ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Reliable Prediction of Complex Phenotypes from a Modular Design in Free Energy Space: An Extensive Exploration of the *lac* Operon

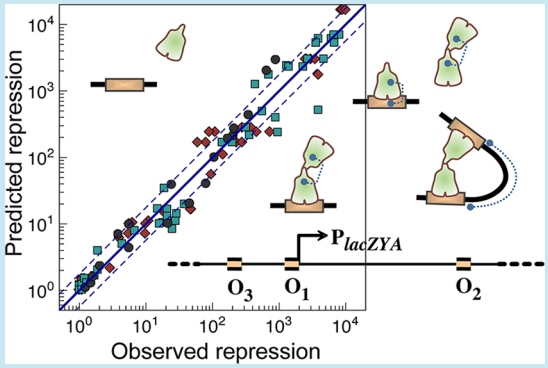
Jose M. G. Vilar*[,†,‡] and Leonor Saiz*[,§]

[†]Biophysics Unit (CSIC-UPV/EHU) and Department of Biochemistry and Molecular Biology, University of the Basque Country, P.O. Box 644, 48080 Bilbao, Spain

[‡]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

[§]Department of Biomedical Engineering, University of California, 451 E. Health Sciences Drive, Davis, California 95616, United States

ABSTRACT: The basic methodology for designing, altering, and constructing biological systems is increasingly relying on well-established engineering principles to move forward from trial and error approaches to reliably predicting the system behavior from the properties of the components and their interactions. The inherent complexity of even the simplest biological systems, however, often precludes achieving such predictive power. A prototypical example is the *lac* operon, one of the best-characterized genetic systems, which still poses serious challenges for understanding the results of combining its parts into novel setups. The reason is the pervasive complex hierarchy of events involved in gene regulation that extend from specific protein−DNA interactions to the combinatorial assembly of nucleoprotein complexes. Here, we integrate such complexity into a few-parameter model to accurately predict gene expression from a few simple rules to connect the parts. The model accurately reproduces the observed transcriptional activity of the *lac* operon over a 10,000-fold range for 21 different operator setups, different repressor concentrations, and tetrameric and dimeric forms of the repressor. Incorporation of the calibrated model into more complex scenarios accurately captures the induction curves for key operator configurations and the temporal evolution of the β-galactosidase activity of cell populations.



KEYWORDS: *lac operon, gene regulation, transcription, repressor, computational modeling, free energy*

Understanding the functioning of biological systems beyond their natural original context is becoming increasingly important for many fundamental applications, stretching from engineering novel molecular functions at the cellular level[1−8] to affecting ecosystem interactions by genetically modifying organisms.[9−13] At the most fundamental level, the rules and general principles needed to obtain a controlled cellular behavior from the interactions of the components are still far from being completely clear. Questions as basic as the effects of targeted mutations, the results of swapping DNA regions, and the consequences of over-expressing key proteins have not yet reliable answers in many occasions.[4,5,14]

The *lac* operon, the *E. coli* genetic system that regulates and produces the enzymes needed to metabolize lactose,[15] provides the proverbial example of such complex behavior resulting from an apparently simple system. The basic model was proposed by Jacob and Monod over 50 years ago in a now classic paper that put forward the very basic principles of gene regulation.[16] They postulated the existence of molecules that bind to specific sites in nucleic acids to control the expression of genes. In the case of the *lac* operon, the original idea of the *lac* repressor preventing transcription upon binding to the operator DNA in the promoter region has been refined over the years[15] to incorporate a complex hierarchy of events that extend from specific protein−DNA interactions to the combinatorial assembly of nucleoprotein complexes.

The complexity of the molecular interactions in the control of the response to lactose is already evident in the mode of functioning of the *lac* repressor,[17] which upon binding to O_1, the main operator, prevents the RNA polymerase from binding to the promoter and transcribing three genes used in lactose metabolism. There are also two auxiliary operators, O_2 and O_3, where the repressor can bind specifically without preventing transcription (Figure 1). These two additional sites are orders of magnitude weaker than the main site and by themselves do not affect transcription substantially. In combination with the main site, however, they can increase repression of transcription by a factor of almost 100.[18−20] This counterintuitive effect results from the *lac* repressor binding simultaneously as a bidentate tetramer to two operators and looping the intervening DNA. Thus, the main operator and at least one auxiliary operator provide a mechanism to form DNA loops

Received: February 14, 2013
Published: May 8, 2013

 © 2013 American Chemical Society

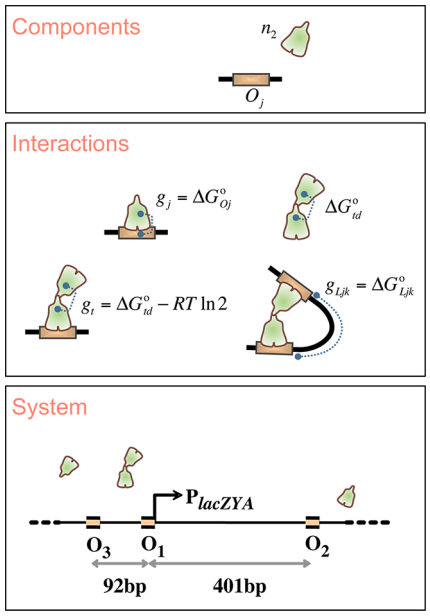


Figure 1. Modular deconstruction of the *lac* operon. Components: the basic modules are the *lac* repressor dimers and operators. Interactions: repressor dimers interact with other dimers and with operator DNA to form tetramers in solution, to bind single DNA sites as dimers or tetramers, and to loop DNA as tetramers bound to two DNA sites. These elementary interactions can be characterized to obtain the system behavior. System: The main, O_1, and the two auxiliary, O_2 and O_3, operators are shown as black rectangles on the black line representing DNA. Binding of the *lac* repressor to O_1 prevents transcription of the three *lacZYA* genes.

that substantially increase the ability of the repressor to bind the main operator.

This type of behavior with oligomeric transcription factors that can bind simultaneously single and multiple DNA sites is a recurrent theme in gene regulation to the extent that transcription regulation through DNA looping is nowadays considered to be the rule rather than the exception.[21−24] It is present in many bacterial operons such as *ara*, *gal*, and *deo* operons[25−27] and in bacteriophages such as phage *λ*.[28,29] DNA looping plays an important role in mediating long-range interactions because it allows proteins bound to nonadjacent DNA sites to come close to each other.[30] This strategy is widely used in eukaryotic enhancers,[31] as in the case of the interactions between enhancers and promoters mediated by androgen[32] and progesterone[33] receptors, to integrate multiple signals into the control of the transcriptional machinery.[31] It is also present in the tumor suppressor p53,[34] the nuclear factor κB (NF-κB),[35] the signal transducers and activators of transcription (STATs),[36] the octamer-binding proteins (Oct),[37,38] and the retinoid nuclear hormone receptor RXR.[23,39]

We focus on the *lac* operon because it embodies the core elements present across many levels of transcription regulation, offers the possibility of considering the actual mode of binding and regulation, and provides a platform with substantial amounts of experimental data to contrast the hypothesis and results of the model.

Current biophysical models of the *lac* operon have taken into account the effects of the tetrameric *lac* repressor's binding to two operators simultaneously[22,40,41] and to the three operators together with the effects of the Catabolite Activator Protein (CAP).[42,43] The specific protein−DNA interactions have been worked out up to atomistic level of detail, both experimentally and computationally,[17,44,45] and the kinetics of the *lac* repressor has been probed at the single-molecule level in living cells.[46−48] There are also phenomenological analyses of the *lac* operon induction[49−51] that can reproduce diverse experimental observations.

Despite all of these efforts, there is no quantitative gene expression model yet that takes into account the coexistence of the different oligomerization states of the *lac* repressor and the effects of adjusting the equilibrium between the tetrameric form, representative of the wild-type (WT) system, and the dimeric form, which is the dominant species in several mutants and repressors tagged with fluorescent proteins. Accounting for the different oligomeric states of the repressor is needed to accurately reproduce the available experimental results of taking apart the different regulatory components of the *lac* operon down to its basic elements and putting them back together with their mutated versions into different arrangements. This task is challenging in several fronts. First, it requires an efficient approach to connect the parts as a system to avoid getting into the combinatorial complexity problem, in which the number of potential states of the system grows exponentially with the number of components. Second, the increase in components leads also to an increase in the number of parameters, but many of these parameters are thermodynamically related to each other. Finally, the values of the parameters might be different under different experimental conditions.

