

***In vivo* and *in vitro* consistency of thermodynamic models for transcription regulation**J. Landman **Van 't Hoff Laboratory for Physical & Colloid Chemistry, 3584 CH Utrecht, the Netherlands
and European Synchrotron Radiation Facility, 38000 Grenoble, France*

R. N. Georgiev

Process & Energy, 3ME, TU Delft, 2628 CB, Delft, the Netherlands

M. Rydenfelt

*Institute of Pathology, Charite Universitätsmedizin Berlin, Chariteplatz 1, 10117 Berlin, Germany*W. K. Kegel †*Van 't Hoff Laboratory for Physical & Colloid Chemistry, 3584 CH Utrecht, the Netherlands*

(Received 30 August 2019; published 12 November 2019)

The ability of cells to respond to external stimuli, by up and down regulation of genes, is a key strategy for survival in changing environmental conditions. Models based on equilibrium statistical mechanics have been able to successfully predict *in vivo* fold changes in gene expression by computing weights of configurational states of the promoter where genes can be expressed. These models are based on the same, perhaps unintuitive, assumption that transcription initiation—an inherently nonequilibrium process—can indeed be effectively described by the equilibrium binding of transcription factors (TFs) and polymerases to the promoter. The few earlier studies that independently test this assumption [P. H. von Hippel *et al.*, *Proc. Natl. Acad. Sci. USA* **71**, 4808 (1974)], were published before much of our modern day understanding of molecular biology was established, and their models fail to explain more recent experimental results. As such, it is not obvious that the original conclusions remain valid. Equilibrium models depend on fitted free energy differences between binding of TFs to specific operator sites versus nonspecific DNA. In this article we compare the fitted binding free energy of the well-studied LacI repressor to equilibrium binding constants measured in independent *in vitro* experiments. To make this comparison we take into account the distribution of binding energies of the transcription factor to the nonspecific DNA, and we adjust LacI binding constants to a common set of physiological conditions. We find that the fitted binding energies of the LacI repressor *in vivo* indeed agree with *in vitro* measured equilibrium binding constants, reestablishing the idea that equilibrium statistical mechanical models of transcriptional regulation should be viewed not merely as mathematical tools, but also as informative physical representations of underlying TF-DNA interactions.

DOI: [10.1103/PhysRevResearch.1.033094](https://doi.org/10.1103/PhysRevResearch.1.033094)**I. INTRODUCTION**

Half a century ago, von Hippel *et al.* [1] measured the binding constant of the transcription factor LacI to isolated DNA fragments on which the *lac* promoter of *Escherichia coli* was either present or not. They found that the ratio of the two equilibrium constants could reproduce fold changes in wild-type *in vivo* activity of the *lac* promoter, based on a simple

equilibrium adsorption model built on the following premise: if the transcription factor is bound to the promoter, the gene stops transcribing. Equilibrium-based models have since been used successfully to quantitatively predict the transcriptional activity of many genes in the presence of transcription factors (TFs) [1–18].

It is important to note that all these models are based on the same cornerstone: the assumption that transcription—an inherently nonequilibrium process—is regulated by equilibrium binding of transcription factors. The transcriptional activity of a gene is taken as proportional to the (equilibrium) probability that its promoter region is occupied by RNA polymerase (RNAP), an assumption which is justified when the formation of the RNAP open complex on the promoter site of a gene is slow in comparison to the binding and unbinding kinetics of TFs over the genome [19–22]. Under these assumptions, equilibrium statistical mechanics is used to calculate the RNAP occupancy. The assumptions needed to treat transcription

*Present address: Laboratory for Physical Chemistry, Eindhoven University of Technology, 5612 AZ Eindhoven, the Netherlands.

†W.K.Kegel@uu.nl

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

regulation as a quasiequilibrium process are subtle (see, e.g., Refs. [23,24] and the Supplemental Material [25]), and there exists a corresponding class of kinetic models which do not require as many assumptions, at the cost of requiring more parameters [26–32]. Thermodynamic models can be used to fit a binding free energy of TFs to their DNA sites from *in vivo* experiments, which yield values that are internally consistent [10,14,16,17,33–35]; that is, different experimental datasets obtained under different conditions can all be described by the same key quantities in a thermodynamic model. While internal consistency is a strong argument for the plausibility of a model, it does not provide a true verification that the model reflects the actual mechanism. It is far more likely that a model is grounded in reality when quantities have been verified by independent experiments, such as the determination of Avogadro's number [36], or the independent verification of many quantities in the standard model of particle physics [37]. As it stands, the possibility remains that the fitted quantity is a kinetic parameter that is erroneously interpreted as an equilibrium binding free energy.

To our knowledge, the seminal work of von Hippel *et al.* [1] provides the only independent evidence that equilibrium binding free energies can be used to predict fold changes in transcriptional activity [38]. However, knowledge of the statistical mechanics of TF binding, as well as of the architecture of the *lac* promoter has increased to such an extent that it is not *a priori* obvious that their conclusions survive. In particular, binding of the TF to non-specific DNA (the reference state) is still DNA sequence specific, so there is a distribution of binding energies. Moreover, the *lac* promoter used by von Hippel has multiple auxiliary binding sites for the transcription factor—which were not known at the time—which allows double, triple, or looped configurations. We will show in this article that the model of von Hippel *et al.* [1] is not able to reproduce more recent experiments because knowledge of the *lac* promoter was incomplete at the time. As such, an independent verification of the quantities in equilibrium thermodynamic models for transcription initiation is missing.

