-1-thio-β-D-galactoside), which is transported by LacY and induces *lac* expression (see Figure 1). Hence, there are two potential phenotypes affecting the fitness of the *lac* operon in an IPTG environment: the rate of *lac* protein production and the rate of IPTG transport into the cell.

We measure the fitness effects of *lac* protein production and activity in thirteen regulatory mutants in the *lac* operon of *Escherichia coli*. Twelve mutant strains have substitutions in the *lac*O1 operator region, which affect expression of the *lac* genes, and one strain has a deletion in the gene of the repressor *lac*I (see Figure 1 and Text S1 for details). We determine the *lac* protein concentration and the fitness of these mutants both with and without substrates of the *lac* permease LacY. Specifically, we compete each mutant strain against a reference strain with deleted *lac* genes. This assay defines the fitness cost of the *lac* pathway as the difference in growth rate, or difference in Malthusian fitness, between reference strain and mutant, $\Delta F = F_r - F$ (see Materials and Methods and Text S1 for details).

Figure 2 summarizes the results of these experiments. They show that fitness always decreases with increasing concentration of *lac* proteins inside the cell, but the form of this dependence depends on presence or absence of the substrate. Without substrate, the fitness cost can be fitted to a linear form, which we associate with *lac* protein production (blue line). When substrate is added, the magnitude of the fitness cost strongly increases and its dependence on concentration becomes nonlinear (purple line). The additional, nonlinear fitness cost in the presence

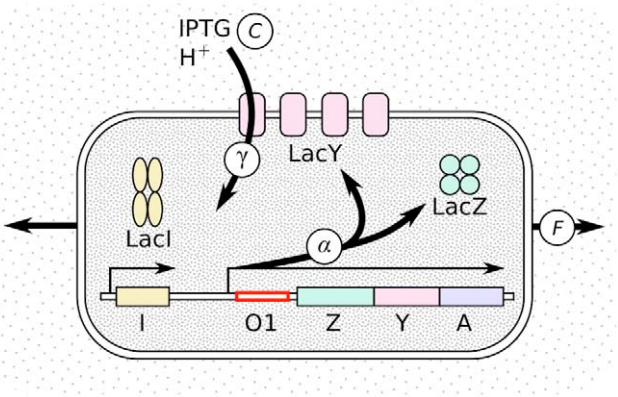

Figure 1. Schematic representation of the *lac* pathway. The *lac* operon is composed of three genes controlled by the same promoter: *lac* Z, *lac* Y, and *lac* A. The *lac* pathway also involves the constitutively expressed repressor LacI. It represses the transcription from the *lac* promoter by attaching to the operator sequence *lac*O1. Inducers, such as IPTG, deactivate the repressor LacI and thus stimulate the synthesis of the gene products LacZ, LacY, and LacA. The rate of production of the three *lac* proteins is denoted by α and it depends on the sequence of *lac* O1, on the presence of inducer inside the cell, and on the growth rate. All three *lac* genes are transcribed with the same rate, hence LacZ can be used as a reporter for the whole operon. LacY transports molecules such as IPTG inside the cell with a rate γ, which depends on the concentration of these molecules. One proton H^+ is transported with each substrate molecule [7]. Growth (measured by the Malthusian fitness F) dilutes the internal molecules, thus lowering their concentrations. The strains used in this study differ by the *lac*O1 sequence and are grown in various IPTG concentrations.
doi:10.1371/journal.pgen.1002160.g001

of IPTG can be associated with the transport activity of the LacY proteins. This is shown by a control mutant with deleted *lac*Y gene, for which we only observe the linear cost of protein production (red dot). Deviations of individual data points from the fit curves can be caused by different sources of noise. Competition assays involve experimental errors, in particular for large fitness differences between strains. For example, there can be slight day-to-day differences in medium composition. Furthermore, some of the strains might have acquired mutations with a fitness effect outside the *lac* operator sequence, although we have controlled for random mutations elsewhere the genome (see Text S1).

As a further experimental step, we test whether these results extend to lactose, which is a natural nutrient of *E. coli*. The sugar used to support cell growth in the above experiments is glycerol, which is a poor carbon source. Lactose supports faster growth and is known to give an advantage to cells which are able to metabolize it. With 1 mM of lactose, the wild type has a *fitness benefit* over the reference strain, which amounts to $-\Delta F = 0.34 \pm 0.04$ (mean of 4 replicates $\pm$ standard error). To assess whether lactose metabolism also involves a cost, we construct a mutant with deleted *lac*Z and *lac*I genes. This mutant cannot use lactose and expresses *lac*Y constitutively. In the presence of lactose, it has a *fitness cost* $\Delta F = 0.22 \pm 0.01$ (mean of 12 replicates $\pm$ standard error) against the reference strain, which indicates that lactose and IPTG cause a similar decrease in fitness in the presence of the *lac* permease (see Materials and Methods for details).

We conclude that both the rate of protein production and the rate of protein activity (intra-cellular transport by LacY) are phenotypes that affect the fitness cost of the *lac* pathway. But what is the cause of the fitness nonlinearity in the presence of substrates, and what are its biological consequences? To address these

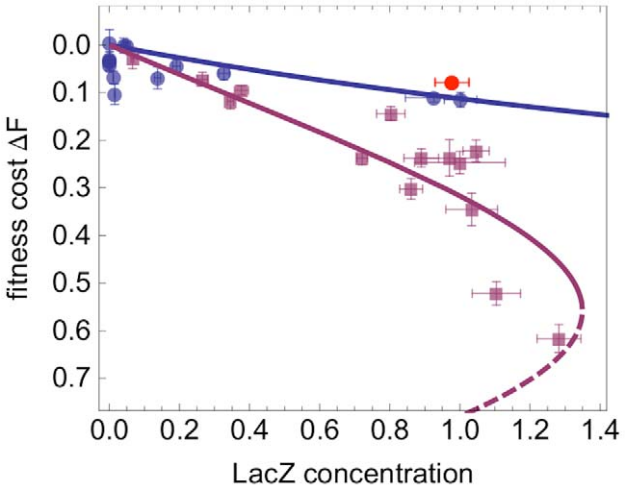

Figure 2. Fitness of *lac* regulatory mutants in different environments. Measured fitness cost of each mutant strain, plotted against LacZ concentration (normalized to the fully induced wild-type value). Measurements are obtained in minimal medium with 0.1% glycerol in the absence of IPTG (blue dots) and in the same medium with 1 mM IPTG (mauve squares). Fitness is measured by competition against a reference strain which has a deletion of the whole *lac* locus and of *lac* I (Δ*lac*IZYA). The fitness cost ΔF of a given strain is defined as the reduction in growth rate (Malthusian fitness) compared to the reference strain (see Materials and Methods for details). In presence of 1 mM IPTG, a control strain with deleted *lac* Y gene (Δ*lac*Y) has an expression level comparable to the wild type, but a fitness close to that of constitutive mutants in absence of IPTG (red dot). All points show the average of 12 replicates for fitness and at least 3 replicates for protein concentration, with error bars giving the standard error. Lines show model predictions (the dashed line represents an unstable solution, see main text).
doi:10.1371/journal.pgen.1002160.g002

questions, we now describe our experiments in terms of a simple biophysical model.

