Ab initio thermodynamic modeling of distal multisite transcription regulation

Leonor Saiz and Jose M. G. Vilar*

Integrative Biological Modeling Laboratory, Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA

ABSTRACT

Transcription regulation typically involves the binding of proteins over long distances on multiple DNA sites that are brought close to each other by the formation of DNA loops. The inherent complexity of assembling regulatory complexes on looped DNA challenges the understanding of even the simplest genetic systems, including the prototypical *lac* operon. Here we implement a scalable approach based on thermodynamic molecular properties to model *ab initio* systems regulated through multiple DNA sites with looping. We show that this approach applied to the *lac* operon accurately predicts the system behavior for a wide range of cellular conditions, which include the transcription rate over five orders of magnitude as a function of the repressor concentration for wild type and all seven combinations of deletions of three operators, as well as the observed induction curves for cells with and without active catabolite activator protein. Our results provide new insights into the detailed functioning of the *lac* operon and reveal an efficient avenue to incorporate the required underlying molecular complexity into fully predictive models of gene regulation.

INTRODUCTION

Control of localized events on DNA by proteins bound at distal sites is intimately linked to DNA looping. DNA flexibility plays an important role in mediating long-range interactions (1), allowing proteins bound to nonadjacent DNA sites to come close to each other. This strategy is widely used in eukaryotic enhancers (2) to integrate multiple signals into the control of the transcriptional machinery (2,3), to the extent that transcription regulation through DNA looping is nowadays considered to be the rule rather than the exception (4,5). DNA looping can also be formed by single-protein complexes, including the regulators of many bacterial operons, such as *ara*, *gal*, and *lac* operons (6), and human proteins involved in cancer, such as retinoic X receptor (RXR) (7) and p53 (8). The presence of DNA looping with single- and multi-protein complexes is important not only for transcription regulation but also for many other cellular processes, including DNA replication (9), recombination (10), nucleosome positioning (11), and telomere maintenance (12).

To date, there is only limited understanding of the factors that drive macromolecular assembly on looped DNA, especially when multiple binding sites and loops are involved (1,13). A notable example is the *lac* operon in *Escherichia coli*, which is still far from being completely understood despite being one of the systems that led to discovery of gene regulation (14).

The *E. coli lac* operon is the genetic system that regulates and produces the enzymes needed to metabolize lactose (14). The response to lactose is controlled by the *lac* repressor (15) that can bind to O_1, the main operator, thus preventing 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, to which the repressor can also bind, but not prevent transcription (Figure 1). Elimination of either one auxiliary operator has only minor effects; yet simultaneous elimination of both of them reduces the repression level by a factor 100 (16). This effect results from the ability of the *lac* repressor, a tetramer with two dimeric DNA-binding domains, to bind simultaneously to two operators and loop the intervening DNA. Thus, the main operator and at least one auxiliary operator suffice to form DNA loops that substantially increase the ability of the repressor to bind the main operator. Beyond increasing the repression level, it is not clear to what extent DNA looping shapes the properties of transcription regulation and the effects that having three instead of two operators has on the behavior of the system. Recent phenomenological analyses of the *lac* operon (17,18) have been able to fit

