UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Fitness landscapes of gene regulation in variable environments

Poelwijk, F.J.

Publication date
2008

Citation for published version (APA):
Poelwijk, F. J. (2008). *Fitness landscapes of gene regulation in variable environments.* [Thesis, externally prepared, Universiteit van Amsterdam].

6

Maintenance and loss of gene regulation in experimental evolution

Degeneration is a much commoner phenomenon than progress.

J.B.S. Haldane,
The Causes of Evolution

The evolution of gene regulation is a major open question in biology. Regulatory systems not only allow organisms to respond to a variable environment, but are themselves shaped by evolution under a variable selective pressure. When adaptation is approached as an optimization process, a variable environment adds many degrees of freedom to the search space compared to adaptation in a constant environment. How environmental variation selects for different modes of regulation, or in which cases other strategies than regulation are favored, such as bet-hedging, is presently under intense debate. These issues cannot be addressed at a theoretical level alone, and require information about the evolutionary plasticity and potential functional constraints of actual biological systems. In the present work we experimentally follow regulatory adaptation starting from a non-optimal regulatory response. We focus on regulation of lactose metabolism in Escherichia coli, which arguably exhibits a near-optimal relation between the amount of lactose in the environment and the level of expression of the lactose metabolic genes. By using separate compounds for induction and metabolism, we dislodged the lac regulatory response from its optimum, and predicted new optima. We followed adaptation in several constant and alternating environments. We find some cases of fast adaptation to the predicted optimum. In a number of instances adaptation occurred, but the predicted optimum was not reached. This may be due to a diminishing return of further optimization, or to the existence of a functional or genetic constraint.

All living organisms are equipped with mechanisms that enable them to sense their environment and respond to it. In many cases the response consists of regulating gene expression. For bacteria the link between sensing and response is often formed by a small network of interacting proteins and regulatory sites on the DNA controlling the expression of downstream genes. Such a gene regulatory system can be characterized by its regulation function (or regulatory profile, or induction curve), that specifies the relation between the environmental signal and the expression level of the regulated genes.

Just as the adaptation of catalytic properties of enzymes or protein expression levels may be viewed as an optimization process [43, 73, 144], so may be the adaptation of regulation. However, in the case of regulation, the system is shaped by varying selective pressures in an environment that is fluctuating, which makes that there are many more potential parameters to optimize. Optimality can, for example, concern the nature of the environmental fluctuations (regular or stochastic), their time scales, the strength of selection in each environmental state, and how variable selective pressures constitute trade-offs experienced by the adapting organism. Moreover, similar regulation profiles can be accomplished by different modes of regulation, for example by employing a repressor (negative control), or an activator (positive control), and several theoretical studies have addressed their optimality in an ecological context [74–76]. Alternatively, in the case of slow and unpredictable environmental variation, it might be optimal not to employ regulatory systems, but to stochastically switch phenotypic states [77]. Such issues have mainly been been approached theoretically and experiments have been lagging behind. As a result, we lack essential information on regulatory plasticity and the potential constraints that hamper reaching an optimal regulatory response.

In this work we will consider the lactose operon of *Escherichia coli* from the viewpoint of optimality and explore adaptation of the regulatory response to new environments by experimental evolution [119].

6.1 Optimality of gene expression

Following ref. [73], we describe the growth rate of a population of *E. coli* cells as a function of expression of metabolic genes and carbon source (here lactose) in the environment in terms of the cost and benefit of gene expression

$$g = g_0 - \eta(Z) + B(Z, L) \tag{6.1}$$

where g_0 is the basal growth rate, set by compounds other than lactose in the environment. $\eta(Z)$ is the decrease of growth rate due to the burden of producing *lac* operon gene products LacZ, LacY, and LacA [134]. $B(Z, L)$ is the growth advantage due to lactose metabolism, which depends on both the expression level of the *lac* gene products (in particular LacZ), and the concentration of lactose in the environment.

This gives rise to an optimal expression level for each concentration of lactose in the environment $Z = Z_{\text{opt}}(L)$. At low levels of lactose the cost term will dominate the

benefit term, and the optimal expression level will be low or zero. Conversely, at high lactose concentrations the optimal expression level will be high.

The purpose of catabolic regulation is to sense the external concentration of the catabolite and to vary the expression level of metabolic genes as a function of this concentration: $Z = Z(L)$. Selection will drive a regulatory system towards the following optimality relation

$$Z_{\text{opt}}(L) = Z(L) \tag{6.2}$$

implying that the system establishes a connection between the catabolic and inductive properties of lactose. Indeed, for *lac* regulation there are strong indications [73, 97] for this relation to hold. It is important to note that this criterion for regulatory optimality only concerns the relation between expression levels and catabolite concentrations. The regulatory system may well have to be optimized for response times, structural architecture, robustness, or otherwise.

In the present work we are interested in the evolutionary plasticity of the regulation profile $Z(L)$. In order to study regulatory adaptation, the system should be dislodged from its optimum, so that selective pressures arise that are directed towards a new optimum. This can be done in several ways. For example, one may change the kinetic parameters or the demands on the downstream regulated genes, which results in different cost and benefit terms and hence in different expression optima. Another approach, which we present here, is to decouple inducer and carbon source and allow the regulatory system to adapt to a new relation between the two, which is imposed by the experimenter. This approach mimics a situation in which an organism is confronted with a novel carbon source that has a different relation between induction and catabolic benefit.

For the *lac* system, a large number of artificial compounds have been synthesized [93], that interact with the gene products in a different way than lactose. The decoupling between *lac* signal and metabolism can be made by using isopropyl-β-D-thiogalactopyranoside (IPTG), and phenyl-β-D-galactoside (Pgal). IPTG is a gratuitous inducer; it binds to the *lac* repressor and relieves repression, but cannot be hydrolyzed by β-galactosidase. Pgal, on the other hand does not induce LacI, but is hydrolyzed by LacZ, releasing galactose (for further metabolism) and phenol. Now the optimality criterion reads

$$Z_{\text{opt}}(P) = Z(I) \tag{6.3}$$

where P and I are independent variables (the Pgal and IPTG concentrations in the medium). In the present work we experimentally determined $Z(I)$ and the growth rate