Here, we elucidate biophysical principles that provide an efficient avenue to integrate the prototypical complex interactions of transcription regulation into a few-parameter model for accurately predicting gene expression from a few simple rules to connect the parts. The key idea is to use a modular design in free energy space and a decomposition of the free energy of the different states into additive contributions of the interactions. In this way, the whole system can be characterized by a few parameters directly connected to the experimental data. We consider as elementary components *lac* repressor dimers and operator sequences. The behavior of the system is obtained starting off from the dimer assembly into tetramers, binding of dimers and tetramers to the different operators, and looping of DNA by the simultaneous binding of a bidentate tetrameric repressor to two operators (Figure 1).

RESULTS AND DISCUSSION

Description of the Components and Their Elementary Interactions. The *lac* repressor is a tetramer consisting of two dimers with one DNA binding domain each. We consider explicitly the concentration of both tetrameric, $[n_4]$, and dimeric, $[n_2]$, species. Their equilibrium values are related to each other through

$$[n_2]^2/[n_4] = K_{td} = e^{\Delta G^\circ_{td}/RT} \quad (1)$$

where RT is the gas constant times the absolute temperature, K_{td} is the tetramer-dimer dissociation constant, and $\Delta G^\circ_{td} = RT \ln K_{td}$ is its corresponding free energy. Because dimerization is very strong, the concentration of free repressor monomers is negligible and the conservation of mass can be written as the total repressor monomer subunit concentration from dimers and tetramers given by $[n_T] = 4[n_4] + 2[n_2]$.

Binding of a dimer to an operator O_j is described by

$$[n_2][O_j]/[n_2O_j] = K_{Oj} = e^{\Delta G^\circ_{Oj}/RT} \quad (2)$$

dx.doi.org/10.1021/sb400013w | ACS Synth. Biol. 2013, 2, 576−586

where n_2O_j represents the dimer−operator complex, and K_{Oj} is the corresponding dissociation constant with its associated free energy $\Delta G^\circ_{Oj} = RT \ln K_{Oj}$.

The presence of a tetramer bound to an operator, forming the complex n_4O_j, can be described in terms of its constituent subunits as the binding of a dimer to a dimer bound to an operator:

$$[n_2][n_2O_j]/[n_4O_j] = K_{td}/2 = e^{(\Delta G^\circ_{td} - RT \ln 2)/RT} \quad (3)$$

Here, we have used the corresponding dissociation constant of $K_{td}/2$ rather than just K_{td} as in eq 1. The difference by a factor 2 reflects that the probability for a bimolecular reaction to happen is proportional to the number of pairs of possible interactions between the molecules present and that the term $[n_2]^2$ in eq 1, a homodimerization reaction, is counting those interactions twice with respect to the term $[n_2][n_2O_j]$ in eq 3, a heterodimerization reaction.[52]

The formation of a DNA loop by a tetramer bound to the operators O_j and O_k can be interpreted as the formation of a tetramer on looped DNA from two dimers already bound to two operators:

$$[n_2O_kn_2O_j]/[O_kn_4O_j] = K_{td}/2[n_{Ljk}] = e^{(\Delta G^\circ_{Ljk} + \Delta G^\circ_{td} - RT \ln 2)/RT} \quad (4)$$

where $[n_2O_kn_2O_j]$ and $[O_kn_4O_j]$ are the concentration of the nonlooped and looped complexes, respectively. Here, $[n_{Ljk}]$ is the local concentration of one dimer with respect to another when they are both bound to the operators.[53] This quantity is also known as j factor[54] and is related to the free energy of looping through $\Delta G^\circ_{Ljk} = -RT \ln[n_{Ljk}]$.[41]

Description of Multiple Interactions. All of these elementary interactions can be combined to take into account the three operators for specific binding of *lac* repressor dimers and tetramers as well as the formation of different DNA loops. We describe the state of the system mathematically using a set s of state variables that take a value of 1 if an interaction is present and of 0 if it is absent.[55,56]

The advantage of using free energies is that the total free energy of a state $\Delta G(s)$ can be obtained as the sum of the contributions of the interactions among the components and that the probability P_s of any of these complex states is determined by its free energy. According to statistical thermodynamics[57,58] both quantities are related through

$$P_s = \frac{e^{-\Delta G(s)/RT}}{Z} \quad (5)$$

where the partition function, $Z = \Sigma_s\, e^{-\Delta G(s)/RT}$, is used as normalization factor. Note that this expression uses free energy changes, $\Delta G(s)$, which in contrast to standard free energy changes, $\Delta G^\circ(s)$, include the dependence on the different species concentration. Equation 5 is equivalently written in terms of standard free energies as $P_s = [n_2]^{d(s)}\, e^{-\Delta G^\circ(s)/RT}/Z$, where $d(s)$ is the number of dimers in the state s, including dimers bound to DNA and dimers forming a tetramer bound to DNA. The thermodynamic approach has long been successfully applied to prokaryotic systems[40,43,58,59] and more recently to eukaryotic systems.[23,60] The widespread success of the thermodynamic approach indicates that very often it is sufficient to consider just key control points that rely on protein−DNA and protein−protein interactions and that other intricate biochemical processes, such as phosphorylation and DNA modifications, are downstream events.

Using this approach with state variables, the free energy of the protein−DNA system is expressed as

$$\begin{aligned}\Delta G(s) &= (m + g_1)s_1 + (m + g_2)s_2 + (m + g_3)s_3 \\ &\quad + (m + g_t)s_1s_{T1} + (m + g_t)s_2s_{T2} + (m + g_t)s_3s_{T3} \\ &\quad + (g_{L12} + g_t)s_{L12} + (g_{L13} + g_t)s_{L13} \\ &\quad + (g_{L23} + g_t)s_{L23} \\ &\quad + \infty C(s)\end{aligned} \quad (6)$$

with $m = -RT \ln[n_2]$ and $g_t = \Delta G^\circ_{td} - RT \ln 2$. The state variables s_1, s_2, and s_3 indicate whether (= 1) or not (= 0) a repressor dimer is bound at the position of O_1, O_2, and O_3, respectively; s_{T1}, s_{T2}, and s_{T3} indicate whether (= 1) or not (= 0) a dimer binds to a dimer bound at the position of O_1, O_2, and O_3, respectively, to form a tetramer; and s_{L12}, s_{L13}, and s_{L23} indicate whether (= 1) or not (= 0) DNA forms the loops O_1-O_2, O_1-O_3, and O_2-O_3, respectively. The subscripts of the different contributions to the free energy have the same meaning as those of the corresponding state variables. For the looping contributions, we always have $g_{Ljk} = \Delta G^\circ_{Ljk}$ because the distance and sequence between operators is never changed. For the dimer-operator binding contributions, the standard free energies of binding g_1, g_2, and g_3 depend on the specific operator sequence that is represent at the position of O_1, O_2, and O_3, respectively. For instance, we have $g_1 = \Delta G^\circ_{O1}$ for the WT case, but we have $g_1 = \Delta G^\circ_{O2}$ for a mutant with the sequence of O_1 replaced by the sequence of O_2.

The infinity in the last term of the free energy, with

$$\begin{aligned}C(s) &= (1 - s_1)s_{T1} + (1 - s_2)s_{T2} + (1 - s_3)s_{T3} \\ &\quad + s_{L12}(1 - s_1s_2) + s_{L13}(1 - s_1s_3) + s_{L23}(1 - s_2s_3) \\ &\quad + s_{L12}(s_{T1} + s_{T2}) + s_{L13}(s_{T1} + s_{T3}) \\ &\quad + s_{L23}(s_{T2} + s_{T3}) \\ &\quad + s_{L12}s_{L13} + s_{L12}s_{L23} + s_{L13}s_{L23}\end{aligned} \quad (7)$$

is used for convenience purposes to assign an infinite free energy value, and therefore zero probability, to the values of the state variables s that are not physically possible.[42] It explicitly implements that a dimer can form a tetramer on DNA only if there is a dimer bound to DNA, that a DNA loop can be formed only if there is a pair of dimers bound to the respective operators and if none of these dimers is bound by another dimer, and that two loops that share one operator cannot be present simultaneously.