*To whom correspondence should be addressed. Tel: +1 646 888 2603; Fax: +1 646 422 0717; Email: vilar@cbio.mskcc.org
Present address:
Leonor Saiz, Department of Biomedical Engineering, University of California, Davis, CA 95616, USA
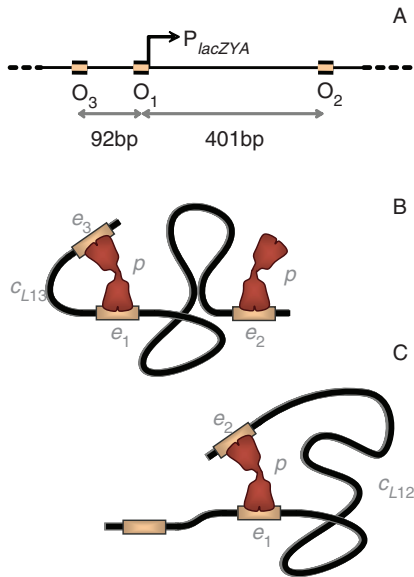
Figure 1. Operators controlling expression of the *lacZ*, *lacY*, and *lacA* genes in the *lac* operon. (**A**) Location of the main (O_1) and the two auxiliary (O_2 and O_3) operators, shown as orange rectangles on the thick black segment representing DNA. Binding of the *lac* repressor to O_1 prevents transcription of the three *lacZYA* genes. (**B**, **C**) The bidentate repressor can bind to any of the three operators and simultaneously to any two of them by looping the intervening DNA, resulting in different protein–DNA complexes. Two of the three possible loops are shown: (B) one *lac* repressor (shown in red) loops DNA by binding simultaneously to O_1 and O_3 (loop *L13*) whereas another repressor binds to O_2; (C) only one repressor is bound to DNA, forming a loop between O_1 and O_2 (loop *L12*). In both cases, the *lacZYA* genes are repressed. The different contributions to the free energy of the *lac* repressor–DNA complexes, which include positional (p), interaction (e_1, e_2, and e_3), and conformational (c_{L12}, c_{L13}, and c_{L23}) contributions, are explicitly indicated in these two cases.

the experimental data, but the type of predictive power that allows one to go from the molecular properties to the observed physiological behavior is still missing.

Here we implement a scalable quantitative computational approach to analyze systems regulated through multiple DNA sites with looping. This approach applied to the *lac* operon allows us to accurately predict *ab initio* the transcription rate over five orders of magnitude for wild type and seven mutants accounting for all the combinations of deletions of the three operators and the observed induction curves of the system for different cellular conditions, such as in the presence or absence of active catabolite activator protein (CAP). In the case of transcription, *ab initio*, or from first principles, refers to the properties of the molecular events upon which gene regulation is built. Here the properties of the molecular events are characterized through previously determined free energies.

METHODS

To incorporate the relevant molecular properties into a quantitative model of transcription regulation, we have used a statistical thermodynamics approach. Following the approach of Ref. (19), we consider a decomposition of the free energy of the protein–DNA complex into positional, interaction, and conformational contributions. In short, the positional free energy, p, accounts for the cost of bringing the *lac* repressor to its DNA-binding site in the protein–DNA complex; interaction free energies, e, arise from the physical contact between the repressor DNA-binding domains and the different operator sites; and conformational free energies, c, account for changes in conformation, including the formation of DNA loops (Figure 1). All these contributions to the free energy can be collected to obtain the free energy $\Delta G(s)$ of a given state s of the protein–DNA complex. The advantage of this approach is that it provides the free energies of a large number of different states from just the individual properties of the interactions and components. Here, different states account for the different ways in which the repressors can bind the three operators.

The free energy of a state s is connected to the equilibrium probability P_s of such state through the statistical thermodynamics relationship $P_s = \frac{1}{Z} e^{-\Delta G(s)/RT}$, where $Z = \sum_s e^{-\Delta G(s)/RT}$ is the partition function, which serves as normalization factor, and RT is the gas constant times the absolute temperature (20).

Straightforward application of the traditional thermodynamic approach (21) in a general framework is of limited use because the number of states that must be considered typically increases exponentially with the number of components (13). For instance, just the binding of the *lac* repressor to three DNA sites would lead to eight states, including states with zero, one, two, and three repressors bound to DNA. If DNA looping is taken into account, the number of states increases to 14. It has become clear recently that it is possible to overcome this limitation and express the free energy of all these states in a compact form by using binary variables (13). By extending the approach of Ref. (13) to consider multiple loops in the *lac* operon, we obtain

$$\Delta G(s) = (p + e_1)s_1 + (p + e_2)s_2 + (p + e_3)s_3$$
$$+ (c_{L12} - ps_1s_2)s_{L12} + (c_{L13} - ps_1s_3)s_{L13}$$
$$+ (c_{L23} - ps_2s_3)s_{L23} + \infty(s_{L12}s_{L13} + s_{L12}s_{L23}$$
$$+ s_{L13}s_{L23}),$$