Fitness model

We use a minimal model of gene expression and inducer transport to disentangle the fitness effects of protein production and activity in a quantitative way. The underlying intra-cellular processes involve transcription and translation, uptake of substrate by active transport, and dilution by cell division. Given the complexity of these processes and their effects on cell growth, our model does not aim at a complete description. However, the model does account for a large part of the fitness variation between strains and between cellular growth conditions. At the same time, it contains only few phenotypes and few parameters, which can be inferred from our fitness measurements.

Within the model, the cost of *lac* protein production is proportional the production rate α, and we infer this rate from our measurements of fitness and LacZ concentration (see Figure 1 and Materials and Methods). The cost of LacY activity has two different potential contributions: the energy consumption of the transport process (direct transport costs) and growth effects of the molecules transported inside the cell (toxicity costs). Direct transport costs can arise from futile transport cycles: LacY transports one proton with every IPTG molecule, and ATP is consumed to pump the excess protons out of the cell. These costs are proportional to the total LacY pumping rate inside the cell, Γ. Toxicity costs are likely to arise from an excess concentration of the transported protons, i.e., a reduction of the intra-cellular pH value [11,12]. The toxicity of IPTG itself appears to be negligible (see Text S1 and [12]). Toxicity costs are proportional to the steady-state concentration of the toxic molecules, which depends on their uptake rate, the rate of dilution by cell divisions, and the cell volume V [18]. The excess concentration of protons is, thus, proportional to Γ/FV. Furthermore, the steady-state cell volume itself depends on the growth rate, $V=V(F)$ [13].

The combined fitness cost of protein production and activity in the *lac* pathway takes the form

$$F_r - F = a\alpha + b\Gamma + c\frac{\Gamma}{FV(F)} \qquad (1)$$

in terms of the pathway phenotypes α and Γ. Here, F_r denotes the fitness of the reference strain with deleted *lac* genes (for which $\alpha=\Gamma=0$). Our model contains a feedback loop: fitness depends on the rates α and Γ, which in turn depend on fitness. This feedback between pathway phenotypes and fitness is illustrated in Figure 3. It has an important consequence: although the cost contributions in Equation 1 are taken to be additive at any given value of F, the resulting dependence of fitness on the pathway phenotypes, $F(\alpha,\Gamma)$, becomes nonlinear.

By calibrating this model to our experimental data, we can infer the amplitudes a, b, and c of the different cost factors. Bayesian analysis shows that there are significant fitness contributions of protein production and steady-state concentration (with maximum-likelihood parameter values $a=0.21$, $b\approx 0$, $c=0.17$), but the data are also compatible with a larger direct cost of transport ($b>0$) (see Materials and Methods and Text S1). As shown in Figure 2, the maximum-likelihood model provides a good fit to the data: the fitness feedback loop quantitatively explains the cost nonlinearity observed in our experiments.

We use Equation 1 to derive two representations of a fitness landscape for the *lac* pathway, which highlight different biological implications of its form. First, we solve this equation to display the dependence of fitness on the pathway phenotypes, $F=F(\alpha,\Gamma)$, as shown in Figure 4. Second, we display the dependence of fitness on the external IPTG concentration, C, and on two genotype summary variables, which depend only on *lac* O1 sequence. As genotype variables, we use the maximal rate of *lac* protein production at a fixed growth rate of one cell division per hour, α_0, and the ratio of repressed to unrepressed protein production rates, ρ [19]. The resulting function $F(\alpha_0,\rho,C)$, which is shown in Figure 5, can be called a genotype-environment-fitness map. We note that the change from the phenotype variables α,Γ to the genotype-environment variables α_0,ρ,C depends itself on fitness. This dependence has two reasons: (i) The LacY pumping rate depends on the production rate, the pumping rate per LacY molecule γ, and fitness, $\Gamma=\alpha\gamma/F$, because LacY molecules are diluted by cell divisions just like the transported molecules. This generic dependence reinforces the basic growth feedback loop by dilution, which also enters Equation 1. (ii) For fixed genotype and environment, the production rate itself can depend on fitness, $\alpha=\alpha(\alpha_0,\rho,C,F)$. This growth effect on gene expression generates an additional feedback between the *lac* pathway and fitness, which is expected under several growth conditions [13,14]. Including this feedback in our model significantly improves the agreement between data and theory (see Materials and Methods and Text S1 for details). The fitness landscapes of Figure 4 and Figure 5 are obtained from our model using maximum-likelihood parameters, but their shape depends only on the presence of a fitness nonlinearity ($c>0$). We now discuss their form and their biological implications in more detail.

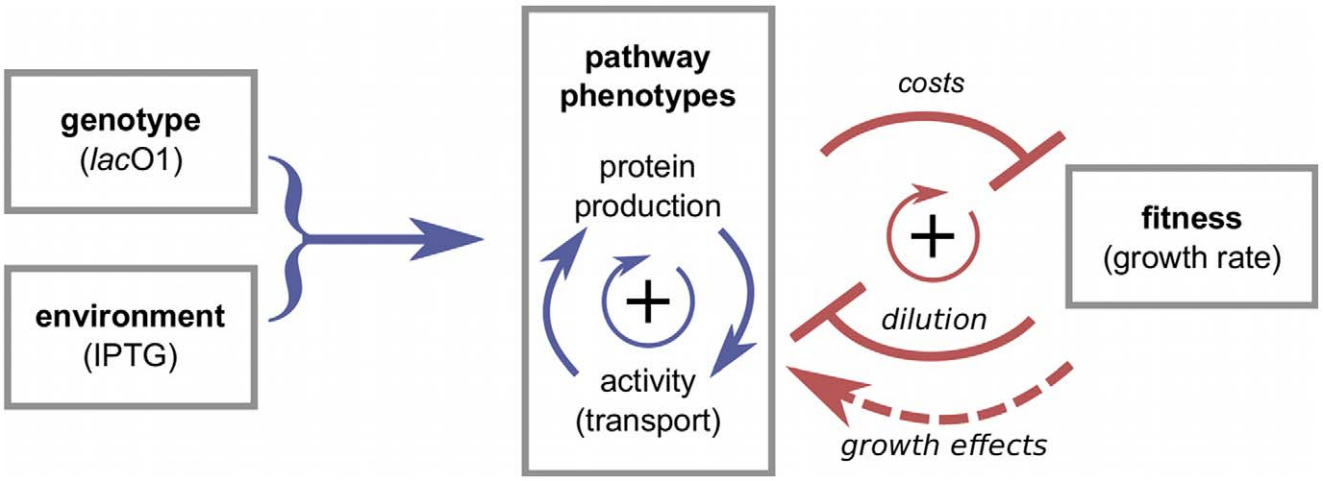

Figure 3. From genotype and environment to pathway phenotypes and fitness. Environment and genotype determine the function of the *lac* pathway, which is described by the two phenotypes of protein production and protein (transport) activity. These phenotypes are coupled by a pathway-specific positive feedback loop (blue circle). The pathway itself is coupled to growth (fitness) by a generic positive feedback loop: protein production and protein activity are fitness costs, and cell growth reduces protein concentration and activity by dilution (red circle). In addition, growth can affect the rate of gene expression [13] (dashed arrow). These feedback loops generate strong nonlinearities in the phenotype-fitness map and the genotype-environment-fitness map; see Figure 4 and Figure 5.
doi:10.1371/journal.pgen.1002160.g003

