

Modeling Complex Biological Systems: Tackling the Parameter Curse Through Evolution



Paulien Hogeweg

Abstract As we all know, “Nothing in biology makes sense except in the light of evolution” Dobzhansky (Am Biol Teach 35(3):125–129, 1973). Among the challenges of modeling complex biological systems is to determine the relevant parameters. The common practice is to extract parameters from the literature, or to determine them from ongoing experiments, or by collectively fitting the parameters to the experimental results the model tries to explain. Doing so ignores, or at least does not exploit, Dobzhansky’s wisdom. In this perspective paper, we argue and demonstrate the importance of using evolutionary methods to derive relevant parameters. We show that by doing so, we can debug experimental and modeling artifacts.

1 Introduction

The holy grail of systems biology is to match experimental and modeling results. In pursuing this lofty goal, one should keep in mind that wet experiments and dry in silico modeling face different opportunities and limitations to the challenge of unraveling complex biological systems. A common heuristic for experiments is to keep conditions as constant as possible, and limiting the variability of the biological material, e.g., by working with clonal populations, or preferring males over females in medical research because of less hormonal variation. This way a simplest “input-output” system is approached, without accounting for the (variable) state, i.e., considering an $\langle I, O, \Omega \rangle$ dynamical system (defined in terms of a set of inputs (I), a set of outputs (O), and a function linking input and output (Ω)) instead of a full $\langle I, O, S, \Omega, \Sigma \rangle$ dynamical system, in which in addition the internal state (S) and internal state changes (Σ) are considered. In contrast, modeling approaches do focus on state changes of the system and use either the full system specification

P. Hogeweg (✉)

Theoretical Biology & Bioinformatics Group, Utrecht University, Utrecht, the Netherlands

e-mail: p.hogeweg@uu.nl

© Springer Nature Switzerland AG 2021

A. Crombach (ed.), *Evolutionary Systems Biology*,

https://doi.org/10.1007/978-3-030-71737-7_2

$\langle I, O, S, \Omega, \Sigma \rangle$ or simplify it to $\langle S, \Sigma \rangle$ considering fixed external conditions (no input beyond initial condition) and observing state changes as outputs.

In silico modeling requires full specification of processes and parameters. An often used heuristic is to compose a very simple model, in which it is possible to survey the full parameter space. This way the result takes the form “the observed *in vivo* or *in vitro* behavior is possible within the proposed model”. In such simple models, the parameters are often composites of potentially measurable quantities and often not validated beyond the fact that they produce the observed behavior. On the other hand, large-scale models try to integrate measurements of many different experiments, often having to add “reasonable” values for unknown parameters, and determine whether these parameters and inferred interactions incorporated in the model indeed produce particular experimental results.

In both cases, one should keep in mind the warning signal put up by James Watson (as quoted by Francis Crick 1988, pp. 59–60 Crick, 1988) “no good model ever accounted for all the facts, since some data was bound to be misleading if not plain wrong. A theory that did fit all the data would have been ‘carpentered’ to do so and would thus be open to suspicion.”

The relevance of this warning signal is preeminently exposed in the history of the study of the lactose operon (*lac* operon). Both models and experiments agreed for a long time that the *lac* operon coded for a bistable switch, (e.g., Griffith, 1968; Novick & Weiner, 1957; Ozbudak et al., 2004), although this notion was challenged early on on theoretical grounds by Savageau (1999). This conclusion is now on theoretical and experimental grounds falsified, (e.g., Afroz et al., 2014; Ozbudak et al., 2004; Rao & Koirala, 2014; Savageau, 2011; Van Hoek & Hogeweg, 2006, 2007; Zander et al., 2017). Here we will relate how evolutionary systems theory contributed to this reversed conclusion.

2 Case Study: The *lac* Operon and Bistability

We will use the *lac* operon to illustrate the power of evolutionary modeling to understand the “how and why” of a particular well-studied regulatory circuit. To this end, we will review an earlier published model and results (Van Hoek & Hogeweg, 2006, 2007), emphasizing the methodology, from a conceptual as well as from a “hands-on” point of view. For details of the model, quantitative results, and mathematical analysis, the reader is referred to the original publications (Van Hoek & Hogeweg, 2006, 2007).

2.1 Background: “State of the Art”

The *lac* operon has been seen for many years as the prototype example of a bistable switch. Indeed the very concept of gene regulation was discovered by Jacob

and Monod (1961) by the observation of population heterogeneity and hysteresis when *E. coli* was grown at different concentrations of an inducer. An artificial, not metabolized, inducer was used, in order to be able to create constant conditions for the experiment. These and subsequent experimental results were soon supported by a simple theoretical model, showing that the positive feedback loop of the import of an inducer on the internal inducer concentration was sufficient to explain the bistability (Novick & Weiner, 1957). This model is taught in many “introduction to biological modeling” courses. Such a so-called mini-model shows that for some parameters, the model can account for the observed behavior and therewith that such a positive feedback is potentially sufficient to explain the observations. For such a compact mini-model, there are only a few parameters; these parameters can be fitted to match the experimental results but cannot be measured in a model-independent way.