where s_1, s_2, and s_3 are binary variables that indicate whether ($s_i = 1$, for $i = 1,2,3$) or not ($s_i = 0$, for $i = 1,2,3$) a repressor is bound to O_1, O_2, and O_3, respectively; and s_{L12}, s_{L13}, and s_{L23} are binary variables that indicate whether ($s_{Lij} = 1$, for $ij = 12,13,23$) or not ($s_{Lij} = 0$, for $ij = 12,13,23$) DNA forms the loops O_1-O_2, O_1-O_3, and O_2-O_3, respectively. For instance, using the notation $s \equiv (s_1, s_2, s_3, s_{L12}, s_{L13}, s_{L23})$ the state shown in Figure 1B is specified by $s = (1,1,1,0,1,0)$ and that of Figure 1C, by $s = (1,1,0,1,0,0)$. The subscripts of the different contributions to the free energy have the same meaning as those of the corresponding binary variables. In this case, with three interaction and three conformational free energies, it is possible to obtain the free energy of 14 states for different repressor concentrations. The dependence on the repressor concentration, n, enters the free energy through the

positional free energy: $p = p^o - RT \ln n$, where p^o is the positional free energy at 1 M concentration. An important advantage of the binary variable description is that it can straightforwardly implement 'logical conditions'. For instance, the infinity in the last term of the free energy implements that two loops that share one operator cannot be present simultaneously by assigning an infinite free energy to those states.

The experimental observations, as summarized in Ref. (22), indicate that in the *lac* operon, transcription is completely abolished when the repressor is bound to O_1. If O_1 is free, transcription takes place either at an activated maximum rate τ_{max} when O_3 is free or at basal reduced rate $\chi\tau_{max}$ when O_3 is occupied. The reduction from the activated state by a factor χ arises because binding of the repressor to O_3 prevents CAP, also known as cAMP Receptor Protein (CRP), from contacting the RNA polymerase and activating transcription (22). Activation is achieved when CAP bound to cyclic AMP (cAMP) binds between O_3 and O_1, contacts the RNA polymerase, and stabilizes its binding to the promoter (23,24). All these experimental facts can be accounted for by the transcription rate $\tau(s)$ expressed in terms of binary variables as

$$\tau(s) = \tau_{max}(1 - s_1)(\chi s_3 + 1 - s_3),$$

which provides a mathematical expression for the observed transcription control (22,25).

The concise mathematical expressions $\Delta G(s)$ and $\tau(s)$ completely specify the thermodynamic and transcriptional properties of the system. In particular, the repression level, defined as the maximum transcription over the actual average transcription rate, is given by $R = \tau_{max}/\bar{\tau}$, where the average transcription rate follows from $\bar{\tau} = \frac{1}{Z}\sum_s \tau(s)e^{-\Delta G(s)/RT}$. Thus, with this approach, it is possible to obtain a compact description for both the DNA–repressor complex and the control of transcription that scales linearly with the number of regulatory elements (13).