Phenotype-fitness map

The phenotype-fitness landscape of the *lac* pathway resulting from our model is shown in Figure 4, together with fitness measurements of different *lac* O1 operator mutants in different inducer environments. The experimental data are plotted as a function of the pathway phenotypes α and Γ inferred from our model; for each mutant, the dependence of these phenotypes on the IPTG concentration is indicated by a red line. Data and model

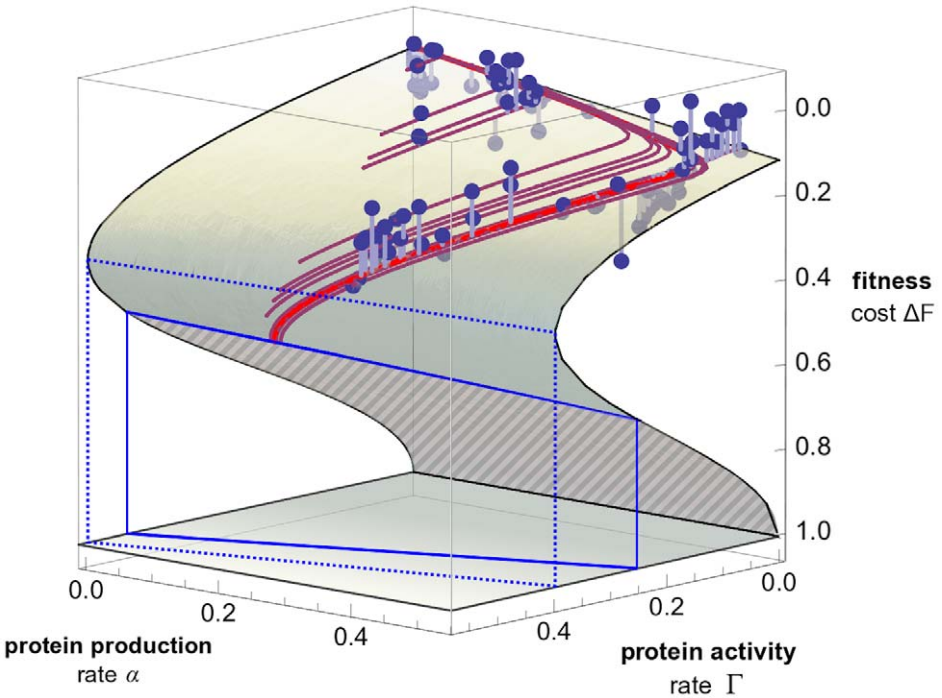

Figure 4. Phenotype-fitness map. The fitness cost ΔF of the *lac* pathway is shown as a function of the protein production rate α and the transport rate Γ. The fitness landscape obtained from our model (shaded surface) is strongly nonlinear and has two branches. The stable part of the landscape (solid shading) ends at a fitness cliff (solid blue line), beyond which populations cannot maintain growth. The remaining part of the lower fitness branch is unstable (striped shading). Protein expression and activity of viable populations are bounded by a barrier (dotted blue line). Model predictions of pathway phenotypes and fitness for individual strains under varying inducer concentrations are shown as a family of red lines (light red: wild type, dark red: operator mutant strains). Experimental fitness values are shown as dots (the offset from the model surface is marked by gray lines).
doi:10.1371/journal.pgen.1002160.g004

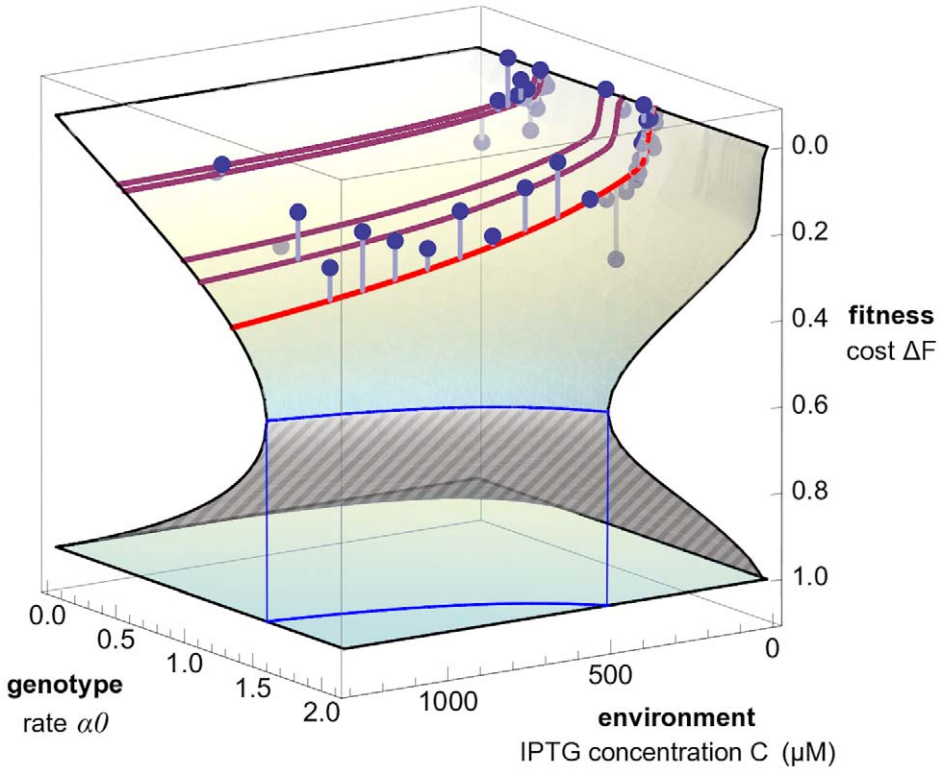

Figure 5. Genotype-environment-fitness map. The fitness cost ΔF of the *lac* pathway is shown as a function of the operator genotype summary variable α_0 (maximum rate of protein production at a growth of 1 cell division/hr, see text) and the external inducer concentration C. The model fitness landscape is again strongly nonlinear: it has a stable upper branch (solid shading) and an unstable lower branch (striped shading) separated by a fitness cliff (blue line), similar to the phenotype-fitness map of Figure 4. Model predictions for individual strains under varying inducer concentrations are shown as a family of red lines (light red: wild type; dark red: operator mutant strains with ρ equal to wild type value, see text). Experimental fitness values are shown as dots (the offset from the model surface is marked by gray lines).
doi:10.1371/journal.pgen.1002160.g005

consistently show that protein production and activity of the *lac* pathway affect fitness in a highly nonlinear way. Our model explains the nonlinearity in terms of the growth feedback mechanism contained in Equation 1.

