LETTERS

Optimality and evolutionary tuning of the expression level of a protein

Erez Dekel[1] & Uri Alon[1]

Different proteins have different expression levels. It is unclear to what extent these expression levels are optimized to their environment. Evolutionary theories suggest that protein expression levels maximize fitness[1–11], but the fitness as a function of protein level has seldom been directly measured. To address this, we studied the *lac* system of *Escherichia coli*, which allows the cell to use the sugar lactose for growth[12]. We experimentally measured the growth burden[13,14] due to production and maintenance of the Lac proteins (cost), as well as the growth advantage (benefit) conferred by the Lac proteins when lactose is present. The fitness function, given by the difference between the benefit and the cost, predicts that for each lactose environment there exists an optimal Lac expression level that maximizes growth rate. We then performed serial dilution evolution experiments at different lactose concentrations. In a few hundred generations, cells evolved to reach the predicted optimal expression levels. Thus, protein expression from the *lac* operon seems to be a solution of a cost–benefit optimization problem, and can be rapidly tuned by evolution to function optimally in new environments.

The expression level of a protein can change over evolutionary timescales by two main processes[15]. The first is neutral evolutionary drift[7], and the second is selection of mutations that increase fitness. The latter is often viewed as an optimization process[1–11,16–18]. Many cases of optimization have been demonstrated on the level of the organism phenotype. For example, bacteria evolve to increase their growth rates in a wide variety of laboratory conditions[1,19,20]. Recently, laboratory evolution experiments showed evolutionary adaptation of *E. coli* towards optimal metabolic fluxes[3,21].

Here we ask whether evolutionary optimization can predict the expression level of a protein in a given environment. We use the lactose operon of *E. coli*, which encodes LacZ, which cleaves the sugar lactose for use as an energy and carbon source, and LacY, which transports lactose into the cell[12,22–24]. Decades of study have provided a quantitative characterization of this system, making it an excellent starting point for a theoretical and experimental study.

The present study had three stages: first, we measured the cost and benefit of Lac protein expression in wild-type *E. coli*. Second, we found that the cost and benefit functions predict that there is an optimal expression level that maximizes growth in a given lactose environment. Third, we monitored the evolution of *E. coli* in different lactose environments and compared its Lac expression to the predicted optimum in each environment (Fig. 1).

To form a cost–benefit theory, we began with direct measurement of the cost and benefit of Lac protein expression in wild-type *E. coli*. The cost is defined as the relative reduction in growth rate due to the burden placed on the cells by production and presence of the Lac

