

RESEARCH ARTICLE

Theoretical analysis of inducer and operator binding for cyclic-AMP receptor protein mutants

Tal Einav¹, Julia Duque², Rob Phillips^{1,3,4*}

1 Department of Physics, California Institute of Technology, Pasadena, CA, United States of America, **2** Department of Physics and the London Centre for Nanotechnology, University College London, London, United Kingdom, **3** Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, United States of America, **4** Department of Applied Physics, California Institute of Technology, Pasadena, CA, United States of America

* phillips@pboc.caltech.edu



OPEN ACCESS

Citation: Einav T, Duque J, Phillips R (2018) Theoretical analysis of inducer and operator binding for cyclic-AMP receptor protein mutants. *PLoS ONE* 13(9): e0204275. <https://doi.org/10.1371/journal.pone.0204275>

Editor: Jordi Garcia-Ojalvo, Universitat Pompeu Fabra, SPAIN

Received: July 5, 2018

Accepted: September 5, 2018

Published: September 26, 2018

Copyright: © 2018 Einav et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported in the RP group by La Fondation Pierre-Gilles de Gennes, the Rosen Center at Caltech (<http://rosen.caltech.edu/>), and the National Institutes of Health (<https://www.nih.gov/>) through DP1 OD000217 (Director's Pioneer Award), R01 GM085286, and 1R35 GM118043-01 (MIRA). We are grateful to the Burroughs-Wellcome Fund (<https://www.bwfund.org/>) for its support of the Physiology Course at the Marine

Abstract

Allosteric transcription factors undergo binding events at inducer binding sites as well as at distinct DNA binding domains, and it is difficult to disentangle the structural and functional consequences of these two classes of interactions. We compare the ability of two statistical mechanical models—the Monod-Wyman-Changeux (MWC) and the Koshland-Némethy-Filmer (KNF) models of protein conformational change—to characterize the multi-step activation mechanism of the broadly acting cyclic-AMP receptor protein (CRP). We first consider the allosteric transition resulting from cyclic-AMP binding to CRP, then analyze how CRP binds to its operator, and finally investigate the ability of CRP to activate gene expression. We use these models to examine a beautiful recent experiment that created a single-chain version of the CRP homodimer, creating six mutants using all possible combinations of the wild type, D53H, and S62F subunits. We demonstrate that the MWC model can explain the behavior of all six mutants using a small, self-consistent set of parameters whose complexity scales with the number of subunits, providing a significant benefit over previous models. In comparison, the KNF model not only leads to a poorer characterization of the available data but also fails to generate parameter values in line with the available structural knowledge of CRP. In addition, we discuss how the conceptual framework developed here for CRP enables us to not merely analyze data retrospectively, but has the predictive power to determine how combinations of mutations will interact, how double mutants will behave, and how each construct would regulate gene expression.

Introduction

Transcriptional regulation lies at the heart of cellular decision making, and understanding how cells modify the myriad of players involved in this process remains challenging. The cyclic-AMP receptor protein (CRP; also known as the catabolite receptor protein, CAP) is an

Biological Laboratory, where part of the work on this work was done, and for a post-course research grant (JD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

allosteric transcription factor that regulates over 150 genes in *Escherichia coli* [1–4]. Upon binding to cyclic-AMP (cAMP), the homodimeric CRP undergoes a conformational change whereby two alpha helices reorient to open a DNA binding domain [5], allowing CRP to bind to DNA and affect transcription [6–8]. While much is known about the molecular details of CRP and how different mutations modify its functionality [9, 10], each new CRP mutant is routinely analyzed in isolation using phenomenological models. We argue that given the hard-won structural insights into the conformational changes of proteins like CRP, it is important to test how well mechanistically motivated models of such proteins can characterize the wealth of available data.

The picture that has emerged from various domains of biology is that allostery involves the interplay of a spectrum of dynamically linked states [11–17]. In some systems, it is straightforward to partition these states into the physiologically relevant categories; for example, CRP naturally divides into the cAMP unbound, singly bound, and doubly bound states as well as the DNA bound and unbound states. Nuclear magnetic resonance (NMR) and isothermal titration calorimetry (ITC) have begun to tease out the precise thermodynamics of the underlying interactions between these states [18, 19]. These methods have demonstrated that allosteric regulation in CRP includes both large structural changes as well as entropic modifications that make the protein more rigid [20, 21]. In this work, we ask whether we can capitalize upon this detailed knowledge of the system to construct a coarse-grained model of the multi-step activation cycle of CRP shown in [Fig 1(A)] using a compact set of parameters. Specifically, we investigate variants of the Monod-Wyman-Changeux (MWC) model, which posits that both CRP subunits fluctuate concurrently between an active and inactive conformational state [22], and the Koshland-Némethy-Filmer (KNF) model, which proposes that each subunit must independently transition from an inactive to active state upon ligand binding [23], adapted for

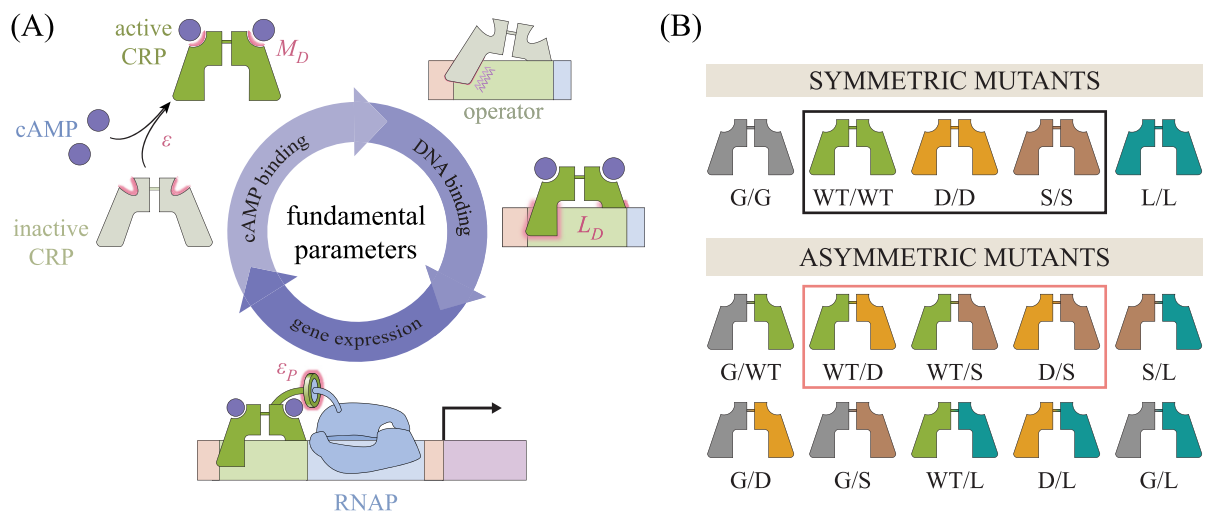


Fig 1. Key parameters governing CRP function. (A) Within the MWC and KNF models, each CRP subunit can assume either an active or an inactive conformation with a free energy difference ϵ between the two states. cAMP can bind to CRP (with a dissociation constant M_D^A in the active state and M_D^I in the inactive state) and promotes the active state ($M_D^A < M_D^I$ in the MWC model; $M_D^I \rightarrow \infty$ in the KNF model). Active CRP has a higher affinity for the operator (L_D^A) than the inactive state (L_D^I). When CRP is bound to DNA, it promotes RNA polymerase binding through an interaction energy ϵ_P , thereby enhancing gene expression. (B) Lanfranco *et al.* constructed a single-chain CRP molecule whose two subunits could be mutated independently. All possible dimers are shown using five mutant subunits: wild type (WT), D (D53H), S (S62F), G (G141Q), and L (L148R). Lanfranco *et al.* constructed the six mutants comprised of WT, D, and S (black and pink boxes) and analyzed each mutant independently.

<https://doi.org/10.1371/journal.pone.0204275.g001>

the CRP system. These two models have been investigated in a wide variety of allosteric systems, and evidence for both models as well as their shortcomings have been extensively analyzed [24–28]. Nevertheless, the simple thermodynamic view provided by the MWC and KNF models provides fertile ground to both verify how well we understand the critical factors governing CRP behavior as well as to explore hypotheses about mutational perturbations to the system.