This form of the phenotype-fitness landscape has two important aspects. First, the nonlinearity of fitness translates into epistatic interactions between the pathway phenotypes: the effect of a change in the production rate α, which is proportional to the slope $\partial F/\partial\alpha$, depends on the transport rate Γ, and vice versa. Second, the fitness landscape is not univalued: for some values of α and Γ, there are two possible fitness values, for others, there is none. Phenotype values in the no-solution regime cannot be attained by a cell population in steady growth. This regime is bounded by a dotted line in the (α,Γ) plane, which marks an expression barrier for the *lac* genes. The barrier occurs at a finite growth rate $F=F_r-\Delta F$ (in contrast to the model of ref. [9]). Double-valued fitness solutions and the existence of an expression barrier for given phenotype values are a direct consequence of the growth feedback loop in Equation 1. The stability analysis described below shows that only the full-shaded part of the landscape describes viable cell populations in stationary growth, whereas the striped part is unstable. Hence, for parameter values between the dotted and the solid lines in the (α,Γ) plane, populations can reach two different steady-state growth rates with the same *lac* pathway phenotypes.

Genotype–environment interactions

We now turn to the dependence of fitness on the *lac* O1 operator sequence and on the external inducer concentration C, which are the quantities we manipulate in our experiments. To display the sequence-dependence, we use the genotype summary variables α_0 (maximal rate of *lac* protein production at a fixed growth rate of one cell division per hour) and ρ (ratio of repressed to unrepressed protein production rates). These variables reflect the double role of the operator sequence: it acts as a binding site for the repressor LacI, but it also affects other processes that lead to changes in protein production [20]. Figure 5 shows the fitness cost as a function of the maximal production rate and the IPTG concentration, $\Delta F(\alpha_0,C)$. The ratio ρ is kept fixed to its wild type value; the figure shows fitness data for the corresponding subset of strains (see Figure S1 and Figure S2 for the full dependence of ΔF on α_0, ρ and C). Again, cell populations in stationary growth cannot exist for some genotype-environment parameters; this regime is bounded by a blue line in the (α_0,C) plane.

The fitness of different mutants as a function of the inducer concentration is again shown as a family of lines. The nonlinearity in the landscape indicates that the inducer environment affects the selective effect of regulatory mutations: higher IPTG concentrations lead to increased fitness differences between mutants. This interaction between genotype and the environment is due to an increase in the pumping rate with increasing IPTG, to the coupling of uptake rate and production rate in the term Γ, and to the growth feedback through dilution. Figure S2 further illustrates this interaction. Genotype-environment interactions in the *lac* operon have been observed previously [21]. Our model shows how such interactions emerge from the basic architecture of metabolic pathways.

Nonlinearities generate extinction thresholds

The fitness landscapes of Figure 4 and Figure 5 have a common feature: over a wide range of parameters, there are two possible fitness values ΔF. This double-valued fitness landscape is partitioned into a stable part (full-shaded) and an unstable part (striped); see Text S1, Figure S3, and Figure S4. The stable part of the landscape describes stationary growth of viable populations; i.e., cells with growth rates close to a point on this surface reach a steady state given by a point on the surface. A large part of the lower surface ΔF_2 is unstable, i.e., cells with fitness cost $F < F_r - \Delta F_2$ are unable to dilute their proteins and transported molecules fast enough to maintain stable growth. These cells will further decline in fitness, whereas cells with $F > F_r - \Delta F_2$ will increase fitness up to the stable value ΔF_1. The stable and the unstable part of the fitness landscape are separated by a fitness cliff, which is shown as a blue line in Figure 4 and 5. The cliff marks an extinction threshold: If a cell population is driven beyond this cliff by mutations or environment changes, it suffers a sudden drop in fitness and cannot maintain a finite growth rate.

Existence and position of the fitness cliff depend on the amount of inducer present (see Figure 2, Figure 5, and Figure S2). For IPTG concentrations used in our and other experiments, the cliff is far from the wild type (Figure S2B). We note, however, that lactose is often used in higher concentrations, and lack of growth due to the presence of lactose (lactose killing) has indeed been observed [22].

Discussion

We have shown that in the presence of an inducer, the fitness cost of the *lac* pathway arises not only from protein production, but also from transport activity of the permease LacY. The cost is governed by a feedback loop, which is the result of two repressive interactions: protein activity results in reduced growth, and growth dilutes proteins as well as transported molecules (see Figure 3). We note that our feedback mechanism does not rely on a limitation of cellular resources to generate a nonlinear relation between *lac* gene expression and growth (in contrast to the model of ref. [9]). This feedback produces a strongly nonlinear dependence of fitness on pathway phenotypes or on genotype and environment, as shown in the fitness landscapes of Figure 4 and Figure 5. Both landscapes contain a fitness cliff, which is an extinction threshold for cell populations. The nonlinearity of fitness is likely to persist for any substrate of the permease LacY and sets an upper bound for its rates of expression and activity. Thus, changes in *lac* permease activity or expression can have strong impacts on fitness. This is consistent with the observation that *lac*Y is under particularly strong selection [23], as reflected notably by its low number of synonymous single-nucleotide polymorphisms [24].

The nonlinearity of fitness and its consequences are expected to hold in the presence of lactose. If the benefit conferred by lactose (or other sugars) also depends on its internal concentration, we expect an effect of diminishing return: the faster a cell grows, the more it will dilute lactose, which leads to a sublinear increase of fitness with lactose concentration. Hence, combining costs and benefits of the *lac* proteins will lead to more complex fitness landscapes; their detailed dependence on pathway phenotypes will be addressed in a future study. Importantly, the full landscapes are expected to have a fitness cliff similar to the cost landscapes derived in this paper. This might explain why induced cells grown in a chemostat die after exposure to high concentrations of lactose, a phenomenon known as lactose killing [22]. Moreover, many other metabolic pathways in microorganisms contain a membrane pump or transporter accumulating substrates inside the cell, which often uses the proton motive force as an energy source. Our results are expected to apply to these pathways as well. In particular, we note the similarity of our fitness landscapes and those of the glucose utilization pathway in yeast [25] (see Figure 5 and Figure S2). Other protein activities such as hydrolysis of substrates can produce the same type of feedback, because they also depend on internal concentrations of molecules.

The shape of the fitness landscape described here has various implications for the genomic evolution of the *lac* pathway. Our fitness model of protein production and activity contains two types of epistasis on the operator *lac* O1. Within the operator, the fitness reduction caused by two mutations that increase expression is larger than the sum of the fitness costs of either one (see Figure 5). Furthermore, the selection pressure on expression depends on the protein activity rate and, hence, on the sequence of the downstream gene *lac* Y. The total pumping rate of the cell also depends on the concentration of LacY substrates in the environment, which generates fitness interactions between the operator genotype and the inducer environment. In a broader context, the costs of gene expression due to protein activity and due to protein production affect the evolution of regulatory systems in a different way. Taking into account only protein synthesis, we expect the length of genes to be the main determinant of the fitness cost of gene expression. Including protein activity, however, the selective pressure against expression of a gene can depend primarily on the coding sequence of functional domains and on the environment. For the *lac* pathway, the cost contributions of protein production and of protein activity are of similar magnitude, and both effects contribute to selection on regulatory sequences.