$$g = g_0 - \eta(I) + B(I, P) \tag{6.4}$$

with which we can make a prediction for the selective pressures on the regulation for combinations of IPTG and Pgal concentrations. Subsequently we determined how regulation changed during experimental evolution, in which growing cultures of *E. coli* were serially passaged in batch cultures for around 800 generations.
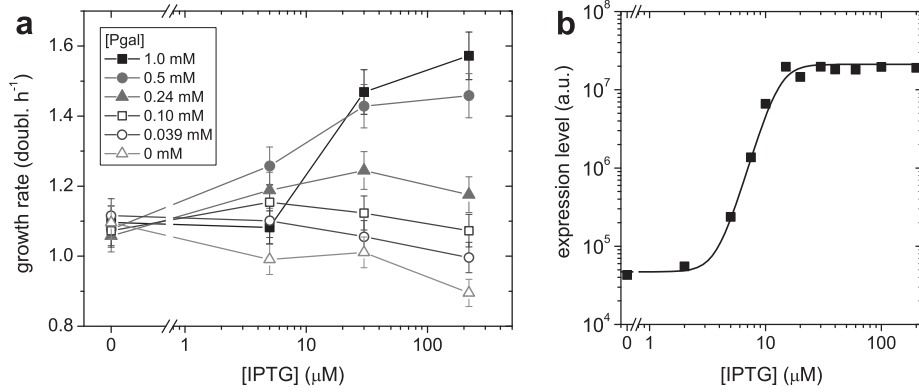
Figure 6.1: a) Measured growth rates as a function of Pgal and IPTG in a minimal M9 medium consisting plus casamino acids (see Materials and methods, section 6.3). At low Pgal concentrations inducing *lac* operon products will yield an expression cost that is larger than the benefit, resulting in a decrease of growth rates. For higher Pgal concentrations the benefit will dominate the cost. b) The induction profile in the absence of Pgal.

We first investigated adaptation of the *lac* system to a constant environment

$$Z_{\text{opt}}(P_1) = Z(I_1) \tag{6.5}$$

which in principle can be attained without regulation, as it only requires the optimization of one expression level.

A selective pressure for a regulatory response can be applied when cells experience multiple environmental conditions that impose opposite expression demands. The system is then confronted with a trade-off that can only be overcome by developing an appropriate regulatory response. To this end we performed evolution experiments in an environment that alternated between two states (P_1, I_1) and (P_2, I_2), so that the optimality criterion reads

$$Z_{\text{opt}}(P_1) = Z(I_1) \qquad \text{and} \qquad Z_{\text{opt}}(P_2) = Z(I_2) \tag{6.6}$$

with different optimal expression levels $Z_{\text{opt}}(P_1) \neq Z_{\text{opt}}(P_2)$.

6.2 Results and discussion

We determined growth rates of *Escherichia coli* MG1655 ('wild-type') cells [188] carrying the *lac* operon, as function of IPTG and Pgal in a minimal medium where casamino acids set the basal growth rate g_0 to be 1.09 generations h^{-1} (see Fig. 6.1). A measured wild-type induction profile is also shown. We observed that when the medium does not contain a carbon source ([Pgal]= 0 mM), induction leads to a cost. The growth rate decrease is 0.20 doublings h^{-1} per hour for full induction.
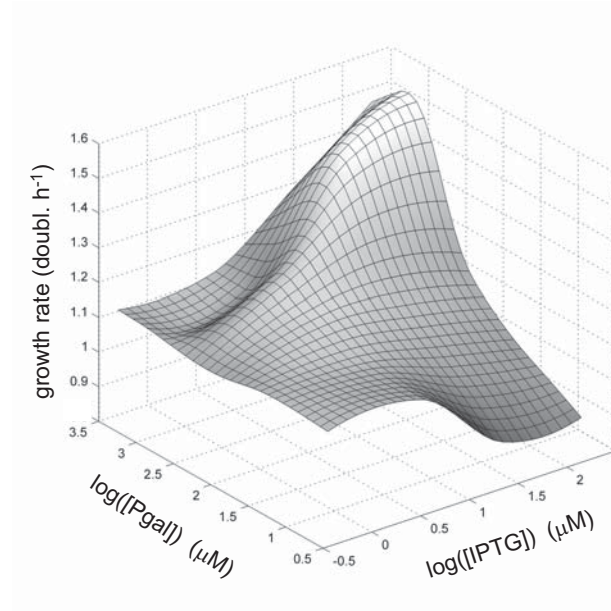
Figure 6.2: Interpolated and smoothened growth data from figure 6.1, providing an impression of the functional form of equation (6.4). The decoupling of inducer and carbon source is visualized: addition of Pgal when not expressing *lac* operon genes (low IPTG concentration) will not result in growth rate increases, and addition of IPTG without Pgal will lower the growth due to an expression cost. The ridge in the landscape is caused by anti-induction of the *lac* repressor by Pgal, when present at a high concentration (see section 6.4). For low Pgal concentrations we used a functional relation for the cost of expression fitted to the data (section 6.4).

At higher concentrations of Pgal cost and benefit are balanced for intermediate inducer concentrations: growth in the presence of 0.10 mM and 0.24 mM Pgal is maximized for IPTG concentrations near 5 μM and 30 μM respectively. For higher Pgal concentrations the maximum observed growth rates lie at inducer levels of 200 μM or higher. We observed that high concentrations of Pgal have an anti-inductive effect due to the competitive binding of IPTG and Pgal to the repressor (see section 6.4). Since the affinity of IPTG for the repressor is much higher than that of Pgal (a K_D of $1 \cdot 10^{-6}$ M versus $1 \cdot 10^{-3}$ M [201]), this effect can be neglected for sufficiently low concentrations of Pgal. We see the effect of anti-induction in the growth data for medium containing 1 mM Pgal, which for 5 μM IPTG has a lower growth rate than medium containing 0.1-0.5 mM Pgal.

In figure 6.2 we show an interpolation and smoothening of the growth data to give an impression of the optimality relations between IPTG and Pgal for the *lac* operon. For low concentrations of IPTG (low expression level) we recover the basal growth rate g_0 independent of the Pgal concentration, while for high IPTG concentrations (high expression level) a benefit only occurs in the presence of a high enough concentra-
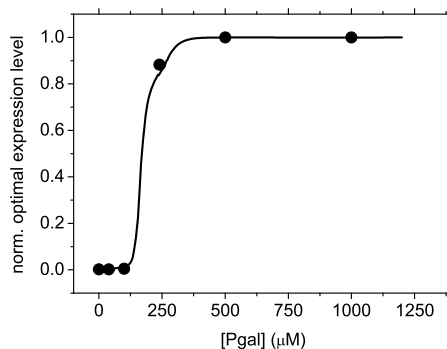
Figure 6.3: Optimal expression levels of *lac* operon genes as a function of Pgal, obtained from the landscape in figure 6.2. The circles represent the optimal expression levels obtained from inspection of the raw growth data in figure 6.1.

tion of Pgal. From these smoothened growth data, we recovered the optimal expression levels for LacZ using the induction profiles that we measured for different concentrations of Pgal (section 6.4). This optimality relation is given in figure 6.3, together with the optimal Pgal concentrations as obtained directly from the growth data in figure 6.1. Although the optimal level shows a very sharp Pgal dependence, this does not mean that the growth difference for optimal and non-optimal expression are necessarily large. We can see from the landscape in fig. 6.2 that for the Pgal concentrations around the inflexion point of fig. 6.3 (~150 μM), the growth rates for high and low expression are very similar. This implies that non-optimality at these Pgal concentrations will not result in high selective pressures.