Our paper is inspired by a recent *in vitro* study of CRP performed by Lanfranco *et al.* who engineered a single-chain CRP molecule whose two subunits are tethered together by an unstructured polypeptide linker [29]. This construct enabled them to mutate each subunit independently, providing a novel setting within which to analyze the combinatorial effects of mutations. Specifically, they took three distinct CRP subunits—the wild type (WT) subunit and the well characterized mutations D53H and S62F (denoted D and S, respectively) originally chosen to perturb the transcription factor's cAMP binding domain [30, 31]—and linked them together in every possible combination to create six CRP mutants as shown in Fig 1(B) (black and pink boxes). Lanfranco *et al.* measured the cAMP-binding and DNA-binding capabilities of these mutants, separating these two key components of transcription factor activation. In this work, we present an analysis of these CRP mutants that demonstrates how their diverse phenotypes are related by their subunit compositions.

More specifically, the effects of mutations are often difficult to interpret, and indeed the results from Lanfranco *et al.* showed no clear pattern. The behavior of each mutant was analyzed *independently* by fitting its binding curve to a second order polynomial [29]. In this work, we propose an alternative framework that bolsters our understanding of the system in two significant ways: (1) we link the response functions of each CRP construct to its subunit composition, closing the gap between structure and function and (2) the number of parameters in our model scales linearly with the number of subunits whereas the number of parameters in the original analysis scaled with the number of CRP mutants (i.e. the square of the number of subunits). The advantage of this scaling behavior grows with the number of subunits. For example, this work focuses on the CRP mutants made by Lanfranco *et al.* using three subunits (black and pink boxes in Fig 1(B)). If we include two additional well-characterized mutants—such as G141Q (G) and L148R (L) [32]—for a total of $N = 5$ subunits, our model would only require $2N = 10$ parameters to describe the $\frac{N(N+1)}{2} = 15$ mutants whereas a model analyzing each mutant independently would require 30 parameters (2 per mutant). With $N = 10$ subunits, we would require 20 parameters to understand 55 mutants while a model characterizing individual mutants would require 110 parameters.

In addition to analyzing the available *in vitro* data for these mutants, we consider how each construct would promote gene expression *in vivo*. Because CRP is a global activator, its activity within the cell is tightly regulated by enzymes that produce, degrade, and actively transport cAMP [7]. We discuss how these processes can either be modeled theoretically or excised experimentally and calibrate our resulting framework for transcription using gene expression measurements for wild type CRP. In this manner, we find a small, self-consistent set of parameters able to characterize each step of CRP activation shown in Fig 1(A).

The remainder of this paper is organized as follows. First, we characterize the interaction between cAMP and CRP for the six CRP mutants created by Lanfranco *et al.* and quantify the key parameters governing this behavior. Next, we analyze the interaction between CRP and DNA and discuss how the inferred parameters align with structural knowledge of the system. Finally, we consider how CRP enhances gene expression and extend the results from Lanfranco *et al.* to predict the activation profiles of the CRP mutants within a cellular environment.

Results

The interaction between CRP and cAMP

In this section, we examine the cAMP-CRP binding process through the lenses of generalized MWC and KNF models which tie each mutant's behavior to its subunit composition. We find that both frameworks can characterize data from a suite of CRP mutants using a compact set of parameters, but only the interpretation of the MWC parameters is consistent with structural knowledge of CRP.

MWC model. We first formulate a description of cAMP-CRP binding using a generalized form of the MWC model, where the two subunits of each CRP molecule fluctuate concurrently between an active and inactive state. The different conformations of CRP binding to cAMP and their corresponding Boltzmann weights are shown in Fig 2(A). We define the free energy difference between inactive CRP and active CRP as 2ϵ (or ϵ per subunit). ϵ will be large and negative since the activator is preferentially inactive in the absence of ligand, which will allow us to simplify the description of the system (see S1 Text Section A). $\beta = \frac{1}{k_B T}$ where k_B is Boltzmann's constant and T represents temperature. The two cAMP binding events are known to be cooperative [26, 33, 34], where the magnitude and the sign of this cooperativity (whether it is favorable or unfavorable) strongly depends upon the conditions of the buffer, mutational perturbations to the system, and whether the full or partial CRP protein is considered [29, 35, 36]. To that end, we introduce two types of cooperativity. First, the classic MWC model is inherently cooperative, as the binding of each ligand alters the probable conformation and hence binding affinity of the other binding site; however, this mode of cooperativity can only be favorable [37]. Because CRP may also exhibit negative cooperativity, we introduce explicit interaction energies ϵ_{int}^A and ϵ_{int}^I between two ligands in the active and inactive CRP states, respectively. For simplicity, and because it will enable us to characterize the CRP collectively rather than requiring a unique parameter for each mutant, we assume that these explicit cooperative interactions are the same across all constructs (see S1 Text Section B where we relax such assumptions).

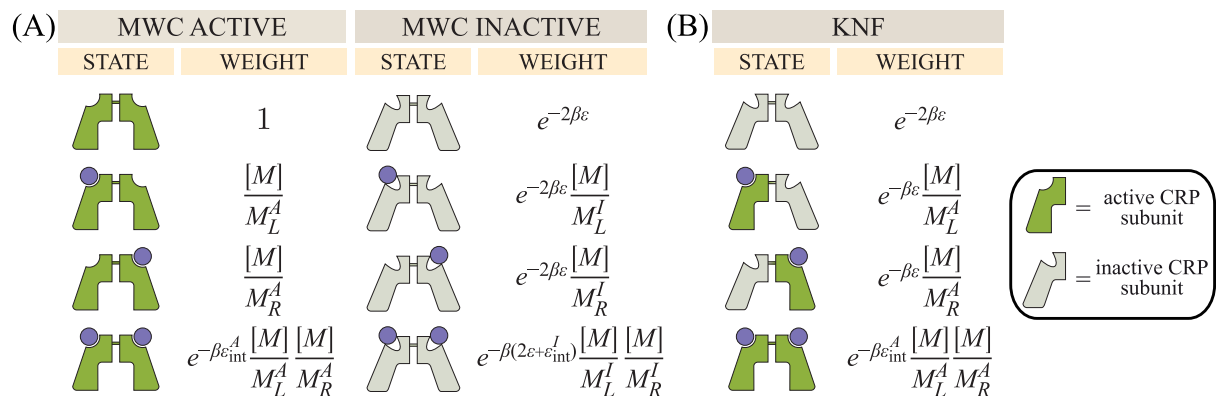


Fig 2. Macroscopic states and Boltzmann weights for cAMP binding to CRP. (A) Within the MWC model, cAMP (purple circles) may bind to a CRP subunit in either the active (dark green) or inactive (light green) state. M_L^A and M_L^I represent the dissociation constants of the left subunit in the active and inactive states, respectively, while M_R^A and M_R^I represent the analogous dissociation constants for the right subunit. $[M]$ denotes the concentration of cAMP and ϵ represents the free energy difference between each subunit's inactive and active states with $\beta = \frac{1}{k_B T}$. ϵ_{int}^A and ϵ_{int}^I represent a cooperative energy when two cAMP are bound to CRP in the active and inactive states, respectively. (B) The KNF model assumes that the two CRP subunits are inactive when unbound to cAMP and transition to the active state immediately upon binding to cAMP. The parameters have the same meaning as in the MWC model, but states where one subunit is active while the other is inactive are allowed.