Generalizing the results of this study, we expect the full landscape of a metabolic network to be filled with cliffs and valleys, whose importance depends on which pathways are more active in a given environment. In addition, a metabolic pathway with growth feedback generates ubiquitous epistasis. For example, any mutation under selection has fitness interactions with mutations in the *lac* operon: In the presence of IPTG, deleterious (beneficial) mutations outside the *lac* pathway affect the protein production rate α and the transport rate Γ, and hence increase (reduce) the fitness cost of *lac* activity-enhancing mutations. Thus, higher-dimensional fitness landscapes including more and more metabolic phenotypes are expected to be increasingly rugged.

Previous experiments have produced fitness landscapes as a function of genotype (see for example [26,27]). This kind of fitness landscapes omits the intermediate level of phenotypes, which describes how genotype changes affect biophysical functions. Here, we record fitness as a function of well-defined phenotypes of a metabolic pathway. These can be connected to a biophysical model, which describes the dependence of fitness on the operator sequence and on the inducer concentration. Phenotype- and model-based fitness landscapes are predictive: Once the model constants are fixed by one set of measurements, the model predicts the outcome of further experiments with different input parameters. In this study, the most striking model prediction is the extinction of populations beyond a fitness cliff.

Our fitness landscape also differs from previous phenotype-fitness maps, perhaps the most popular of which is Fisher's geometric model [28]. Fitting this model to fitness data is a method to infer distributions of fitness effects of mutations and of epistasic effects between mutations [29,30]. Fisher's geometric model contains an *a priori* arbitrary number of unknown molecular phenotypes. In contrast, our model contains a small number of known phenotypes associated to a specific pathway, which are shown to capture salient features of fitness variation between

populations (clearly, this does not rule out further phenotypes of this pathway affecting fitness). In the classical geometric model, the fitness landscape is assumed to be smooth, and different phenotypes to contribute additively to fitness. Our fitness landscape contradicts both of these assumptions: there is strong epistasis and ruggedness. These features have been extensively analyzed for genotype-fitness maps (a well-known example is the NK model [31]), but the dependence of fitness on quantitative phenotypes is generally assumed to be smoother. Our study shows that strong epistasis and ruggedness can persist in phenotype-fitness landscapes. It calls for new statistical models of such landscapes, which address their broad consequences for speed and constraints of molecular evolution. An interesting example is a recent extension of the geometric model, which contains epistasis and a fitness cliff [32].

In summary, our measurements and modeling show that the *lac* pathway of *E. coli* is governed by a strongly nonlinear fitness landscape depending on phenotypes of protein production and activity. These phenotypes, in turn, depend on the *lac* operon genotype and on environmental parameters in a coupled way. Fitness nonlinearities and genotype-environment interactions are not specific to the system studied here, but are likely to be general features of metabolic pathways. Thus, the fitness landscape of a metabolic network is much more than a simple superposition of the cost of protein production and the benefit of protein activity. It describes the entire network as a unit of natural selection. Such system-level fitness landscapes emerge already at simplest level of cell growth and metabolism.

Materials and Methods

Strain construction

The background of all strains used in this study is *Escherichia coli* BW30270 (K12 MG1655 rph^+). The *lac*O1 mutant strains (summarized in Table S1 and Table S2) are constructed as described in [33]. First, the complete *lac* promoter is deleted and replaced with the chloramphenicol resistance cassette from plasmid pKD3 (see Table S3 for a list of plasmids used in this study). This yields strain S4146 which is $\mathrm{Cm^R}$, $\mathrm{Amp^S}$ and lac^-. The full *lac* promoter and 5′UTR of wild-type *Escherichia coli* are amplified and cloned (see Table S4 for a list of oligonucleotides used in this study). Specific *lac* O1 mutations are inserted using PCR mediated mutagenesis [34], and the mutant sequences are cloned in a high-copy-number plasmid (derived from pUC12). The same gene replacement method [33] is then used to replace the chloramphenicol resistance cassette in strain S4146 with the chosen *lac* promoter and O1 operator. The strains produced in this way are all $\mathrm{Cm^S}$, $\mathrm{Amp^S}$ and lac^+. We noticed that these strains have a general lower fitness than strain BW30270 that cannot be explained by the inserted mutations (see Figure S5) so we use T4GT7 mediated transduction [35] to transfer the *lac* mutations back to the parent background (BW30270). First, the resistance cassette from strain S4146 is transduced to BW30270, producing strain T218. Then, the mutated *lac* operon is transduced from each *lac*O1 mutant to T218. The *lac* promoter and O1 operator are then sequenced to confirm the correct insertion of the *lac* operator allele. As a control for the transduction, a wild type construct is obtained in the same way (T273). It has the same fitness as BW30270. The reference strain for the competition (S4085 − Δ*lac*IZYA), the *lac* permease and the *lac* repressor mutants (T407 − Δ*lac*Y and T523 − Δ*lac*I) are constructed as described in [33]. Strain Δ*lac*IΔ*lac*Z is constructed by first deleting *lac* Z following [36] and then deleting *lac* I following [33].

Media and growth conditions

Unless stated otherwise, all measurements are made in M9 minimal medium with glycerol (0.1% v/v) as carbon source. To distinguish strains in competition, tetrazolium lactose (TL) medium (1% bacto-tryptone, 0.1% yeast extract, 0.5% NaCl, 1% lactose, 0.005% tetrazolium chloride and 1.5% agar) is used. $\mathrm{Lac^+}$ colonies are white and $\mathrm{Lac^-}$ colonies are red in TL plates [37]. Δ*lac*Y is also white on TL plates. To distinguish this strain from Δ*lac*IZYA and Δ*lac*IΔ*lac*Z, LB-XGal-IPTG plates are used (1% bacto-tryptone, 0.5% yeast extract, 0.5% NaCl, 1 mM isopropyl-1-thio-β-D-galactoside (IPTG), $40\,\mu\mathrm{g/ml}$ 5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) and 1.5% agar).

Protein expression

Protein concentration is estimated using a β-galactosidase (LacZ) activity assay [38]. Since all our mutants have the same coding sequence for this protein, changes in activity reflect changes in protein concentration inside the cell. The LacZ assays are performed as described in [38]. Overnight cultures are diluted in fresh medium to an optical density at 600 nm (OD600) of 0.05 and harvested after growth in the indicated media at $37^\circ\mathrm{C}$ to an OD600 of 0.3. IPTG was added to the overnight culture and to the test cultures in the concentrations mentioned in the text. The enzyme activities are determined from at least three independent cultures. Figure S6 shows the measured LacZ levels for all strains used in this study, in absence and presence of IPTG.

Fitness measurements

Fitness is measured in head to head competition as described in [39]. Briefly, frozen cultures (stored at $-80^\circ\mathrm{C}$) are streaked on a Luria broth agar plate and grown over night at $37^\circ\mathrm{C}$. An isolated colony is randomly selected and grown overnight in 3 ml of the same medium used in the competition, in particular with the same amount of IPTG. Both the reference strain (Δ*lac*IZYA, unless stated otherwise) and assay strains are treated in this way separately. The strains are then mixed and diluted in saline solution (10 mM $\mathrm{MgSO_4}$ and 0.85% NaCl), such that about 50,000 colony forming units (CFUs) of each strain are used to start the competition. The mixed dilutions are also used to count the starting titer. Cultures are grown for 24 h on 96 deep-well plates in 1 ml of medium, shaken at 150 RPM, reaching saturation ($\sim 10^8$ CFUs). They are then diluted and plated on TL or LB-XGal-IPTG medium.