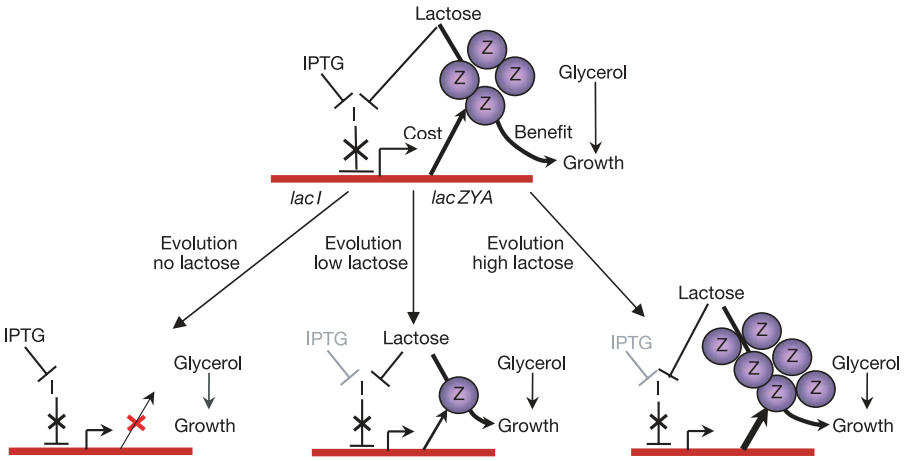

Figure 1 | The *lac* operon of *E. coli* and the experimental design in the present study. The *lac* operon encodes the enzyme LacZ, denoted Z, which uses the sugar lactose, L, to increase growth rate. In the experiments, we consider environments with constant concentrations of lactose. The repressor of the *lac* system LacI, denoted I, was deactivated by the presence of lactose. To examine evolution at zero lactose under the same conditions, we added the non-metabolized inducer IPTG, which also deactivates LacI, and glycerol as an additional carbon source, in all cases. Measurements on the wild-type strain provided estimates of the reduction in growth due to the burden of Lac protein production or maintenance (cost η) and the increase in growth generated by Lac proteins in the presence of lactose (benefit B). Evolution experiments were performed by serial dilution of cells growing in tubes for several hundred generations. Cells undergoing evolution in the absence of lactose were predicted to lose *lac* expression. Cells undergoing evolution at low (high) lactose concentrations were predicted to evolve to optimal low (high) Lac protein levels relative to wild type.

[1]Department of Molecular Cell Biology and Department of Physics of Complex Systems, The Weizmann Institute of Science, Rehovot 76100, Israel.

proteins[13,14]. To measure the cost of Lac expression, we measured the growth rate at various concentrations of the inducer isopropyl-β-D-thiogalactoside (IPTG; Fig. 2a) in the absence of lactose. IPTG induces the *lac* system, but no benefit is gained because IPTG is not metabolized. We found a nonlinear, concave dependence of the cost on the level of expression (Fig. 2a). The cost per protein seems to increase with the amount of proteins produced.

The nonlinear cost function may reflect the fact that the cell has limited resources, so that production of Lac proteins reduces the capacity to produce other important proteins. For example, rapidly growing cells are known to devote most of their ribosomal capacity to synthesizing ribosomal proteins and other translation factors, and most of their transcription capacity to produce ribosomal related RNA[25]. To describe the cost as a function of protein level we used two nonlinear functions. The first is a quadratic function

$$\eta_1(Z) = \eta_0 Z + \eta_0' Z^2 \quad (1)$$

where Z is the expression level of the Lac proteins and η_0 and η_0' are parameters. The second cost function assumes that there is an effective maximal capacity M for producing non-essential proteins. Production of Lac proteins at levels that approach M would inhibit production of essential systems, so that cell growth would be significantly slowed down. Perhaps the simplest phenomenological cost function that increases strongly as an effective limit M is approached is (see Methods for a derivation of this function):

$$\eta_2(Z) = \frac{\eta_0 Z}{1 - Z/M} \quad (2)$$

The parameters of cost function 2 that describe the measured reduction in growth rate are $\eta_0 Z_{WT} = 0.02 \pm 0.003$, where Z_{WT} is the fully induced expression of the wild-type Lac proteins and $M = (1.8 \pm 0.3) Z_{WT}$. We note that the parameter M is expected to depend on the growth rate of the cells: at slower growth rates, more resources are known to be free for protein expression[25] and M is higher (see Supplementary Information for more details). We also note that a linear cost function does not fit the direct cost measurement of the wild-type strain (Fig. 2a) and is not able to explain the results presented below. We find that the cost of fully induced wild-type expression of the *lac* system is about 4.5%.

In addition to the cost of Lac proteins, we also directly measured their benefit. For this purpose, we provided saturating IPTG for full induction and various concentrations of the sugar lactose. The non-metabolized inducer IPTG kept the Lac protein levels constant, and hence the cost constant, allowing measurement of the benefit due to use of lactose. The growth rate of the cells increased with lactose concentration, reflecting the benefit gained by the action of the Lac proteins that transport and use this sugar (Fig. 2b).

We find that the form of the benefit function is well described by the established transport and catabolism kinetics of this system (see