<https://doi.org/10.1371/journal.pone.0204275.g002>

For each cAMP-CRP dissociation constant M_X^Y , the subscript denotes which CRP subunit it describes—either the left (L) or right (R) subunit—while the superscript denotes the active (A) or inactive (I) state of CRP. Note that the left and right subunits may be different (see Fig 1(B)). Given a cAMP concentration $[M]$, the fraction of occupied cAMP binding sites is given by

$$\text{fractional CRP occupancy}([M]) = \frac{\frac{1}{2} \left(\frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} \right) + e^{-\beta\epsilon_{\text{int}}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + e^{-2\beta\epsilon} \left(\frac{1}{2} \left(\frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} \right) + e^{-\beta\epsilon_{\text{int}}^I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I} \right)}{1 + \frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} + e^{-\beta\epsilon_{\text{int}}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + e^{-2\beta\epsilon} \left(1 + \frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} + e^{-\beta\epsilon_{\text{int}}^I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I} \right)} \quad (1)$$

Here, the fractional occupancy of CRP bound to zero, one, or two cAMP equals 0, 1/2, and 1, respectively. Experimentally, the fractional occupancy was measured *in vitro* in the absence of DNA using ANS fluorescence which utilizes a fluorescent probe triggered by the conformational change of cAMP binding to CRP [29].

Lanfranco *et al.* considered CRP subunits with either the D53H or S62F point mutations (hereafter denoted by D and S, respectively), with the D subunit binding more strongly to cAMP than the wild type while the S subunit binds more weakly as shown in Fig 3(A). While we could characterize the dose-response curves of each CRP mutant independently—for example, by using Eq (1) to extract a set of parameters for each mutant—such an analysis lacks a direct connection between the subunit composition and the corresponding binding behavior. Instead, we assume that the cAMP binding affinity for each subunit should be uniquely dictated by that subunit’s identity as either the WT, D, or S subunit. To that end, we represent the fractional occupancy of CRP_{D/WT} using Eq (1) with one D subunit ($M_L^A = M_D^A$, $M_L^I = M_D^I$) and one WT subunit ($M_R^A = M_{\text{WT}}^A$, $M_R^I = M_{\text{WT}}^I$). The equations for the remaining CRP mutants follow analogously, tying the behavior of each mutant to its subunit composition. For simplicity, we will assume that the D and S mutations do not alter the cAMP interaction energies ϵ_{int}^A and ϵ_{int}^I .

One difficulty in inferring parameter values from Eq (1) is that degenerate sets of parameters may produce equivalent binding curves. For example, in S1 Text Section A, we demonstrate how the same cAMP-CRP binding curves can arise from an arbitrarily large and negative free energy difference ($\epsilon \rightarrow -\infty$) provided that the dissociation constants scale appropriately. In that same supporting information section, we demonstrate how this degeneracy can be excised so that Eq (1) is well approximated by the following form,

$$\text{fractional CRP occupancy}([M]) \approx \frac{\frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + \frac{1}{2} \left(\frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} \right) + e^{-\beta\epsilon_{\text{int}}^I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I}}{\frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + \left(1 + \frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} + e^{-\beta\epsilon_{\text{int}}^I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I} \right)}, \quad (2)$$

where we have neglected the unbound and singly-cAMP-bound active CRP states and defined the effective dissociation constants

$$\tilde{M}_L^A = e^{-\beta\epsilon} e^{\beta\epsilon_{\text{int}}^A/2} M_L^A \quad (3)$$

and

$$\tilde{M}_R^A = e^{-\beta\epsilon} e^{\beta\epsilon_{\text{int}}^A/2} M_R^A. \quad (4)$$

Using Eq (2), we can extract the set of effective dissociation constants for the WT, D, and S subunits that determine the behavior of all six CRP mutants. The resulting parameters (shown in Table 1) give rise to the cAMP-CRP binding curves in Fig 3(A) and 3(B). Note that in

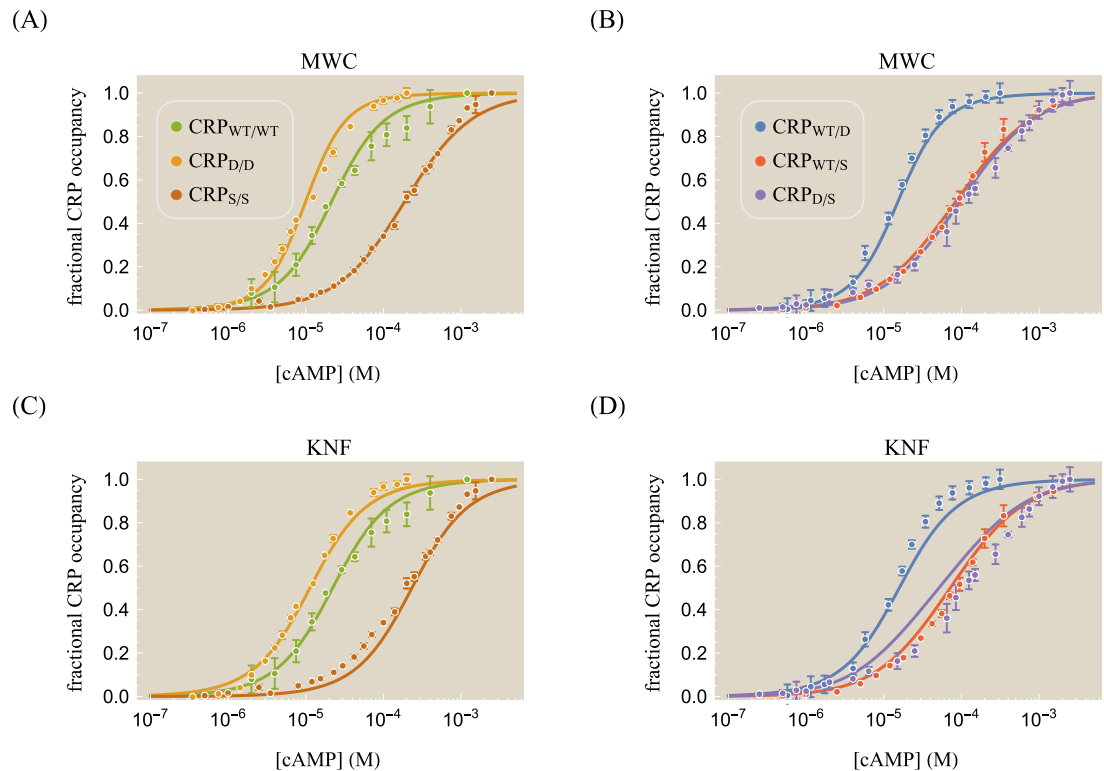


Fig 3. cAMP binding for different CRP mutants. In addition to the wild type CRP subunit (denoted WT), the mutation D53H (denoted D) and the mutation S62F (denoted S) can be applied to either subunit as indicated by the subscripts in the legend. (A) Curves were characterized using the MWC model, Eq (1). The D subunit increases CRP's affinity while the S subunit decreases this affinity. (B) Asymmetrically mutating the two subunits results in distinct cAMP binding curves. The data for the WT/D mutant lies between the WT/WT and D/D data in Panel A, and analogous statements apply for the WT/S and D/S mutants. (C) The fraction of CRP in the active state. Within the MWC model, mutants with an S subunit will be inactive even in the limit of saturating cAMP. (D) The symmetric and (E) asymmetric mutants can also be analyzed using the KNF model, Eq (6), resulting in curves that are similar to those found by the MWC model. (F) The KNF model predicts that all CRP mutants will be completely active in the limit of saturating cAMP. The (corrected) sample standard deviation $\sqrt{\frac{1}{n-1} \sum_{j=1}^n (y_{theory}^{(j)} - y_{data}^{(j)})^2}$ equals 0.03 for the MWC model and 0.05 for the KNF model, and the best-fit parameters for both models are given in Table 1. Data reproduced from Ref. [29].

<https://doi.org/10.1371/journal.pone.0204275.g003>

removing the parameter degeneracy using Eqs (3) and (4), we can no longer determine the individual values of ϵ , ϵ_{int}^A , and the active state dissociation constants M_X^A , but rather only the parameter combinations \tilde{M}_X^A . On the other hand, the inactive state cooperativity energy ϵ_{int}^I can be unambiguously determined to be negligible. The effective dissociation constant of the S

Table 1. Parameters for cAMP binding to CRP. The data in Fig 3 can be characterized using a single set of dissociation constants for the WT, D, and S subunits whose values and standard errors are shown. To excise parameter degeneracy, the active-inactive free energy difference ϵ and the cAMP interaction energy in the active state ϵ_{int}^A are absorbed into the active state dissociation constants in the MWC model (Eqs (2) and (3)). Similarly, ϵ is absorbed into the KNF dissociation constants (Eqs (6) and (7)).