Transcriptional Control. Transcription in the *lac* operon proceeds at three different rates depending on the occupancy of the operators. It is completely abolished when the repressor is bound to O_1. If O_1 and O_3 are free, transcription takes place at an activated maximum rate, Γ_{max}. Activation results from CAP, bound between O_1 and O_3, enhancing the binding of the RNA polymerase to the promoter. If O_1 is free but the repressor occupies O_3, transcription takes place at a basal reduced rate, $\chi\Gamma_{max}$. This reduction by a factor χ arises because binding of the repressor to O_3 prevents CAP from contacting the RNA polymerase and activating transcription.[15,42]

dx.doi.org/10.1021/sb400013w | ACS Synth. Biol. 2013, 2, 576−586

The observed transcription rate can be expressed in terms of state variables as

$$\Gamma(s) = \Gamma_{max}(1 - s_1)(\chi s_3 + (1 - s_3)) \quad (8)$$

With this approach, the effective transcription rate, $\bar{\Gamma}$, is obtained by computing the thermodynamic average over all of the representative states; namely, by performing the sum above over all possible combinations of values of s:

$$\bar{\Gamma} = \sum_s \Gamma(s) P_S = \frac{1}{Z} \sum_s \Gamma(s)\, e^{-\Delta G(s)/RT} \quad (9)$$

The repression level, Ω, is defined as the maximum transcription over the effective transcription, which in mathematical terms is given by $\Omega = \Gamma_{max}/\bar{\Gamma}$.

Induction. Induction of the *lac* operon takes place when inducers, such as allolactose and isopropyl-β-D-thiogalactoside (IPTG), bind the repressor and significantly decrease its specific binding for the operators. The inducer binds to a pocket in each monomeric subunit of the repressor, resulting in a coexisting population of repressor dimers without inducer, n_2^0, with one inducer molecule, n_2^1, and with two inducer molecules, n_2^2. Because the inducer's binding to one monomeric subunit does not affect its binding to the other one, the different dimer concentrations are related to each other through $[n_2^0][I]/[n_2^1] = K_I/2$ and $[n_2^1][I]/[n_2^2] = K_I$, where $[I]$ is the inducer concentration and K_I is repressor-inducer dissociation constant. Binding of these dimers to an operator O_j is described as usual by $[n_2^0][O_j]/[n_2^0 O_j] = K_{Oj}$, $[n_2^1][O_j]/[n_2^1 O_j] = K_{Oj1}$, and $[n_2^2][O_j]/[n_2^2 O_j] = K_{Oj2}$. Here K_{Oj1} and K_{Oj2} are the dissociation constants for dimers with one and two inducer molecules, respectively, which are much larger than K_{Oj} for typical inducers.

The occupancy of the operator can be expressed in terms of the total concentration of dimers, $[n_2] = [n_2^0] + [n_2^1] + [n_2^2]$, and dimer-operator complexes, $[n_2 O_j] = [n_2^0 O_j] + [n_2^1 O_j] + [n_2^2 O_j]$, as

$$\frac{[n_2][O_j]}{[n_2 O_j]} = \frac{(1 + 2[I]/K_I + [I]^2/K_I^2)}{(1/K_{Oj} + 2[I]/K_{Oj1}K_I + [I]^2/K_{Oj2}K_I^2)} \equiv K_{Oj}([I]) = e^{\Delta G^\circ_{Oj}([I])/RT} \quad (10)$$

where $K_{Oj}([I])$ is an inducer-dependent effective dissociation constant, and $\Delta G^\circ_{Oj}([I]) = RT \ln K_{Oj}([I])$ is its corresponding free energy. Thus, the effects of the inducer can be fully taken into account just by changing one element of the modular approach.

Model Parameters. The parameters of the overall model include *lac* repressor-operator binding free energies (as many as different operator sequences), three free energies of looping (one for each possible loop), the extent of transcription activation by CAP, and the *lac* repressor tetramer-dimer dissociation constant. In principle, it is possible to consider all of these parameters explicitly and fit the model to the available experimental data. Here we will follow a more effective approach that reduces the number of free parameters to 6 and obtains their values from just 6 experimental data points without fitting. The key idea is to consider specific cases in which the values of these parameters can be obtained explicitly as a function of the experimental repression level values by inverting the model. The same approach is applied subsequently to infer the repressor-inducer dissociation constant, which is the only additional parameter needed to accurately predict the induction curves.

Standard Free Energies of Dimer-Operator Binding. Experimental data is available for setups with the two auxiliary operators deleted for strains with the dimeric form of the repressor. Under these conditions, the standard free energy of binding can be obtained straightforwardly by inverting the expression of the repression level, which leads to

$$g_1 = -RT \ln \frac{2(\Omega - 1)}{[n_T]} \quad (11)$$

This expression results from the general model by considering the limit of the two free energies of binding of the deletions and the free energy of tetramerization going to infinity ($g_2 \to \infty$, $g_3 \to \infty$, and $\Delta G^\circ_{td} \to \infty$).

The experimental data included cases with the WT sequence of O_1 as well as with the replacement of O_1 by the sequence of O_2 and O_3. Therefore, eq 11 can be used to obtain the free energies of binding of the repressor dimer to the three wild type operators.

We use the values for O_1 and O_3 to estimate the standard free energies of mutated operators, ΔG°_{Oj}, from their sequence using their position weight matrix (PWM) scores, S_{Oj}. Explicitly,

$$\Delta G^\circ_{Oj} = \frac{\Delta G^\circ_{O1} - \Delta G^\circ_{O3}}{S_{O1} - S_{O3}}(S_{Oj} - S_{O1}) + \Delta G^\circ_{O1} \quad (12)$$

This expression assumes that scores and free energies are linearly related and that the values obtained with both approaches are the same for O_1 and O_3. The PWM was obtained from the sequences of the three WT operators O_1, O_2, and O_3 as described in ref 61. The scores correctly ranked the three WT operators according to their measured strength and consistently ranked all of the deletions below all of the WT operators (Table 1).

Table 1. Operator Sequences and Their Binding Properties[a]

O_j	Sequence	ΔG°_{Oj}	K_{Oj}
O_1	AATTGTGAGCGGATAACAATT	**−12.50**	**0.90**
O_1-a	gATTGTtAGCGGAgAAgAATT	−6.37	2.4×10^4
O_1-b	gAactacAtCctccgctAggT	12.42	9.8×10^{17}
O_1-c	AATTGTTAGCGGAGAAGAATT	−6.95	9.3×10^3
O_2	AAaTGTGAGCGagTAACAAcc	**−11.07**	**9.7**
O_2-a	gAatGTtAatGaATAgCAccc	−2.23	2.4×10^7
O_2-b	GAAGGTTAATGAATAGCACCC	−0.66	3.3×10^8
O_3	ggcaGTGAGCGcAacgCAATT	**−9.01**	**299**
O_3-a	AAcctcGAGCtcAacgCAATT	−3.22	4.6×10^6
O_3-b	TCGATCGAGCTCAACGCAATT	−0.07	8.4×10^8

[a]The standard free energy of binding ΔG°_{Oj} (in kcal/mol) of the *lac* repressor dimer to the operator O_j was computed using eq 11 for O_1, O_2, and O_3 and using eq 12 for the operator deletions. The corresponding dissociation constant K_{Oj} (in nM) is shown for convenience.

Activation by CAP. Similarly, when O_1 and O_2 are deleted, we can obtain the extent of activation by CAP from just the repression level and the free energy of binding: $1/\chi = [n_T]\Omega/([n_T] - 2(\Omega - 1)\, e^{g_3/RT})$, which results from the general model by considering the limit of the two free energies of binding of the deletions going to infinity. Assuming that in addition the repressor concentration is high, it simplifies to

dx.doi.org/10.1021/sb400013w | ACS Synth. Biol. 2013, 2, 576−586

$$\chi = \frac{1}{\Omega} \tag{13}$$

Therefore, the extent of activation by CAP, $1/\chi$, is just the repression level Ω for the case in which $g_2 \to \infty$, $g_3 \to \infty$, and $[n_T] \to \infty$.

Free Energy of Looping. The free energies of looping are obtained from the repression level with just one operator deleted and the parameters obtained previously by considering the general expression with the appropriate limit for the case with just the tetrameric form of the repressor ($\Delta G^\circ_{td} \to -\infty$).

The case of the O_1-O_2 loop is obtained in the limit of the free energy of binding to O_3 going to infinity ($g_3 \to \infty$) as

$$g_{L12} = -RT \ln \frac{(2\, e^{g_2/RT} + [n_T])(2(\Omega - 1)\, e^{g_1/RT} - [n_T])}{2[n_T]} \tag{14}$$

The case of the O_1-O_3 loop is slightly more complex because binding of the repressor to O_3 affects the ability of CAP to activate transcription. The limit of the free energy of binding to O_2 going to infinity ($g_2 \to \infty$) leads to

$$\begin{aligned} g_{L13} = -RT \ln ((&4(\Omega - 1)\, e^{(g_1+g_3)/RT} \\ &+ 2[n_T]\, e^{g_1/RT}(\Omega\chi - 1) \\ &- 2[n_T]\, e^{g_3/RT} - [n_T]^2)/2[n_T]) \end{aligned} \tag{15}$$

This value includes the contribution of the stabilization of the O_1-O_3 loop by CAP.[62,63]

The value for the O_2-O_3 loop is obtained using the observed scaling of the free energy for long loops, $\Delta G_L = \Delta G_{L_0} + 1.24RT \ln(L/L_0)$, where ΔG_{L_0} is the free energy for a loop of length L_0:[30]

$$g_{L23} = g_{L12} + 1.24RT \ln \frac{493}{401} \tag{16}$$

which uses that the length of the O_2-O_3 and O_1-O_2 loops are 493 bp and 401 bp, respectively.[30,64,65] The advantage of this approach is that the scaling law provides the free energy of looping for any loop length in its domain of applicability. In general, however, the free energy of looping depends, among other factors, on the length of the loop, the specific sequence connecting the operators, the DNA supercoiling state, and the presence of DNA architectural proteins. This dependence, which is specially marked for short loops between ~50 bp and ~150 bp,[64,66,67] should be parametrized explicitly for each specific loop and experimental condition.