In this paper we aim to systematically and critically evaluate the idea that equilibrium binding of transcription factors is the mechanism behind the regulation of transcription initiation. To this end, we compare the free energy of TF binding, as fitted by thermodynamic equilibrium models, to multiple independent *in vitro* experiments. Here, we choose to focus on the binding of the *lac* repressor LacI of *Escherichia coli*—a well-studied model architecture with general applicability—and vary the binding target on the DNA and thereby change the binding affinity of the TF to the substrate. The comparison needs to be independent: any quantitative experimental determinations should not be based on the assumption that fold changes in transcriptional activity are proportional to the equilibrium binding probability of RNAP to the promoter. Furthermore, the binding free energies that are found need to correspond to single sites only and not be complicated by cooperative binding of the transcription factor. This way, the binding free energies can be directly compared without introducing additional assumptions. Finally, we must be assured that binding free energies are compared to the appropriate reference energy, in both the thermodynamic model and the independent experiment.

As a rule, interactions between the molecular building blocks of living cells are strongly influenced by their surroundings [39]. For this reason, in order to independently determine the strength of a biological interaction, experiments *in vivo* would be preferred if they were available. However, *in vivo* measurements are limited to few well-designed experiments and not all parameters are accessible with current experimental methods. Direct comparison between *in vivo* and *in vitro* binding free energies is not *a priori* straightforward because of the presence of many different cellular components, as well as differences in pH, salt concentrations, and temperature. These differences are expected to significantly affect the affinity of TFs for their targets on DNA. However, the transcription factors that we consider have a very high affinity for DNA, even outside of their specific binding sites [1,40,41]. Consequently, these transcription factors are hardly ever found in solution and are overwhelmingly more likely to be bound to nonspecific DNA. As such, the relevant reference state of the TFs considered here is not the solution state, but rather the situation where the transcription factor is bound to nonspecific sequences on DNA. As a consequence, the influence of the crowded cell environment is expected to cancel.

The affinity of transcription factors for DNA depends on sequence-specific interactions such as hydrogen bonding [42–44], and is therefore dependent on the local nucleotide sequence [13,45,46]. As such, there is no single binding free energy for a transcription factor to nonspecific DNA. Rather, there is a distribution of binding free energies. In this article, we start by calculating a single *effective* binding free energy for a distribution of binding sites, inspired by previous work by von Hippel and Berg [47] and Gerland *et al.* [48]. The effective binding free energy is related to the equilibrium constant and can be measured *in vitro*. Proceeding, we systematically gather LacI binding constants from many different *in vitro* equilibrium binding experiments in the literature, to both nonspecific DNA and to the four known operator sites, each with their own TF binding affinity. Taking care that the *in vitro* experiments were performed on single operator sites, we calculate the *in vitro* difference in LacI binding energy between specific and nonspecific DNA. In addition, we recalculate observed binding constants to a common set of conditions that are relevant *in vivo*. We find that the fitted energy differences from *in vivo* experiments using thermodynamic models closely match the independently measured *in vitro* binding free energy differences, providing a strong case that the quantity that governs transcriptional activity is indeed a true equilibrium binding free energy, and not an effective kinetic parameter.

II. METHODS

We will use the formalism of the grand canonical ensemble, a natural ensemble to work in when dealing with multichemical binding. We take the binding of TFs to different DNA sites as uncorrelated, similar to [34].

The observed (*in vitro*) binding constant of a protein binding to a DNA site is directly related to the binding free energy between these objects. We consider a protein P that binds to

DNA sites D in the following equilibrium:



with a binding constant

$$K = \frac{[DP]}{[D][P]} = \frac{\theta}{1 - \theta} \frac{1}{[P]}, \quad (2)$$

where θ is the fraction of DNA sites occupied by P. The free concentration $[P]$ is related to the chemical potential of P, μ , through the well-known relation (see, e.g., Ref. [49])

$$\mu = \mu^0 + k_B T \log x_P. \quad (3)$$

Here, x_P is the mole fraction of P, which for dilute solutions is related to $[P]$ by

$$[P] = \frac{n_P}{V} = \frac{1}{v_w} \frac{n_P}{n_w} \simeq \frac{x_P}{v_w}, \quad (n_w \gg n_P) \quad (4)$$

with v_w the molecular volume of water. The implied volume scaling will later drop out and will not affect the final result, as we will eventually compare ratios of binding constants.

A. Identical DNA binding sites

If the protein P can bind (either *in vivo* or *in vitro*) to a specific site, s , we write down the grand canonical partition function of that site as

$$\Xi = \sum_{p=0}^1 \lambda^p Z(p) = 1 + \lambda e^{-\beta \epsilon_s}, \quad (5)$$

with $\lambda = e^{\beta \mu}$ the fugacity of a protein P that can adsorb to the lattice site, $\beta = (k_B T)^{-1}$, p the occupancy of the site, and $Z(p)$ the relevant part of the canonical partition function. The first term corresponds to the state where the DNA binding site is free and is therefore given the weight 1. The second term corresponds to a state that has a single molecule of P adsorbed to the binding site. The relevant part of the canonical partition function for the occupied state is the Boltzmann exponent $\exp(-\beta \epsilon_s)$, of the binding (free) energy ϵ_s of the protein P to the (specific) binding site, s . A system of N independent copies of this binding site has a grand canonical partition function of $\Xi_N = \Xi^N$. We can obtain the occupancy θ from the partition function by taking the partial derivative with respect to λ . It follows that the occupancy θ_s is given by the Langmuir isotherm

$$\theta_s = \frac{1}{N} \frac{\lambda}{\Xi_N} \left(\frac{\partial \Xi_N}{\partial \lambda} \right) = \frac{\lambda e^{-\beta \epsilon_s}}{1 + \lambda e^{-\beta \epsilon_s}}. \quad (6)$$

Using this adsorption isotherm and the relation between chemical potential and protein concentration in (3), we can express the binding constant K_s in Eq. (2) as

$$K_s = v_w e^{-\beta(\epsilon_s - \mu^0)}. \quad (7)$$

The binding constant reflects the equilibrium between the protein P bound to its specific site on the DNA and P in solution.