Subsequent experimental results elucidated many details of the structure of the *lac* operon and its regulation by a combination of the inducer (lactose or an artificial substitute (IPTG or TMG)) and the preferred resource glucose (see scheme in Fig. 1). A shorthand description is the *lac* operon is an AND gate: it is *ON* when there is lactose and no glucose and *OFF* otherwise.

Subsequent large-scale modeling (e.g., Wong et al., 1997; Yildirim & Mackey, 2003) incorporated this accumulated experimental knowledge. Consequently these models contain many parameters, which were taken from the literature or estimated as “reasonable”. Also these models concluded that yes indeed the *lac* operon of *E. coli* encodes a bistable switch. However, close scrutiny of the parameters used revealed large differences between those used in different models. Moreover some parameters were adjusted in order to ensure bistability.

Finally, the notion that the *lac* operon encoded a bistable switch was also reinforced by an evolutionary mini-model which showed bistability to be advantageous (e.g., Thattai & Van Oudenaarden, 2004).

Although the agreement between models, experiments, and optimization consideration may seem conclusive, the parameter uncertainties and their adjustments to match experimental outcomes suggest that we should heed Watson’s warning quoted above.

Our research was triggered by a then recent paper of (Setty et al., 2003) which reported direct measurements of the transcription rate of the operon (by coupling a GFP reporter to the operon) for many combinations of the artificial inducer IPTG and cAMP (high cAMP concentrations correspond to low glucose concentration; see scheme of the lactose operon in Fig. 1). The resulting promoter function is not a simple AND gate (*ON* for high inducer and high cAMP (i.e., low glucose) and *OFF* otherwise) but shows distinct (non-zero) expression levels for, respectively, low inducer high cAMP, high inducer low cAMP, and low inducer low cAMP (see Fig. 3b).

They fitted the data to a phenomenological promoter function (see Fig. 1) and obtained a good fit. However, they also showed that this function is quite sensitive to its parameters. They concluded that “the promoter is selected to perform an elaborate computation in setting the transcription rate” (Setty et al., 2003).

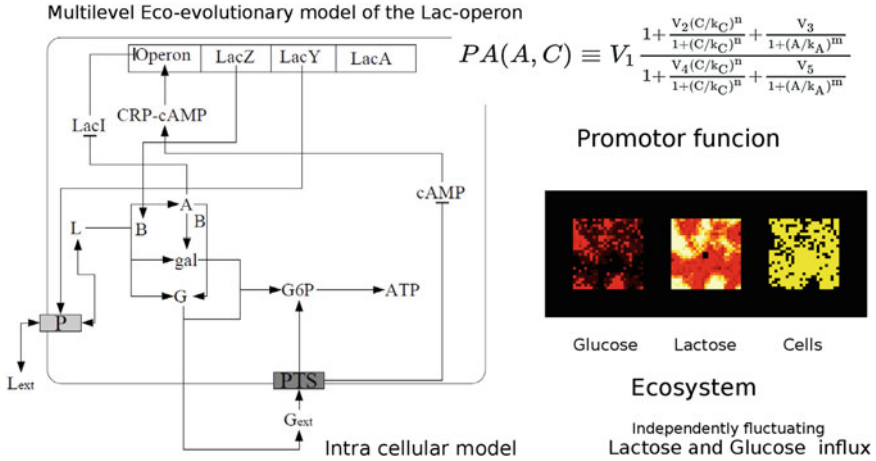


Fig. 1 Overview of the model, which includes the intracellular as well as the intercellular interactions. On the left, the intracellular metabolism and gene regulation related to lactose and glucose utilization are shown as implemented in the model. Expression of the *lac* operon is regulated by the concentration of allolactose (A) and cAMP (C). Allolactose is derived from lactose, and cAMP is inhibited by the influx of glucose in the cell. The expression of the operon is needed for the production of allolactose. It codes for the pump to get lactose into the cell, as well as the enzyme β -galactosidase which transforms lactose to allolactose. Because allolactose induces the *lac* operon, by inhibiting LacI (which inhibits the operon), and needs the expression of the operon to be produced, there is a positive feedback loop which might lead to bistability. On the upper right, the form of the promoter function, dependent on allolactose (A) and cAMP (C), is given as fitted to experimental data by Setty et al. (2003). The V parameters are functions of the following physiological parameters: RNA polymerase and its dissociation constant for binding to the free promoter site and to the site when occupied by CRP (the cAMP-associated transcription factor), as well as its transcription rate dependent on the site occupancy (α and β); the “leakage” of the promoter (γ), i.e., its expression when not induced; the concentration of LacI and its dissociation constant; and the CRP concentration and its dissociation constant. It are these (more physiological) parameters which are subject to mutation and selection in the model; see main text. Finally, in the lower right, the “ecosystem” is depicted, showing the local variation of the external concentration of glucose and lactose and the presence of the cells at arbitrary point in time. For details, see Van Hoek and Hogeweg (2006)

Heeding Dobzhansky’s dictum (Dobzhansky, 1973) we wondered if we could “make sense” of the form of the promoter function from an evolutionary point of view, i.e.:

1. Should we expect such a promoter function to evolve given the known and/or hypothesized details of the metabolic pathways involved?
2. What is the functionality that is in fact being selected?