Free Energy of *lac* Repressor Tetramerization. Detailed gene expression data is available mainly for strains with WT repressors, which exist basically in the tetrameric form, and mutant repressors that do not tetramerize.[18,20] As in the experimental setups, we consider the two limiting cases of no tetramerization ($\Delta G^\circ_{td} \to \infty$) and high tetramerization ($\Delta G^\circ_{td} \to -\infty$), which correspond to $K_{td} \to \infty$ and $K_{td} \to 0$, respectively.

Repressor-Inducer Dissociation Constant. To infer the repressor-inducer dissociation constant we consider experimental data with the two auxiliary operators deleted for strains with the tetrameric form of the repressor. Under these conditions, proceeding as in the deduction of eq 11 but using eq 10 for the inducer-dependent effective free energy of binding and $\Delta G^\circ_{td} \to -\infty$ leads to

$$\Delta G^\circ_{Oj}([I]) = -RT \ln \frac{2(\Omega([I]) - 1)}{[n_T]} \tag{17}$$

where $\Omega([I])$ is the observed repression level at the inducer concentration $[I]$. The value of K_I can be obtained explicitly for low inducer concentration taking into account that K_{Oj1} and K_{Oj2} are much larger than K_{Oj} as

$$K_I = [I]/(\sqrt{[n_T]\, e^{-\Delta G_{Oj}/RT}/2(\Omega([I]) - 1)} - 1) \tag{18}$$

Repression Data. To calibrate and test the model, we focus on an extensive collection of classic gene expression experiments on the *lac* operon in strains covering diverse combinations of operator swapping and deletions,[20] including all eight possible combinations of operator deletions.[18] All of the operator setups considered different repressor concentrations for both WT *lac* repressor, which under physiological conditions exists primarily as a tetramer, and a mutant repressor that does not tetramerize. In total, there are 92 experimental data points spanning more than 4 orders of magnitude of the repression level. So far, there has not been a model able to capture all this classic gene expression data in full.

System Behavior for Dimeric Species. The case of mutant repressors that do not tetramerize ($\Delta G^\circ_{td} \to \infty$) is the natural starting point to test the accuracy of the modular design in free energy space. Under such conditions, the behavior of the system does not depend on the looping parameters because dimers have only a single DNA-binding domain.

We took just four experimental data points to calibrate the model without looping and then used the calibrated model to predict the transcriptional activity for all of the other operator configurations (Figure 2). One data point was used for each free energy of binding to O_1, O_2, and O_3, and another one for the extent of CAP activation. The values of the remaining parameters were obtained using PWM scores and thermodynamic relationships. The results show that without any free parameter this approach is able to accurately predict the observed gene expression data for the remaining 38 data points over a 10,000-fold range for an almost exhaustive range of operator configurations.

System Behavior for Tetrameric Species. Regulation by *lac* repressor tetramers ($\Delta G^\circ_{td} \to -\infty$) brings about DNA looping. This increase in complexity needs just two additional experimental data points to calibrate the looping components of the model (Figure 3). With these 2 additional data points, the model is able to accurately predict the observed gene expression behavior for regulation by *lac* repressor tetramers for the remaining 48 experimental data points in 21 strains involving different operator configurations.

The results show that the behavior of single operator setups for both dimeric and tetrameric repressors is very similar. The presence of multiple operators with the possibility of looping, in contrast, leads to marked differences between the two oligomeric forms of the repressor, with the tetrameric form exhibiting much higher repression levels for low repressor concentrations.

Effects of Parameter Inference. The calibration of the model with just a few experimental data points makes the model predictions dependent on the experimental variability of the particular points used to infer the parameter values. In our analysis, the most marked difference compared with previous estimations of the parameters is the value of the inverse of the extent of activation by CAP, χ. The value we inferred using

dx.doi.org/10.1021/sb400013w | ACS Synth. Biol. 2013, 2, 576−586

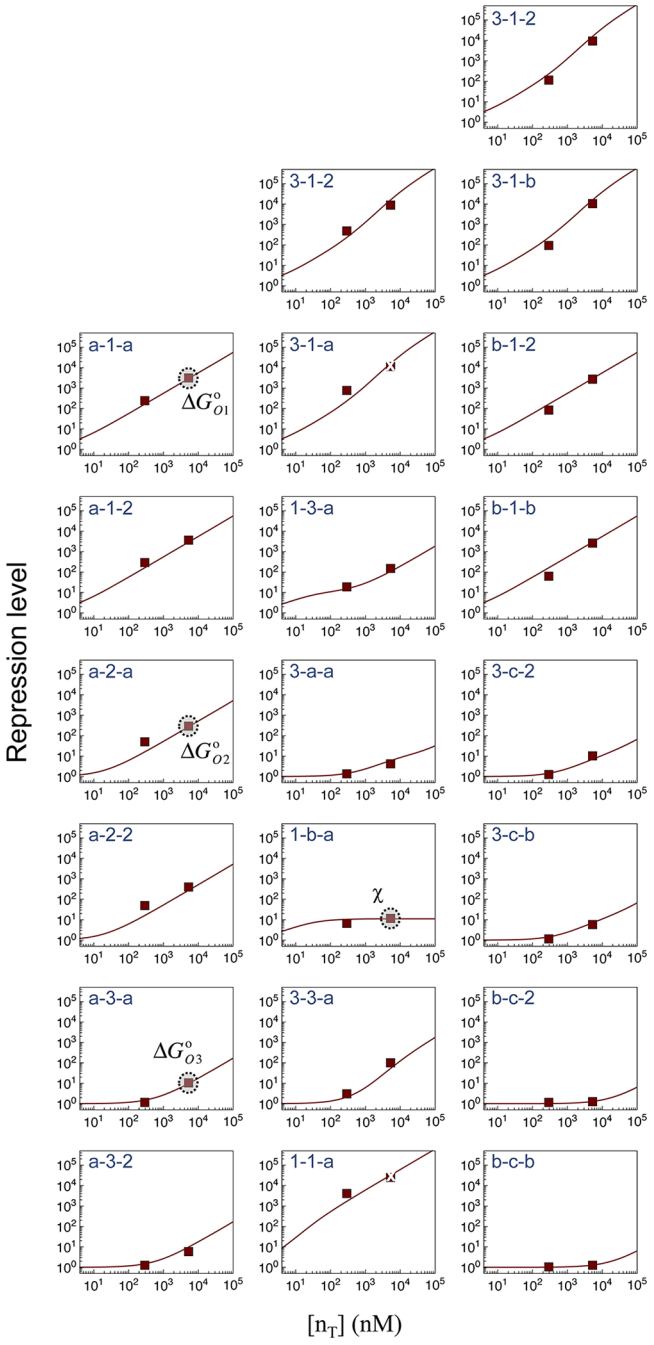


Figure 2. Accurate prediction of transcriptional activity for 21 strains regulated by mutant dimeric *lac* repressor from just four calibration points. The repression level, $\Omega = \Gamma_{max}/\Gamma$, was obtained for operator configurations including WT, all of the combinations of deletions of the three operators, and different combinations of operator deletions and swapping. For each of the 21 cases, the results of the model (continuous lines) as a function of the repressor subunit concentration are compared with the experimental data of Oehler et al.[18,20] (square symbols). The broken symbols indicate the experimental data points that could measure only a lower bound of the repression level. We have used the equivalence that 1 molecule per *E. coli* cell, as reported in the experimental data, corresponds to 1.5 nM. The operator setup is indicated for each case following the genomic order O_3-O_1-O_2 with a number indicating the sequence of the corresponding WT operator and a letter indicating the corresponding deletion. For instance 3-1-2 corresponds to WT *lac* operon (O_3-O_1-O_2), b-c-b corresponds to all operators deleted (O_{3^-b}-O_{1^-c}-O_{2^-b}), and 1-3-a corresponds to swapping the sequences of O_1 and O_3 and deleting O_2 with the O_{2^-a} sequence (O_3-O_1-O_{2^-a}). Only four experimental points (indicated by large gray circles) are used to obtain the values of the parameters $\Delta G^\circ_{O1} = -12.50$ kcal/mol, $\Delta G^\circ_{O2} = -11.07$ kcal/mol, $\Delta G^\circ_{O3} = -9.01$ kcal/mol, and $\chi = 0.091$. The free energy of binding of the repressor to the mutated operators was obtained from the PWM scores, as shown in Table 1.

experiments with dimeric repressors (Figure 2) is about a factor of three larger than the one, $\chi = 0.03$, previously reported in ref 20. We inferred the value of this parameter also from the experiments with tetrameric repressors and observed that in this case its value is just slightly larger than previous estimates (Figure 4). Comparison of the model results for both parameter values indicates that both of them lead to accurate predictions over the 4 orders of magnitude of the experimental data, with the parameter closer to previously reported estimates[20] producing generally even better results (Figure 4).