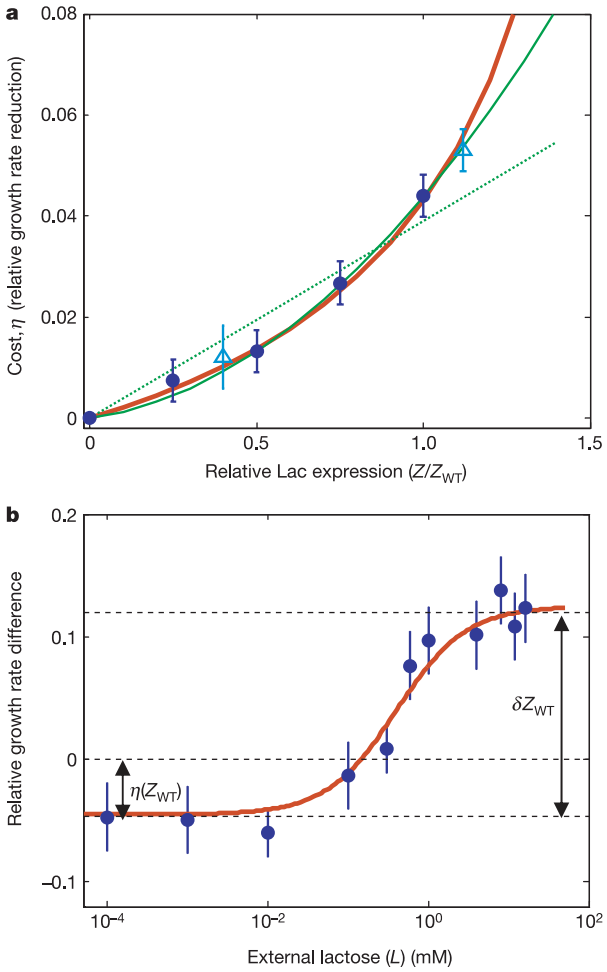

Figure 2 | Cost and benefit functions of *lac* expression in wild-type *E. coli*. **a**, Cost, defined as relative reduction in growth of *E. coli* wild-type cells grown in defined glycerol medium with varying amounts of IPTG relative to cells grown with no IPTG. The *x* axis is LacZ protein level relative to LacZ protein level at saturating IPTG (Z_{WT}). Also shown are the cost of strains evolved at 0.2 mM lactose for 530 generations (data point at the 0.4 point of the *x* axis; open triangle) and 5 mM lactose for 400 generations (data point at the 1.12 point of the *x* axis; open triangle). The green solid line represents cost function 1 (equation (1) with $\eta_0 Z_{WT} = 0.09 \pm 0.01$ and $\eta_0' Z_{WT}^2 = 3 \pm 0.1$). The red line represents cost function 2 (equation (2)). The green dotted line represents a linear cost function. **b**, Benefit of Lac proteins as a function of lactose. Cells were grown with saturating levels of IPTG and varying levels of lactose. The growth rate difference is shown relative to cells grown with no IPTG or lactose. $\eta(Z_{WT})$ is the cost of the *lac* system at zero lactose, and δZ_{WT} is the benefit of *lac* induction at saturating lactose levels. The red line indicates the theoretical growth rate, using $g = -\eta(Z_{WT}) + \delta Z_{WT} L/(K_Y + L)$, with $K_Y = 0.4$ mM, $\eta(Z_{WT}) = 0.044$ and $\delta Z_{WT} = 0.17$ (equation (5)). Error bars are the experimental standard errors.

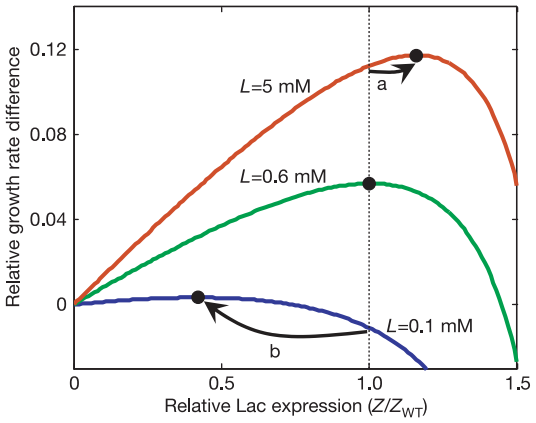

Figure 3 | 3 Predicted relative growth rate of cells (the fitness function) as a function of Lac protein expression. The fitness function, given by the difference of cost and benefit, is shown for different concentrations of lactose. The *x* axis is the ratio of protein level to the fully induced wild-type protein level, Z/Z_{WT}. Shown are relative growth rate differences with respect to the uninduced wild-type strain for environments with lactose levels $L = 0.1$ mM (blue line), $L = 0.6$ mM (green line) and $L = 5$ mM (red line), according to equation (5). The dot on each line is the predicted optimal expression level that provides maximal growth (equation (6)). Cells grown in lactose levels above 0.6 mM are predicted to evolve to increased Lac protein expression (arrow a), whereas cells grown at lactose levels lower than 0.6 mM are predicted to evolve to decreased Lac protein expression (arrow b).