To answer these questions, we used an evolutionary systems biology approach.

3 Eco-evolutionary Model of the *lac* Operon

We constructed a multilevel agent-based eco-evolutionary model. The model includes the within-cell physiological dynamics related to the *lac* operon, cell growth, and reproduction and competition between cells in a spatially explicit environment, which is modified by the cell metabolism (see Fig. 1).

The model for cell-level physiological dynamics is adapted from the model of (Wong et al., 1997), using their parameter values. However, the promoter function of (Setty et al., 2003) was incorporated, and its parameters were subjected to evolution. The rationale for using fixed parameters for all processes except those of the promoter function was that we wanted to study how the promoter function evolved, given the constraints set by the rest of the system. Important, for example, are the relatively slow protein dynamics.

Thus the model includes the following components (Fig. 1):

- A promoter function. We used the same (phenomenological) function that (Setty et al., 2003) used to (successfully) fit their experimental data (see Fig. 1). The parameters of the function were subject to evolution, i.e., subjected to mutation and selection. Importantly, after initial trials in which “nothing happened” (i.e., no evolutionary adaptation was observed), we realized we should not use the dimension (parameter) reduction used to simplify the model fitting, but the underlying binding reactions instead. This increases the number of parameters from the 7 shown in Fig. 1 to 11 more physiological parameters (see the legend of Fig. 1). Thus we create a redundant genotype-to-phenotype (GP) mapping. Such a redundant GP mapping has been shown to strongly improve evolutionary search. (For a recent extensive review on the role of GP maps in evolution, see Manrubia et al., 2020.)
- Intracellular molecular interactions, including protein expression and degradation, transport into the cell of lactose and glucose, and ATP production, as modeled by (Wong et al., 1997). In addition, the cells grow as a function of ATP production, causing dilution of the protein concentrations. When a cell reaches a certain predefined size, it can divide.
- Ecology: the cells are embedded on a spatial grid. Resources, i.e., lactose and glucose, flux into the medium and are taken up by the cells. The cells compete for the resources as well as empty grid cells. Cells divide after reaching a certain size and die with a probability which depends on the global cell density.

The aim of this evolutionary model is to alleviate the “parameter curse,” inherent in detailed models. Paradoxically, but unavoidably, extra (semi-arbitrary) parameters have to be set in the evolutionary model, in this case, for example, the cost (in terms of ATP) of protein expression and the definition of the environment in which the evolution takes place. The latter involves relative changes in external and internal resource concentration when resources are consumed, as well as the temporal changes of the influx of the resources into the environment. Fortunately, for the environmental parameters, we could use an “adequacy” criterion, i.e., in order

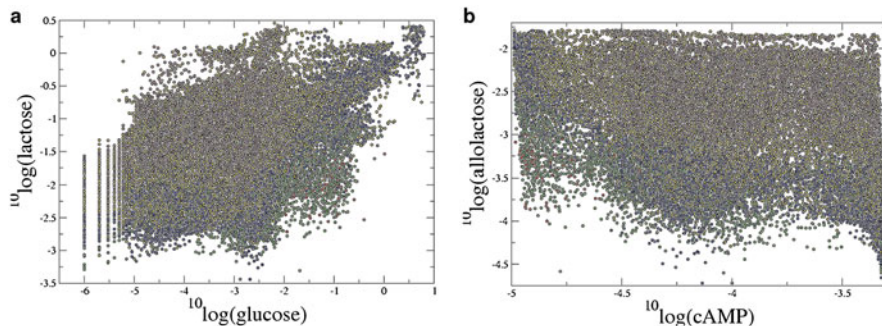


Fig. 2 Coverage of the state space, i.e., the conditions the cells experience over time. On the left, the external conditions, i.e., the concentrations of glucose and lactose, encountered. On the right, the realized internal conditions which directly impact on the expression of the *lac* operon, i.e., the concentrations of cAMP and allolactose. Because of the dynamics of the model, all these concentrations cannot be directly manipulated as input but are the result of the model dynamics. Glucose and lactose were influxed in independently Poisson-distributed blocks of certain duration and concentration. We tuned frequency and amount of influx in such a way that all circumstances were encountered regularly, as shown in the pictures

to select for the full operon function, all combinations of concentrations of glucose and lactose as well as the resulting internal concentrations of allolactose and cAMP should be regularly encountered by the cells. As these concentrations are not directly imposed, but result from the consumption and metabolism, we tuned the timing and amount of influx of glucose and lactose in such a way that this requirement is met (see Fig. 2 for the resulting coverage of the state space).