RESULTS

Our model accurately reproduces the observed behavior of the *lac* operon (22) in quantitative detail over five orders of magnitude of the repression level for three repressor concentrations and eight strains with all the possible combinations of operator deletions (Figure 2). The values of the parameters used are the same as those previously reported (1) except for two small variations. One of the changes is a 0.35 kcal/mol shift in the affinity of the repressor for the operators. This difference falls within the typical variation for different experimental conditions. The other change is a decrease of 0.9 kcal/mol in the free energy of forming the O_1-O_3 loop. This decrease is consistent with stabilization of the loop by binding of CAP to its DNA site between O_1 and O_3, which has been estimated to be within −0.8 kcal/mol and −1.4 kcal/mol with an average value of −1.1 kcal/mol (26) and is similar to the values obtained for other proteins that bend DNA, such as the HU protein (27). Note that the experiments
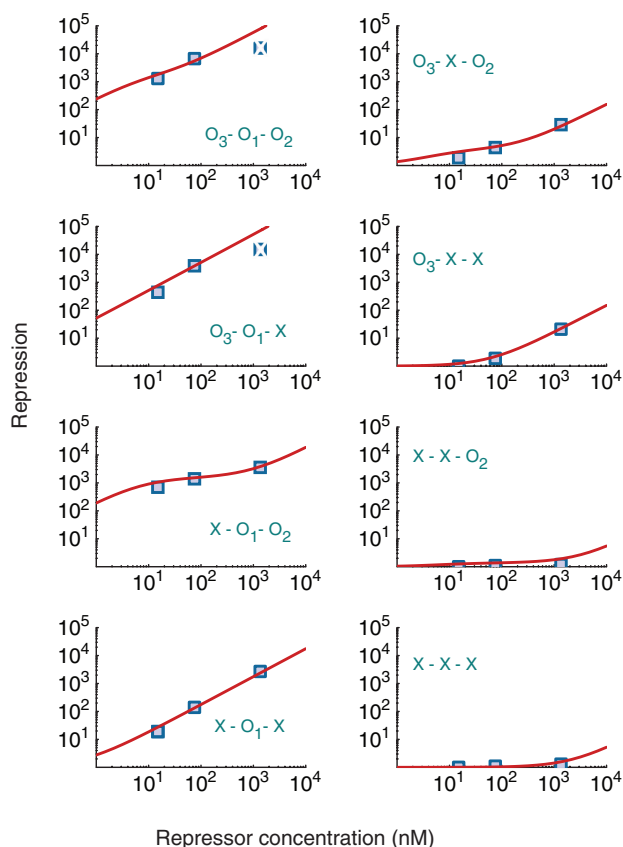
Figure 2. Repression level as a function of the repressor concentration for WT and all seven combinations of deletions of three operators. The repression levels in the presence of active CAP were obtained for WT and seven mutants accounting for all the combinations of deletions of the three operators. For each of the eight cases, the results of the model (red curves) as a function of the repressor concentration are compared with the experimental data (shaded blue squares) (22) available for three concentrations corresponding to 10 (WT cells), 50, and 900 repressors per cell. The particular set of WT or deleted (X) operators is indicated for each curve; for instance, O_3-O_1-O_2 corresponds to WT *lac* operon and X-X-X, to the mutant with all three operators deleted. The excellent agreement indicates that the model not only captures the repression values quantitatively but also the shapes of the curves, which are very different depending on the mutant. The values of the parameters used are: $e_1 = -27.8$ kcal/mol, $e_2 = -26.3$ kcal/mol, $e_3 = -24.1$ kcal/mol, $c_{L12} = 23.35$ kcal/mol, $c_{L13} = 22.05$ kcal/mol, $c_{L23} = 23.50$ kcal/mol, $p^\circ = 15$ kcal/mol, and $\chi = 0.03$. A deleted operator is modeled by increasing its free energy by 5 kcal/mol. The two repression levels represented by broken squares in the graphs for O3-O1-O2 and O3-O1-X were not accurately measured in the experiments, as described by the authors in a subsequent publication (16).

used to infer the free energies of looping (28,29) were performed with artificial constructs that did not contain the CAP-binding site within the loop. We have also observed that the experimental repression levels are better accounted for if the deletion of O_1 is not considered to be complete. Explicitly, we infer that the binding of the repressor is not completely abolished but reduced by ~5 kcal/mol, which is consistent with the mutation of just three base pairs of the operator (22). This strong reduction of affinity would be virtually indistinguishable from a complete deletion in a single operator setup, but it is not so in a multi-loop configuration.
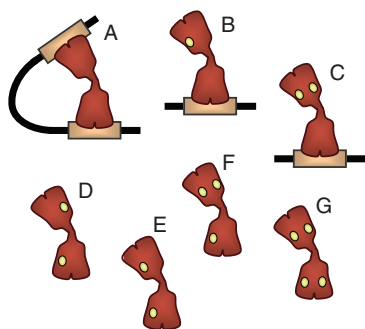
Figure 3. Inducer binding to the *lac* repressor. An inducer molecule, such as IPTG, can bind independently to each of the four identical monomeric units that form the *lac* repressor and impair its function. Without IPTG bound, each of the two dimers of the tetrameric repressor (in red) has a functional DNA-binding domain and can bind strongly to the operators (orange rectangle) and loop the intervening DNA (**A**). With one or two IPTG molecules (small yellow circles) bound to the same domain, the repressor can bind strongly to one operator but cannot loop DNA (**B,C**). With two or more IPTG molecules bound to different domains of the repressor, strong operator binding is abolished (**D,E,F,G**).