Effects of Operator Deletion Residual Binding. The predictions of the free energies of binding for the deletions using PWM scores show that some deletions, especially those affecting the main operator O_1 are not complete deletions because they involve the mutation of just a few DNA base pairs. These mutations substantially increase the free energy of binding but still leave their values relatively close to the free energy of binding to O_3, the weakest WT operator (Table 1). The results of the model show that such small residual binding makes these mutations distinguishable from a complete deletion at the phenotypic level (Figure 5).

From Dimer to Tetramer Binding. To compare with the experimental data, we have assumed that the *lac* repressor is present either as a tetramer or a dimer, which provides excellent results. It is generally assumed that WT *lac* repressor exists mainly as a tetramer under physiological conditions. The *in vitro* estimates of its ΔG°_{td} in the literature range from −13.2 kcal/mol[68] to −10.6 kcal/mol,[69] which correspond to K_{td} = 0.28 nM and K_{td} = 21.2 nM, respectively. The approach we have developed allows the computation of the repression level for any value of K_{td}. The results indicate that the range of the experimentally reported values of K_{td} provide very similar repression curves for WT (∼60 nM repressor monomeric subunits) and higher repressor concentrations (Figure 6).

Experimental Variability and Model Accuracy. There are several experimental setups that are equivalent in terms of the dependence of the transcriptional activity on the total number of repressor monomeric subunits. A prime example involves the configuration with the two auxiliary operators deleted. In this case, the repression curves would depend on the strength of the main operator but are expected to be the same for the tetrameric and dimeric forms of the repressor and for different promoters. It is important to compare these situations with each other to have a sense of the level of experimental variability one would encounter in new experiments.

We have explicitly compared data for O_1, O_2, and O_3 as main operators for tetrameric and dimeric repressors, including additional data for repression of the *lac*UV5 promoter,[70] a mutated version of the WT *lac* promoter that does not need CAP to be active. In the case of dimeric repressors, we have also included the data for the configurations with the auxiliary operator O_2, which does not affect transcription in the absence of DNA looping (Figure 7). The results indicate that there is substantial experimental variability, reaching up to a 5-fold difference between the data of ref 18 and ref 20 for the dimeric

dx.doi.org/10.1021/sb400013w | *ACS Synth. Biol.* 2013, 2, 576−586

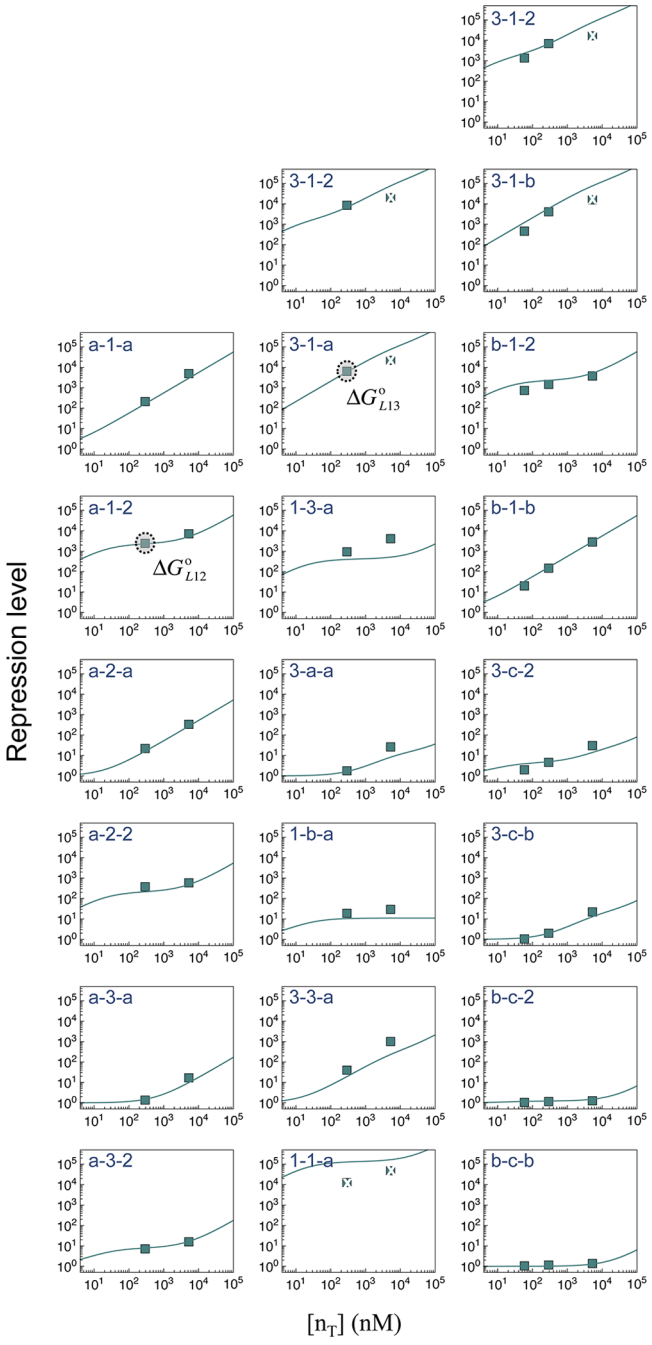


Figure 3. Accurate prediction of transcriptional activity for 21 strains regulated by WT *lac* repressor from just 2 additional calibration points. The repression level from the model (continuous lines) is shown for the same operator configurations as in Figure 2 as a function of the concentration of repressor units and is compared with the experimental data of Oehler et al.[18,20] (square symbols). The broken symbols indicate the experimental data points that could measure only a lower bound of the repression level. Only two additional experimental points (indicated by large gray circles) are used to obtain the values of the parameters ΔG°_{L12} = 7.86 kcal/mol and ΔG°_{L13} = 6.85 kcal/mol. The free energy of looping between O_2 and O_3 was obtained as ΔG°_{L23} = 8.01 kcal/mol.

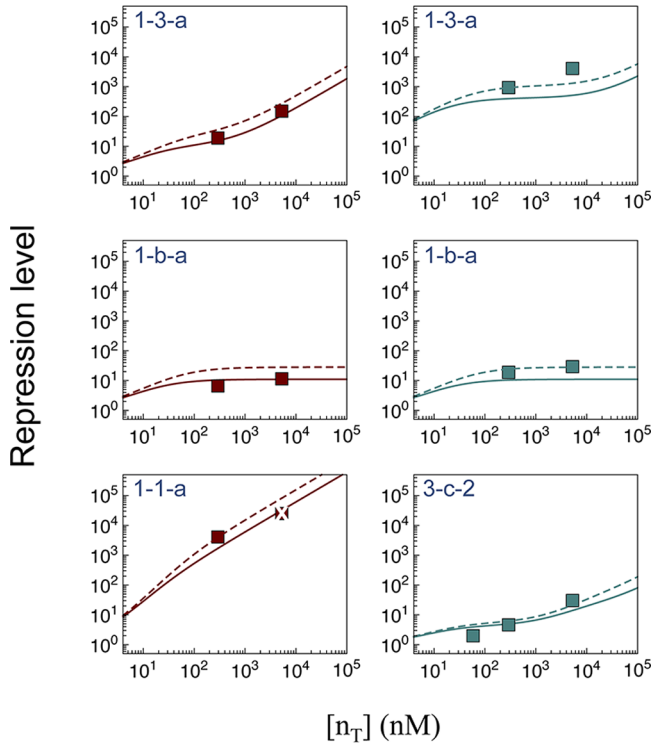


Figure 4. Effects of parameter inference. The repression level from the model is shown for representative operator configurations for the parameter values of Figures 2 and 3 (continuous line) and for the same parameter values except for χ = 0.035 (discontinuous line). Experimental data of Oehler et al.[18,20] (square symbols) are shown for comparison.