3.1 Analysis of the Eco-evolutionary Dynamics of the Model

Darwin distinguished “natural selection” from “artificial selection”, where the latter referred to selection by breeders for certain properties preferred by them. The above described eco-evolutionary model of the *lac* operon (artificial as it is) incorporates in this sense “natural selection”: no a priori fitness criterion is defined. Instead the environmental conditions are constantly shifting, not only due to fluctuating external influx of glucose and lactose but importantly also through the current population of cells and the variation of the promoter functions of neighboring cells which defines their uptake of the resources, and therefore the local resource conditions. (Indeed recent experiments have shown the importance of micro-scale gradients in the functioning and evolution of bacterial colonies Dal Co et al., 2019; Van Vliet et al., 2017). These local conditions determine the immediate fitness. Long-term integration of immediate fitness will determine, in the long run, what evolves. This is indeed what evolution is about. However, it makes life harder for the modeler, because there is not one obvious observable (fitness) to evaluate whether or not the

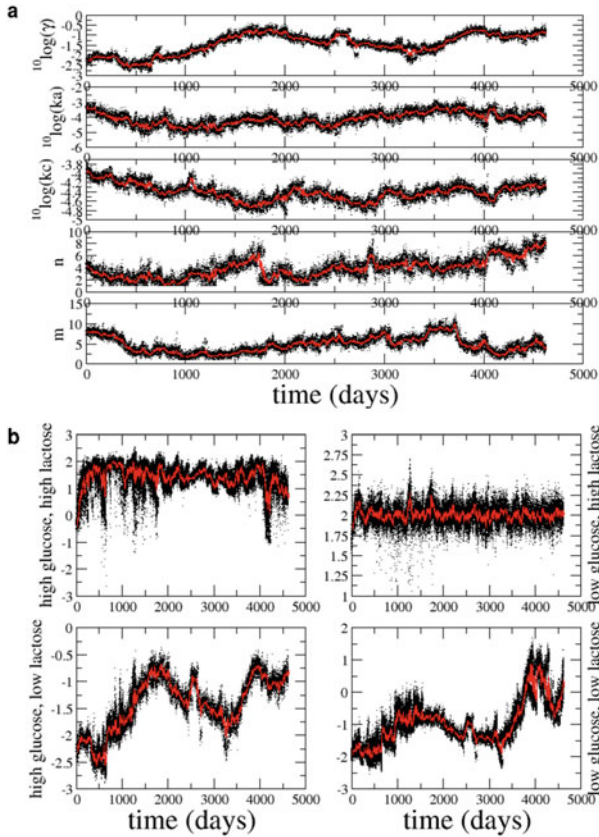


Fig. 3 First observations on the evolutionary dynamics of the model. (a) The evolution through time of individual parameters does not show any clear trend. (b) Depicts the evolution over time of some phenotypic features, namely, the promoter activity the evolving promoter function would have when encountering the four extremes of high and low cAMP and allolactose concentration (although the circumstances in which it finds itself are different). The phenotypic features show a somewhat clearer evolutionary trend, although they also do not convincingly show that anything other than drift occurs

model is actually evolving something. As shown in Fig. 3a, b, looking at changes in the parameter values over time is hardly informative, although looking at the change over time of some selected phenotypic features indicates something beyond neutral drift might be happening. However, further analysis and experiments with the model are needed to establish this, as discussed below.

There is at all times plenty of variation in the population. To get a more detailed understanding of what is evolving, we extracted the last common ancestor of the population at the end of the simulation. This cell obviously was most successful in producing surviving offspring, thus, in hindsight, being per definition the fittest. This works quite well (as shown below), but one should keep in mind that the success of

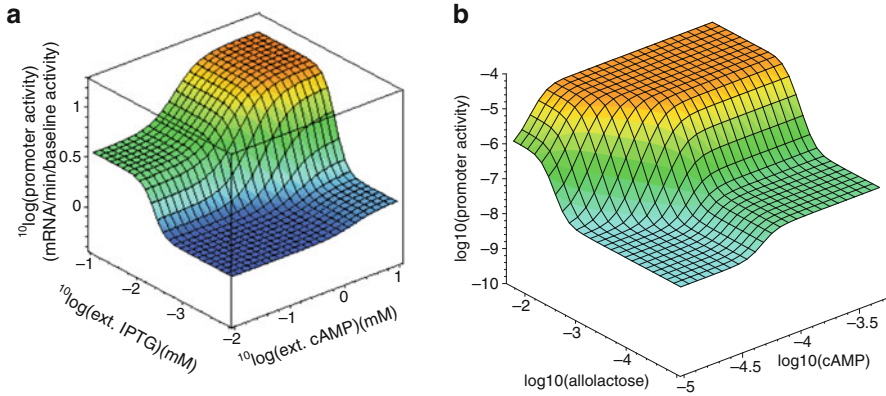


Fig. 4 Similarity of the shape of the promoter function of the best evolved common ancestor (**b**) with the promoter function of Setty et al. (2003); see panel (**a**). Shown is the activity in the state space as experienced by the cells (compare Fig. 3), which is not the same in both cases

the last common ancestor could be caused simply by chance or importantly by later occurring mutations. Several independent evolutionary runs were performed. To select the “best” evolved promoter function, we pairwise competed the last common ancestors of the various runs and selected the one which won most often. This best promoter function is depicted in Fig. 4 alongside the promoter function of (Setty et al., 2003) as fitted to their measured data. The similarity is striking, especially realizing that *no* fitting was involved in setting up the model.