Methods and Supplementary Information). The increase in growth rate is proportional to the rate of lactose use, and hence to the enzymatic rate of LacZ:

$$B(Z) = \delta[ZL_{\mathrm{in}}] \quad (3)$$

where L_{in} is the concentration of lactose in the cell, $[ZL_{\mathrm{in}}]$ is the concentration of LacZ bound to lactose, and δ is the relative growth advantage per LacZ molecule at saturating lactose concentration. Using the known biochemical kinetic parameters of LacY and LacZ yields good agreement with our measured growth rates, and provides an experimental estimate of the benefit parameter δ under the present conditions: $\delta = 0.17\ Z_{\mathrm{WT}}^{-1}$. Thus, full induction of the *lac* system confers a 17% growth advantage at saturating lactose relative to cells with no lactose.

These experiments defined the cost and benefit as a function of Lac protein levels in the wild-type bacterium. The growth rate relative to the growth rate with no Lac expression is given by the difference between the cost and benefit functions

$$g = -\eta(Z) + B(Z) \quad (4)$$

We find that the cost and benefit functions are such that at each level of lactose in the environment there exists a well-defined optimal protein expression level that maximizes growth. The reason for this optimum is that benefit increases linearly with expression level, whereas cost increases nonlinearly and thus high expression levels become unfavourable. The optimal protein expression level depends on the concentration of lactose in the environment, L (Fig. 3). To compute the optimum we calculated the protein level Z_{opt}, which maximizes growth (see equation (6) in Methods for an analytical solution); that is, where $\mathrm{d}g/\mathrm{d}Z = 0$.

We find that the predicted optimal protein level Z_{opt} increases with the lactose level in the environment (Fig. 3). In the absence of lactose, the optimal expression is $Z_{\mathrm{opt}} = 0$, because the Lac proteins bear only a cost and no benefit. At high lactose levels, Z_{opt} is greater than the wild-type expression level Z_{WT}, because increased *lac* expression brings additional growth benefit to the cells. The cost–benefit analysis predicts that the wild-type unrepressed expression level is optimal ($Z_{\mathrm{opt}} = Z_{\mathrm{WT}}$) at a lactose concentration of $L_0 = 0.6 \pm 0.1$ mM.

Does *E. coli* actually evolve to these optimal expression levels if supplied with a constant lactose environment? To test this, we performed laboratory evolution experiments using serial dilution protocols[19]. Cells were grown in 10-ml cultures in a defined medium supplemented with a given lactose concentration. Each day, the cells were diluted 100-fold into a new tube with fresh medium, resulting in $\log_2 100 = 6.6$ generations of growth per day. Experiments at seven lactose concentrations were performed in parallel for over 500 generations.

At different times during the experiment we measured LacZ activity using an enzymatic assay (ONPG assay; see Supplementary Information) and LacZ protein level using quantitative gel electrophoresis (see Supplementary Information). We found that, in all cases, the activity per LacZ enzyme did not measurably change over the course of the evolutionary experiments. Hence, changes in LacZ activity were proportional to changes in LacZ protein level (Supplementary Fig. 6).

We found that LacZ activity and protein level of the cells changed heritably during the course of the experiment, and approached a new adapted state after 300–500 generations (Fig. 4). Cells grown at low lactose concentrations ($L = 0$, 0.1 mM and 0.2 mM) showed a decrease in LacZ activity and expression (Fig. 4). Cells grown in the absence of lactose lost enzyme expression and activity altogether due to mutations such as a 764-base-pair deletion that included the entire *lac* promoter (see Supplementary Information). Cells growing at $L = 0.5$ mM showed little change in LacZ activity and expression. Cells growing at lactose concentrations of 1 mM, 2 mM and 5 mM showed an increase in LacZ activity and expression, although the increase saturated at about 12% of the wild-type level.