We measure the Malthusian fitness F, i.e., the growth rate, of each strain in units of the growth rate of the reference strain (such that $F_r = 1$). The fitness value of a mutant is inferred from a competition experiment with the reference strain,

$$F = \frac{\ln(N_f/N_i)}{\ln(N_{f,r}/N_{i,r})},$$

where N_f, N_i are the final and initial number of mutant CFUs after and before the competition, and $N_{f,r}$,$N_{i,r}$ are the corresponding numbers for the reference strain. The growth rate of the reference strain is not affected by IPTG (see Figure S7). Thus, the doubling time of the reference strain is a fixed time unit and fitness measurements across environments are directly comparable. We report the fitness cost of a mutant compared to the reference strain, $\Delta F = F_r - F$ (which is proportional to its selection coefficient measured in units of doubling time, $\Delta F = -s/\ln 2$ [40]).

The Δ*lac*IΔ*lac*Z strain has the same phenotype as the reference strain (both are red on TL plates and white on LB-XGal-IPTG

plates), so the two cannot be competed directly. Instead, we measure the fitness of this strain by competing it with Δ*lac*Y. Δ*lac*Y has the same fitness as the reference strain in competition in glycerol minimal medium with 1 mM lactose ($\Delta F = 0.00 \pm 0.02$).

Dependence of pathway phenotypes on genotype and environment

As explained in the Results section, the protein production rate α and the transport rate Γ can be expressed in terms of genotypic and environmental parameters, and fitness. This map relates the fitness landscapes of Figure 4 and Figure 5 and can be obtained as follows.

The first phenotype of the *lac* pathway, the protein production rate α, has two main components: one is independent of the lac repressor (LacI) and the other depends on the probability of the repressor to bind the operator. The independent component is given by the direct effect of the operator sequence (quantified by the first genetic component α_0) and by the growth rate F (through a function f_G specified below). The LacI-dependent component of α depends on the affinity of the operator sequence (measured by the second genetic component ρ) and on the concentration of inducer C in the environment. The dependence on C has the form of a Hill function $f_I(C) = C^n/(K_z^n + C^n)$ with parameters K_z (the half saturation constant, taken to be $20\,\mu\mathrm{M}$) and n (the Hill coefficient, taken to be 4.5) [18]. The protein production rate is then $\alpha(\alpha_0,\rho,C,F) = \alpha_0 f_G(F)[\rho + (1-\rho) f_I(C)]$. We now derive the form of f_G, and estimate α_0 and ρ.

As mentioned before, f_G is the dependence of the production rate on the growth rate F. Following [13,14], f_G is expressed relative to the fitness of a strain growing at the rate of 1 doubling/hour, F_0, such that $f_G(F_0) = 1$. Note that the reference strain has a growth rate of $0.008\,\mathrm{min}^{-1}$, so $F_0 = 1.5$. The parameter f_G reflects the following observation: When the growth rate changes due to nutrient quality, there is a linear inverse correlation between protein concentration (C_z) and growth rate [14], $C_z(C,F) = C_z(C,F_0)(1/3)[4-(F/F_0)]$ (see Figure S8A). This relationship can be extended to the protein production rate α, because $C_z(C,F) = \alpha(\alpha_0,\rho,C,F)/V(F)F$ at steady state. We choose a linear dependence of the cell volume V on the fitness, $V(F) = F/F_0$; see Text S1 and Figure S8B. Using the dependences inferred above and assuming ρ to be independent of F, the dependence of protein production on growth rate can be estimated: $f_G(F) = (1/3)[4-(F/F_0)](F/F_0)V(F)$. We have verified that including f_G significantly improves the agreement between model and data (see Text S1), although it is not obvious *a priori* that a correlation between C_z and growth rate is relevant in the context of our experiments.

The two genetic components, α_0 (the maximal protein production rate at fixed growth rate) and ρ (the ratio of repressed to unrepressed protein production rates), depend only on the genotype and were calculated for each strain separately. α_0 can be derived from the protein concentration and fitness measured at a concentration $C_1 = 1\,\mathrm{mM}$ IPTG, where the LacI proteins cannot bind DNA ($f_I(C_1) = 1$). As explained above, the cell volume V, the growth rate F, and the effects of growth on expression f_G affect C_z, such that $\alpha_0 = C_z(C_1,F_1)V(F_1)F_1[1/f_G(F_1)]$, where F_1 is the measured growth rate at 1 mM of IPTG. Similarly, ρ can be estimated using C_z and F measured at 0 mM of IPTG: $\rho\alpha_0 = C_z(0,F_\varnothing)V(F_\varnothing)F_\varnothing[1/f_G(F_\varnothing)]$, where $F_\varnothing$ is fitness in the absence of IPTG. Both α_0 and ρ are independent of the model in Equation 1 and of the growth-dependence of the volume. Inferred values of α_0 and ρ are shown in Figure S9 and Table S1. The parameter ρ is related to the "repression level" R defined by Müller-Hill and co-workers as the ratio of LacZ activity between strains differing only by the presence/absence of the *lac* repressor, $R = C_z(\mathrm{lacI}^-)/C_z(\mathrm{lacI}^+)$ [19]. Neglecting the growth difference between both strains, these quantities are inversely related, $R \approx 1/\rho$.

The second phenotype of the *lac* pathway, the total transport rate Γ, is the product of the number of LacY molecules in the cell and the transport rate per LacY molecule, $\Gamma = N_y\gamma$. The number N_y is equal to α/F, with α the protein production rate and F the growth rate, because LacY molecules are diluted by cell divisions. Note that α is measured for LacZ, but all proteins of the operon are produced proportionally. The ratio of LacY molecules per LacZ molecule, which is close to 3 [10], and other numerical constants are absorbed in the coefficients a, b and c. The transport rate γ depends on the external IPTG concentration, C, and on the half-saturation constant for inducer uptake, $K_p = 420\,\mu\mathrm{M}$ [41]. An expression for γ can be derived from the known functioning of the permease [42], with efflux neglected (see Text S1). We obtain $\gamma(C) = (C/C_1)[(K_p + C_1)/(K_p + C)]$, normalizing γ to $C_1 = 1\,\mathrm{mM}$ of IPTG.

The uncertainties on α_0 and ρ are obtained by standard error propagation, assuming independent experimental errors on F and C_z (see Text S1). A possible error in the IPTG concentration C is not considered, because it is expected to be small.

The coefficients a, b and c in Equation 1 are obtained by likelihood analysis of our model and the experimental data. This analysis is based on the dependence $F(\alpha_0,\rho,C)$, where α_0 and ρ are inferred for each mutant as described above. The fitting procedure and score-based model comparisons are detailed in Text S1 (see also Figure S10 and Figure S11).