The approach we have developed can straightforwardly be extended to study the induction curves of the *lac* operon, which provide the steady-state transcription rate as a function of the inducer concentration. Inducers, such as allolactose in the natural environment and isopropyl-β-D-thiogalactoside (IPTG) in typical experimental setups, bind to the repressor and impair its function (Figure 3). Specifically, each of the four identical monomeric components of the repressor can bind independently an inducer molecule. Each of the two dimers that form the tetrameric *lac* repressor has a DNA-binding domain. Binding of either one or two inducer molecules to a dimeric unit strongly reduces its affinity for operator DNA. Therefore, with one or two inducer molecules in the same dimeric domain, the repressor can bind strongly to one operator but cannot loop DNA; with two or more inducer molecules in different domains, strong operator binding is also abolished (Figure 3).

Our model accurately reproduces the observed *in vivo* induction properties of the *lac* operon when we consider as functional repressors only those that do not have any inducer bound. This approximation is very good because the simultaneous binding of a repressor to two operators on looped DNA is much stronger (∼600 times) than the binding to a single operator and because the repressor concentration is very low, about ten repressors per cell (see below for a more general treatment). Thus, the concentration of functional repressors is given by $n = n_T P_F^4$, where n_T is the total repressor concentration, P_F is the probability that a monomer of the tetrameric repressor is not bound to IPTG, and P_F^4 is the probability that no IPTG molecule is bound to the repressor. The probability for a repressor monomer to be free of IPTG follows from $P_F = \frac{1}{1+[I]/K_i}$, where $[I]$ is the intracellular IPTG concentration and K_i is the inducer-repressor equilibrium dissociation constant (30). It is important to note that *in vitro* studies have found that IPTG can bind cooperatively when the operator
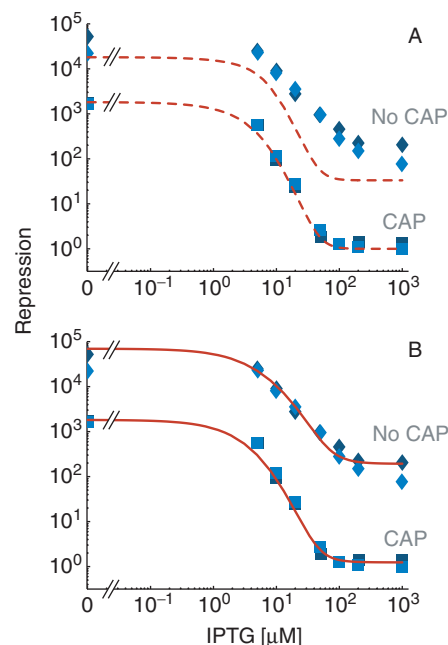
Figure 4. Induction curves of the *lac* operon for cellular conditions with and without active CAP. (**A**) The repression level in the presence of active CAP (lower curves labeled 'CAP') was obtained as a function of the concentration of the inducer IPTG for the model given by $\Delta G(s)$, $\tau(s)$, and $n = n_T P_F^4$ (dashed red lines) and compared with the experimental data from Ref. (17) (blue squares). The values of the parameters are the same as in Figure 2, with additional parameters $K_I = 6.9\,\mu M$ and $n_T = 15\,nM$. In the absence of active CAP (upper curves, labeled 'No CAP'), the model (dashed red lines) does not accurately reproduce the experimental data from Ref. (17) (blue diamonds). We consider that without active CAP, in addition to a reduced transcription $\tau(s) = \frac{1}{Z}\sum_s \tau_{\max}\chi(1-s_1)e^{-\Delta G(s)/RT}$, the formation of the O_1-O_3 loop is 0.9 kcal/mol more costly ($c'_{L13} = 22.95$ kcal/mol) than with CAP. (**B**) The excellent agreement with induction experiments in the absence of active CAP is recovered when the model is generalized to include an arbitrary number of repressor molecules per cell through the expressions $\Delta G_F(s)$, $\tau_F(s)$, n, n_d, and n_i, with same values of the parameters as in panel A, with the additional parameter $r = 2.8$ kcal/mol and 200 repressors per cell ($n_T = 300\,nM$) (upper continuous red line). In the presence of active CAP, the generalized model (lower continuous red line) recovers the results in panel A for low number of repressors ($n_T = 15\,nM$).