The evolved cells reached LacZ expression levels close to the predicted optimal levels (Fig. 4). All cost functions give good predictions at low to intermediate lactose levels ($L < 1$ mM) (Fig. 4b), predicting, for example, that the wild-type protein levels are optimal in an environment with $L \approx 0.5$ mM, as observed. The cost function that best explains the data at high lactose levels is cost function 2 (equation (2)). This cost function shows a large cost per protein at high expression levels, and thus limits the optimal protein levels in high lactose environments. Hence, the cost and benefit fitness function measured in the wild-type strain can be used to evaluate the optimal protein level reached by evolutionary selection in different environments.

The rate at which the cells converged to their adapted expression level is shown in Fig. 4a. The adapted expression is reached within

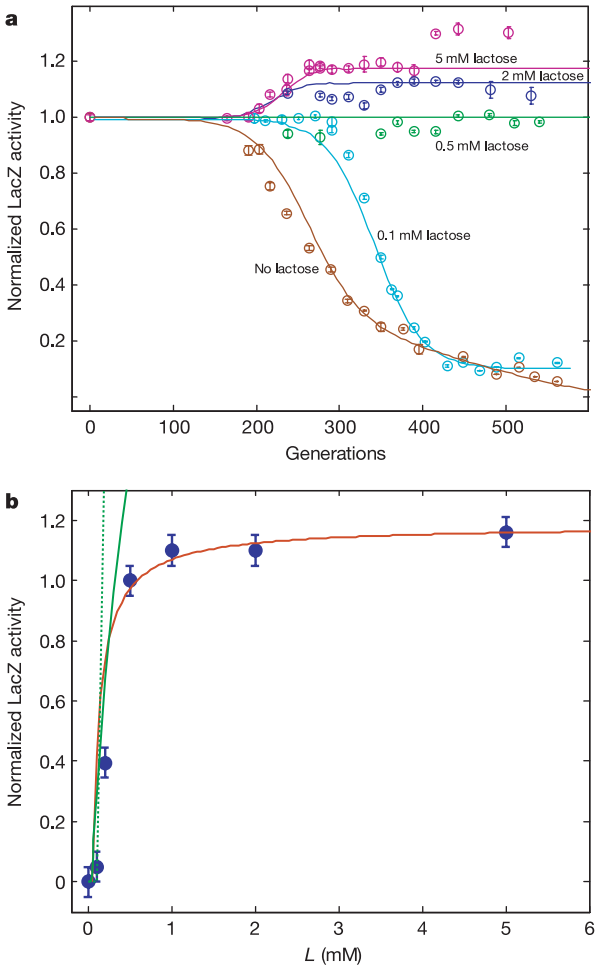

Figure 4 | Experimental evolutionary adaptation of *E. coli* cells to different concentrations of lactose. **a**, LacZ activity, relative to wild-type cells, of cells grown for 530 generations in serial dilution experiments with different lactose levels is shown as a function of generation number. Cells were grown in 0, 0.1 mM, 0.5 mM, 2 mM and 5 mM lactose in a glycerol minimal medium supplemented with 0.15 mM IPTG. Lines are population genetics simulations. The only fitting parameter in these simulations is the probability per generation per cell of a mutation that yields the predicted optimal LacZ level (see Methods). **b**, Adapted LacZ activity of cells in serial dilution experiments as a function of lactose concentration, L, relative to wild-type cells. Data are for more than 530 generations, except for the data point at 5 mM lactose, which is at generation 400. The red line indicates the theoretical prediction for optimal expression level using cost function 2 (equation (6)). The green solid line indicates the quadratic cost function. The green dotted line indicates the linear cost function. Error bars are the experimental standard errors.