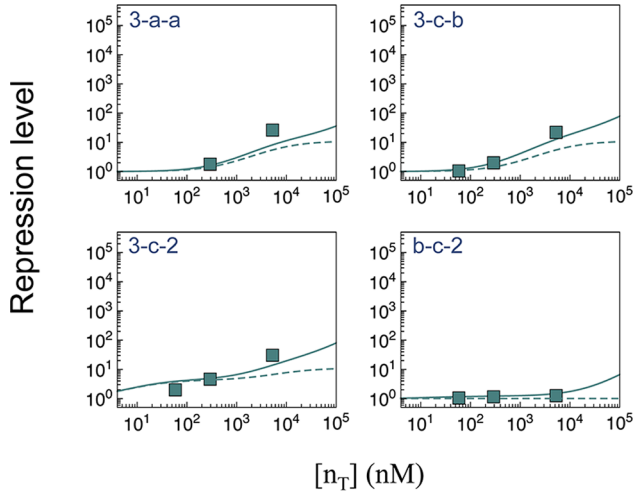


Figure 5. Deletions versus weak binding. The repression level from the model is shown for representative operator configurations for the parameter values of Figures 2 and 3 (continuous line) and for the same parameter values except that the free energy of binding to the mutant operators is infinite as in a complete deletion (discontinuous line). Experimental data of Oehler et al.[18,20] (square symbols) are shown for comparison.

repressors at 300 nM and between the data of ref 20 and ref 70 for the tetrameric repressors at 5.4 μM when O_1 is used as the main operator. In all of these cases, the results of the modular approach are consistent with the experimental variability and on average the model predicts the experimental data with 1.7-fold accuracy over a 10,000-fold variation of the repression level (Figure 7).

Induction Curves and Activity Kinetics. To further ascertain the applicability of the modular approach to more complex scenarios, we consider the transcriptional activity as a

dx.doi.org/10.1021/sb400013w | *ACS Synth. Biol.* 2013, 2, 576–586

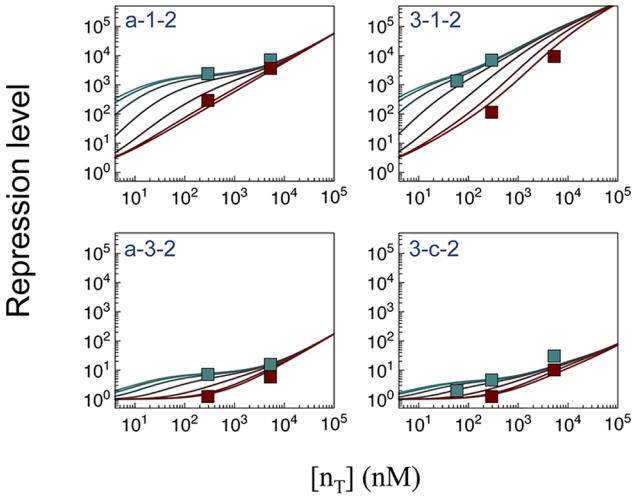


Figure 6. Effect of repressor tetramerization on transcription. For each panel, the continuous lines represent the model results for the tetramer-dimer dissociation constant that takes the values $K_{td} = 10^{-1}$, 1, 10, 10^2, 10^3, 10^4, and 10^5 nM from top (green) to bottom (red). The values of the other parameters are same as in Figures 2 and 3. Experimental data of Oehler et al.[18,20] (square symbols) for tetramer (green) and dimer (red) forms of the repressor are shown for comparison.

function of the IPTG inducer concentration for key operator configurations. Just a single experimental point is needed to infer the repressor-inducer dissociation constant using eq 18, which takes into account that K_{Oj1} and K_{Oj2} are much larger than K_{Oj} and can be assumed to be infinite. With this additional value, and without any change in the six previously inferred parameter values, the model accurately captures the experimentally observed induction curves[49,71,72] for a single operator with 50 tetrameric repressors per cell and for the wild type operator configuration with 100 dimeric, 50 tetrameric, and 10 tetrameric repressors per cell (Figure 8).

The modular approach can also be extended to follow the kinetics of induction in terms of the normalized β-galactosidase activity, C_β. For instance, in strains lacking the *lac* permease, the inducer concentration inside the cell is the same as in the media and the kinetics of this process is concisely described by

$$dC_\beta/dt = r(1/\Omega([I]) - C_\beta) \quad (19)$$

where r is the growth rate of the cell population. In this way, incorporation of the transcriptional module described by $\Omega([I])$ into the cell-population growth dynamics given by eq 19 allows the modular approach to capture the induction kinetics. The resulting kinetic model captures in detail the experimental time courses of β-galactosidase activity[72] without any additional molecular parameter, just by using the measured growth rate of the cell population, for a wide range of inducer concentrations (Figure 8).

Conclusions. A common theme in synthetic biology is the level of detail needed to capture the system behavior from the properties of its components. Very often, the underlying molecular processes do not have a description that fits into the prototypical simple chemical kinetics schemes. In these cases, networks of molecular interactions that extend beyond simple binding events need to be considered explicitly.[55] A key issue is, therefore, to understand how the underlying molecular complexity shapes the system and controls the cellular behavior.

Our results have identified key biophysical principles to efficiently integrate the prototypical complexity of the control of gene expression into a few-parameter model to accurately predict system behavior from a few simple rules to connect the parts. A central feature is a decomposition of the free energy of the different molecular states into additive contributions of the interactions. In the case of the *lac* operon, the resulting modular design in free energy space makes it possible to use just six experimental points to calibrate the model without fitting, which allowed us to capture for the first time the full range of regulation of the *lac* operon by dimeric and tetrameric repressors in a comprehensive set of operator mutations and swapping.

The cases analyzed covered extensively the diverse phenomenology of the *lac* operon, including ablations of looping by mutations of both auxiliary operators, different strengths of the main operator with and without one of the two auxiliary operators, incomplete operator deletions with residual

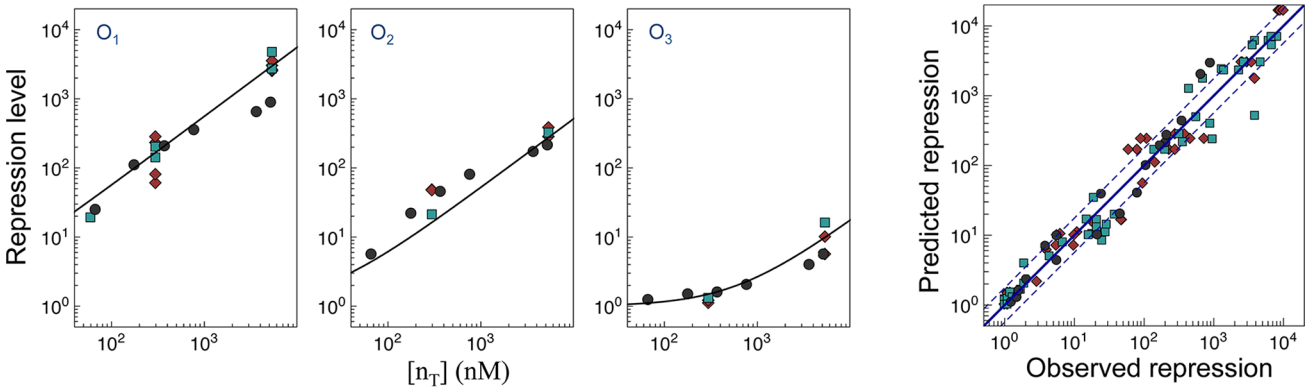


Figure 7. Experimental variability and model accuracy. The dependence of the repression level on the total repressor concentration from the model (continuous lines) and experimental data (symbols) is shown for the configurations without both auxiliary operators for the case of tetrameric repressors and without O_3 for the case of dimeric repressors in separate panels for each of the setups with O_1, O_2, and O_3 as main operators (three leftmost panels). The experimental data correspond to strains with the WT *lac* promoter for tetrameric (squares) and dimeric (diamonds) repressors from Oehler et al.[18,20] and with the *lac*UV5 promoter for tetrameric repressor (circles) from Garcia and Phillips.[70] The predicted vs the observed repression level is plotted for all available configurations, repressor oligomeric forms, and promoters (rightmost panel). The dashed lines represent a factor of 1.7 higher and 1/1.7 lower than the perfect prediction indicated by the continuous line.