We modeled the cost and benefit aspects of our system given by equation (6.4) in a similar fashion to what was done in refs. [202] and [73]. Since in our system induction and catabolite are separated, we included IPTG induction and anti-induction for high concentrations of Pgal (section 6.4) in the model, based on independent measurements of the expression levels of LacZ. Although we obtained a clear qualitative agreement, a quantitative agreement was reached only for higher concentrations of IPTG. Data and model for 220 μM IPTG are shown in figure 6.4, and model predictions other IPTG concentrations are given in section 6.4. The observed discrepancies point at an interesting issue: whereas expression levels only rise marginally for IPTG concentrations up to 5 μM, cost and benefit already diversified the obtained growth rates (see figure 6.1). Adjustment of the model is required (section 6.4).
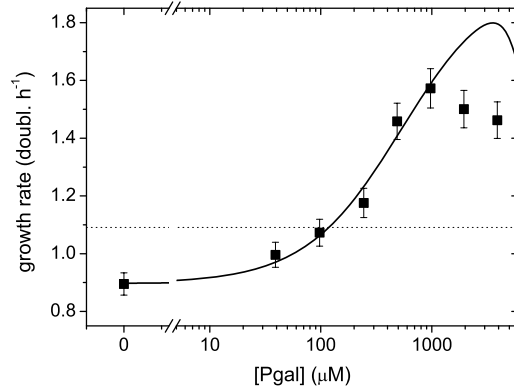
Figure 6.4: Fit of growth data at high induction (220 μM IPTG) using a reaction kinetics model (section 6.4). The dotted horizontal line indicates the growth rate in the absence of Pgal and IPTG. The growth difference between this line and the data point at [Pgal] = 0 represents the cost of protein expression. For higher Pgal concentrations this cost is compensated and eventually dominated by the benefit of Pgal metabolism. Cost and benefit are balanced for a Pgal concentration of $1.2 \cdot 10^2$ μM.

6.2.1 Evolution in constant environments

We performed a serial dilution experiment in a number of constant environments with different concentrations of IPTG and Pgal, as indicated schematically in figure 6.5. For each condition, a 10 ml culture was grown and diluted twice daily 300-500 fold for a total of ~800 generations. Each week a sample of each culture was stored at -80°C to preserve snapshots of its evolutionary history. Afterwards, the LacZ activity[1] of the adapting populations was determined for different time points during the experiment.

The history traces in figure 6.6 indeed show that the expression level of LacZ changes during the adaptation experiment. As expected on the basis of the optimality curve in figure 6.3, cultures grown in the presence of 350 μM Pgal, but at IPTG concentrations that do not fully induce expression (fig. 6.6a and e), increase their uninduced expression levels. In these cases we observed a loss of repression. For the population grown without IPTG it takes ~200 generations before the uninduced levels resemble the induced levels. Notably, two replicate experiments performed at this condition (squares and triangles in fig. 6.6a), are indistinguishable.

---

[1]The measured LacZ activity in principle is determined by both the expression level and the kinetic parameters of LacZ. Therefore the LacZ activity is indicative for the expression level if we assume that the kinetic parameters of LacZ remain unchanged, as we do in this chapter. We argue that this assumption reasonable since we do not see an increase in maximum LacZ activity for adaptation lines that select for higher expression (fig. 6.6).
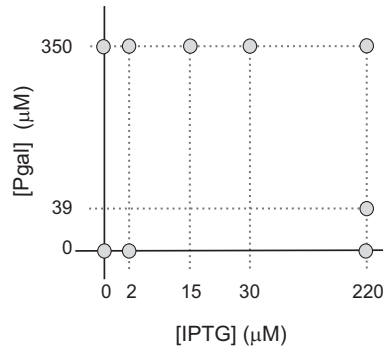
Figure 6.5: Overview of Pgal and IPTG concentrations of the constant environments in which adaptation experiments were performed.

Comparing the history traces of the cultures without IPTG (fig. 6.6a) to the trace of the culture grown at 2 μM IPTG (fig. 6.6e), we observe that the rates of adaptation are markedly different. If both traces are fitted with a simple competition model (assuming a single mutant fixation and a sufficiently high mutation rate to be able to neglect stochasticity due to bottlenecking the population, see section 6.4), we find that the selection coefficient of the population growing without IPTG is more than 4 times larger than that of the population at 2 μM IPTG ($s = 0.055$ versus 0.013)[2]. Although we would indeed expect the selection coefficient to decrease for increasing concentrations of IPTG, the observed large difference between 0 and 2 μM IPTG is remarkable in the face of the small expression differences between these IPTG concentrations in wild-type cells (see fig. 6.1b). On the other hand, from fig. 6.1a, we can see that wild-type cells for a Pgal concentration of 0.5 mM already realize more than half of their expression benefit at 5 μM IPTG. Consequently, at this IPTG concentration the additional selective advantage of abolishing regulation is decreased considerably compared to cells growing in the absence of IPTG. We further note that the absolute values of the selection coefficients found from the history traces are lower than expected on the basis of the wild-type growth rates. On the basis of the landscape in figure 6.2, we would expect a selection coefficient on the order of 0.2 for adaptation in the absence of IPTG.

In figure 6.6b we show the evolutionary trace of a culture grown under conditions of high amounts of carbon source ($[\text{Pgal}] = 350$ μM) and high induction ($[\text{IPTG}] = 220$ μM). No significant adaptation is observed, and as these conditions provide near optimal growth rates, this is as expected. Unchanged expression levels were similarly found for the culture grown at 350 μM Pgal and 30 μM IPTG (not shown). When fully induced, the regulatory system is in principle free to lose regulation by neutral drift: mutations that deactivate the repressor do not affect the growth rate. Since mutations that restore