Supporting Information

Figure S1 Fitness cost ΔF as a function of the IPTG concentration for five strains: (A) the wild type, (B) T274, (C) T320, (D) T523−Δ*lac*I, (E) T275. The full lines are model predictions; dots show the experimental data (the error bars represent the standard error of the mean on 4 replicates, on 12 replicates for data at 0 and 1 mM IPTG). See Table S1 for a list of strains and the corresponding values of α_0 and ρ.
(TIF)

Figure S2 Fitness cost ΔF as a function of α_0 and ρ, at fixed external IPTG concentration (A) $C = 0$, (B) $C = 30\,\mu\mathrm{M}$ and (C) $C = 1000\,\mu\mathrm{M}$; as a function of α_0 and C, at fixed (D) $\rho = 0$, (E) $\rho = 0.32$ and (F) $\rho = 1$. The dots are the experimental data, the grey vertical bars show the distance between data and model prediction. Strains shown: (A) and (C) all operator mutants, wild type, T523−Δ*lac*I (D) wild type, T319, T320, T378, T379; (B) wild type, T274, T275, T320, T523−Δ*lac*I; (E) T323; (F) T275, T523−Δ*lac*I. The light-green surfaces show the stable solutions, the dark gray the unstable one. The blue line marks their boundary: when α_0 or the external IPTG concentration is increased beyond this "cliff", the population falls on the no-growth solution, and thus goes to extinction. Panel D is identical to Figure 5 of the main text.
(TIF)

Figure S3 Dynamical analysis. (A) $H(F)$ for the wild type at different IPTG concentrations: 0.1 mM (red), 1 mM (blue) and 100 mM (green). The steady-state solutions $F = H(F)$ lie at the intersection of H with the first bisecting line (black line); if it crosses it from above, the solution is stable (full dot), otherwise it is unstable (empty dot). (B) Time evolution of the growth rate of the wild type in 1 mM IPTG obtained by a discrete process (Equation 8 of Text S1, with a time step $\ln 2/F_n$ between step $n+1$ and n; dots) and a continuous-time description (Equation 10 of Text S1;

lines), for various initial growth rates. The generation time in minutes is $(1/\mu_r)(\ln 2/F_n)$, with μ_r the growth rate of the reference strain measured to be $0.008\,\mathrm{min}^{-1}$. The full black line shows the stable steady state, the dashed line the unstable one. See Text S1 for definitions.
(TIF)

Figure S4 Fitness cost ΔF as a function (A) of the protein synthesis rate α, (B) of the protein concentration C_z. The dots show the measured fitness cost for different strains, in absence of IPTG (blue circles) and in 1 mM IPTG (mauve squares). The red dot shows the fitness cost measured for $\mathrm{T407}-\Delta\mathit{lac}\mathrm{Y}$ in 1 mM IPTG. The data shown in panel B are the same as those shown in Figure 2 of the main text. Error bars represent the standard error of the mean. The lines are the theoretical prediction, in absence of IPTG (blue) and in 1 mM IPTG (mauve). The dashed lines show the unstable solutions. The gray lines show the correlation of α and C_z with ΔF due to growth effects (see Text S1), for different values of α_0. Starting from an initial selection coefficient (e.g., upon a change of medium), a given strain moves along a gray line toward the stable steady-state solution, and away from the unstable one.
(TIF)

Figure S5 Comparison of protein expression *(left)* and fitness cost ΔF *(right)* on control strains. BW30270 is the wild type strain, T45 is a direct Datsenko-Wanner wild type construction and T273 is a transduction wild-type construction which went through the same procedures as all the *lac* operon mutants. Measurements were made in glycerol minimal medium without IPTG (white) and with 1 mM of IPTG (blue). Fitness was measured in competition against $\Delta\mathit{lac}\mathrm{IZYA}$. See Materials and Methods of the main text for a description of the strain constructions and competition experiments. The error bars represent the standard error of the mean.
(TIF)

Figure S6 Expression levels of the different *lac* operator mutants. Protein expression was measured as described in Materials and Methods of the main text without IPTG (white) and with 1 mM of IPTG (blue). The error bars represent the standard error of the mean, with at least three replicates in each condition.
(TIF)

Figure S7 Growth rate, in min^{-1}, measured in the same conditions as described for the competition experiments, except each strain was grown separately. Every hour, for 10 hours, 10 µl of the culture was taken and diluted appropriately, then plated on LB plates. Their mean lag phase was about 2 hours, therefore points 0, 1 hour and 2 hours were not used to estimate the growth rate. The growth rate μ was estimated as the slope of the regression of $\ln N(t)$ on time t, where $N(t)$ is the population size, such that: $N(t)=N(0)e^{\mu t}$. The error bars represent the standard error of the mean of 3 independent replicates. The Malthusian fitness F defined in Materials and Methods of the main text is equal to μ/μ_r, with the growth rate of the reference strain $\mu_r=0.008\,\mathrm{min}^{-1}$.
(TIF)

Figure S8 Growth effects on gene expression and cell volume. (A) The protein concentration C_z of a constitutively expressed gene has been proposed to correlate linearly with the growth rate F (red line), instead of the hyperbolic dependence dilution alone would induce (black line) [14]. (B) The cell volume V also correlates with F; dots show experimental data taken from [13]; we choose to represent this correlation via a simple proportional dependence (red line). (C) Both correlations lead to a dependence $f_G(F)$ of the rate of protein synthesis on the growth rate F (red line; see Materials and Methods of the main text). Following [13,14], the dependences are shown relative to the values at a growth rate $F_0=(1\text{ doubling/hour})/\mu_r=(0.012\,\mathrm{min}^{-1})/(0.008\,\mathrm{min}^{-1})=1.5$. The highlighted area $F\le F_r=1\approx 0.7F_0$ shows the range of growth rates relevant in this study.
(TIF)

Figure S9 Estimated maximal rate of expression at 1 doubling/hour α_0 and ratio of repressed to unrepressed rates ρ (see Materials and Methods of the main text), for all mutants used in this study. The wild type (red) and $\mathrm{T407}-\Delta\mathit{lac}\mathrm{Y}$ (orange) are barely distinguishable, as expected. In purple, the mutants which have a ρ value very close to that of the wild type and are shown on Figure 5 of the main text (these are strains T319, T320, T378 and T379). In yellow, the whole operator mutants (T274, T275, T318). In green, the strain $\mathrm{T523}-\Delta\mathit{lac}\mathrm{I}$. The values of α_0 and ρ are reported in Table S1. Errors were computed as described in Text S1.
(TIF)

Figure S10 Statistical score of the model for a range of coefficients b and c, with a fixed at its fitted value. The higher the score, the lighter the shading color. The contours are drawn at scores -420, -450, -500, -550, -600, -650, -700, and -750. The highest score -417 is obtained for $b=0.0026$ and $c=0.17$ (red dot), significantly better than the best model with $c=0$ (which has score -426).
(TIF)

Figure S11 (A) Fitness cost ΔF as a function of $\rho\cdot\alpha_0$, in absence of IPTG. (B) Fitness cost ΔF as a function of α_0, in 1 mM IPTG. Dots show the selection coefficient measured for different strains (error bars represent the standard error of the mean), lines are model predictions. In presence of IPTG, the stable solution shown as a full line in panel B was used to compute the score S and fit the data. α_0 and ρ are estimated for each strain as explained in Materials and Methods of the main text. Errors were computed as described in Text S1.
(TIF)