dx.doi.org/10.1021/sb400013w | *ACS Synth. Biol.* 2013, 2, 576–586

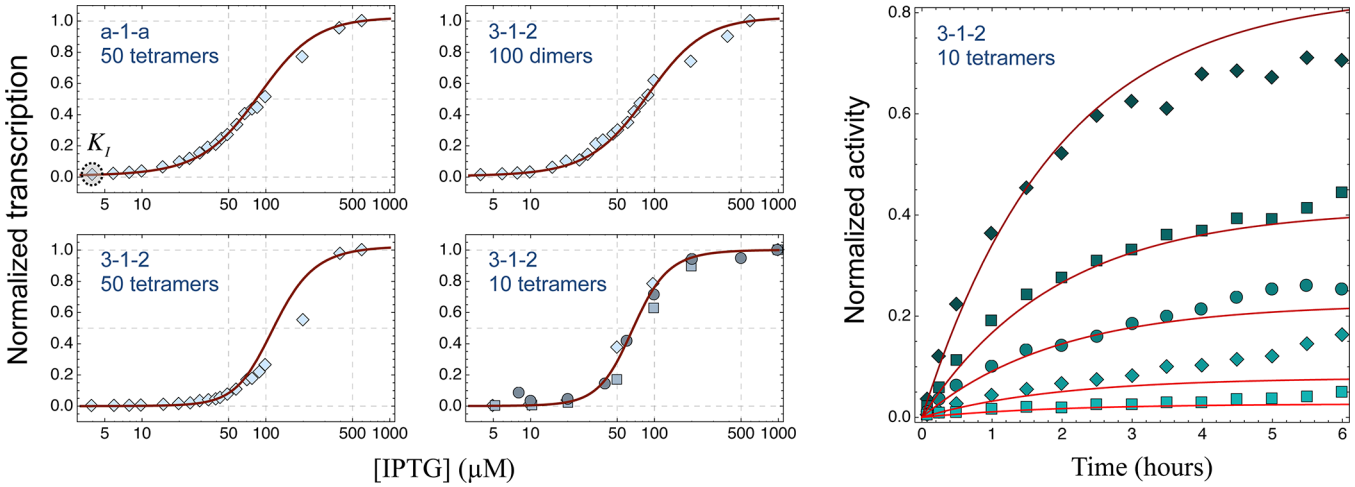


Figure 8. Induction curves and activity kinetics. The normalized transcriptional activity $1/\Omega([I])$ as a function of the IPTG inducer concentration $[I]$ = [IPTG] from the model (continuous lines) and experimental data (symbols) is shown for a single operator (a-1-a) with 50 tetrameric repressors per cell and for the WT operator configuration (3-1-2) with 100 dimeric and 50 tetrameric repressors per cell [experimental data from Oehler et al.[71]] and for the WT operator configuration (3-1-2) with 10 tetrameric repressors per cell [experimental data from Kuhlman et al.[49] (squares and diamonds) and Laurent et al.[72] (circles)]. Just a single experimental point (indicated by a large gray circle) is used to infer the value of the repressor-inducer dissociation constant K_I = 7.16 μM. The previously inferred values of the six remaining parameters are as in Figures 2 and 3. The normalized β-galactosidase activity, C_β, as a function of time after inducer addition for the WT operator configuration (3-1-2) with 10 tetrameric repressors per cell from the model (solid line) and from experimental data (symbols) by Marbach and Bettenbrock[73] is shown, from bottom to top, for 20, 30, 45, 60, and 120 μM IPTG. The growth rate of the cell population is r = 0.52 h^{-1}.

binding, responses to changes in repressor concentration with the main operator deleted, and interactions of the repressor with CAP. All of these cases were tested for different concentrations of both dimeric and tetrameric forms of the repressor using always the same values of the parameters inferred from just six experimental points and the thermodynamic relationships between them.

The low number of experimental points needed for model calibration results not only from modularity but also from the use of thermodynamic principles that establish relationships between parameters of different processes, such as tetramer vs dimer binding and tetramerization on operator DNA vs tetramerization outside operator DNA.

The versatility of the modular approach is evident in its ability to capture even more complex situations just by generalizing the key elements without the need to recalibrate the model. Explicitly, we have shown that changing just one element of the modular approach and a single additional data point are needed to accurately recapitulate in detail the experimentally observed induction curves for key operator configurations involving different concentrations of tetrameric and dimeric forms of the repressor. We have also shown that incorporation of the transcriptional module into the cell-population growth dynamics allows the modular approach to track in detail the induction kinetics over time for a wide range of inducer concentrations without any additional molecular parameter.

Fully predictive computational approaches able to efficiently address multiple complex interactions, like the one we have implemented here, are becoming increasingly important in designing synthetic genetic networks with multiple components. In many instances, it is possible to use a modular structure at the level of promoters so that each promoter can be characterized experimentally to predict the system outcome.[1,2] Our results bring this type of predictive capabilities within the structure of the promoter with multiple sites, interactions among regulators, and long-range interactions through DNA looping.

METHODS

Model Implementation. The effective transcription rate, as defined by eq 9, is computed analytically using the software package CplexA[56] with the expressions of the free energy and the transcription rate given by eqs 6, 7, and 8. The result is expressed in terms of the total repressor monomer subunit concentration, $[n_T]$, using $[n_2] = ((K_{td}^2 + 4K_{td}[n_T])^{1/2} - K_{td})/4$, which is obtained from eq 1 and the conservation of mass.

AUTHOR INFORMATION

Corresponding Author

*E-mail: j.vilar@ikerbasque.org; lsaiz@ucdavis.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the MINECO under grants FIS2009-10352 and FIS2012-38105 (J.M.G.V.) and the University of California, Davis (L.S.).

REFERENCES

(1) Vilar, J. M. G. (2006) Modularizing gene regulation. *Mol. Syst. Biol. 2*, 2006.0016.

(2) Guido, N. J., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C. R., Elston, T. C., and Collins, J. J. (2006) A bottom-up approach to gene regulation. *Nature 439*, 856−860.

(3) Elowitz, M. B., and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature 403*, 335−338.

(4) Guet, C. C., Elowitz, M. B., Hsing, W., and Leibler, S. (2002) Combinatorial synthesis of genetic networks. *Science 296*, 1466−1470.

(5) Kinkhabwala, A., and Guet, C. C. (2008) Uncovering cis regulatory codes using synthetic promoter shuffling. *PLoS One 3*, e2030.