---

[2]The selection coefficients determine the rate at which a mutant is fixed in the population, which in a history trace corresponds to the steepness of the curve (see section 6.4).
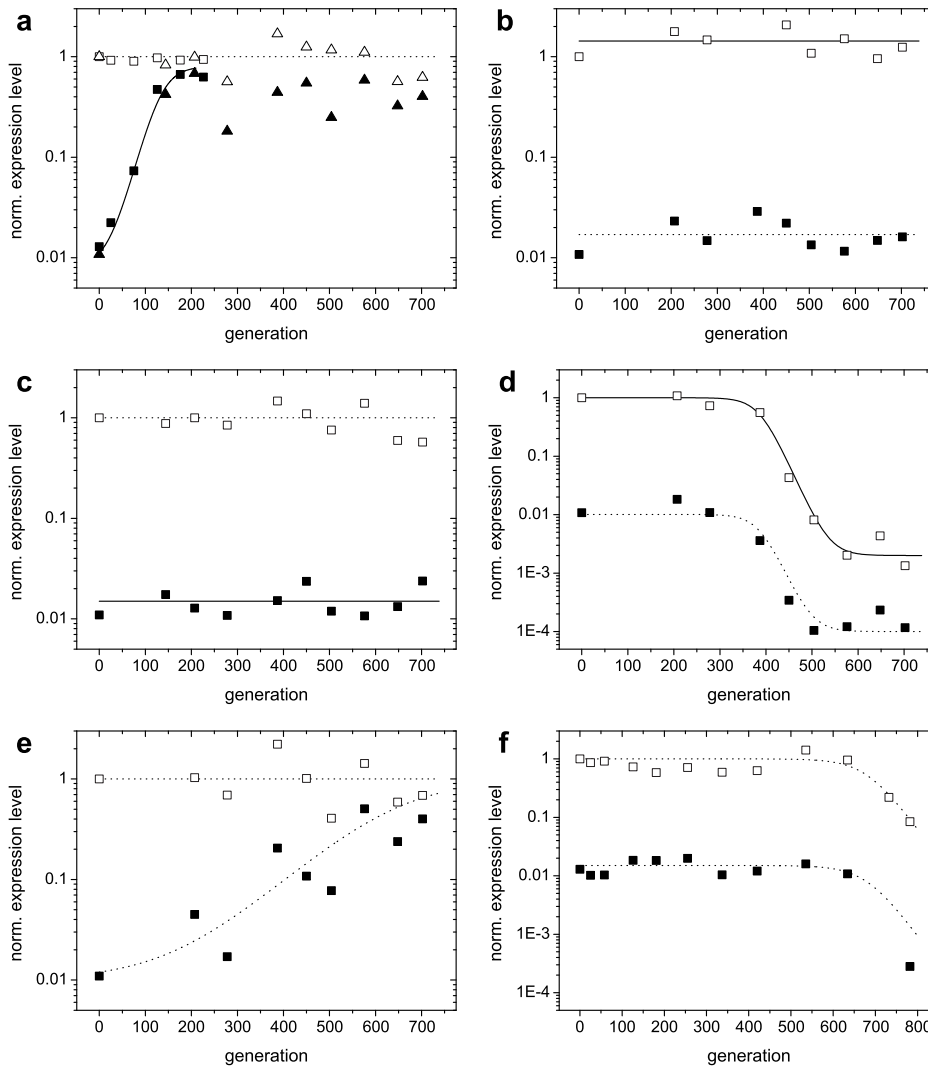
Figure 6.6: History traces of expression levels (population averages) for a subset of populations adapting in constant environments. Open symbols represent induced expression levels (at 220 μM IPTG), solid symbols are uninduced expression levels. Curves are fits based on growth rate differences under exponential growth (section 6.4). Where induction levels are the same as in the environment to which the populations adapted, the curves are solid. a) 0 μM IPTG, 350 μM Pgal. Two populations evolved in parallel are shown to yield the same adaptation dynamics (triangles and squares). b) 220 μM IPTG, 350 μM Pgal. c) 0 μM IPTG, 0 μM Pgal. d) 220 μM IPTG, 0 μM Pgal. e) 2 μM IPTG, 350 μM Pgal. f) 2 μM IPTG, 0 μM Pgal.

repressor function are in general much less likely to occur, in the long run repressor null mutants will fix in the population. However, the expected rate at which this would occur is on the order of $1/\mu$ generations [203], where μ is the mutation rate towards $lacI^-$ mutants, being $\sim 1 \cdot 10^{-6}$ [204]. If repressor deactivation is neutral, fixation would only be expected after $1 \cdot 10^6$ generations. Interestingly, a null mutation in the promoter controlling the transcription of the repressor may actually be selectively favored, since it should reduce the cost associated with the production of repressor protein. On the basis of the low amount of repressor protein compared to the other *lac* gene products, we expect that the selection coefficient associated with the loss of repressor production are too low to be observable within the time course of the experiments performed here.

Figure 6.6c also shows an unchanged regulation. Here the medium contains no IPTG and no Pgal, and expression of *lac* operon products would only incur a cost. A similar argument as above for the neutral loss, now of downstream operon products (LacZ, LacY, and LacA) could be developed. Indeed no change is expected within the 800 generations followed here.

Expression is drastically reduced during growth on 200 μM IPTG and 0 μM Pgal (fig. 6.6d). The rate at which the expression decreases in the population suggests a selection coefficient of around 0.067, comparable with the loss of repression observed in fig. 6.6a. However, the fact that fixation occurs at later generations indicates that this type of mutants occur less frequently than the repressor null mutants.

Figure 6.6f shows the evolutionary history of a population growing without Pgal, but with 2 μM IPTG. Here again expression of downstream genes is decreased. We observe a selection coefficient that is roughly half of that in fig. 6.6d. Apparently, 2 μM IPTG increases expression of downstream genes enough to be selected against in the absence of carbon source.

Two conditions indicated in fig. 6.5 remain. For medium containing 39 μM Pgal and 220 μM IPTG we observed no significant change in expression levels, while we found (fig. 6.3) that the optimal expression level at this Pgal concentration would be zero. However, from our landscape we would predict the selection coefficient for this condition to be around 2.5 times lower than for medium containing no Pgal and 220 μM IPTG. Assuming a similar mutation rate towards low expression for the medium condition discussed here, we do not expect observable changes in expression levels before generation ~900.

For the constant environment at 15 μM IPTG and 350 μM Pgal we found an altered induction profile (figure 6.7), such that the expression level at 15 μM IPTG remained similar to wild-type expression (taking into account Pgal anti-induction). From the data shown in our landscape in fig. 6.2 we can infer only a marginal difference in fitness between the expression level as induced with 15 μM IPTG compared to that at 220 μM IPTG. Interestingly, in this population a mutant was fixed that did not abolish repression altogether. Once fixed, the fitness advantage of a constitutive mutant would be minimal, and hence its chance to invade the population would be small.
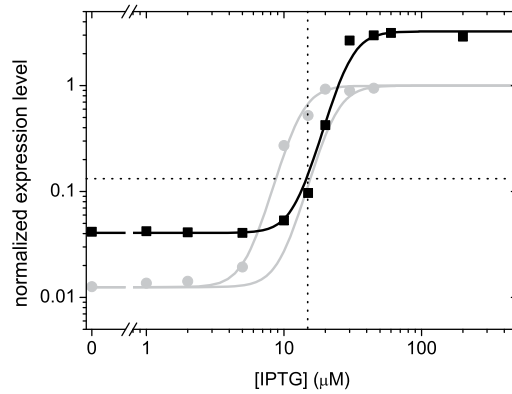
Figure 6.7: Induction profile (black squares and curve) of a population evolved for approximately 700 generations in the presence of 15 μM IPTG and 350 μM Pgal. Also shown is the wild-type induction profile (grey data points and fit), as well as the profile for wild-type incorporating Pgal anti-induction (grey curve shifted to the right). The expression levels at 15 μM IPTG for wild-type (with anti-induction) and evolved strain are very similar.