B. Distribution of binding sites

The binding affinity of proteins P to nonspecific DNA varies with the sequence of the DNA. We consider a system of (nonspecific) DNA binding sites with a distribution in

the binding free energy of P. The grand canonical partition function of adsorption onto a distribution of N_{ns} binding sites is given by

$$\Xi_{ns} = \prod_{i=1}^{N_{ns}} (1 + \lambda e^{-\beta \epsilon_i}), \quad (8)$$

where $\lambda = e^{\beta \mu}$ is again the fugacity of the TF, and ϵ_i the binding free energy of site i . If we take the logarithm of this expression and isolate the factor N_{ns} , the resulting sum $\sum_{i=1}^{N_{ns}} N_{ns}^{-1} \log [1 + \lambda \exp(-\beta \epsilon_i)]$ can be interpreted as an ensemble average. Consequently, we can write

$$\log \Xi_{ns} = N_{ns} (\log(1 + \lambda e^{-\beta \epsilon})). \quad (9)$$

When the distribution is sufficiently narrow, around $\sigma \leq 2 k_B T$ for biological relevant parameters (see Supplemental Material [50] and [51,52]), we can approximate (9) with

$$\log \Xi_{ns} \simeq N_{ns} \log(1 + \lambda \langle e^{-\beta \epsilon} \rangle). \quad (10)$$

This approximation also makes the factor that needs to be averaged independent of TF fugacity. The average $\langle \exp(-\beta \epsilon) \rangle$ can be expanded into the series

$$\langle e^{-\beta \epsilon} \rangle = \left\langle \sum_{n=0}^{\infty} \frac{(-\beta \epsilon)^n}{n!} \right\rangle = \sum_{n=0}^{\infty} (-\beta)^n \frac{\langle \epsilon^n \rangle}{n!} = M_\epsilon(-\beta), \quad (11)$$

where $\langle \epsilon^n \rangle$ is the n th raw moment of the distribution. This series is known as the moment-generating function of the distribution, $M_\epsilon(-\beta)$. It is convenient to introduce the cumulant-generating function $K_\epsilon(-\beta) = \log M_\epsilon(-\beta)$. The cumulant-generating function can also be expressed as a series expansion [53],

$$K_\epsilon(-\beta) = \sum_{n=1}^{\infty} \kappa_n \frac{(-\beta)^n}{n!} = -\beta \langle \epsilon \rangle + \frac{\beta^2 \sigma^2}{2} - \frac{\beta^3 \gamma_1 \sigma^3}{6} + \dots, \quad (12)$$

where κ_n is the n th cumulant of the distribution. The advantage of using the cumulant-generating function is that cumulants are directly related to observable quantities of the distribution, such as the mean $\langle \epsilon \rangle$, variance σ^2 , and skewness γ_1 . If we express (10) in terms of the cumulant-generating function, we obtain

$$\Xi_{ns} \simeq \left[1 + \lambda \exp \left(-\beta \langle \epsilon \rangle + \frac{\beta^2 \sigma^2}{2} - \frac{\beta^3 \gamma_1 \sigma^3}{6} + \dots \right) \right]^{N_{ns}}, \quad (13)$$

where we have also taken the exponent of both sides. We can define an effective energy as the sum of cumulants in the expansion

$$\epsilon_{\text{eff}} \equiv \langle \epsilon \rangle - \frac{\beta \sigma^2}{2} + \frac{\beta^2 \gamma_1 \sigma^3}{6} - \dots \quad (14)$$

so that the expression for the partition function in (13) becomes

$$\Xi_{ns} \simeq (1 + \lambda e^{-\beta \epsilon_{\text{eff}}})^{N_{ns}}. \quad (15)$$

The resulting partition function for a system with a distribution of binding sites is isomorphic to the partition function of a system with identical sites with a binding free energy equal to ϵ_{eff} . The effective free energy is smaller (more negative) than the mean binding free energy of the distribution, since at finite temperatures the TFs favor binding at the lower energy sites.

TABLE I. Binding constants of LacI to nonspecific DNA. The nonspecific DNA was isolated from calf thymus [55,56,59], phage λ [57], or T7 bacteriophage [58]

Author	Year	K_{ns} (M ⁻¹)	T (K)
DeHaseth <i>et al.</i> [55]	1977	2.22×10^3	294
DeHaseth <i>et al.</i> [56]	1977	6.54×10^3	293
Revzin and von Hippel [57]	1977	1.30×10^4	277
Lohman <i>et al.</i> [58]	1980	5.76×10^3	293
Ha <i>et al.</i> [59]	1992	1.87×10^4	277

The effective energy greatly simplifies the adsorption isotherm corresponding to that system. The properties of the distributions are essentially condensed into a single parameter. Since in general we do not know the energy distribution of binding sites, the effective energy is a very useful quantity that implicitly carries the information of the distribution.

We can calculate the occupancy of the nonspecific sites, θ_{ns} , from the grand canonical partition function using Eq. (6),

$$\theta_{ns} = \frac{1}{N_{ns}} \frac{\lambda}{\Xi_{ns}} \left(\frac{\partial \Xi_{ns}}{\partial \lambda} \right) = \frac{\lambda e^{-\beta \epsilon_{eff}}}{1 + \lambda e^{-\beta \epsilon_{eff}}}. \quad (16)$$

With this adsorption isotherm, combined with the expression for the binding constant in Eq. (2), we write down the nonspecific equilibrium constant K_{ns}

$$K_{ns} \simeq v_w e^{-\beta(\epsilon_{eff} - \mu^0)}, \quad (17)$$

and we find an expression very similar to Eq. (7).