So far so good: apparently the shape of the promoter function as determined in the experiment, with the “fine-tuned parameter values” noted by Setty et al. (2003), is explained by the “natural selection” in our eco-evolutionary model, given the background metabolic processes as modeled previously (Wong et al., 1997). But why? We can now study its behavior in different external resource concentrations. The results are given in Fig. 5a for various concentrations of external lactose, either with high or low glucose concentrations. In contrast to the common expectation that the natural promoter function of the *lac* operon of *E. coli* codes for a bistable switch, the promoter function evolved in our eco-evolutionary model does not, despite its similarity to the measured promoter function. What is wrong?

3.2 Internal Validation of the Model

It turns out that nothing is wrong with the model and that, indeed, the *lac* operon of *E. coli* does not encode a functional bistable switch. This insight is first of all obtained from the model itself. Realizing that experiments were almost always done with artificial inducers (IPTG or TMG), which are not metabolized, we tested our promoter function by stimulating it with IPTG, adjusting the model accordingly. As

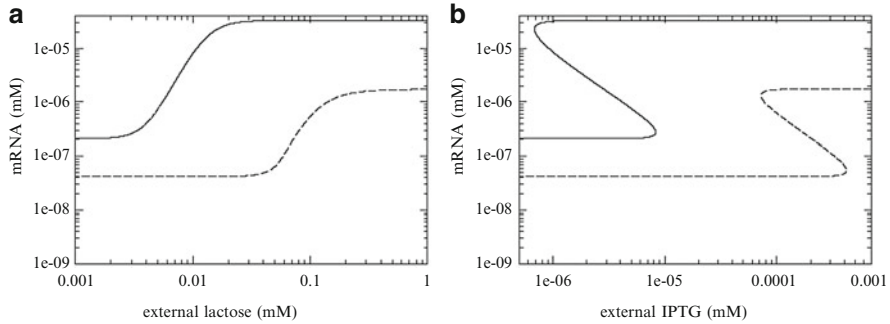


Fig. 5 The best evolved promoter function does not code for a bistable switch in its natural environment and the model, i.e., when it is induced by lactose (a). However, when studied with an artificial inducer (IPTG) as commonly done in the lab, it does code for a bistable switch (b). Solid line, low glucose concentration; dotted line, high glucose concentration (Reproduced from Van Hoek and Hogeweg, 2006)

seen in Fig. 5b, the model with the evolved promoter function in that case recovers a strong bistable switch, both for low and for high concentrations of external glucose. The difference in behavior is due to the fact that artificial inducers are not metabolized, whereas lactose is. This is in fact the advantage of using these artificial inducers in experiments, as it allows to control the conditions. However, because they are not metabolized, the positive feedback loop is strengthened, causing bistability under a much wider set of circumstances than is the case for lactose, which is metabolized (Díaz-Hernández & Santillán, 2010). In other words, we conclude that the common notion that the *lac* operon is coding a bistable switch is an experimental artifact, derived from the preference to do experiments in controlled conditions. As mentioned above, in our eco-evolutionary model, as in nature, conditions are extremely non-controlled.

3.3 Experimental Validation of the Model Results

Currently, the consensus opinion has shifted away from considering the *lac* operon as a bistable switch. Although the model results described above, in my opinion, support this conclusion strongly, the *communis opinio* is based on more recent, more conventional systems biology experiments as well as wet evolutionary experiments. Strikingly, the paper entitled “Multistability in the lactose utilization network of *Escherichia coli*” mentions in passing in the supplementary material “During induction with lactose, as opposed to IPTG, TMG. . . . the steady state distribution after 4 hours of growth is always uni-modal, and we never observe hysteresis” (Ozbudak et al., 2004). Strikingly, despite its title, this paper is frequently cited as evidence for the gradual response instead of bistability.

A combined experimental and modeling paper (Zander et al., 2017) confirmed in their carefully parameterized model that the *lac* operon is not bistable. However, they showed in the model and in experiments that overexpression of LacI (the repressor of the operon which is repressed by the inducer (allolactose or artificial inducer)) does induce bistability. In fact, their results show that the wild-type promoter function is only just not bistable. Similarly, in our model, we see that over evolutionary time some individuals of the variable population do show bistability for lactose. Moreover, we see that when, as we did, evolution starts off with a bistable promoter, it evolves away from bistability by increasing the default expression of the operon without induction (the γ parameter (see Fig. 3a top panel)). Note that decreasing LacI expression implies less repression of the operon and therewith an increase of leakage (i.e., γ). Indeed we showed analytically that the occurrence of bistability (i.e., a Hopf bifurcation) depends primarily on a low enough value of γ (Van Hoek & Hogeweg, 2006). These results suggest that long-term evolution avoids bistability but minimizes the (costly) expression of the operon when not induced as much as possible without becoming bistable.

Interestingly, LacI is in fact itself also regulated by the *lac* operon (Semsey et al., 2013). This autoregulation leads to a further smoothing of the response. This is another indication that avoidance of bistability is an evolved feature.