MWC Parameter	Best-Fit Value	KNF Parameter	Best-Fit Value
$\tilde{M}_{WT}^A, M_{WT}^I$	$\{25 \pm 1, 40 \pm 3\} \times 10^{-6} \text{ M}$	\bar{M}_{WT}^A	$(30 \pm 2) \times 10^{-6} \text{ M}$
\tilde{M}_D^A, M_D^I	$\{10 \pm 1, 50 \pm 5\} \times 10^{-6} \text{ M}$	\bar{M}_D^A	$(20 \pm 1) \times 10^{-6} \text{ M}$
\tilde{M}_S^A, M_S^I	$\{\geq 1000, 200 \pm 10\} \times 10^{-6} \text{ M}$	\bar{M}_S^A	$(350 \pm 10) \times 10^{-6} \text{ M}$
ϵ_{int}^I	$0.0 \pm 0.2 k_B T$	ϵ_{int}^A	$-0.8 \pm 0.2 k_B T$

<https://doi.org/10.1371/journal.pone.0204275.t001>

subunit in the MWC model can only be bounded from below as $\tilde{M}_S^A \geq 1000 \times 10^{-6}$ M. However, NMR measurements reported that in the limit of saturating cAMP, the S/S mutant will be inactive state 98% of the time (see Fig 3(C) and S1 Text Section B) which corresponds to a value of $\tilde{M}_S^A \approx 1300 \times 10^{-6}$ M [20].

In S1 Text Section B, we demonstrate that the symmetric CRP mutants in Fig 3(A) provide sufficient information to approximate the behavior of the asymmetric mutants in Fig 3(B). We further show that fitting each CRP data set individually to the MWC or KNF models without constraining the WT, D, and S subunits to a single unified set of dissociation constants results in only a marginal improvement over the constrained fitting. Finally, we analyze the slope of each cAMP binding response and explain why they are nearly identical for the six CRP mutants. In S1 Text Section C, we investigate the effects of the double mutation D+S on a single subunit by comparing its CRP occupancy data supposing that the change in free energy from both mutations is additive and independent. Within this epistasis-free model, we can similarly predict the behavior of other double mutants including CRP_{D/D+S}, CRP_{S/D+S}, and CRP_{D+S/D+S}.

Lastly, we reiterate that the MWC model presented here provides a coarse-grained model of the system. For example, experiments have revealed that the first cAMP binding does not alter the conformation of the second subunit, although it does drastically diminish its protein motions [34]. In the MWC model, these effects are captured both by the inherent cooperativity [37] as well as by the explicit interaction energies ϵ_{int}^A and ϵ_{int}^I , since within this model the binding of one cAMP can induce the other CRP subunit to change (e.g. changing the unbound inactive state into the active singly-bound state). In light of these results, we next consider an alternative model of the system which explicitly assumes that each subunit only becomes active upon ligand binding.

KNF model. We now turn to a KNF analysis of CRP, where the two subunits are individually inactive when not bound to cAMP and become active upon binding as shown in Fig 2(B). Some studies have claimed that cAMP binding to one CRP subunit does not affect the state of the other subunit, in support of the KNF model [34]. Other studies, meanwhile, have reported that a fraction of CRP molecules are active even in the absence of cAMP, thereby favoring an MWC interpretation [9, 38]. To determine whether either model can accurately represent the system, we explore some of the consequences of a KNF interpretation of CRP.

Using the statistical mechanical states of the system in Fig 2(B), the occupancy of CRP is given by

$$\text{fractional CRP occupancy}([M]) = \frac{\frac{e^{-\beta\epsilon}}{2} \left(\frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} \right) + e^{-\beta\epsilon_{int}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A}}{e^{-2\beta\epsilon} + e^{-\beta\epsilon} \left(\frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} \right) + e^{-\beta\epsilon_{int}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A}}. \tag{5}$$

where the parameters have the same meaning as in the MWC model. Multiplying the numerator and denominator by $e^{2\beta\epsilon}$, we obtain the form

$$\text{fractional CRP occupancy}([M]) = \frac{\frac{1}{2} \left(\frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} \right) + e^{-\beta\epsilon_{int}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A}}{1 + \frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} + e^{-\beta\epsilon_{int}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A}} \tag{6}$$

where, similar to the MWC model effective dissociation constants Eqs (3) and (4), we have defined

$$\bar{M}_L^A = e^{-\beta\epsilon} M_L^A \tag{7}$$

and

$$\bar{M}_R^A = e^{-\beta\epsilon} M_R^A. \tag{8}$$

This simplification occurs because within the KNF model, a CRP monomer only switches from the inactive to active state upon cAMP binding. As a result, the free energy of cAMP binding to CRP and the free energy of the CRP undergoing its inactive-to-active state conformational always occur concurrently and may be combined into the effective dissociation constants \bar{M}_L^A and \bar{M}_R^A .

As shown in Fig 3(D) and 3(E), the KNF model can approximately characterize the six mutant CRP binding curves, although the S/S and WT/D responses lie slightly below the data while the D/S curve deviates above the data. These discrepancies could potentially be alleviated by letting the interaction energy ϵ_{int}^A vary with each mutant, although doing so would significantly increase the number of parameters in the model (which would then scale with the number of mutants rather than the number of subunits). However, a greater failing of the KNF model is that it predicts that at saturating cAMP concentrations the protein will always be completely active, even though the S/S mutant is 98% inactive in this limit (Fig 3(F)) [20]. These results suggest that a more complex variant of the KNF model should be used to quantitatively dissect the CRP system.

The interaction between CRP and DNA

We now turn to the second binding interaction experienced by CRP, namely, that between CRP and DNA. Since the preceding analysis demonstrated that the KNF model considered here cannot characterize the existing data, we proceed by only analyzing the MWC model.

Consider a concentration $[L]$ of CRP whose subunits either assume an active state (where they tightly bind to DNA with a dissociation constant L_A) or in an inactive state (characterized by weaker DNA binding with dissociation constant L_I satisfying $L_I > L_A$). The states and weights of this system within the generalized MWC model are shown in Fig 4.

Lanfranco *et al.* fluorescently labeled a short, 32 bp DNA sequence which binds to CRP. Using a spectrometer, they measured the anisotropy of this fluorescence when different concentrations of CRP and cAMP were added *in vitro* [29]. The data are shown in Fig 5(A) for CRP_{D/S} for various concentrations of the receptor and effector. When CRP binds, it slows the random tumbling of the DNA so that over very short time scales the fluorescence is oriented along a particular axis, resulting in a larger anisotropy readout. Unbound DNA is defined as having anisotropy equal to 1 while DNA-bound CRP with 0, 1, or 2 bound cAMP have higher anisotropies of $1 + r_0$, $1 + r_1$, and $1 + r_2$, respectively. Thus, the total anisotropy within the model is given by the weighted sum of each species [39], namely,

$$\text{anisotropy} = 1 + r_0 p_0 + r_1 p_1 + r_2 p_2. \tag{9}$$

Here, p_0 , p_1 , and p_2 represent the probabilities that DNA-bound CRP will be bound to 0, 1, and 2 cAMP molecules, respectively. Using the effective dissociation constants (Eqs (3) and (4)) and neglecting all terms proportional to the small quantity $e^{\beta\epsilon}$, we can write these probabilities as

$$p_0 = \frac{e^{2\beta\epsilon} \frac{[L]}{L_A} + \frac{[L]}{L_I}}{Z} \approx \frac{[L]}{Z}, \tag{10}$$

		MWC ACTIVE		MWC INACTIVE	
		STATE	WEIGHT	STATE	WEIGHT
DNA unbound			1		$e^{-2\beta\epsilon}$
			$\frac{[M]}{M_L^A}$		$e^{-2\beta\epsilon} \frac{[M]}{M_L^I}$
			$\frac{[M]}{M_R^A}$		$e^{-2\beta\epsilon} \frac{[M]}{M_R^I}$
			$e^{-\beta\epsilon_{int}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A}$		$e^{-\beta(2\epsilon + \epsilon_{int}^I)} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I}$
DNA bound			$\frac{[L]}{L_A}$		$e^{-2\beta\epsilon} \frac{[L]}{L_I}$
			$\frac{[L][M]}{L_A M_L^A}$		$e^{-2\beta\epsilon} \frac{[L][M]}{L_I M_L^I}$
			$\frac{[L][M]}{L_A M_R^A}$		$e^{-2\beta\epsilon} \frac{[L][M]}{L_I M_R^I}$
			$e^{-\beta\epsilon_{int}^A} \frac{[L][M]}{L_A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A}$		$e^{-\beta(2\epsilon + \epsilon_{int}^I)} \frac{[L][M]}{L_I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I}$

Fig 4. States and weights for CRP binding to DNA. The DNA unbound states from Fig 2 are shown together with the DNA bound states. The Boltzmann weight of each DNA bound state is proportional to the concentration [L] of CRP and inversely proportional to the CRP-DNA dissociation constants L_A or L_I for the active and inactive states, respectively.