From 8 clonal isolates after the serial dilution experiment we sequenced the chromosomal region consisting of the *lac* repressor, the *lac* promoter (upstream of *lacZ*), until 420 base pairs into the *lacZ* coding sequence (see fig. 1.4 in chapter 1). Compared to the reference GenBank nucleotide sequence of the *lac* operon (accession number J01636.1), all isolates contain a often occurring *lacI* polymorphism (C857T) that does not affect LacI function, and a silent mutation in the coding sequence of *lacZ*. From earlier work we know that C857T pre-existed in the MG1655 strain, and we assume that the *lacZ* mutation did also. Apart from these mutations, three clones isolated from the population adapted to 350 μM Pgal, 0 μM IPTG all showed a known hotspot frameshift deletion of four base pairs from a triply repeated TGGC (nucleotides 593-604 of the *lacI* coding sequence) [204]. This frameshift leads to complete inactivation of the repressor [204], which corresponds to our observation. One clone sequenced from adaptation on 350 μM Pgal, 220 μM IPTG and another from 0 μM Pgal, 0 μM IPTG, which retained wild-type induction characteristics, did not reveal any mutations. Remarkably, three clones sequenced from the population that adapted to 220 μM IPTG, 0 μM Pgal, also showed the hotspot frameshift. These isolates do not show a constitutive expression, but instead a greatly reduced expression, which means that they must carry another mutation. However, since these isolates did not contain mutations in the promoter controlling *lacZ* expression, no cause for the observed loss of LacZ activity (which originated from selection against expression cost, not against activity) can be identified at present.

109
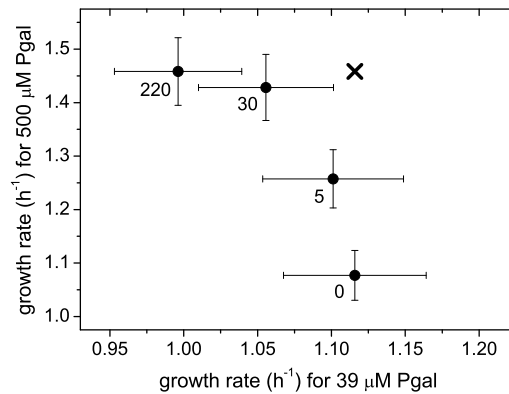
Figure 6.8: Example of trade-offs experienced when expression is not regulated in an environment that alternates between a low Pgal concentration (39 μM) and a high Pgal concentration (350 μM). Data points denote the growth rate in each environment for a certain constant expression level (at the indicated concentrations of IPTG in μM). This trade-off data is directly obtained from figure 6.1. Low expression levels (0 μM IPTG) yield optimal growth in medium with low Pgal concentrations, but non-optimal growth in medium containing high Pgal concentrations, and *vice versa* for high expression levels. Only when expression is regulated (low in low Pgal conditions and high in high Pgal conditions), overall growth over both environments can be optimal (indicated by the black cross).

6.2.2 Evolution in alternating environments

Regulation is favorable when an organism is confronted with a fitness trade-off, which occurs when optimizing the expression level in one state decreases the fitness in the other state. Using the growth data in figure 6.1 we can visualize such trade-offs. In figure 6.8, we plotted the growth rate in an environment with a high Pgal concentration (500 μM) versus the growth rate in an environment with a low Pgal concentration (39 μM). The IPTG concentrations to which these data points belong are given in the figure (in μM). The point of maximum growth rate under both conditions is marked with a cross. The figure suggests that there is no intermediate inducer concentration (hence expression level) that provides optimal growth under both conditions. From the concave shape of the curve we would expect that in an environment that alternates between these states in equal periods, an unregulated expression level would be evolutionary adjusted towards an intermediate level between low and high expression (see chapter 4).

We performed a number of serial dilution experiments in which the environment was alternating between two states (fig. 6.9). A change of environment was accomplished once or twice daily (see section 6.3). For 4 out of 6 experiments (marked with
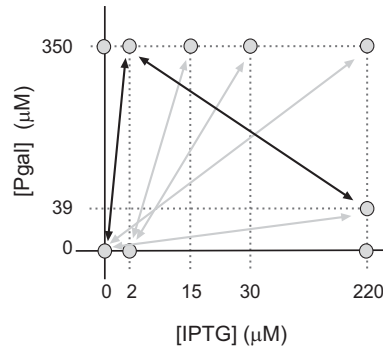
Figure 6.9: Overview of Pgal and IPTG concentrations of the alternating environments in which adaptation experiments were performed. Grey arrows indicate conditions which resulted in unaltered induction profiles. Black arrows did result in adapted profiles (see text).

grey arrows in the figure) we found no significant change of the induction profile. Interestingly, from the experiments in a constant environment we know that there is a selective pressure to decrease expression levels for a population grown at 2 μM IPTG and 0 μM Pgal. However, for the cultures alternating between this condition and high Pgal (350 μM) plus moderately high IPTG (15 and 30 μM), we found no response to decrease the expression level at low IPTG concentrations. Growth in the high Pgal condition prevents loss of expression, so that it would be advantageous here to increase the ratio between low and high expression levels. Since this was not observed, this may indicate a functional constraint in the system. For the environment alternating between no IPTG, no Pgal and 220 μM IPTG, 39 μM Pgal, we would expect an overall decrease of expression on the basis of the optimality curve in fig. 6.3, but evolution in a constant environment of 220 μM IPTG, 39 μM Pgal already showed that the selection coefficients are probably too small to see adaptation here within 800 generations.

In two alternating environments the induction profile did change. First, alternating between 2 μM IPTG, 350 μM Pgal and 220 μM IPTG, 39 μM Pgal almost fully abolished regulation and acquired a high constitutive expression. These conditions were intended to elicit an inverted regulatory response to IPTG (as was accomplished in chapter 4), but the selective pressure to decrease expression in the presence of 39 μM Pgal was not strong enough to lower the expression level at high concentration of IPTG (at least not within the time course of this experiment). The recovery of a constitutive expression suggests that the inverted response is genetically less accessible. Whether continuation of the adaptation experiment would result in the optimal inverted response, or whether the fixation of an inactivated repressor constitutes an evolutionary dead-end cannot be decided at the moment.