about 350 generations for high lactose levels and within 400–500 generations at low lactose levels. These rates are well described by a population genetics model in which cells grow exponentially and produce, with probability p, mutants that have the predicted protein level and the measured adapted growth rate. The only free parameter in the theoretical curves shown in Fig. 4a is the probability per cell per generation, p, of producing a mutant, which ranged between 10^{-7} and 3×10^{-7}. The simulations suggest that the mutation occurred and was lost during dilution many times during the experiment until it was fixed in the population, and that the dynamics can be explained by a single mutation that changed *lac* expression rather than a combination of several mutations[11]. The curves in every run of the stochastic simulation are nearly identical, showing that the selection dynamics can be described by a deterministic approximation (see Supplementary Information). Because the mutation rate in *E. coli* is on the order of 10^{-9} per base pair per generation, the observed mutant probabilities, $p \approx 10^{-7}$, suggest that there is an effective 'mutational target' on the order of 100 different base pairs that can give rise to each of the optimal phenotypes.

At the highest level of lactose, 5 mM, cells first converged to the predicted optimal expression level, and then, around generation 400, showed a second jump[1] in LacZ expression. This jump may represent mutations in other systems that provide an advantage in the present growth conditions, or a change in the parameter M in cost function 2. Indeed, direct measurements of the cost function of this mutant suggest a threefold increase in the parameter M (Supplementary Fig. 8). In contrast, direct measurement of the cost functions of strains that evolved at lower lactose levels shows a cost function that is the same as that of the wild-type strain (Fig. 2a).

This study provides evidence that cells can rapidly evolve towards protein expression levels that are optimal solutions of a cost–benefit problem. The cost and benefit fitness function measured in the wild-type strain was found to predict optimal protein levels in different lactose environments, which were then reached in direct evolutionary experiments. It would be intriguing to test whether cost–benefit principles apply to more complex gene regulation circuits, to attempt to form and test a theory of optimal gene circuit design[6,8,26–28]. This could explain the observed convergence of different organisms to similar network motifs[16,28,29]. More generally, this study suggests that small expression differences between microorganisms can be due to selected biological functionality rather than random drift. Optimality principles might contribute to a fundamental description of the interaction between biological circuitry, evolution and the environment.

METHODS

Strains and media. *E. coli* MG1655 (*E. coli* genetic stock centre) was used. All experiments were in M9 defined medium consisting of M9 salts, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.05% casamino acids, 0.1% glycerol, 0.15 mM IPTG and specified concentrations of D-lactose (Sigma). This concentration of IPTG had no measurable effect on growth of a strain deleted for the *lac* operon (data not shown). Use of glycerol in the medium allowed study of evolution also in the absence of lactose. The bacteria used lactose and glycerol simultaneously throughout growth (Supplementary Fig. 1). Previous studies have shown that cells grown in a chemostat with limiting lactose evolve to greatly increase the expression of the Lac proteins[20]. In those experiments lactose was the sole carbon source and hence the *lac* system affected growth rate much more strongly than in the present experiments.

Growth rate measurements. The exponential growth rate difference of two strains was measured by comparing 48 cultures of each strain grown in a checkerboard pattern on a 96-well plate, yielding an accuracy of about 0.8% (see Supplementary Information for more details). Multiple assays were averaged in cases where lower standard errors were required. Exponential growth rate in the 96-well plate is the same as the growth rate in 50-ml tubes (Supplementary Figs 3 and 4).

Serial dilution experiments. Ten-millilitre cultures were grown in 50-ml tubes shaken at 220 r.p.m. at 37 °C. After a day of growth, cells were diluted 1:100 into a fresh tube. Samples were frozen (−80 °C) every 3 days. Lactose levels in five of the tubes (5 mM, 2 mM, 1 mM, 0.5 mM, 0.2 mM) were high enough to cause full induction of the *lac* system (Supplementary Fig. 7). To make induction equal and maximal in all tubes, including those with 0 mM and 0.1 mM lactose, each tube also contained a saturating level of IPTG.

β-Galactosidase activity and protein level measurements. A substrate of β-galactosidase (LacZ) that gives optical readout (ONPG) was used in a multi-well plate reader to obtain high accuracy measurements of activity (see Supplementary Information for details). LacZ protein levels were quantified using SDS gel electrophoresis, and relative expression was quantified by comparison to control lanes with appropriate mixtures of induced and uninduced wild-type cell lysates (Supplementary Fig. 6).