<https://doi.org/10.1371/journal.pone.0204275.g004>

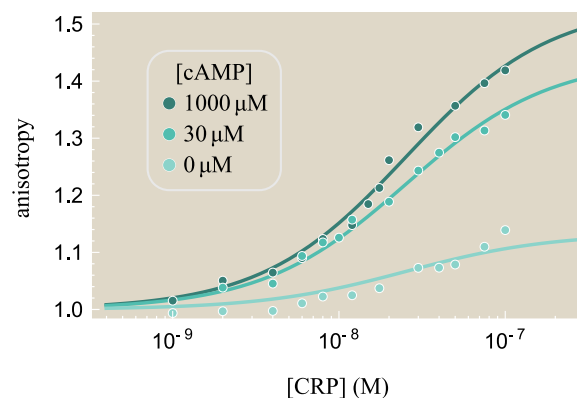


Fig 5. The interaction between CRP and DNA. Anisotropy of 32-bp fluorescein-labeled *lac* promoter binding to CRP_{D/S} at different concentrations of cAMP. An anisotropy of 1 corresponds to unbound DNA while higher values imply that DNA is bound to CRP. In the presence of cAMP, more CRP subunits will be active, and hence there will be greater anisotropy for any given concentration of CRP. The sample standard deviation $\sqrt{\frac{1}{n-1} \sum_{j=1}^n (y_{theory}^{(j)} - y_{data}^{(j)})^2}$ is 0.01, with the corresponding parameters given in Tables 1 and 2. Data reproduced from Ref. [29].

<https://doi.org/10.1371/journal.pone.0204275.g005>

$$p_1 = \frac{e^{2\beta\epsilon} \frac{[L]}{L_A} \left(\frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} \right) + \frac{[L]}{L_I} \left(\frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} \right)}{Z} \approx \frac{\frac{[L]}{L_I} \left(\frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} \right)}{Z}, \tag{11}$$

and

$$p_2 = \frac{e^{2\beta\epsilon} e^{-\beta\epsilon^A} \frac{[L]}{L_A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + e^{-\beta\epsilon^I} \frac{[L]}{L_I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I}}{Z} \approx \frac{\frac{[L]}{L_A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + e^{-\beta\epsilon^I} \frac{[L]}{L_I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I}}{Z} \tag{12}$$

with

$$\begin{aligned} Z &= e^{2\beta\epsilon} \left(1 + \frac{[L]}{L_A} \right) \left(1 + \frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} + e^{-\beta\epsilon^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A} \right) \\ &\quad + \left(1 + \frac{[L]}{L_I} \right) \left(1 + \frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} + e^{-\beta\epsilon^I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I} \right) \\ &\approx \left(1 + \frac{[L]}{L_A} \right) \frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + \left(1 + \frac{[L]}{L_I} \right) \left(1 + \frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} + e^{-\beta\epsilon^I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I} \right). \end{aligned} \tag{13}$$

In making these approximations, we have assumed the stricter conditions $e^{2\beta\epsilon} \frac{L_I}{L_A} \ll 1$ and $e^{2\beta\epsilon} \frac{L_I M_L^I}{L_A M_L^A} \ll 1$ for the WT, D, and S subunits, all of which are valid assumptions for this system (see S1 Text Section A).

Fig 5 shows the resulting best-fit curves for the anisotropy data, with the corresponding CRP_{D/S} DNA dissociation constants given in Table 2. Since $1 + r_0 \approx 1$, cAMP-unbound CRP binds poorly to DNA, in accordance with the inactive state crystal structure whose DNA recognition helices are buried inside the protein [10]. Additionally, the anisotropy $1 + r_1 = 1.7$ of the DNA-CRP-cAMP complex is larger than that of both the cAMP-unbound state and the doubly bound state DNA-CRP-(cAMP)₂ with $1 + r_2 = 1.4$; this suggests that CRP-(cAMP)₂ binds more weakly to DNA than CRP-cAMP. However, we note that these results depend upon the anisotropy values for the three CRP states (r_j in Table 2); Lanfranco *et al.* assumed that difference between the singly-cAMP-bound CRP state and the unbound CRP state should be the same as the difference between the doubly- and singly-cAMP-bound states and subsequently determined that the singly- and doubly-cAMP bound CRP states bind with roughly the same affinity to DNA. That said, previous studies have supported the claim that the singly-cAMP bound state binds tightest to DNA using multiple experimental methods including proteolytic digestion by subtilisin, chemical modification of Cys-178, and fluorescence measurements [40–42]. Given the ability of the MWC model to characterize the cAMP-binding and DNA-binding data of Lanfranco *et al.*, we next consider the final step in the CRP activation cycle, namely, how well CRP can enhance gene expression.

Table 2. Parameters for CRP binding to DNA. The anisotropy data for CRP_{D/S} characterized using Eq (9), as shown in Fig 5. Each value is given as a mean ± standard error. The uncertainty in the M_S^A parameter (shown in Table 1) leads to a corresponding uncertainty in the active CRP dissociation constant L_A .

MWC Parameter	Best-Fit Value
r_0, r_1, r_2	{0.1, 0.8, 0.5} ± 0.1
L_A, L_I	{≤ 30, 30 ± 10} × 10 ⁻⁹ M

<https://doi.org/10.1371/journal.pone.0204275.t002>

Implications of mutations for *in vivo* systems

Since CRP is a global transcriptional activator that governs many metabolic genes in *E. coli* [8], introducing mutations *in vivo* may vastly change cell behavior. Nevertheless, because the framework introduced above is very generic, it can be readily applied to other transcriptional activators that regulate a more limited number of genes. In that spirit, we briefly explore how the CRP mutants characterized in the Lanfranco *et al.* experiments would behave *in vivo* assuming that they only affect a single gene.

Simple activation. Consider a cell with cAMP concentration $[M]$ and CRP concentration $[L]$ where the population of CRP is split between an active $[L_A]$ and an inactive $[L_I]$ conformation. Suppose the cell has a concentration $[P]$ of RNA polymerase (RNAP) which have a dissociation constant P_D with a promoter of interest. The thermodynamic states of the system are shown in Fig 6, where the activator can bind to and recruit RNAP via an interaction energy ϵ_{P,L_A} between active CRP and RNAP with a weaker interaction ϵ_{P,L_I} between inactive CRP and RNAP. Without these two interaction energies ($\epsilon_{P,L_A} = \epsilon_{P,L_I} = 0$), the RNAP and CRP binding events would be independent and there would be no activation. Moreover, if the two activation energies were the same ($\epsilon_{P,L_A} = \epsilon_{P,L_I}$), the system could not exhibit the level of activation seen in the data (see S1 Text Section B).