By defining the effective energy, we have a reference point for the binding free energy of TFs to nonspecific DNA, which is less sensitive to offsets in free energy due to the presence of other solutes, and which can be used even when the distribution of binding free energies is unknown. Moreover, the difference between the specific binding free energy ϵ_s and the effective binding free energy of TFs to nonspecific DNA can immediately be compared to the experiments from the ratio of the observed binding constants

$$\frac{K_s}{K_{ns}} = e^{-\beta(\epsilon_s - \epsilon_{eff})}. \quad (18)$$

III. RESULTS AND DISCUSSION

We obtained from literature the experimental binding constants from *in vitro* binding assays of the wild-type repressor transcription factor LacI (wild-type *E. coli*), to the operator sequences O1, O2, and O3, the symmetrical operator Oid, and nonspecific DNA.

Table I lists observed binding constants of LacI to *E. coli* nonspecific DNA. The binding constants have been measured at various ionic strengths. Here we modify binding constants to the value at a physiological salt concentration of 200 mM of NaCl, room temperature, and pH 7.5 (see Supplemental Material [54]). Applying the relation between observed binding constant and the binding free energy in Eq. (17), we can calculate the effective binding free energy ϵ_{eff} of LacI to nonregulatory DNA.

In Table II we list observed binding constants of LacI to the *lac* operon wild-type operators and to the symmetric operator Oid. We recalculated the reported value of the bind-

TABLE II. Binding constants of LacI to the wild-type operators O1, O2, and O3 of the *lac* operon and to the symmetric operator Oid.

Site	Author	Year	K_s (M ⁻¹)	T (K)
Oid	Ha <i>et al.</i> [59]	1992	3.54×10^{11}	296
	Frank <i>et al.</i> [60]	1997	7.54×10^{10}	297
	Tsodikov <i>et al.</i> [61]	1999	1.54×10^{11}	297
O1	O’Gorman <i>et al.</i> [62]	1980	1.75×10^{10}	293
	Whitson <i>et al.</i> [63]	1986	1.46×10^{10}	298
	Spotts <i>et al.</i> [64]	1991	2.19×10^{10}	293
	Chakerian and Matthews [65]	1991	9.78×10^9	293
	Zhang and Gottlieb [66]	1993	2.56×10^{10}	298
	Schlax <i>et al.</i> [67]	1995	1.55×10^{10}	310
	Swint-Kruse <i>et al.</i> [68]	2005	2.46×10^{10}	293
	Winter and von Hippel [69]	1981	3.96×10^{10}	293
O2	Winter and von Hippel [69]	1981	3.07×10^{10}	298
	Romanuka <i>et al.</i> [70]	2009	9.76×10^9	293
O3	Winter and von Hippel [69]	1981	1.73×10^7	298
	Romanuka <i>et al.</i> [70]	2009	9.76×10^6	293

ing constant to physiological conditions. In the Supplemental Material [54] we provide more details on these corrections.

In the work of Garcia and Phillips [14] LacI binding energies were fit to *in vivo* measurements of *E. coli* under minimal growth conditions, using a thermodynamic model. We see the model in [14] as representative of the class of thermodynamic models for transcription regulation. We show the values they report in Table III, together with *in vitro* data calculated from the ratio of the binding constants according to Eq. (18). Figure 1 (blue) shows a graphic representation of these results. Especially for the stronger binding operator sites, the correspondence between *in vivo* and *in vitro* data is convincing, within $1 k_B T$. For the weaker binding O3 operator there is a larger mismatch between *in vivo* and *in vitro* data. This is likely a reflection of the scarcity of data published on the auxiliary operator, combined with a large uncertainty in determining low-affinity binding [69,70].

A. Earlier work

In the seminal work of von Hippel *et al.* [1], estimates from *in vitro* experiments for the binding constants of tetrameric LacI to its cognate site on the *lac* operon, as well as to nonspecific DNA, were used to calculate *in vivo* fold changes in

TABLE III. Comparison between *in vitro* and *in vivo* binding free energies of LacI to operator sequences O1, O2, O3, and Oid, offset by the binding free energy to nonspecific DNA. The *in vivo* binding affinities were obtained by Garcia and Phillips [14] (row 2) and Vilar and Saiz [16] (rows 3 and 4). The binding free energies in row 3 were rescaled to the size of the nonspecific genome in row 4.

$\Delta\epsilon/k_B T$	Oid	O1	O2	O3
<i>In vitro</i>	-16.9	-14.8	-14.7	-7.5
Garcia 2011	-17.0	-15.3	-13.9	-9.7
Vilar 2013		-20.8	-18.5	-15.0
Vilar 2013, rescaled		-15.9	-13.6	-10.1