Also interesting are the results of the evolutionary experiments of (Quan et al., 2012). They evolved *E. coli* cells on four different media, only glucose, only lactose, glucose and lactose, and alternating glucose and lactose, and studied the bistability of the evolved *lac* operons, using artificial inducer (TMG). Even with artificial inducer, they only observed bistability consistently in the glucose-only medium and in a subset of cases in the lactose-only medium. These results, counterintuitive as they may seem at first sight, can be understood in terms of the above discussion. On glucose-only medium, the operon should never be expressed. Therefore, the “leakage” expression without inducer should be low. When this is low enough, bistability is even seen with lactose as inducer (see also Fig. 6b).

Conversely, if they evolved on lactose only, the operon should be active all the time. Whether or not it is bistable under other circumstances is irrelevant. Hence, in some replicates, it remains bistable for artificial inducer as it was the initial wild type, and in other cases, bistability is lost by neutral evolution. In the other two cases, a graded response is observed even for TMG. In contrast to our eco-evolutionary model, where we tuned the parameters so that all environmental conditions were experienced regularly, these evolutionary experiments severely limited the environmental conditions experienced by the cells. Therefore, the cells adapted quickly to the subset of conditions encountered. Likewise when we varied the environmental conditions or internal parameters (e.g., cost of gene expression) and only a subset of conditions occurred, different promoter functions evolved (Van Hoek & Hogeweg, 2006).

Finally I mention the study of (Afroz et al., 2014), who studied bistability for a number of carbon sources. They report no bistability for the *lac* operon, but do find bistability for others, e.g., L-arabinose. It would be interesting to see whether the modeling methodology we used here would for L-arabinose indeed predict bistability.

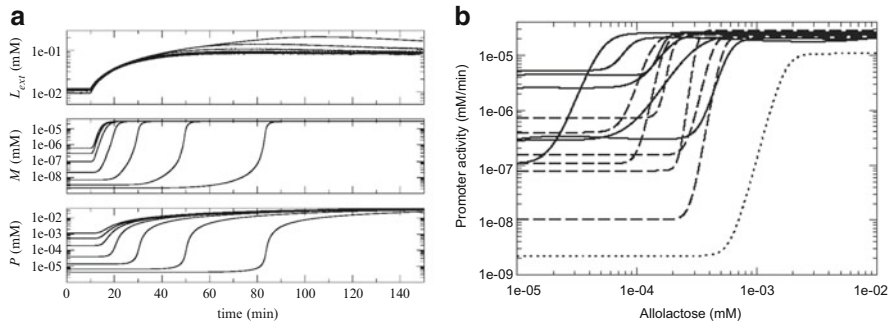


Fig. 6 (a) Delay in activation of the lac operon when external lactose becomes available, dependent on un-induced activity level: delays longer than average division time of *E. coli* for un-induced values compatible with bistability. (b) Evolved promoter activity dependent on internal allolactose in various replicates. Dotted line, initial bistable promoter. Dashed and solid line, evolved promoters in, respectively, the deterministic and the stochastic model. The stochastic model evolves even farther from bistability by increased un-induced expression (Reproduced from Van Hoek and Hogeweg, 2006, 2007)

3.4 Why Avoid Bistability?

The theoretical and experimental results discussed so far show that bistability is evolutionary avoided. But why? When some phenotypic feature evolves, this does not automatically imply that it has an adaptive benefit, as it may be produced by neutral drift. Indeed neutral drift can generically lead to well-defined, apparently nonrandom phenomena which are attractors of the stochastic dynamical system defined by the mutational operators employed, a striking example of which is shown in (Cordero & Hogeweg, 2006). The fact that an evolutionary model and empirical data converge to the same outcome, as is the case in both (Cordero & Hogeweg, 2006) and the current model, also does not preclude a neutral explanation. For example, the avoidance of bistability simply could be due to the fact that a larger part of parameter space generates a graded response, rather than bistability, which is indeed so for the natural system with lactose as inducer, whereas the parameter space leading to bistability is much larger in the case of artificial inducer (as shown here and argued in Savageau, 1999, 2011). Apart from neutral drift, another non-adaptive explanation of an evolved phenotypic feature might be that it is a side effect of the positive selection acting on an apparently unrelated feature, when the same mutations affect both. The above mentioned bistability of the *lac* operon when evolved on glucose medium is a nice example of this. Important for the discussion here is that whether or not the studied feature is generated by adaptive or neutral evolution, or as a side effect, does not affect the main conclusion of this paper, i.e., that an evolutionary perspective is very helpful to debug matching theoretical and experimental results.

In the present case, we see a clearly adaptive signature in the outcome. First of all, competition experiments mentioned above clearly show the fitness advantage of the evolved promoter function, in the type of environment in which it evolved (but note that in competition experiments, the environment unavoidably differs from the native environment because of the presence of the competitor). Another indication of the adaptive relevance of the evolved promoter function is that under different environmental circumstances, clearly different promoter functions evolve, as discussed above in this model as well as in experiments (Quan et al., 2012).