We assume that gene expression is equal to the product of the RNAP transcription rate r_{trans} and the probability that RNAP is bound to the promoter of interest, namely,

$$\text{activity} = r_{trans} \frac{\frac{[P]}{P_D} \left(1 + \frac{[L_I]}{L_I} e^{-\beta\epsilon_{P,L_I}} + \frac{[L_A]}{L_A} e^{-\beta\epsilon_{P,L_A}} \right)}{\frac{[P]}{P_D} \left(1 + \frac{[L_I]}{L_I} e^{-\beta\epsilon_{P,L_I}} + \frac{[L_A]}{L_A} e^{-\beta\epsilon_{P,L_A}} \right) + 1 + \frac{[L_I]}{L_I} + \frac{[L_A]}{L_A}} \tag{14}$$

Several additional factors influence gene expression *in vivo*. First, cAMP is synthesized endogenously by *cyaA* and degraded by *cpdA*, although both of these genes have been knocked out for the data set shown in Fig 7(A) (see Methods and Ref. [7]). Furthermore, cAMP is

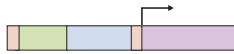
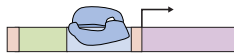

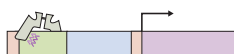
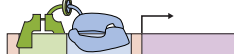

DESCRIPTION	STATE	WEIGHT
empty promoter		1
RNAP bound		$\frac{[P]}{P_D}$
active CRP bound		$\frac{[L_A]}{L_A}$
inactive CRP bound		$\frac{[L_I]}{L_I}$
RNAP and active CRP bound		$\frac{[P]}{P_D} \frac{[L_A]}{L_A} e^{-\beta\epsilon_{P,L_A}}$
RNAP and inactive CRP bound		$\frac{[P]}{P_D} \frac{[L_I]}{L_I} e^{-\beta\epsilon_{P,L_I}}$

Fig 6. States and weights for a simple activation motif. Binding of RNAP (blue) to a promoter is facilitated by the binding of the activator CRP. Simultaneous binding of RNAP and CRP is facilitated by an interaction energy ϵ_{P,L_A} for active CRP (dark green) and ϵ_{P,L_I} for inactive CRP (light green). cAMP (not drawn) influences the concentration of active and inactive CRP as shown in Fig 4.

<https://doi.org/10.1371/journal.pone.0204275.g006>

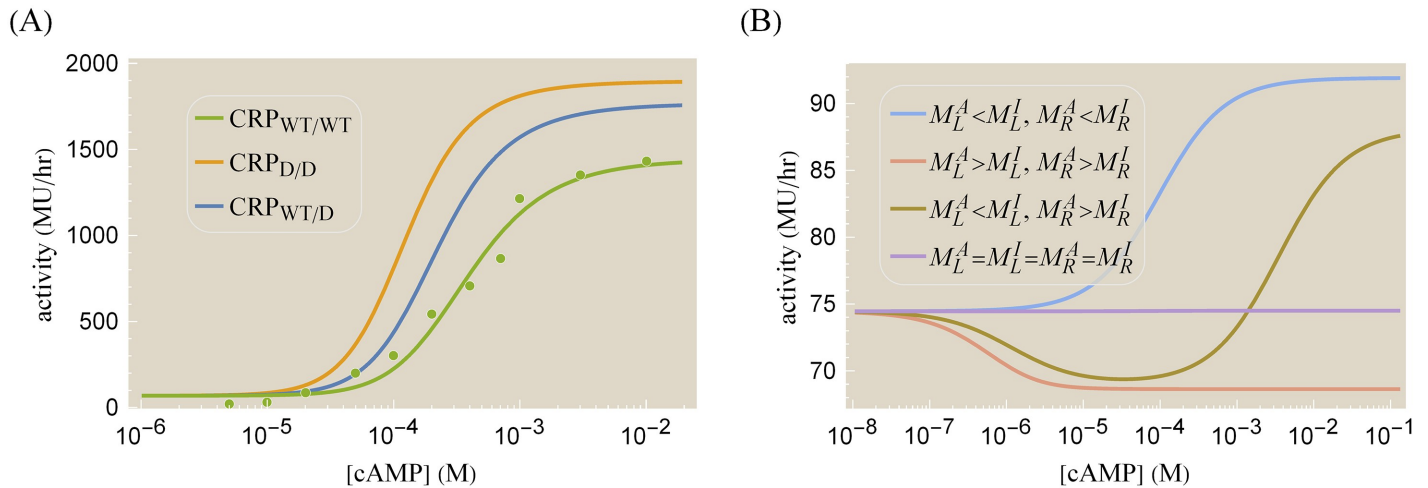


Fig 7. Predicted gene expression profiles for a simple activation architecture. (A) Gene expression for wild type CRP (green dots from Ref. [7]), where 1 Miller Unit (MU) represents a standardized amount of β -galactosidase activity. This data was used to determine the relevant parameters in Eq (14) for the promoter in the presence of $[L] = 1.5 \mu\text{M}$ of CRP [45]. The predicted behavior of the CRP mutants is shown using their corresponding cAMP dissociation constants. (B) The spectrum of possible gene expression profiles can be categorized based upon the cAMP-CRP binding affinity in each subunit. In all cases, we assumed $M_L^A = M_R^A = 3 \times 10^{-6} \text{ M}$ and $e^{-\beta\epsilon_{\text{int}}^A} = 0$. The activation response (blue) was generated using $M_L^I = M_R^I = 6 \times 10^{-6} \text{ M}$. The repression response (orange) used $M_L^I = M_R^I = 10^{-7} \text{ M}$. The peaked response (gold) used $M_L^I = 10^{-7} \text{ M}$ and $M_R^I = 300 \times 10^{-6} \text{ M}$. The flat response used $M_L^I = M_R^I = 3 \times 10^{-6} \text{ M}$. The remaining parameters in both plots were $\frac{[p]}{p_D} = 130 \times 10^{-6}$, $r_{\text{trans}} = 5 \times 10^5 \frac{\text{MU}}{\text{hr}}$, $\gamma = 0.1$, $\epsilon_{p,L_A} = -3k_B T$, $\epsilon_{p,L_I} = 0k_B T$, $\epsilon = -3k_B T$, and those shown in Tables 1 and 2.

<https://doi.org/10.1371/journal.pone.0204275.g007>

actively transported out of a cell leading to a smaller concentration of intracellular cAMP. Following Kuhlman *et al.*, we will assume that the intracellular cAMP concentration is proportional to the extracellular concentration, namely, $\gamma[M]$ (with $0 < \gamma < 1$) [43, 44]. Hence, the concentration of active CRP satisfies $\frac{[L_A]}{[L]} = p_{\text{act}}^L(\gamma[M])$ where the fraction of active CRP p_{act}^L is given by Fig 2(A) as

$$p_{\text{act}}^L([M]) = \frac{1 + \frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} + e^{-\beta\epsilon_{\text{int}}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A}}{1 + \frac{[M]}{M_L^A} + \frac{[M]}{M_R^A} + e^{-\beta\epsilon_{\text{int}}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + e^{-2\beta\epsilon} \left(1 + \frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} + e^{-\beta\epsilon_{\text{int}}^I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I} \right)} \quad (15)$$

$$\approx \frac{e^{-\beta\epsilon_{\text{int}}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A}}{e^{-\beta\epsilon_{\text{int}}^A} \frac{[M]}{M_L^A} \frac{[M]}{M_R^A} + \left(1 + \frac{[M]}{M_L^I} + \frac{[M]}{M_R^I} + e^{-\beta\epsilon_{\text{int}}^I} \frac{[M]}{M_L^I} \frac{[M]}{M_R^I} \right)}$$

In the last step, we have again introduced the effective dissociation constants from Eqs (3) and (4) and dropped any terms proportional to $e^{\beta\epsilon}$. In addition to these considerations, proteins *in vivo* may experience crowding, additional forms of modification, and competition by other promoters. However, since our primary goal is to understand how CRP mutations will affect gene expression, we proceed with the simplest model and neglect the effects of crowding, modification, and competition.