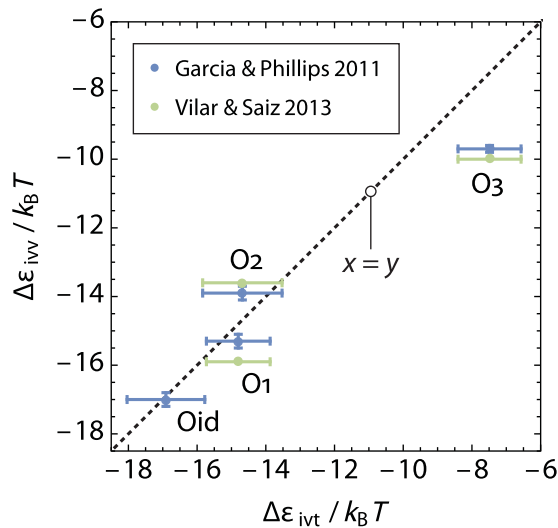


FIG. 1. Comparison of *in vivo* and *in vitro* determinations of the binding free energy of the operator sites of the *lac* operon. The *in vivo* data was taken from Garcia and Phillips [14] (blue points) and Vilar and Saiz [16] [green points, recalculated according to Eq. (22)]. The dotted line denotes the line $x = y$. Data are displayed as mean \pm the standard error of the mean (SEM).

transcriptional activity. They also found that the ratio between the *in vitro* binding constants is sufficient to quantitatively predict the basal transcription level of wild-type *E. coli*. This insight is an important step in the physical interpretation of thermodynamic models. However, with today’s knowledge of the promoter architecture, it is not obvious that their conclusions are justified.

First, the effect of the auxiliary *lac* operators in the promoter neighborhood was unknown at the time of publication of von Hippel *et al.* [1]. The presence of auxiliary operator

sites allow conformations where the DNA is looped between two operator sites simultaneously bound to tetrameric LacI, with a significantly increased free energy. The *in vitro* binding constant used by von Hippel *et al.* [1] consequently is an effective parameter due to the presence of auxiliary sites.

The work in [1] also considered the nonspecific sites on the DNA as a set of identical sites, without taking into account the sequence-specific interaction between the TF and DNA. As we have shown here, a distribution with a finite variance causes a shift in the effective free energy of the reservoir, but the same shift also affects the *in vitro* (measured) binding constant.

In wild-type *E. coli* cells, only ~ 10 copies of LacI exist. The *in vivo* measurements from Jacob and Monod [71] that was used by von Hippel *et al.* [1] were therefore also taken in this limit of low LacI fugacity. Figure 2 shows a comparison of the simple repression model used in [1] with a model taking into account the full promoter architecture. It is for the following reason that a match was found between the calculated and observed fold change in transcriptional activity: in the low LacI fugacity limit the behavior of the *lac* promoter can be approximated with a simple repression model, where LacI always binds in a looped conformation as shown in Fig. 2(b). However, such a model fails when the LacI copy number increases. Had the gene activity measurements used in [1] been measured at a higher LacI copy number, their model would have overestimated the degree of repression by up to an order of magnitude.

Essentially, the work of von Hippel *et al.* [1] was far ahead of its time: for the *lac* promoter architecture, transcriptional activity is regulated by the equilibrium binding of LacI. Reexamining their work almost half a century later, their conclusions are based on incomplete knowledge about the system. However, the systematic match between *in vitro* and *in vivo* binding free energy that can be seen from the

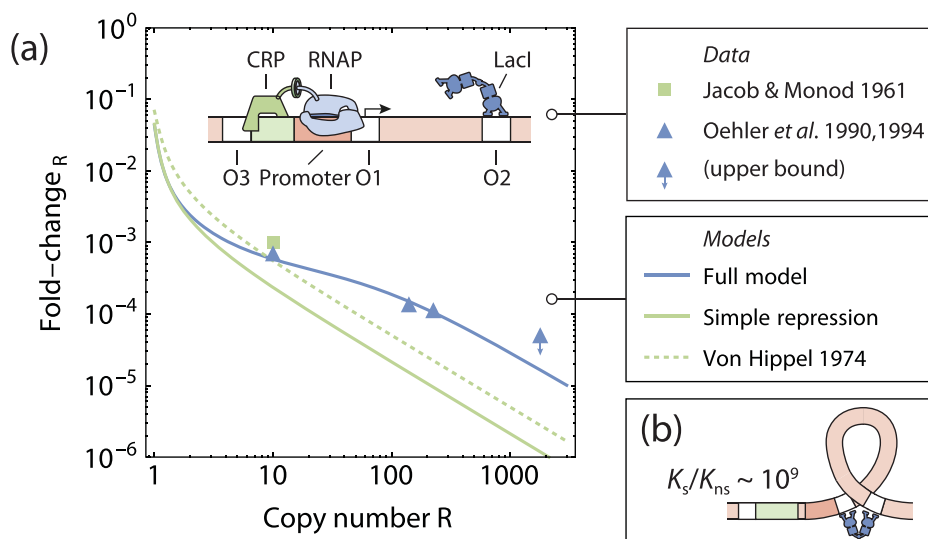


FIG. 2. (a) Fold change of the wild-type *lac* promoter as function of LacI copy number R , compared to two different thermodynamic models. Data are from Jacob and Monod [71] and Oehler *et al.* [72,73]. In blue we show our full model for the intact *lac* promoter architecture (see Landman *et al.* [74]), with operator binding free energies taken from Garcia and Phillips [14] and other quantities taken from thermodynamic fits to other data from [72,73]. The green dotted line is the simple repression model used by von Hippel *et al.* [1] using the *in vitro* ratio of binding constants. The solid green line is the same model, but with the binding free energy calculated using the model in [74] (see the Supplemental Material for more details [75]). (b) Cartoon of the dominant conformation of LacI bound to the promoter architecture.

literature published since shows that, at the core, the quantities in thermodynamic models can be verified from independent experiments.