Moreover we can pinpoint why a sufficiently high value of γ and therewith the avoidance of bistability give an evolutionary advantage. In Fig. 6a, we show the onset of mRNA and protein production when external lactose becomes available. For promoters with very low activity when no external lactose is available, the delays are very long, in fact longer than the average division time of *E. coli*. Obviously such long delays are detrimental. Cells which avoid such delays consume the external lactose earlier, leaving less resource for cells with longer delays. Slow protein dynamics plays an important role in these long delays. This underscores the importance of using the large-scale parameterized model for the cell metabolism in our eco-evolutionary model. Note that bistability, and therewith hysteresis, would even further aggravate the delays. We conclude that it is the transient, non-equilibrium situation which determines the long-term evolutionary outcome.

In contrast, an earlier evolutionary model “explained” the advantages of bistability (Thattai & Van Oudenaarden, 2004). In their mini-model, instantaneous switching was assumed. In addition, their model was stochastic. Bistability ensured heterogeneity in the population, so that some cells were pre-adapted to a changing environment. This raises the question whether it is the lack of stochasticity in gene expression in our model which prevents evolution to exploit the advantages of bistability. We modified the model to incorporate stochastic gene expression (Van Hoek & Hogeweg, 2007) and conducted a similar set of experiments in the stochastic model. Figure 6b shows that the stochastic model evolved even farther away from bistability by increasing the expression in the absence of lactose (i.e., γ). Indeed, again the explanation is in terms of delays, which are even more severe in the stochastic model (Van Hoek & Hogeweg, 2007). Moreover, the stochasticity only marginally increases the heterogeneity of the population, relative to the genetic and environmental heterogeneity prevalent in the eco-evolutionary model (Van Hoek & Hogeweg, 2007). Likewise through metagenomic analysis, extreme heterogeneity is commonly observed in natural bacterial populations at a micro-scale (e.g., Preheim et al., 2011; Vetsigian et al., 2011).

4 Discussion

The evolutionary systems biology approach discussed in this paper proved to be surprisingly powerful. We showed that the measured promoter function (Setty et al., 2003) was evolutionary favored, which was our original aim. The modeling

was moreover rich in “results++”. i.e., in unexpected novel insights. Foremost the insight emerged that the assumption of bistability of the *lac* operon, which was supported by experiments, mini-models, large-scale models, and evolutionary optimization models, is in fact incorrect. How could this false notion be sustained for so many years? The need for well-defined conditions in experiments, and therewith the use of artificial inducer, clearly was the primary cause, combined with the construction of models and the setting of parameters such that the model results match the experimental results. In contrast, we used an unsupervised modeling approach and observed a striking match to some experimental results, namely, the shape of the promoter function together with the totally unexpected evolutionary trend away from bistability. Moreover, we could determine that bistability was avoided in order to avoid delays in activation and therewith gain a competitive edge. Our results also indicate that examining monomorphic, clonal populations in experiments or models may lead to artifacts, in the sense that it does not reflect what happens in natural populations.

Having argued that the *lac* operon does not encode a bistable switch, we should reflect on what we mean with such a statement. Bacteria, including *E. coli*, adapt to a prevailing environment very quickly, as shown in evolutionary experiments, e.g., those of (Quan et al., 2012) discussed above, and stressed by (e.g., Dekel and Alon, 2005). This we also see in our model: if the environment switches too fast, regulation is largely lost, as only an average environment is experienced by the cells. As another example, when cost of protein expression is set very high, bistability may evolve but occurs at very high glucose concentrations, which were very seldom if ever encountered. In such a case, like in the glucose-only environment of (Quan et al., 2012), bistability occurs as a side effect which does not harm the system. Thus, indeed as stated by Setty et. al., the promoter function can be fine-tuned easily. It is therefore even more remarkable that, given that a full set of environmental conditions is encountered (which is not the case in the abovementioned examples), evolution of wild-type *E. coli* and the model converge to an unequivocal solution.

In the eco-evolutionary model discussed here, we only evolved some of the large number of parameter values which needed to be specified, and for which the experimental evidence is not unequivocal. However, because of the evolution of the parameters determining the phenomena of interest, their precise value might not matter too much and certainly was not tuned/fitted for the results obtained. In that sense, the parameter curse which encumbers large-scale models was somewhat alleviated. This was enough to debug the results obtained from models in which parameters were fitted or tuned to match the experimental results.

Finally I like to note that the general approach advocated here, i.e., non-supervised, multilevel eco-evolutionary modeling, can be generalized beyond evolving parameters in a fixed model structure as done here. Giving the models many degrees of freedoms to adjust model structure, we have repeatedly seen surprising convergence to biological systems, leading to novel insights in their functioning as well as novel insights in evolution itself (e.g., Cuypers & Hogeweg, 2012, 2014; van Dijk et al., 2019; Van Hoek & Hogeweg, 2009).