Because of the uncertainty in the dissociation constant L_A between active CRP and DNA (see Table 2), it is impossible to unambiguously determine the transcription parameters from the single data set for wild type CRP shown in Fig 7(A). Instead, we select one possible set of parameters ($\frac{[p]}{p_D} = 130 \times 10^{-6}$, $r_{\text{trans}} = 5 \times 10^5 \frac{\text{MU}}{\text{hr}}$, $\gamma = 0.1$, $\epsilon_{p,L_A} = -3k_B T$, and $\epsilon_{p,L_I} = 0k_B T$) that is consistent with the wild type data. Next, we inserted the other cAMP-CRP dissociation constants (given in Table 1) into Eq (14) to predict the gene expression profiles of the CRP

mutants. Fig 7(A) show the possible behavior of the CRP_{D/D} and CRP_{WT/D} mutants. As expected, replacing a WT subunit with a D subunit shifts the gene expression profile leftwards since the D subunit has a higher cAMP affinity (see Fig 3(A)). Interestingly, the substitution of WT with D subunits comes with a concomitant increase in the maximum gene expression because at saturating cAMP concentrations, a larger fraction of CRP_{D/D} is active compared to CRP_{WT/WT} (96% and 68%, respectively) as seen by using Eq (15) and the parameters in Table 1. Note that we cannot predict the behavior of any of the CRP mutants with S subunits due to the large uncertainty in \tilde{M}_S^A .

Lastly, we probe the full spectrum of phenotypes that could arise from the activity function provided in Eq (14) for any CRP mutant by considering all possible values of the cAMP-CRP dissociation constants M_L^A , M_L^I , M_R^A , and M_R^I in Eq (15). In particular, we relax our assumption that cAMP binding promotes the CRP's active state, as a CRP mutation may exist whose inactive state binds more tightly to cAMP than its active state. Fig 7(B) demonstrates that given such a mutation, a variety of novel phenotypes may arise. The standard sigmoidal activation response is achieved when cAMP binding promotes the active state in both CRP subunits ($M_L^A < M_L^I$, $M_R^A < M_R^I$). A repression phenotype is achieved in the opposite extreme when cAMP binding favors the inactive CRP state ($M_L^A > M_L^I$, $M_R^A > M_R^I$); we note that the ability to switch between a repressing and activating phenotype was achieved in the Lac repressor with as few as three mutations (see the R^c phenotypes in Ref. [46]). When one subunit is activated and the other is repressed by cAMP ($M_L^A < M_L^I$, $M_R^A > M_R^I$ or $M_L^A > M_L^I$, $M_R^A < M_R^I$), a peaked response can form. If the CRP subunits have the same affinity for cAMP in the active and inactive states ($M_L^A = M_L^I = M_R^A = M_R^I$), then CRP will behave identically for all concentrations of CRP, generating a flat-line response. It will be interesting to see whether these phenotypes can be achieved experimentally.

Discussion

The recent work of Lanfranco *et al.* provides a window into the different facets of gene regulation through activation [29]. Using insights from their *in vitro* experiments, we can break down the process of activation into its key steps, namely: (1) the binding of cAMP to make the activator CRP competent to bind DNA (Fig 3); (2) the binding of CRP to DNA (Fig 5); and (3) the recruitment of RNAP to promote gene expression (Fig 7(A)). In this work, we generalized the classic MWC and KNF models to include a cAMP interaction energy as well as different DNA-binding affinities for the various cAMP-CRP bound states, allowing us to globally analyze the CRP binding data. Whereas biological research relishes the unique nuances in each system, the physical sciences suggest that common motifs—such as the prevalence of systems adopting an MWC-like description—lead to equally profound insights into the underlying principles governing systems.

By concurrently modeling the multi-step process of activation, we begin to unravel relationships and set strict limits for the binding energies and dissociation constants governing these systems. One hurdle to precisely fixing these values for CRP has been that many different sets of parameters produce the same degenerate responses (see S1 Text Section A). This parameter degeneracy is surprisingly common when modeling biological systems [47, 48], and we discuss how to account for it within the MWC and KNF models of CRP. A key feature of our analysis is that it permits us to identify the relevant parameter combinations for the system, quantify how well we can infer their values, and suggest which future experiments should be pursued to best constrain the behavior of the system.

Lanfranco *et al.* further explored how mutations in one or both subunits of CRP would influence its behavior. Specifically, they used three distinct subunits (WT, D, and S) to create

the six CRP mutants shown in Fig 1(B) (black and pink boxes). In this work, we showed that the effects of these mutations can be naturally understood through simple thermodynamic models so that each mutation need not be analyzed individually as if it had no relation to any other mutant. Instead, a compact set of parameters characterizing each subunit (see Table 1) could self-consistently characterize the cAMP-binding of all six mutants. The MWC model was shown to successfully describe the CRP activation data for all mutants whereas the KNF model led to a poorer characterization of the data and moreover incorrectly predicted the inhomogeneous population of CRP in the absence and presence of saturating cAMP. Even though an MWC description of the system was sufficient for the data set considered here, the full CRP system exhibits richer behavior that may require more generalized models that include the ensemble of different states seen by NMR [34, 49]. Nevertheless, it remains a useful exercise to understand how much of a system's behavior can be successfully captured by such simple models [50].

The models presented here suggest several avenues to further our understanding of CRP. First, we note that both the MWC and KNF models can serve as a springboard for more complex descriptions of CRP or other regulatory architectures [51]. However, a key advantage of simple frameworks lies in their ability to *predict* how different CRP subunits combine. For example, in S1 Text Section B we demonstrate how the data from the three symmetric CRP mutants in Fig 3(A) can be used to coarsely predict the asymmetric mutant responses in Fig 3(B). It would be interesting to see whether such predictions continue to hold as more mutant subunits are characterized, such as for the expanded suite of mutants shown in Fig 1(B). This framework has the potential to harness the combinatorial complexity of oligomeric proteins and presents a possible step towards systematically probing the space of mutations. In addition, any deviations in these predictions will provide further information on how allostery propagates in this system.

Second, several groups have proposed that multiple CRP mutations (K52N, T127, S128, G141K, G141Q, A144T, L148K, H159L from Refs. [9, 32, 52]) only affect the free energy difference ϵ between the CRP subunit's active and inactive states while leaving the cAMP-CRP dissociation constants unchanged. Our model predicts a narrow spectrum of phenotypes for such mutants, since the dependence of the ϵ parameter is solely confined to the effective dissociation constants (see Eqs (3) and (4)).

Finally, the framework considered here can be used to predict how the CRP mutants generated by Lanfranco *et al.* would behave *in vivo*. We calibrated the CRP_{WT/WT} gene expression profile using data from Ref. [7] and suggested how the remaining CRP mutants may function within a simple activation regulatory architecture given the currently available data (see Fig 7). It would be interesting to measure such constructs—or better yet, similar activators that regulate very few genes—within the cell and test the intersection of our *in vivo* and *in vitro* understanding both in the realm of the multi-step binding events of transcription factors as well as in quantifying the effects of mutations.

Methods

As described in Ref. [29], the fractional CRP occupancy data in Fig 3 was measured *in vitro* using 8-anilino-1-naphthalenesulfonic acid (ANS) fluorescence which is triggered by the conformational change of cAMP binding to CRP. Experiments were conducted in 20mM Tris, 50mM NaCl, 1mM EDTA, pH 7.8, and at 25°C. The CRP-DNA anisotropy data in Fig 5 was measured *in vitro* by tagging the end of a 32bp *lac* promoter with a fluorescein molecule and measuring its anisotropy with a spectrometer. When CRP is bound to DNA, anisotropy arises from two sources: the fast bending of the flanking DNA sequence and the slower rotation of

the CRP-DNA complex. Sources of error include oligomerization of CRP, the bending of the flanking DNA, and nonspecific binding of CRP to the DNA.

The *in vivo* gene expression data was taken from Kuhlman *et al.* using the *lac* operon *E. coli* strain TK310 [7]. This strain had two genes knocked out: *cyaA* (a gene encoding adenylate cyclase, which endogenously synthesizes cAMP) and *cpdA* (encoding cAMP-phosphodiesterase, which degrades cAMP within the cell). Experiments were done at saturating concentrations of inducer ([IPTG] = 1mM) so that Lac repressor negligibly binds to the operator [53]. In this limit, the only transcription factor affecting gene expression is the activator CRP. Gene expression was measured using β -galactosidase activity.