B. DNA availability

An important point to address here is the influence of the available number of nonspecific sites *in vivo*. In the analysis of the *in vivo* measurements by Garcia and Phillips [14] it is assumed that the whole bacterial genome is (immediately) available for the transcription factors. However, nucleoid associated proteins and supercoiling of the genome are expected to influence the effective number of nonspecific sites. Fixing the number of nonspecific sites in a model will influence the effective binding free energy as extracted from experiments. Garcia and Phillips [14] fit fold-change data for a given number of repressor molecules to the (canonical) expression for the fold change of a gene regulated by a simple repressor [4,8]:

$$\text{Fold change} = \frac{1}{1 + \frac{R}{N_{\text{ns}}} \exp(-\beta \Delta \epsilon_s)}, \quad (19)$$

with R is the number of repressor molecules and N_{ns} the total number of base pairs in a cell. Now let us define $N_{\text{ns}}^* \leq N_{\text{ns}}$ as the the number of nonspecific sites accessible for transcription factors, and not supercoiled or otherwise compacted by nucleoid-associated proteins. Taking into account an effective number of available sites implies we have to replace N_{ns} by N_{ns}^* in (19). Not taking the effect into account leads to an error in the binding free energy $\delta \epsilon_s$ of

$$\delta \epsilon_s = -k_B T \log \frac{N_{\text{ns}}^*}{N_{\text{ns}}}, \quad (20)$$

where $N_{\text{ns}}^*/N_{\text{ns}}$ denotes the fraction of available sites. When the effective number of available sites is within a factor of 2–3 lower than the total number of sites, the error does not exceed the uncertainty range of the *in vitro* binding free energies. A difference in availability of a full order of magnitude should be noticeable as a reduction of the binding free energy of approximately $2k_B T$. This is relevant because DNA is usually significantly compacted [39,41], and TF binding to compacted DNA may be inhibited [76].

C. Nonspecific binding is the appropriate reference state

Vilar and Saiz [16] also report binding energies, based on fits to the experiments of Oehler *et al.* [72,73], but in contrast to the work of Garcia and Phillips [14], their model assumes that LacI is present in solution when not bound to its cognate site. Consequently, the binding free energies reported in their work differ significantly from both the binding free energies of [14] and the *in vitro* experiments (see Table III, row 3). Vilar and Saiz [16] use the experimentally determined relation that one molecule per cell corresponds to a cellular concentration of 1.5 nM, to express the number of repressor molecules in the cell given in [72,73] into units of concentration. They then use a similar expression to determine the binding free energy

as in Eq. (19),

Fold change

$$= \frac{1}{1 + [R] \exp(-\beta \Delta \epsilon_s)} \quad [\text{see Ref. [16], Eq. (11)}, \quad (21)$$

with the factor $[R]$ the concentration of LacI, divided implicitly by a reference concentration of 1 M, replacing R/N_{ns} in (19). This substitution essentially rescales their result to a different reference state, introducing a shift in the binding free energy of

$$\delta \epsilon_s = -k_B T \log \frac{R/N_{\text{ns}}}{[R]} = -k_B T \log \frac{10^3 N_A V_{\text{cell}}}{N_{\text{ns}}}, \quad (22)$$

with V_{cell} the cell volume and N_A Avogadro's constant. The factor 10^3 follows from converting the units of the reported dissociation constant from M to mol m^{-3} . Vilar and Saiz use the relation that 1.5 nM corresponds to one molecule per cell, fixing the size of the cell at $V_{\text{cell}} = (N_A \times 1.5 \text{ nM/molecule})^{-1} = 1.1 \mu\text{m}^3$.

In Table III (row 4) and in Fig. 1 (green) we show the binding free energy of the operator sites of the *lac* operon, after recalculation to the size of the nonspecific genome. Contrary to the originally reported quantities, we see that there is a convincing match between the data from Vilar and Saiz [16] and Garcia and Phillips [14], and the *in vitro* data. This provides additional evidence that it is the nonspecific genome that acts as the relevant reference state for LacI.

IV. CONCLUSION

The power of thermodynamic equilibrium models for transcription regulation has already been extensively demonstrated in the existing literature. However, the quantity governing the fold change in transcriptional activity is interpreted in these models as an equilibrium free energy, and while its value is internally consistent between different thermodynamic equilibrium models, the possibility remains that it is a hidden kinetic parameter. The justification needed to treat transcription regulation as an equilibrium process has so far been provided by the experiments of von Hippel *et al.* [1], yet we found that their conclusions are based on models that cannot account for current-day knowledge of transcription.

In this work, we have investigated the validity of thermodynamic equilibrium models for gene regulation beyond internal consistency by comparing the fitted binding free energy of transcription factors to independent *in vitro* experiments. We find that the agreement between *in vivo* and *in vitro* experiments is quantitative within the error range of the experiments. Our work provides evidence that the quantity that governs transcriptional activity is indeed a true equilibrium binding free energy, and not an effective kinetic parameter. This new evidence confirms what has essentially been (in hindsight) a conjecture: that equilibrium binding can be seen as an informative physical representation of TF action.

This result not only provides significant additional plausibility for thermodynamic models of gene regulation, but also points to a large fraction (being more than roughly one-third of the total genome size) of the nonspecific part of the genome being accessible for transcription factors to bind.

ACKNOWLEDGMENTS

We thank Rob Phillips, Misha Sheinman, and Rutger Hermesen for very pleasant and inspiring discussions on the topic of nonspecific adsorption. J.L. acknowledges financial support was provided by the Netherlands Organisation for Scientific Research (NWO) under Grant No. 022.004.016,

and by the European Synchrotron Radiation Facility (ESRF), Grenoble, France.

J.L. and R.N.G. contributed equally to this work. J.L., R.N.G. and W.K.K. developed methods, J.L. and R.N.G. analyzed data, J.L. wrote the paper, J.L., R.N.G., M.R., W.K.K. revised the paper, and W.K.K. supervised research.