References

- Afroz, T., Biliouris, K., Kaznessis, Y., & Beisel, C. L. (2014). Bacterial sugar utilization gives rise to distinct single-cell behaviours. *Molecular Microbiology*, 93(6), 1093–1103.
- Cordero, O. X., & Hogeweg, P. (2006). Feed-forward loop circuits as a side effect of genome evolution. *Molecular Biology and Evolution*, 23(10), 1931–1936.
- Crick, F. (1988). What mad pursuit: a personal view of scientific discovery. *New York: Basic Books*, 13(182), 8.
- Cuypers, T. D., & Hogeweg, P. (2012). Virtual genomes in flux: an interplay of neutrality and adaptability explains genome expansion and streamlining. *Genome Biology and Evolution*, 4(3), 212–229.
- Cuypers, T. D., & Hogeweg, P. (2014). A synergism between adaptive effects and evolvability drives whole genome duplication to fixation. *PLoS Computational Biology*, 10(4), e1003547.
- Dal Co, A., Van Vliet, S., & Ackermann, M. (2019). Emergent microscale gradients give rise to metabolic cross-feeding and antibiotic tolerance in clonal bacterial populations. *Philosophical Transactions of the Royal Society B*, 374(1786), 20190080.
- Dekel, E., & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050), 588–592.
- Díaz-Hernández, O., & Santillán, M. (2010). Bistable behavior of the lac operon in *E. coli* when induced with a mixture of lactose and TMG. *Frontiers in Physiology*, 1, 158.
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, 35(3), 125–129.
- Griffith, J. (1968). Mathematics of cellular control processes I. Negative feedback to one gene. *Journal of Theoretical Biology*, 20(2), 202–208.
- Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3), 318–356.
- Manrubia, S., Cuesta, J. A., Aguirre, J., Ahnert, S. E., Altenberg, L., Cano, A. V., Catalán, P., Díaz-Uriarte, R., Elena, S. F., García-Martín, J. A., et al. (2020). From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. arXiv preprint arXiv:2002.00363.
- Novick, A., & Weiner, M. (1957). Enzyme induction as an all-or-none phenomenon. *Proceedings of the National Academy of Sciences of the United States of America*, 43(7), 553.
- Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., & Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, 427(6976), 737.
- Preheim, S. P., Timberlake, S., & Polz, M. F. (2011). Merging taxonomy with ecological population prediction in a case study of vibronaceae. *Applied and Environmental Microbiology*, 77(20), 7195–7206.
- Quan, S., Ray, J. C. J., Kwota, Z., Duong, T., Balázs, G., Cooper, T. F., & Monds, R. D. (2012). Adaptive evolution of the lactose utilization network in experimentally evolved populations of *Escherichia coli*. *PLoS Genetics*, 8(1), e1002444.
- Rao, C. V., & Koirala, S. (2014). Black and white with some shades of grey: the diverse responses of inducible metabolic pathways in *Escherichia coli*. *Molecular Microbiology*, 93(6), 1079–1083.
- Savageau, M. (1999). Design of gene circuitry by natural selection: analysis of the lactose catabolic system in *Escherichia coli*. *Biochemical Society Transactions*, 27(2), 264–270.
- Savageau, M. A. (2011). Design of the lac gene circuit revisited. *Mathematical Biosciences*, 231(1), 19–38.
- Semsey, S., Jauffred, L., Csiszovszki, Z., Erdőssy, J., Stéger, V., Hansen, S., & Krishna, S. (2013). The effect of lacI autoregulation on the performance of the lactose utilization system in *Escherichia coli*. *Nucleic Acids Research*, 41(13), 6381–6390.
- Setty, Y., Mayo, A. E., Surette, M. G., & Alon, U. (2003). Detailed map of a CIS-regulatory input function. *Proceedings of the National Academy of Sciences*, 100(13), 7702–7707.

- Thattai, M., & Van Oudenaarden, A. (2004). Stochastic gene expression in fluctuating environments. *Genetics*, *167*(1), 523–530.
- van Dijk, B., Meijer, J., Cuypers, T. D., & Hogeweg, P. (2019). Trusting the hand that feeds: Microbes evolve to anticipate a serial transfer protocol as individuals or collectives. *BMC Evolutionary Biology*, *19*(1), 201.
- Van Hoek, M., & Hogeweg, P. (2006). In silico evolved lac operons exhibit bistability for artificial inducers, but not for lactose. *Biophysical Journal*, *91*(8), 2833–2843.
- Van Hoek, M., & Hogeweg, P. (2007). The effect of stochasticity on the lac operon: An evolutionary perspective. *PLoS Computational Biology*, *3*(6), e111.
- Van Hoek, M. J., & Hogeweg, P. (2009). Metabolic adaptation after whole genome duplication. *Molecular Biology and Evolution*, *26*(11), 2441–2453.
- Van Vliet, S., Winkler, A. R., Spriewald, S., Stecher, B., Ackermann, M., et al. (2017). Local interactions lead to spatially correlated gene expression levels in bacterial groups. *BioRxiv* 109991.
- Vetsigian, K., Jajoo, R., & Kishony, R. (2011). Structure and evolution of *Streptomyces* interaction networks in soil and in silico. *PLoS Biology*, *9*(10), e1001184.
- Wong, P., Gladney, S., & Keasling, J. D. (1997). Mathematical model of the lac operon: Inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnology Progress*, *13*(2), 132–143.
- Yildirim, N., & Mackey, M. C. (2003). Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data. *Biophysical Journal*, *84*(5), 2841–2851.
- Zander, D., Samaga, D., Straube, R., & Bettenbrock, K. (2017). Bistability and nonmonotonic induction of the lac operon in the natural lactose uptake system. *Biophysical Journal*, *112*(9), 1984–1996.