Table S1 List of the strain studied, their *lac* O1 alleles and sequences (starting at the +1 site; underlined: mutations with respect to the wild type). The estimated values for the maximum rate of protein synthesis at 1 doubling/hour α_0 and ratio of repressed to unrepressed rates ρ are also shown (see Materials and Methods of the main text). Errors were computed as described in Text S1.
(PDF)

Table S2 List of the strains used in this study, their genotype and the way they were constructed.
(PDF)

Table S3 List of the plasmids used in this study, their relevant traits and the way they were obtained.
(PDF)

Table S4 List of the oligonucleotides used in this study, their sequence and the strain for the construction of which they were used.
(PDF)

Text S1 Supplementary Material.
(PDF)

Acknowledgments

We thank I. Bouchara and T. Stratmann for strains S4085 and S4197; O. Tenaillon, T. Hwa and an anonymous reviewer for helpful comments.

Author Contributions

Conceived and designed the experiments: LP KS ML. Performed the experiments: LP. Analyzed the data: LP SG KS JB ML. Wrote the paper: LP SG KS JB ML. Designed the model: LP JB ML. Performed model analysis: SG.

References

1. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134: 25–36.
2. Zamenhof S, Eichhorn HH (1967) Study of microbial evolution through loss of biosynthetic functions: establishment of 'defective' mutants. Nature 216: 456–458.
3. Andrews KJ, Hegeman GD (1976) Selective disadvantage of non-functional protein synthesis in *Escherichia coli*. J Mol Evol 8: 317–328.
4. Koch AL (1983) The protein burden of *lac* operon products. J Mol Evol 19: 455–462.
5. Vind J, Sørensen MA, Rasmussen MD, Pedersen S (1993) Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. Expression from reporter genes does not always reect functional mRNA levels. J Mol Biol 231: 678–688.
6. Dong H, Nilsson L, Kurland CG (1995) Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J Bacteriol 177: 1497–1504.
7. Müller-Hill B (1996) The *lac* operon: a short history of a genetic paradigm Walter de Gruyter.
8. Novick A, Weiner M (1957) Enzyme induction as an all-or-none phenomenon. Proc Natl Acad Sci U S A 43: 553–566.
9. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. Nature 436: 588–592.
10. Stoebel DM, Dean AM, Dykhuizen DE (2008) The cost of expression of *Escherichia coli lac* operon proteins is in the process, not in the products. Genetics 178: 1653–1660.
11. v Hofsten B (1961) The inhibitory effect of galactosides on the growth of *Escherichia coli*. Biochimica et Biophysica Acta 48: 164–171.
12. Wilson DM, Putzrath RM, Wilson TH (1981) Inhibition of growth of *Escherichia coli* by lactose and other galactosides. Biochim Biophys Acta (BBA) - Biomembranes 649: 377–384.
13. Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. Cell 139: 1366–1375.
14. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of Cell Growth and Gene Expression: Origins and Consequences. Science 330: 1099–1102.
15. Elf J, Nilsson K, Tenson T, Ehrenberg M (2006) Bistable bacterial growth rate in response to antibiotics with low membrane permeability. Phys Rev Lett 97: 258104.
16. Fange D, Nilsson K, Tenson T, Ehrenberg M (2009) Drug efflux pump deficiency and drug target resistance masking in growing bacteria. Proc Natl Acad Sci U S A 106: 8215–8220.
17. Tan C, Marguet P, You L (2009) Emergent bistability by a growth-modulating positive feedback circuit. Nat Chem Biol 5: 842–848.
18. Kuhlman T, Zhang Z, Saier MH, Hwa T (2007) Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. Proc Natl Acad Sci U S A 104: 6043–6048.
19. Lehming N, Sartorius J, Niemöller M, Genenger G, von Wilcken-Bergmann B, et al. (1987) The interaction of the recognition helix of *lac* repressor with *lac* operator. EMBO J 6: 3145–3153.
20. Yarchuk O, Jacques N, Guillerez J, Dreyfus M (1992) Interdependence of translation, transcription and mRNA degradation in the *lacZ* genes. J Mol Biol 226: 581–596.
21. Dean AM (1995) A molecular investigation of genotype by environment interactions. Genetics 139: 19–33.
22. Dykhuizen D, Hartl D (1978) Transport by the lactose permease of *Escherichia coli* as the basis of lactose killing. J Bacteriol 135: 876–882.
23. Dykhuizen DE, Dean AM, Hartl DL (1987) Metabolic ux and fitness. Genetics 115: 25–31.
24. Wagner RR, Riley MA (1996) Low synonymous site variation at the *lacY* locus in *Escherichia coli* suggests the action of positive selection. J Mol Evol 42: 79–84.
25. Youk H, van Oudenaarden A (2009) Growth landscape formed by perception and import of glucose in yeast. Nature 462: 875–879.
26. Weinreich DM, Delaney NF, Depristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312: 111–114.
27. de Visser JAGM, Park SC, Krug J (2009) Exploring the effect of sex on empirical fitness landscapes. Am Nat 174 Suppl 1: S15–S30.
28. Fisher RA (1930) The Genetical Theory of Natural Selection Oxford University Press.
29. Orr HA (2006) The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. J Theor Biol 238: 279–285.
30. Martin G, Elena SF, Lenormand T (2007) Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat Genet 39: 555–560.
31. Kauffman S, Levin S (1987) Towards a general theory of adaptive walks on rugged landscapes. J Theor Biol 128: 11–45.
32. Gros PA, Nagard HL, Tenaillon O (2009) The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. Genetics 182: 277–293.
33. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640–6645.
34. Sambrook J, Russell D (2001) Molecular cloning: a laboratory manual Cold Spring Harbor Laboratory Press.
35. Wilson GG, Young KY, Edlin GJ, Konigsberg W (1979) High-frequency generalised transduction by bacteriophage T4. Nature 280: 80–82.
36. Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR (1989) New method for generating deletions and gene replacements in *Escherichia coli*. J Bacteriol 171: 4617–4622.
37. Nguyen TN, Phan QG, Duong LP, Bertrand KP, Lenski RE (1989) Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of *Escherichia coli* K12. Mol Biol Evol 6: 213–225.
38. Miller JH (1992) A short course in bacterial genetics. A laboratory manual and handbook for *Escherichia coli* and related bacteria Cold Spring Harbor Laboratory Press.
39. Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am Nat 138: 1315–1341.
40. Chevin LM (2011) On measuring selection in experimental evolution. Biol Lett 7: 210–213.
41. Cheng B, Fournier RL, Relue PA, Schisler J (2001) An experimental and theoretical study of the inhibition of *Escherichia coli lac* operon gene expression by antigene oligonucleotides. Biotechnology and bioengineering 74: 220–229.
42. Noel JT, Pilyugin SS, Narang A (2009) The diffusive inux and carrier efflux have a strong effect on the bistability of the *lac operon in Escherichia coli*. J Theor Biol 256: 14–28.