-
- [1] P. H. von Hippel, A. Revzin, C. A. Gross, and A. C. Wang, *Proc. Natl. Acad. Sci. USA* **71**, 4808 (1974).
- [2] G. K. Ackers, A. D. Johnson, and M. A. Shea, *Proc. Natl. Acad. Sci. USA* **79**, 1129 (1982).
- [3] M. A. Shea and G. K. Ackers, *J. Mol. Biol.* **181**, 211 (1985).
- [4] J. M. Vilar and S. Leibler, *J. Mol. Biol.* **331**, 981 (2003).
- [5] N. E. Buchler, U. Gerland, and T. Hwa, *Proc. Natl. Acad. Sci. USA* **100**, 5136 (2003).
- [6] J. M. Vilar and L. Saiz, *Curr. Opin. Genet. Dev.* **15**, 136 (2005).
- [7] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, *Curr. Opin. Genet. Dev.* **15**, 125 (2005).
- [8] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, *Curr. Opin. Genet. Dev.* **15**, 116 (2005).
- [9] Y. Zhang, A. E. McEwen, D. M. Crothers, and S. D. Levene, *PLoS One* **1**, e136 (2006).
- [10] T. Kuhlman, Z. Zhang, J. Saier M. H., and T. Hwa, *Proc. Natl. Acad. Sci. USA* **104**, 6043 (2007).
- [11] E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, and U. Gaul, *Nature* **451**, 535 (2008).
- [12] E. Segal and J. Widom, *Nat. Rev. Genet.* **10**, 443 (2009).
- [13] J. B. Kinney, A. Murugan, C. G. Callan, and E. C. Cox, *Proc. Natl. Acad. Sci. USA* **107**, 9158 (2010).
- [14] H. G. Garcia and R. Phillips, *Proc. Natl. Acad. Sci. USA* **108**, 12174 (2011).
- [15] L. Keren, O. Zackay, M. Lotan-Pompan, U. Barenholz, E. Dekel, V. Sasson, G. Aidelberg, A. Bren, D. Zeevi, A. Weinberger, U. Alon, R. Milo, and E. Segal, *Mol. Syst. Biol.* **9**, 701 (2013).
- [16] J. M. G. Vilar and L. Saiz, *ACS Synth. Biol.* **2**, 576 (2013).
- [17] R. C. Brewster, F. M. Weinert, H. G. Garcia, D. Song, M. Rydenfelt, and R. Phillips, *Cell* **156**, P1312 (2014).
- [18] M. Rydenfelt, R. S. Cox, H. Garcia, and R. Phillips, *Phys. Rev. E* **89**, 012702 (2014).
- [19] D. K. Hawley and W. R. McClure, *J. Mol. Biol.* **157**, 493 (1982).
- [20] H. Buc and W. R. McClure, *Biochemistry* **24**, 2712 (1985).
- [21] N. Mitarai, I. B. Dodd, M. T. Crooks, and K. Sneppen, *PLoS Comput. Biol.* **4**, e1000109 (2008).
- [22] N. Mitarai, S. Semsey, and K. Sneppen, *Phys. Rev. E* **92**, 022710 (2015).
- [23] J. Elf, G. W. Li, and X. S. Xie, *Science* **316**, 1191 (2007).
- [24] A. Sanchez, M. L. Osborne, L. J. Friedman, J. Kondev, and J. Gelles, *EMBO J.* **30**, 3940 (2011).
- [25] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.1.033094> for a discussion about the assumptions behind the application of equilibrium-based models.
- [26] M. S. H. Ko, *J. Theor. Biol.* **153**, 181 (1991).
- [27] J. Peccoud and B. Ycart, *Theor. Popul. Biol.* **48**, 222 (1995).
- [28] M. T. Record Jr., W. S. Reznikoff, M. L. Craig, K. L. McQuade, and P. J. Schlax, in *In Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by N. F. C. *et al.* (ASM Press, Washington, 1996) pp. 792–821.
- [29] T. B. Kepler and T. C. Elston, *Biophys. J.* **81**, 3116 (2001).
- [30] A. Sanchez and J. Kondev, *Proc. Natl. Acad. Sci. USA* **105**, 5081 (2008).
- [31] D. Michel, *Prog. Biophys. Mol. Biol.* **102**, 16 (2010).
- [32] R. Phillips, *Annu. Rev. Condens. Matter. Phys.* **6**, 85 (2015).
- [33] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt, *Nature* **491**, 249 (2012).
- [34] F. M. Weinert, R. C. Brewster, M. Rydenfelt, R. Phillips, and W. K. Kegel, *Phys. Rev. Lett.* **113**, 258101 (2014).
- [35] L. A. Sepúlveda, H. Xu, J. Zhang, M. Wang, and I. Golding, *Science* **351**, 1218 (2016).
- [36] P. Becker, *Rep. Prog. Phys.* **64**, 1945 (2001).
- [37] K. A. Olive and Particle Data Group, *Chin. Phys. C* **38**, 090001 (2014), [arXiv:gr-qc/0402007](https://arxiv.org/abs/gr-qc/0402007).
- [38] By independent we mean a verification based on an independent experiment using a model that is not based on the same assumption that the fold change in transcription activity is proportional to the probability RNAP binds to the promoter. Virtually all of the models mentioned above take as input parameters a fitted value based on similar equilibrium models for transcription initiation and therefore fail this test of independency. Some models mentioned above do take independent *in vitro* binding data into account. The work of Ackers *et al.* [2] and Shea and Ackers [3] is based mostly on *in vitro* binding free energy but still includes a fitting parameter for the binding free energy of nonspecific sites. The model by von Hippel *et al.* [1] is fully based on *in vitro* binding free energies but as detailed further below and in the Supplemental Material [77], their model is based on an incomplete picture of the binding mechanism of LacI and as such cannot explain more recent measurements.
- [39] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell*, 5th ed. (Garland Science, New York, 2008).
- [40] G. D. Stormo and D. S. Fields, *Trends Biochem. Sci.* **23**, 109 (1998).
- [41] R. Phillips, J. Kondev, J. Theriot, H. G. Garcia, and N. Orme, *Physical Biology of the Cell*, 2nd ed. (Garland Science, New York, 2012).
- [42] M. T. Record, C. F. Anderson, and T. M. Lohman, *Q. Rev. Biophys.* **11**, 103 (1978).
- [43] M. D. Barkley, P. a. Lewis, and G. E. Sullivan, *Biochemistry* **20**, 3842 (1981).
- [44] C. G. Kalodimos, N. Biris, A. M. Bonvin, M. M. Levandoski, M. Guennegues, R. Boelens, and R. Kaptein, *Science* **305**, 386 (2004).