Second, when the environment alternates between no IPTG, no Pgal and 2 μM IPTG, 350 μM Pgal, a constitutive expression results. An optimal regulatory strategy would have been here to change the inflexion point of the induction curve to lower
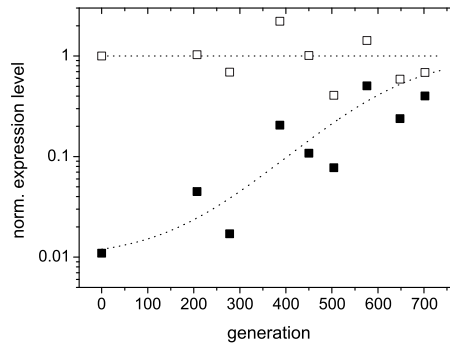
Figure 6.10: History trace of expression levels for the population evolved in an environment alternating between no IPTG, no Pgal and 2 μM IPTG, 350 μM Pgal. The population evolves towards a constitutive expression.

IPTG concentrations, which could result from a higher affinity of the repressor for IPTG. It is likely, however, that genetic changes that accomplish this are not easily accessible. The adaptation that occurred here maximizes growth in the environmental state with Pgal. However, the predicted fitness loss due to spurious expression in the state without Pgal is comparable. The fact that the mutation fixes rapidly in the population (fig. 6.10) on the other hand implies a considerable selection coefficient.

6.2.3 Conclusions and outlook

The decoupling of inducer and carbon source presented in this work provides a framework for studying the evolutionary plasticity of gene regulation. In most cases within the course of a few hundreds of generations a predicted optimal expression level was reached. In some cases optimal expression levels under the same conditions were not reached by populations that evolved in an alternating environment, when the other environmental state imposes a different optimal expression level. This implies that the lack of evolutionary change in these cases is not due to a lack of selective pressure, but points at functional constraints of the gene regulatory system.

We observed that mutations inactivating the repressor are more easily accessible than mutations that abolish expression, in cases where the selective advantages for such mutations are comparable. The high accessibility of the hotspot frameshift mutation [204], leading to *lac*⁻ phenotypes at a frequency that is an order of magnitude higher than would be expected from the genomic mutation rate[3], is interesting in the

---

[3]The genomic mutation rate being $5.4 \cdot 10^{-10}$ per base pair per replication for E. coli [205], we would expect mutations to occur in the *lac* coding sequence (1080 base pairs) at a rate of ~$6 \cdot 10^{-7}$. It is known that roughly a quarter of substitutions is synonymous, and that around half of the amino acid substitutions in the *lac* repressor are deleterious [177]. This yields an expected rate of *lacI*⁻ mutations of ~$2 \cdot 10^{-7}$. The observed rate is $2 \cdot 10^{-6}$ [204]

light of regulatory evolution. In fact, both deletions and additions of the 4 base pair repeat are observed at high frequency [204], which implies that reversals of hotspot mutations will also be more likely than reversals of e.g. base pair substitution that deactivate *lacI*. Together with the observation that the *lacI* coding sequence surrounding the hotspot is highly structured (palindromic), which elevates the mutation frequency due to slippage of a replicating DNA polymerase [206], it seems as if the *lac* repressor also has an in-built mutational regulation.

We obtained one instance (evolved at 350 μM Pgal, 15 μM IPTG) where the induction profile was not changed towards either constitutive expression or loss off expression. In some cases evolving towards constitutive expression was not predicted to be optimal on the basis of the measured growth rates from fig. 6.1. For example the culture alternating between 350 μM Pgal, 2 μM IPTG and 39 μM Pgal, 220 μM IPTG, would be optimal when the response to inducer would be inverted. However, it acquired a mutation which abolished repression first, after which the additional fitness gain of lowering expression for the high IPTG condition was diminished, and probably not enough to adapt within the course of 800 generations (even apart from whether such phenotype is still genetically accessible after the first mutation). The combination of fitness effect and mutational accessibility will determine which mutation will fix in the population first. This initial fixation is likely to lead to a 'law of diminishing returns' for subsequent mutations [27], and hence longer fixation times.

As a final remark, we note that the superior mutant would cheat our decoupling between inducer and carbon source and become responsive to Pgal, so that

$$Z_{\mathrm{opt}}(P) = Z(P) \tag{6.7}$$

None of the evolved populations here was found to be induced by Pgal. Although this is an interesting issue, we expect that for this type of mutation to occur, a range of environments and an amount of generations need to be surveyed that is not easily accessible in this type of laboratory evolution experiments. Adaptation of inducer specificity is a subject probably more effectively studied at a higher level of control over the system, as was achieved in chapter 4.

In this work we presented an experimental approach to explore the adaptation of a gene regulatory response. Due to the many ways an environment can be variable, the survey of environmental conditions and fluctuations was necessarily limited. To obtain a high level of understanding of the evolutionary aspects of gene regulation will require a major effort, but which is necessary, as it increasingly appears that evolution is more strongly driven by regulatory changes than by modifications in non-regulatory proteins [165].

6.3   Materials and methods

Strains and media
We used *Escherichia coli* strain MG1655 [188]. All experiments were performed in M9

minimal medium, consisting of M9 salts (Sigma-Aldrich), supplemented with 0.1 mM CaCl$_2$ (Merck Eurolab), 1 mM MgSO$_4$ (Merck Eurolab), and 5 g/l casamino acids (BD Biosciences). When indicated, media contained isopropyl-β-D-thiogalactopyranoside (IPTG) and phenyl-β-D-galactoside (Pgal), both obtained from Sigma-Aldrich. All cultures were grown at 37°C.

Determination of growth rates

Growth rate determinations were performed after overnight growth in the medium described above, without IPTG or Pgal, followed by at least 3 hours growth in medium with the appropriate concentrations of Pgal and IPTG. Subsequently the cultures were diluted to an optical density of ~5·10^{-4} and transferred to a pre-warmed flat bottom 96 well microtiter plate (VWR 351172), at 200 μl per well. Optical density at 600 nm was recorded in a Perkin & Elmer Victor[3] plate reader every 4 minutes, and every 29 minutes 9 μl sterile water was added to each well to counteract evaporation. When not measuring, the plate reader was shaking the plate at double orbit with a diameter of 2 mm. All presented growth values are averages of 3 independent measurements. From measurements in which all 96 wells were inoculated with wild-type MG1655, we determined the error margins on our averaged growth data to be 4.3%.

Determination of β-galactosidase activity

To determine the activity of β-galactosidase (LacZ) we used essentially the same method as described in section 4.1. Before transfer to a 96 well plate, cultures were grown overnight without IPTG and Pgal, and then diluted to an optical density of ~5·10^{-4}. When expression levels were high so that overnight passage through stationary phase resulted in 'superinduced' LacZ activity levels (see section 6.4), growth times before fluorescence determination were prolonged.