(6) Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature 403*, 339−342.
(7) Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012) Genetic programs constructed from layered logic gates in single cells. *Nature 491*, 249−253.
(8) Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. (2008) A fast, robust and tunable synthetic gene oscillator. *Nature 456*, 516−519.
(9) Shou, W., Ram, S., and Vilar, J. M. G. (2007) Synthetic cooperation in engineered yeast populations. *Proc. Natl. Acad. Sci. U.S.A. 104*, 1877−1882.
(10) Chuang, J. S., Rivoire, O., and Leibler, S. (2009) Simpson's paradox in a synthetic microbial system. *Science 323*, 272−275.
(11) Chuang, J. S., Rivoire, O., and Leibler, S. (2010) Cooperation and Hamilton's rule in a simple synthetic microbial system. *Mol. Syst. Biol. 6*, 398.
(12) Gore, J., Youk, H., and van Oudenaarden, A. (2009) Snowdrift game dynamics and facultative cheating in yeast. *Nature 459*, 253−256.
(13) Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellen, J., Peeters, T., Hohmann, S., de Nadal, E., Posas, F., and Sole, R. (2011) Distributed biological computation with multicellular engineered networks. *Nature 469*, 207−211.
(14) Slusarczyk, A. L., Lin, A., and Weiss, R. (2012) Foundations for the design and implementation of synthetic genetic circuits. *Nat. Rev. Genet. 13*, 406−420.
(15) Müller-Hill, B. (1996) *The lac Operon: A Short History of a Genetic Paradigm*, Walter de Gruyter, Berlin, New York.
(16) Jacob, F., and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol. 3*, 318−356.
(17) Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science 271*, 1247−1254.
(18) Oehler, S., Eismann, E. R., Kramer, H., and Muller-Hill, B. (1990) The three operators of the lac operon cooperate in repression. *EMBO J. 9*, 973−979.
(19) Mossing, M. C., and Record, M. T., Jr. (1986) Upstream operators enhance repression of the lac promoter. *Science 233*, 889−892.
(20) Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., and Muller-Hill, B. (1994) Quality and position of the three lac operators of E. coli define efficiency of repression. *EMBO J. 13*, 3348−3355.
(21) Blackwood, E. M., and Kadonaga, J. T. (1998) Going the distance: a current view of enhancer action. *Science 281*, 60−63.
(22) Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008) *Molecular Biology of the Cell*, 5th ed., Garland Science, New York.
(23) Vilar, J. M. G., and Saiz, L. (2011) Control of gene expression by modulated self-assembly. *Nucleic Acids Res. 39*, 6854−6863.
(24) Saiz, L. (2012) The physics of protein-DNA interaction networks in the control of gene expression. *J. Phys.: Condens. Matter 24*, 193102.
(25) Schleif, R. (1992) DNA looping. *Annu. Rev. Biochem. 61*, 199−223.
(26) Matthews, K. S. (1992) DNA looping. *Microbiol. Rev. 56*, 123−136.
(27) Cournac, A., and Plumbridge, J. (2013) DNA looping in prokaryotes: experimental and theoretical approaches. *J. Bacteriol. 195*, 1109−1119.
(28) Revet, B., von Wilcken-Bergmann, B., Bessert, H., Barker, A., and Muller-Hill, B. (1999) Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. *Curr. Biol. 9*, 151−154.
(29) Dodd, I. B., Shearwin, K. E., Perkins, A. J., Burr, T., Hochschild, A., and Egan, J. B. (2004) Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev. 18*, 344−354.
(30) Saiz, L., and Vilar, J. M. G. (2006) DNA looping: the consequences and its control. *Curr. Opin. Struct. Biol. 16*, 344−350.
(31) Maston, G. A., Evans, S. K., and Green, M. R. (2006) Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet. 7*, 29−59.
(32) Wang, Q., Carroll, J. S., and Brown, M. (2005) Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol. Cell 19*, 631−642.
(33) Hewetson, A., and Chilton, B. S. (2008) Progesterone-dependent deoxyribonucleic acid looping between RUSH/SMARCA3 and Egr-1 mediates repression by c-Rel. *Mol. Endocrinol. 22*, 813−822.
(34) Stenger, J. E., Tegtmeyer, P., Mayr, G. A., Reed, M., Wang, Y., Wang, P., Hough, P. V., and Mastrangelo, I. A. (1994) p53 oligomerization and DNA looping are linked with transcriptional activation. *EMBO J. 13*, 6011−6020.
(35) Phelps, C. B., Sengchanthalangsy, L. L., Malek, S., and Ghosh, G. (2000) Mechanism of kappa B DNA binding by Rel/NF-kappa B dimers. *J. Biol. Chem. 275*, 24392−24399.
(36) Zhang, X., and Darnell, J. E., Jr. (2001) Functional importance of Stat3 tetramerization in activation of the alpha 2-macroglobulin gene. *J. Biol. Chem. 276*, 33576−33581.
(37) Tomilin, A., Remenyi, A., Lins, K., Bak, H., Leidel, S., Vriend, G., Wilmanns, M., and Scholer, H. R. (2000) Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell 103*, 853−864.
(38) Kang, J., Gemberling, M., Nakamura, M., Whitby, F. G., Handa, H., Fairbrother, W. G., and Tantin, D. (2009) A general mechanism for transcription regulation by Oct1 and Oct4 in response to genotoxic and oxidative stress. *Genes Dev. 23*, 208−222.
(39) Yasmin, R., Yeung, K. T., Chung, R. H., Gaczynska, M. E., Osmulski, P. A., and Noy, N. (2004) DNA-looping by RXR tetramers permits transcriptional regulation "at a distance". *J. Mol. Biol. 343*, 327−338.
(40) Vilar, J. M. G., and Leibler, S. (2003) DNA looping and physical constraints on transcription regulation. *J. Mol. Biol. 331*, 981−989.
(41) Zhang, Y., McEwen, A. E., Crothers, D. M., and Levene, S. D. (2006) Analysis of in-vivo LacR-mediated gene repression based on the mechanics of DNA looping. *PloS One 1*, e136.
(42) Saiz, L., and Vilar, J. M. G. (2008) Ab initio thermodynamic modeling of distal multisite transcription regulation. *Nucleic Acids Res. 36*, 726−731.
(43) Vilar, J. M. G., and Saiz, L. (2005) DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev. 15*, 136−144.
(44) Swigon, D., Coleman, B. D., and Olson, W. K. (2006) Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. *Proc. Natl. Acad. Sci. U.S.A. 103*, 9879−9884.
(45) Villa, E., Balaeff, A., and Schulten, K. (2005) Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proc. Natl. Acad. Sci. U.S.A. 102*, 6783−6788.
(46) Elf, J., Li, G. W., and Xie, X. S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science 316*, 1191−1194.
(47) Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G., and Elf, J. (2012) The lac repressor displays facilitated diffusion in living cells. *Science 336*, 1595−1598.
(48) Kuhlman, T. E., and Cox, E. C. (2012) Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Mol. Syst. Biol. 8*, 610.
(49) Kuhlman, T., Zhang, Z., Saier, M. H., Jr., and Hwa, T. (2007) Combinatorial transcriptional control of the lactose operon of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A. 104*, 6043−6048.
(50) Narang, A. (2007) Effect of DNA looping on the induction kinetics of the lac operon. *J. Theor. Biol. 247*, 695−712.
(51) Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., and Van Oudenaarden, A. (2004) Multistability in the lactose utilization network of Escherichia coli. *Nature 427*, 737−740.

(52) Wilkinson, D. J. (2006) *Stochastic Modelling for Systems Biology*, Taylor & Francis, Boca Raton, FL.
(53) Oehler, S., and Muller-Hill, B. (2010) High local concentration: a fundamental strategy of life. *J. Mol. Biol. 395*, 242−253.
(54) Jacobson, H., and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys. 18*, 1600−1606.
(55) Saiz, L., and Vilar, J. M. G. (2006) Stochastic dynamics of macromolecular-assembly networks. *Mol. Syst. Biol. 2*, 2006.0024.
(56) Vilar, J. M. G., and Saiz, L. (2010) CplexA: a Mathematica package to study macromolecular-assembly control of gene expression. *Bioinformatics 26*, 2060−2061.
(57) Hill, T. L. (1960) *An Introduction to Statistical Thermodynamics*, Addison-Wesley Pub. Co., Reading, MA.
(58) Shea, M. A., and Ackers, G. K. (1985) The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J. Mol. Biol. 181*, 211−230.
(59) Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005) Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev. 15*, 125−135.
(60) Gertz, J., Siggia, E. D., and Cohen, B. A. (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature 457*, 215−218.
(61) Vilar, J. M. G. (2010) Accurate prediction of gene expression by integration of DNA sequence statistics with detailed modeling of transcription regulation. *Biophys. J. 99*, 2408−2413.
(62) Hudson, J. M., and Fried, M. G. (1990) Co-operative interactions between the catabolite gene activator protein and the lac repressor at the lactose promoter. *J. Mol. Biol. 214*, 381−396.
(63) Saiz, L., and Vilar, J. M. G. (2007) Multilevel deconstruction of the in vivo behavior of looped DNA-protein complexes. *PLoS One 2*, e355.
(64) Muller, J., Oehler, S., and Muller-Hill, B. (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J. Mol. Biol. 257*, 21−29.
(65) Saiz, L., Rubi, J. M., and Vilar, J. M. G. (2005) Inferring the in vivo looping properties of DNA. *Proc. Natl. Acad. Sci. U.S.A. 102*, 17642−17645.
(66) Boedicker, J. Q., Garcia, H. G., and Phillips, R. (2013) Theoretical and experimental dissection of DNA loop-mediated repression. *Phys. Rev. Lett. 110*, 018101.
(67) Becker, N. A., Peters, J. P., and Maher, L. J., 3rd. (2013) Mechanism of promoter repression by Lac repressor-DNA loops. *Nucleic Acids Res. 41*, 156−166.
(68) Brenowitz, M., Pickar, A., and Jamison, E. (1991) Stability of a Lac repressor mediated "looped complex". *Biochemistry 30*, 5986−5998.
(69) Royer, C. A., Chakerian, A. E., and Matthews, K. S. (1990) Macromolecular binding equilibria in the lac repressor system: studies using high-pressure fluorescence spectroscopy. *Biochemistry 29*, 4959−4966.
(70) Garcia, H. G., and Phillips, R. (2011) Quantitative dissection of the simple repression input-output function. *Proc. Natl. Acad. Sci. U.S.A. 108*, 12173−12178.
(71) Oehler, S., Alberti, S., and Muller-Hill, B. (2006) Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Res. 34*, 606−612.
(72) Laurent, M., Charvin, G., and Guespin-Michel, J. (2005) Bistability and hysteresis in epigenetic regulation of the lactose operon. Since Delbruck, a long series of ignored models. *Cell. Mol. Biol. (Noisy-le-grand) 51*, 583−594.
(73) Marbach, A., and Bettenbrock, K. (2012) lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J. Biotechnol. 157*, 82−88.