- [45] J. M. Vilar, *Biophys. J.* **99**, 2408 (2010).
- [46] R. C. Brewster, D. L. Jones, and R. Phillips, *PLoS Comput. Biol.* **8**, e1002811 (2012).
- [47] P. H. von Hippel and O. G. Berg, *Proc. Natl. Acad. Sci.* **83**, 1608 (1986).
- [48] U. Gerland, J. D. Moroz, and T. Hwa, *Proc. Natl. Acad. Sci. USA* **99**, 12015 (2002).
- [49] J. Israelachvili, *Intermolecular and Surface Forces*, 3rd ed. (Academic Press, New York, 2011).
- [50] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.1.033094> for a discussion about when the distribution is sufficiently narrow to apply this approximation.
- [51] M. Slutsky and L. A. Mirny, *Biophys. J.* **87**, 4021 (2004).
- [52] E. G. Marklund, A. Mahmutovic, O. G. Berg, P. Hammar, D. van der Spoel, D. Fange, and J. Elf, *Proc. Natl. Acad. Sci.* **110**, 19796 (2013).
- [53] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, 2nd ed. (Griffin, London, 1963).
- [54] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.1.033094> for details about the recalculations needed to set the measured binding constants to a common physiological standard.
- [55] P. L. DeHaseth, T. M. Lohman, and M. T. Record, *Biochemistry* **16**, 4783 (1977).
- [56] P. L. DeHaseth, C. A. Gross, R. R. Burgess, and M. T. Record Jr., *Biochemistry* **16**, 4777 (1977).
- [57] A. Revzin and P. H. von Hippel, *Biochemistry* **16**, 4769 (1977).
- [58] T. M. Lohman, C. G. Wensley, J. Cina, R. R. Burgess, and M. T. Record, *Biochemistry* **19**, 3516 (1980).
- [59] J. H. Ha, M. W. Capp, M. D. Hohenwarter, M. Baskerville, and M. T. Record, *J. Mol. Biol.* **228**, 252 (1992).
- [60] D. E. Frank, R. M. Saecker, J. P. Bond, M. W. Capp, O. V. Tsodikov, S. E. Melcher, M. M. Levandoski, and M. T. Record Jr., *J. Mol. Biol.* **267**, 1186 (1997).
- [61] O. V. Tsodikov, R. M. Saecker, S. E. Melcher, M. M. Levandoski, D. E. Frank, M. W. Capp, and M. T. Record, *J. Mol. Biol.* **294**, 639 (1999).
- [62] R. B. O’Gorman, M. Dunaway, and K. S. Matthews, *J. Biol. Chem.* **255**, 10100 (1980).
- [63] P. A. Whitson, J. S. Olson, and K. S. Matthews, *Biochemistry* **25**, 3852 (1986).
- [64] R. O. Spotts, A. E. Chakerian, and K. S. Matthews, *J. Biol. Chem.* **266**, 22998 (1991).
- [65] A. E. Chakerian and K. S. Matthews, *J. Biol. Chem.* **266**, 22206 (1991).
- [66] X. Zhang and P. A. Gottlieb, *Biochemistry* **32**, 11374 (1993).
- [67] P. J. Schlux, M. W. Capp, and M. T. Record Jr., *J. Mol. Biol.* **245**, 331 (1995).
- [68] L. Swint-Kruse, H. Zhan, and K. S. Matthews, *Biochemistry* **44**, 11201 (2005).
- [69] R. B. Winter and P. H. von Hippel, *Biochemistry* **20**, 6948 (1981).
- [70] J. Romanuka, G. E. Folkers, N. Biris, E. Tishchenko, H. Wienk, A. M. J. J. Bonvin, R. Kaptein, and R. Boelens, *J. Mol. Biol.* **390**, 478 (2009).
- [71] F. Jacob and J. Monod, *J. Mol. Biol.* **3**, 318 (1961).
- [72] S. Oehler, E. R. Eismann, H. Kramer, and B. Muller-Hill, *EMBO J.* **9**, 973 (1990).
- [73] S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, and B. Müller-Hill, *EMBO J.* **13**, 3348 (1994).
- [74] J. Landman, R. C. Brewster, F. M. Weinert, R. Phillips, and W. K. Kegel, *PLoS One* **12**, e0179235 (2017).
- [75] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.1.033094> for more details about the *lac* promoter model used.
- [76] A. Travers and G. Muskhelishvili, *Nat. Rev. Microbiol.* **3**, 157 (2005).
- [77] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.1.033094> for an extensive discussion why the conclusions of von Hippel *et al.* [1] are based on an incomplete picture of the binding mechanism.