Derivation of cost function 2. A simple derivation of cost function 2 is based on a growth rate that is a saturated function[30] of an internal resource R, given by $g = \beta R/(K + R)$ where β is the maximal growth rate and K is the resource level for half maximal growth. Production of a unit of protein Z reduces the resource R by ε. Hence, $g(Z) = \beta(R - \varepsilon Z)/(K + R - \varepsilon Z)$. This yields a cost $\eta(Z) = (g(0) - g(Z))/g(0) = \eta_0 Z/(1 - Z/M)$, as in equation (2), where $\eta_0 = \varepsilon K/R(K + R)$ and $M = (K + R)/\varepsilon$. Note that the cost function can not diverge but rather reaches a maximum value of $\eta(Z) = 1$ when $\varepsilon Z = R$, resulting in $g = 0$.

Analytical approximation to the optimal expression level. A detailed transport model of the *lac* system[22–24] (see Supplementary Information) suggests that, to a good approximation, the rate of LacZ action is equal to the rate of lactose influx through the LacY permease at all but the highest lactose concentrations:

$$V_z[ZL_{in}] \approx V_Y[YL] \approx V_Y \frac{YL}{K_Y + L}$$

where V_z is the velocity of LacZ, V_Y is the velocity of LacY and $K_Y = 0.4$ mM is the Michaelis constant of LacY[24]. This yields a cost–benefit fitness function, using cost function 2 and using Y proportional to Z (so that the measured benefit parameter δ includes the ratio Y/Z as well as V_Y and V_z):

$$g = -\frac{\eta_0 Z}{1 - Z/M} + \delta \frac{ZL}{K_Y + L} \quad (5)$$

The optimum of this, $dg/dZ = 0$, occurs at an optimal protein level of

$$Z_{opt} = M\left(1 - \sqrt{\frac{\eta_0}{\delta}\frac{L + K_Y}{L}}\right) \quad (6)$$

Note that $Z_{opt} = 0$ for lactose levels lower than $L_c = K_Y(\delta/\eta_0 - 1)^{-1} \approx 0.057$ mM because cost exceeds benefit. The parameters are $M = 1.8\,Z_{WT}$, $\eta_0/\delta = 0.02/0.17 = 0.12$ and $K_Y = 0.4$ mM.

Similar results for Z_{opt} were found by numerically solving the detailed transport model and finding the optimal expression by varying LacY and LacZ levels (Supplementary Fig. 9). Note that the analytical approximation (equations (5) and (6)) underestimates the growth advantage of the adapted strains relative to the wild-type strain in the same environment, except near $L = 0$ and $L = 0.6$ mM, where predicted growth advantage matches the experimental results. The detailed simulation gives better predictions for the growth advantage of the adapted strains (Supplementary Fig. 9) at all lactose levels.

Simulations of evolution rate. Population genetics simulations were performed as follows[7] (see Supplementary Information for details): a population of wild-type cells grew exponentially at growth rate g_0, growing each simulated day from $N_0 = 10^8$ cells to $N_f = 10^{10}$ cells. Mutants were formed with a probability p per generation per cell. The mutants grew at rate $g_0 + \Delta g$, with relative LacZ expression ΔA. The parameter Δg was set equal to the measured growth rates of the adapted strains (Supplementary Fig. 9), and ΔA was used from the optimum of the cost and benefit fitness function (equations (5) and (6)). At the end of each simulated day, 1/100 of the population was passed to the next simulated day, of which the fraction of mutants was determined by a random binomial process. The resulting dynamics show that the mutants eventually take over the population. The simulations have only one free parameter, p, which was fitted to the data of Fig. 4a, resulting in $p = 6.5 \times 10^{-6} \pm 2 \times 10^{-6}$, $3 \times 10^{-7} \pm 1 \times 10^{-7}$, $3 \times 10^{-7} \pm 1 \times 10^{-7}$ and $3 \times 10^{-7} \pm 1 \times 10^{-7}$ for $L = 0$ mM, 0.1 mM, 2 mM and 5 mM, respectively. Simulations suggested that adaptation was due to a single mutation, except at $L = 0$, in which two mutations occurred: one that affected LacZ protein level and the other that increased growth rate in glycerol[3], in agreement with measurements on this adapted strain (Supplementary Fig. 11).