DNA concentration is at least 100 nM (31). *In vivo*, however, as in the experiments considered here, such high operator concentrations are never reached (one operator per *E. coli* cell corresponds to a concentration of ∼1.5 nM).

For the reported value of K_i (30) and the previous values of the parameters, without any free parameter, our model reproduces exactly the experimentally observed induction curves (17) of cells with active CAP (Figure 4A; lower dashed red curve labeled 'CAP'). The model, however, does not accurately reproduce the experimental data (Figure 4A; upper dashed red curve labeled 'No CAP') for induction of the *lac* operon without active CAP. We modeled this condition with the inferred value of the free energy of forming the O_1-O_3 loop in the absence of active CAP and with $\tau(s) = \frac{1}{Z}\sum_s \tau_{\max}\chi (1-s_1)e^{-\Delta G(s)/RT}$, which considers that there is no activated transcription.

Previous studies based on phenomenological models have shown that, it is possible to fit the experimental results by assuming that the absence of active CAP reduces the effects of looping in repression by a factor ~10 (17). However, such a factor appears to be extremely high because half of the contribution to repression comes from the O_1-O_2 loop, which is not affected by CAP. Therefore, within the current molecular view of the *lac* operon, the absence of CAP could only reduce the effects of looping by a factor ~2 at most. Another possibility not considered so far is to assume a different repressor concentration, which has been found to depend on the binding of CAP in the regulatory region through the post-transcriptional effects of the SsrA system (32).

In the general case that accounts for arbitrary repressor concentrations, we need to consider also the concentration of repressors with one, $n_d = n_T(4P_F^3(1 - P_F) + 2P_F^2(1 - P_F)^2)$, and two, $n_i = n_T - n - n_d$, inducer-impaired DNA-binding domains (Figure 3). The concentration n_i needs to be considered explicitly because the inducer substantially reduces the binding of the repressor to operator DNA but it does not abolish it completely (33). This 'residual' binding is taken into account by an offset r in the free energy of interaction with the operators. The expression of the free energy for the binding of the three repressor types is given by

$$\Delta G_F(s) = \Delta G(s) + (p_d + e_1)s_{1d} + (p_d + e_2)s_{2d}$$
$$+ (p_d + e_3)s_{3d} + (p_i + e_1 + r)s_{1i}$$
$$+ (p_i + e_2 + r)s_{2i} + (p_i + e_3 + r)s_{3i}$$
$$+ \infty(s_1 s_{1d} + s_2 s_{2d} + s_3 s_{3d} + s_1 s_{1i} + s_2 s_{2i} + s_3 s_{3i}$$
$$+ s_{1d}s_{1i} + s_{2d}s_{2i} + s_{3d}s_{3i}),$$

where d and i subscripts refer to the binding of repressors with one and two impaired DNA-binding domains, respectively. The corresponding positional free energies are given by $p_d = p^o - RT\ln n_d + RT\ln 2$ and $p_i = p^o - RT\ln n_i$. The term $RT\ln2$ in the expression of p_d takes into account that only one of the two DNA-binding domains is available for binding. The infinity in the last term implements that an operator cannot be occupied simultaneously by more than one repressor, which can be a repressor with zero, one, or two inducer-impaired DNA-binding domains. Note that other possible states, such as those accounting for looping of a repressor with inducer bound, are not taken into account because they are energetically very unfavorable for the observed large values of r.

Now, $s \equiv (s_1, s_2, s_3, s_{L12}, s_{L13}, s_{L23}, s_{1d}, s_{2d}, s_{3d}, s_{1i}, s_{2i}, s_{3i})$ specifies the state of the protein–DNA complex through the values of 12 binary variables. The number of states has increased from 14 to 76 but the only additional parameter is r, which has been measured experimentally ($r = 2.8$ kcal/mol) using operator fragment binding analysis (33).