As some of the expression levels in the current work were one or two orders of magnitude lower than the lowest expression levels in chapter 4, determination of the initial slope of these FDG curves was inaccurate due to experimental noise. In these cases we used the maximum slope at long timescales, which is proportional to the slope at t=0 (by a factor ~30, see section 4.1).

As before, during the assay of hydrolysis concentrations of IPTG and Pgal in each sample are made equal, to prevent unfair comparison due to competitive inhibition of LacZ by IPTG or Pgal.

Serial dilution experiments

10 ml cultures were grown in 50 ml flasks in a 37°C water bath under vigorous shaking (200 rpm). Cultures were diluted 300-500x twice daily in fresh medium. As stationary cultures contain ~10^9 cells ml^{-1}, this implies bottleneck sizes of ~10^7 cells (for 10 ml total culture volume). The alternating conditions were either switched twice daily (for the cultures alternating between 2 μM IPTG, 0 μM Pgal and 15/30 μM IPTG, 350 μM Pgal, see fig. 6.9), or once daily (for the remaining conditions). When switching from a higher concentration of IPTG or Pgal to a lower one, cultures were washed 3x in minimal medium. Each four days a sample of the cultures was frozen at -80°C. Re-inoculation
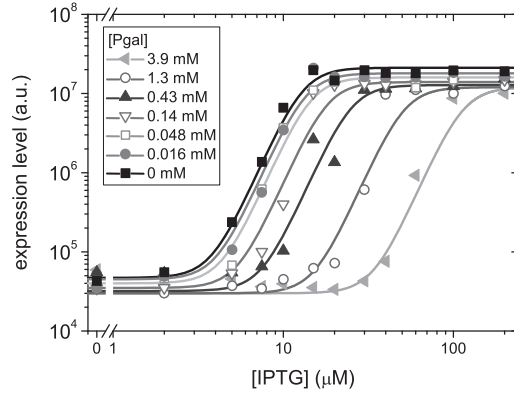
Figure 6.11: Induction profiles for wild-type cultures grown at different Pgal concentrations. The data is fitted with a Hill function that incorporates anti-induction by Pgal.

occurred after thawing and 3x washing in minimal medium.

6.4 Supplementary information

Analysis of induction curves and competitive inhibition by Pgal
The *lac* repressor does not only bind IPTG, it also has a affinity for Pgal. Pgal does not induce the repressor, but does competitively prevent IPTG from binding and thus effectively anti-induces the repressor. Since the equilibrium dissociation constants differ by three orders of magnitude ($K_D = 1 \cdot 10^{-6}$ for IPTG and $1 \cdot 10^{-3}$ for Pgal [201]), this effect is only noticeable when the Pgal concentration is much higher than the IPTG concentration. In other cases IPTG and Pgal can be considered to be decoupled with respect to induction of the *lac* repressor. We measured the effect of Pgal anti-induction by growing cultures under different concentrations of Pgal and IPTG. Immediately before the LacZ assay, all concentrations of IPTG and Pgal were equalized to prevent unequal inhibition at the level of LacZ. The results are given in figure 6.11.

The data for [Pgal] = 0 was fitted with a general Hill function (compare ref. [207])

$$\alpha_{\text{IPTG}} = \alpha_0 \frac{1 + F\left([\text{IPTG}]/C_{\text{IPTG}}\right)^m}{1 + \left([\text{IPTG}]/C_{\text{IPTG}}\right)^m} \tag{6.8}$$

where α_0 is a parameter relating the number of LacZ molecules to the measured LacZ activity, F is the ratio between induced and uninduced LacZ activity, C_{IPTG} is a dissociation constant associated with the affinity of IPTG to the repressor, and m is a phenomenological Hill exponent incorporating non-linear behavior of the *lac* induction (due to e.g. cooperative binding of the repressor to multiple operators).

115

The curves for higher Pgal concentrations in fig. 6.11 were obtained by incorporating Pgal anti-induction into the Hill description of the system. This is done in a similar way as in chapter 4, where we described and measured the competitive binding of FDG and IPTG to LacZ. Pgal anti-induction increases the effective equilibrium dissociation constant of IPTG to the *lac* repressor, (or equivalently lowers the effective IPTG concentration).

$$C_{\text{IPTG,eff}} = C_{\text{IPTG}}(1 + \frac{[\text{Pgal}]}{K_P}) \tag{6.9}$$

where K_P is the apparent equilibrium dissociation constant of Pgal binding to the repressor (which incorporates a potential difference in internal and external Pgal concentrations).

All curves in fig. 6.11 could be fitted using a K_P of 0.45 mM. A minor vertical offset of the curves was observed due to the different growth rates under the different (Pgal, IPTG) conditions, as can be expected on the basis of a slightly different dilution rate (see e.g. equation (7.1) for the relation between growth rate and expression level).

Comparison to reaction kinetics model for transport and degradation

Here we modeled our system according to the cost-benefit analysis for a fixed expression level as reported in [73]. We modify this model by including induction by IPTG and anti-induction by Pgal. A comparison of this model to the obtained growth data in figure 6.1 is made.

The proposed [73] functional form for relative growth due to the cost and benefit of *lac* operon gene expression is

$$\Delta g = -\eta(Z) + B(Z, L) = -\frac{\eta_0 \cdot Z}{1 - Z/M} + \delta \frac{Z \cdot L}{K_Y + L} \tag{6.10}$$

where η is the cost term and B the benefit term, that depend on the concentration of LacZ (Z), lactose (L), and the equilibrium dissociation constant of the LacY permease and lactose (K_Y). This expression was derived under the assumption of low lactose concentrations (< 1 mM). This assumption assures that the rate limiting step in lactose metabolism is the import of lactose into the cell by LacY. In our system this condition is also fulfilled, as the equilibrium dissociation constant of LacY for Pgal is of the same order of magnitude as that for lactose, being 1.3 mM [208]. Pgal has a ~10 times higher affinity for LacZ [209], which justifies the assumption for Pgal.

We thus modified the expression for relative growth as follows

$$\Delta g = -\eta(Z(I, P)) + B(Z(I, P), P) = -\frac{\eta_0 \cdot Z(I, P)}{1 - Z(I, P)/M} + \delta \frac{Z(I, P)) \cdot P}{K_Y + P} \tag{6.11}$$

where the LacZ expression $Z(I, P)$ depends on IPTG as well as on Pgal concentration (the latter being of influence only at high concentrations).