Received 11 April; accepted 19 May 2005.

1. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev. Genet.* **4**, 457–469 (2003).

2. Orr, H. A. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* **6,** 119–127 (2005).
3. Ibarra, R. U., Edwards, J. S. & Palsson, B. O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420,** 186–189 (2002).
4. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402,** C47–C52 (1999).
5. Rosen, R. *Optimality Principles in Biology* (Butterworths, London, 1967).
6. Savageau, M. A. *Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology* (Addison-Wesley, Reading, Massachusetts, 1976).
7. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer, Sunderland, Massachusetts, 1997).
8. Heinrich, R. & Schuster, S. *The Regulation of Cellular Systems* (Chapman and Hall, New York, 1996).
9. Maynard Smith, J. & Szathmary, E. *The Major Transitions in Evolution* (Oxford Univ. Press, Oxford, 1997).
10. Hartl, D. L. & Dykhuizen, D. E. The population genetics of *Escherichia coli*. *Annu. Rev. Genet.* **18,** 31–68 (1984).
11. Liebermeister, W., Klipp, E., Schuster, S. & Heinrich, R. A theory of optimal differential gene expression. *Biosystems* **76,** 261–278 (2004).
12. Muller-Hill, B. *The lac Operon: a Short History of a Genetic Paradigm* (Walter de Gruyter, New York, 1996).
13. Koch, A. L. The protein burden of lac operon products. *J. Mol. Evol.* **19,** 455–462 (1983).
14. Nguyen, T. N., Phan, Q. G., Duong, L. P., Bertrand, K. P. & Lenski, R. E. Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of *Escherichia coli* K12. *Mol. Biol. Evol.* **6,** 213–225 (1989).
15. Fay, J. C., McCullough, H. L., Sniegowski, P. D. & Eisen, M. B. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* **5,** R26 (2004).
16. Conant, G. C. & Wagner, A. Convergent evolution of gene circuits. *Nature Genet.* **34,** 264–266 (2003).
17. Stephanopoulos, G. & Kelleher, J. Biochemistry. How to make a superior cell. *Science* **292,** 2024–2025 (2001).
18. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99,** 15112–15117 (2002).
19. Cooper, T. F., Rosen, D. E. & Lenski, R. E. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **100,** 1072–1077 (2003).
20. Dykhuizen, D. E., Dean, A. M. & Hartl, D. L. Metabolic flux and fitness. *Genetics* **115,** 25–31 (1987).
21. Honisch, C., Raghunathan, A., Cantor, C. R., Palsson, B. O. & van den Boom, D. High-throughput mutation detection underlying adaptive evolution of *Escherichia coli*-K12. *Genome Res.* **14,** 2495–2502 (2004).
22. Kremling, A. *et al.* The organization of metabolic reaction networks. III. Application for diauxic growth on glucose and lactose. *Metab. Eng.* **3,** 362–379 (2001).
23. Wong, P., Gladney, S. & Keasling, J. D. Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol. Prog.* **13,** 132–143 (1997).
24. Yildirim, N., Santillan, M., Horike, D. & Mackey, M. C. Dynamics and bistability in a reduced model of the lac operon. *Chaos* **14,** 279–292 (2004).
25. Bremer, H. & Dennis, P. P. in *Escherichia coli* and *Salmonella* (ed. Neidhardt, F. C.) 1553 (American Society for Microbiology, Washington DC, 1996).
26. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99,** 16587–16591 (2002).
27. Endy, D., You, L., Yin, J. & Molineux, I. J. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl Acad. Sci. USA* **97,** 5375–5380 (2000).
28. Dekel, E., Mangan, S. & Alon, U. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys. Biol.* **2,** 81–88 (2005).
29. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298,** 824–827 (2002).
30. Monod, J. The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3,** 371–394 (1949).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Elowitz, R. Kishony, G. Sela, B. Shraiman and all members of our laboratory for discussions. We thank the NIH, ISF and Minerva for support. E.D. thanks the Clore postdoctoral fellowship for support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to U.A. (urialon@weizmann.ac.il).