The control of transcription is as follows: occupancy of O_1 by any repressor type always prevents transcription, and occupancy of O_3 prevents activated transcription when active CAP is present (if active CAP is absent, there is never activated transcription).

Therefore, the mathematical expression of the transcription rate in the presence of active CAP is

$$\tau_F(s) = \tau_{\max}(1 - (s_1 + s_{1d} + s_{1i}))$$
$$(\chi(s_3 + s_{3d} + s_{3i}) + 1 - (s_3 + s_{3d} + s_{3i}))$$

and in the absence of active CAP is

$$\tau_F(s) = \tau_{\max}\chi(1 - (s_1 + s_{1d} + s_{1i})).$$

Note that these expressions use the fact that only one of the variables s_1, s_{1d}, s_{1i} and only one of the variables s_3, s_{3d}, s_{3i} can be one at a given time.

The experimental data in the absence of active CAP is reproduced in detail just by setting the repressor concentration to 300 nM, equivalent to 200 repressors per cell (Figure 4B; upper full red curve labeled 'No CAP'). This repressor concentration is 10 times higher than that of wild type with active CAP but 4.5 times smaller than that of the mutants studied in Ref. (22) that express 900 repressors per cell. In the presence of active CAP, the results obtained are in excellent agreement with both the experimental data (Figure 4B; lower full red curve labeled 'CAP') and the low-repressor concentration model (Figure 4A; lower dashed red curve labeled 'CAP'). Thus, the general model accurately reproduces the experimental observations in the whole range of IPTG concentrations and cellular conditions with and without active CAP.

DISCUSSION

Regulation of gene expression by the assembly of protein–DNA complexes is strongly dependent on networks of molecular interactions that extend beyond simple binding events. A key issue is, therefore, to understand how the underlying molecular complexity shapes gene expression and controls the cellular behavior. Here we have shown that, in the case of the *lac* operon, incorporation of the molecular details—such as the presence of multiple DNA loops, the stabilization of loops by DNA-binding proteins, and residual binding of inducer-inactivated repressors and deleted sites—is crucial for the development of fully predictive models. Our results illustrate how to accurately infer the effects of multiple distal DNA-binding sites from the thermodynamic properties of the molecular components and provide new insights into the detailed functioning of the *lac* operon, including the role of CAP and the effects of the inducer.

Two key features of the thermodynamic approach we have implemented are the decomposition of the free energy of protein–DNA complexes into different modular contributions and the description of the different molecular states through binary variables. These two features taken together lead to concise mathematical expressions that completely specify the thermodynamic and transcriptional properties of all the states of the protein–DNA complex from just a few molecular parameters. In this way, we are able to circumvent one of the major issues that have precluded the development of computational approaches to study complex gene regulation systems;

namely, the individual specification of the properties of all the potential states of the complex, which typically results in large itemized lists that increase exponentially with the number of interactions and components (13,34,35). Here we have shown explicitly with the *lac* operon that our approach can successfully be scaled up to include multiple molecular species, binding sites, and DNA loops.

Fully predictive computational approaches able to efficiently address multiple complex interactions, like the one we have implemented here, are becoming increasingly important not only for understanding gene regulation in natural systems but also for designing synthetic genetic networks with multiple components (36,37). This type of computational approaches is especially important because complex systems can reveal the effects of molecular details that were hidden in simpler setups used to characterize the components (38).