Using this model we fitted the growth at high IPTG concentrations (220 μM), which works well for Pgal concentrations up until 1 mM (fig. 6.4). Indeed for higher concentrations the assumptions of the model may be violated.
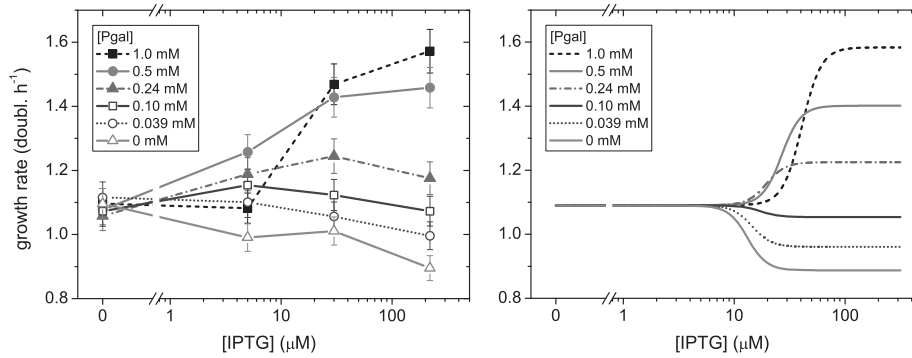
Figure 6.12: Comparison of growth data in the presence of various concentrations of IPTG and Pgal (left), with a model incorporating transport and catalysis of Pgal, as well as induction by IPTG and anti-induction by Pgal (right). The qualitative trend and the predictions at higher IPTG concentrations correspond well. For lower IPTG concentration there is a discrepancy: cost and benefit of expression occur at lower induction levels than is predicted by the model, on the basis of the measured induction profiles (fig. 6.11).

When we compare model predictions using the obtained parameters from the fit for 200 µM IPTG to reproduce the data at lower IPTG concentrations, we observe a qualitative correspondence only (fig. 6.12). The major difference is the occurrence of a cost and a benefit at very low IPTG concentrations (e.g. 5 µM), while LacZ expression levels have only increased marginally (from fig. 6.11 we see that they are still a factor ~100 below fully induced levels). These observations suggest that the cost and benefit terms may exhibit a steeper dependence on operon expression levels than assumed in the model. Alternatively, the model might need to incorporate competition between Pgal and IPTG for LacZ. However this would imply that Pgal import by LacY is not rate limiting, and hence violate the assumptions underlying the present model, which precludes an analytical solution. A numerical description of the system is ongoing work.

Non-stochastic competition model

Evolutionary traces were fitted (dotted and solid curves in expression history graphs) using a non-stochastic model for the change in expression when a mutant fixes in the population. It is known that the fate of mutants in a population that is periodically bottlenecked is influenced by 'sampling noise' when the mutation is initially only present in a few individuals [210]. However, when the mutation rate is such that the expected number of mutants after bottlenecking is significantly larger than 1 ($\mu b \gg 1$, where b is the bottleneck size and μ the mutation rate), these stochastic effects can be ignored. This seems to hold in our case at least for the hotspot mutations that occur at a rate of ~$1 \cdot 10^{-6}$, while our bottleneck size is ~10^7. Moreover, the selection coefficients are estimated from the rate of the fixation process, which is independent of the mutation

rate (unless the population is very small or the mutation rate very high).

As both the wild-type and the mutant population grow exponentially in between the bottlenecks we have for their numbers

$$\text{wild-type} \quad N(t) = N_0 e^{\ln 2 g t}$$

$$\text{mutant} \quad N^*(t) = N_0^* e^{\ln 2 g (1+s) t} \tag{6.12}$$

where g is the wild-type growth rate, s is the selection coefficient. On the basis of these numbers of individuals, we have for the expression levels of a population average

$$E_{\text{ave}}(t) = \frac{E^* N^*(t) + E N(t)}{N^*(t) + N(t)} = E \frac{E_r R_0 e^{\ln 2 g s t} + 1}{R_0 e^{\ln 2 g s t} + 1} \tag{6.13}$$

where $R_0 = N_0^* / N_0$ is the initial ratio of mutants, and $E_r = E^* / E$ the ratio of the expression levels of the mutant and the wild-type. This expression assumes that a mutant arises close to the beginning of the experiment and further does not address consecutive mutations.

Enzyme dilution

Important for both the correct determination of expression levels, as well as important to take into account when setting up an experiment with alternating medium conditions, is the fact that we observed a 'superinduced' LacZ activity for cells after spending a stationary phase at high expression levels. We found that the expression levels of induced cells as determined immediately after they leave stationary phase, can be up to a factor of 10 higher that the expression during exponential growth. If this happens, it can take very long before LacZ molecules are diluted out by cellular division, even when their production is low. To demonstrate this effect, a culture of wild-type cells was grown overnight at full induction (200 μM IPTG). The next morning the culture was washed and grown in fresh medium without IPTG. At specific time points samples were taken and frozen at -80°C. Afterwards the expression levels for these samples was determined (figure 6.13). In the figure induced and uninduced levels of expression for an exponentially growing population are given as dotted lines. We indeed observe that the cells initially have a much higher LacZ expression than exponentially growing induced cells. The expression levels decrease over time, which corresponds to the observed growth rate of the cells. Remarkably, even after 8 hours of growth the expression level of exponentially growing uninduced cells has not yet been reached.

In our determination of expression levels (see above), we have taken into account the long times it may take to be able to determine the expression levels associated with exponential growth. Importantly, for evolution experiments under alternating conditions, it is essential to take into account enzyme dilution effects on the response times of the regulatory system.
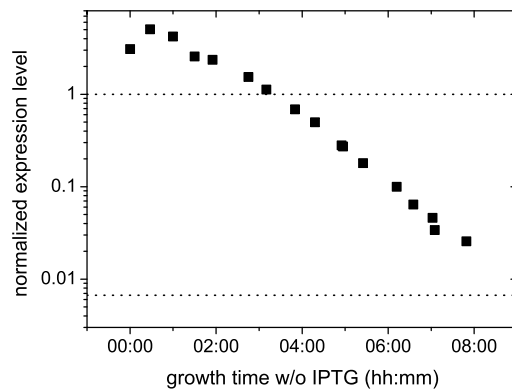
Figure 6.13: Enzyme dilution by cellular division, visualized by decreasing expression levels. At t=0 a stationary overnight culture of induced (200 µM) wild-type cells is inoculated in fresh medium. At the indicated time points, samples are taken from this culture of which the expression is determined. Dotted lines are uninduced (lower) and fully induced (upper) expression levels of an exponentially growing wild-type population. On the logarithmic vertical scale we observe a near linear decay, corresponding to exponential dilution of the enzyme. From the decay rates a slowly increasing decay rate $(1/t_{1/2})$ was determined, starting at 0.92 h^{-1} and ending at 1.3 h^{-1}. The initial rate corresponds well with the growth rate of the population at full induction, whereas the end rate is somewhat higher than that of an uninduced population, which might be caused by intrinsic degradation of LacZ.