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Saiz,L. and Vilar,J.M.G. (2006) DNA looping: the consequences and its control. *Curr. Opin. Struct. Biol.*, **16**, 344–350.
2. 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.
3. Tolhuis,B., Palstra,R.J., Splinter,E., Grosveld,F. and de Laat,W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*, **10**, 1453–1465.
4. Alberts,B. (2002) *Molecular Biology of the Cell, 4th edn.* Garland Science, New York.
5. Blackwood,E.M. and Kadonaga,J.T. (1998) Going the distance: a current view of enhancer action. *Science*, **281**, 60–63.
6. Schleif,R. (1992) DNA looping. *Annu. Rev. Biochem.*, **61**, 223–199.
7. 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.
8. 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.
9. Su,W., Middleton,T., Sugden,B. and Echols,H. (1991) DNA looping between the origin of replication of Epstein-Barr virus and its enhancer site: stabilization of an origin complex with Epstein-Barr nuclear antigen 1. *Proc. Natl Acad. Sci. USA*, **88**, 10870–10874.
10. Grindley,N.D., Whiteson,K.L. and Rice,P.A. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem.*, **75**, 567–605.
11. Segal,E., Fondufe-Mittendorf,Y., Chen,L., Thastrom,A., Field,Y., Moore,I.K., Wang,J.P. and Widom,J. (2006) A genomic code for nucleosome positioning. *Nature*, **442**, 772–778.
12. de Lange,T. (2004) T-loops and the origin of telomeres. *Nat. Rev. Mol. Cell Biol.*, **5**, 323–329.
13. Saiz,L. and Vilar,J.M.G. (2006) Stochastic dynamics of macromolecular-assembly networks. *Mol. Syst. Biol.*, **2**, 2006.0024.
14. Müller-Hill,B. (1996) *The lac Operon: A Short History of a Genetic Paradigm* Walter de Gruyter, Berlin, New York.
15. 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.
16. 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.
17. 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. USA*, **104**, 6043–6048.
18. Narang,A. (2007) Effect of DNA looping on the induction kinetics of the lac operon. *J. Theor. Biol.*, **247**, 695–712.
19. 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.
20. Hill,T.L. (1960) *An Introduction to Statistical Thermodynamics* Addison-Wesley, Reading, Massachusetts.
21. Ackers,G.K., Johnson,A.D. and Shea,M.A. (1982) Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl Acad. Sci. USA*, **79**, 1129–1133.
22. 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.
23. Benoff,B., Yang,H., Lawson,C.L., Parkinson,G., Liu,J., Blatter,E., Ebright,Y.W., Berman,H.M. and Ebright,R.H. (2002) Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science*, **297**, 1562–1566.
24. Busby,S. and Ebright,R.H. (1999) Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.*, **293**, 199–213.
25. Setty,Y., Mayo,A.E., Surette,M.G. and Alon,U. (2003) Detailed map of a cis-regulatory input function. *Proc. Natl Acad. Sci. USA*, **100**, 7702–7707.
26. 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.
27. Saiz,L. and Vilar,J.M.G. (2007) Multilevel deconstruction of the in vivo behavior of looped DNA-protein complexes. *PLoS ONE*, **2**, e355.
28. 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.
29. Saiz,L., Rubi,J.M. and Vilar,J.M.G. (2005) Inferring the in vivo looping properties of DNA. *Proc. Natl Acad. Sci. USA*, **102**, 17642–17645.
30. 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.
31. Swint-Kruse,L. and Matthews,K.S. (2004) Thermodynamics, protein modification, and molecular dynamics in characterizing lactose repressor protein: strategies for complex analyses of protein structure-function. *Meth. enzymol.*, **379**, 188–209.
32. Abo,T., Inada,T., Ogawa,K. and Aiba,H. (2000) SsrA-mediated tagging and proteolysis of LacI and its role in the regulation of lac operon. *EMBO J.*, **19**, 3762–3769.
33. Horton,N., Lewis,M. and Lu,P. (1997) Escherichia coli lac repressor-lac operator interaction and the influence of allosteric effectors. *J. Mol. Biol.*, **265**, 1–7.
34. Borisov,N.M., Markevich,N.I., Hoek,J.B. and Kholodenko,B.N. (2005) Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. *Biophys. J.*, **89**, 951–966.
35. Bray,D. and Duke,T. (2004) Conformational spread: the propagation of allosteric states in large multiprotein complexes. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 53–73.
36. Guet,C.C., Elowitz,M.B., Hsing,W. and Leibler,S. (2002) Combinatorial synthesis of genetic networks. *Science*, **296**, 1466–1470.
37. 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.
38. Vilar,J.M.G. (2006) Modularizing gene regulation. *Mol. Syst. Biol.*, **2**, 2006.0016.