Supporting information

S1 Text. Aforementioned derivations and discussions.

(PDF)

S1 File. A Mathematica notebook that contains all of the data, reproduces the fitting (using both nonlinear regression and MCMC), and generates the plots in this paper.

(PDF)

Acknowledgments

We thank Lacramioara Bintu for bringing the recent developments on CRP to our attention as well as Terry Hwa, Tom Kuhlman, and Michael Manhart for helpful discussions. All plots were made entirely in *Mathematica* using the CustomTicks package [54] with data obtained from the authors or using WebPlotDigitizer [55]. This work was supported in the RP group by La Fondation Pierre-Gilles de Gennes, the Rosen Center at Caltech, and the National Institutes of Health through DP1 OD000217 (Director's Pioneer Award), R01 GM085286, and 1R35 GM118043-01 (MIRA). We are grateful to the Burroughs-Wellcome Fund for its support of the Physiology Course at the Marine Biological Laboratory, where part of the work on this work was done, and for a post-course research grant (JD).

Author Contributions

Conceptualization: Tal Einav, Julia Duque, Rob Phillips.

Formal analysis: Tal Einav, Julia Duque, Rob Phillips.

Writing – original draft: Tal Einav, Rob Phillips.

References

1. Martínez-Antonio A, Collado-Vides J. Identifying Global Regulators in Transcriptional Regulatory Networks in Bacteria. *Current Opinion in Microbiology*. 2003; 6(5):482–489. <https://doi.org/10.1016/j.mib.2003.09.002> PMID: 14572541
2. You C, Okano H, Hui S, Zhang Z, Kim M, Gunderson CW, et al. Coordination of Bacterial Proteome with Metabolism by Cyclic AMP Signalling. *Nature*. 2013; 500(7462):301–306. <https://doi.org/10.1038/nature12446> PMID: 23925119
3. Vilar JMG, Saiz L. Reliable Prediction of Complex Phenotypes from a Modular Design in Free Energy Space: An Extensive Exploration of the *lac* Operon. *ACS Synthetic Biology*. 2013; 2(10):576–586. <https://doi.org/10.1021/sb400013w> PMID: 23654358
4. Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeda D, Muñiz-Rascado L, García-Sotelo JS, et al. RegulonDB Version 9.0: High-level Integration of Gene Regulation, Coexpression, Motif Clustering and Beyond. *Nucleic Acids Research*. 2016; 44(D1):D133–D143. <https://doi.org/10.1093/nar/gkv1156> PMID: 26527724

5. Popovych N, Tzeng SR, Tonelli M, Ebricht RH, Kalodimos CG. Structural Basis for cAMP-mediated Allosteric Control of the Catabolite Activator Protein. *Proceedings of the National Academy of Sciences*. 2009; 106(17):6927–6932. <https://doi.org/10.1073/pnas.0900595106>
6. Hudson JM, Fried MG. Co-operative Interactions Between the Catabolite Gene Activator Protein and the Lac Repressor at the Lactose Promoter. *Journal of Molecular Biology*. 1990; 214(2):381–396. [https://doi.org/10.1016/0022-2836\(90\)90188-R](https://doi.org/10.1016/0022-2836(90)90188-R) PMID: 2166165
7. Kuhlman T, Zhang Z, Saier MH, Hwa T. Combinatorial Transcriptional Control of the Lactose Operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences*. 2007; 104(14):6043–6048. <https://doi.org/10.1073/pnas.0606717104>
8. Kochanowski K, Gerosa L, Brunner SF, Christodoulou D, Nikolaev YV, Sauer U. Few Regulatory Metabolites Coordinate Expression of Central Metabolic Genes in *Escherichia coli*. *Molecular Systems Biology*. 2017; 13(1):903. <https://doi.org/10.15252/msb.20167402> PMID: 28049137
9. Youn H, Koh J, Roberts GP. Two-state Allosteric Modeling Suggests Protein Equilibrium as an Integral Component for Cyclic AMP (cAMP) Specificity in the cAMP Receptor Protein of *Escherichia coli*. *Journal of Bacteriology*. 2008; 190(13):4532–4540. <https://doi.org/10.1128/JB.00074-08> PMID: 18456811
10. Sharma H, Yu S, Kong J, Wang J, Steitz TA. Structure of Apo-CAP Reveals That Large Conformational Changes Are Necessary for DNA Binding. *Proceedings of the National Academy of Sciences*. 2009; 106(39):16604–9. <https://doi.org/10.1073/pnas.0908380106>
11. Gunasekaran K, Ma B, Nussinov R. Is Allostery an Intrinsic Property of All Dynamic Proteins? *Proteins: Structure, Function and Genetics*. 2004; 57(3):433–443. <https://doi.org/10.1002/prot.20232>
12. Won HS, Lee YS, Lee SH, Lee BJ. Structural Overview on the Allosteric Activation of Cyclic AMP Receptor Protein. *Biochimica et Biophysica Acta*. 2009; 1794(9):1299–308. <https://doi.org/10.1016/j.bbapap.2009.04.015> PMID: 19439203
13. Changeux JP. Allostery and the Monod-Wyman-Changeux Model After 50 Years. *Annual Review of Biophysics*. 2012; 41:103–33. <https://doi.org/10.1146/annurev-biophys-050511-102222> PMID: 22224598
14. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The Ensemble Nature of Allostery. *Nature*. 2014; 508(7496):331–9. <https://doi.org/10.1038/nature13001> PMID: 24740064
15. Tsai CJ, Nussinov R. A Unified View of “How Allostery Works”. *PLoS Computational Biology*. 2014; 10(2):e1003394. <https://doi.org/10.1371/journal.pcbi.1003394> PMID: 24516370
16. Liu J, Nussinov R. Energetic Redistribution in Allostery to Execute Protein Function. *Proceedings of the National Academy of Sciences*. 2017; 114(29):7480–7482. <https://doi.org/10.1073/pnas.1709071114>
17. White JT, Li J, Grasso E, Wrabl JO, Hilser VJ. Ensemble Allosteric Model: Energetic Frustration Within the Intrinsically Disordered Glucocorticoid Receptor. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*. 2018; 373(1749):20170175. <https://doi.org/10.1098/rstb.2017.0175> PMID: 29735729
18. Baldwin AJ, Kay LE. NMR Spectroscopy Brings Invisible Protein States into Focus. *Nature Chemical Biology*. 2009; 5(11):808–814. <https://doi.org/10.1038/nchembio.238> PMID: 19841630
19. Freiburger L, Auclair K, Mittermaier A. Global ITC Fitting Methods in Studies of Protein Allostery. *Methods*. 2015; 76:149–161. <https://doi.org/10.1016/j.ymeth.2014.12.018> PMID: 25573261
20. Tzeng SR, Kalodimos CG. Dynamic Activation of an Allosteric Regulatory Protein. *Nature*. 2009; 462(7271):368–372. <https://doi.org/10.1038/nature08560> PMID: 19924217
21. Grutsch S, Brüschweiler S, Tollinger M. NMR Methods to Study Dynamic Allostery. *PLoS Computational Biology*. 2016; 12(3):e1004620. <https://doi.org/10.1371/journal.pcbi.1004620> PMID: 26964042
22. Monod J, Wyman J, Changeux JP. On the Nature of Allosteric Transitions: A Plausible Model. *Journal of Molecular Biology*. 1965; 12:88–118. [https://doi.org/10.1016/S0022-2836\(65\)80285-6](https://doi.org/10.1016/S0022-2836(65)80285-6) PMID: 14343300
23. Koshland DE, Némethy G, Filmer D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry*. 1966; 5(1):365–385. <https://doi.org/10.1021/bi00865a047> PMID: 5938952
24. James LC, Roversi P, Tawfik DS. Antibody Multispecificity Mediated by Conformational Diversity. *Science*. 2003; 299(5611):1362–1367. <https://doi.org/10.1126/science.1079731> PMID: 12610298
25. Bahar I, Chennubhotla C, Tobi D. Intrinsic Enzyme Dynamics in the Unbound State and Relation to Allosteric Regulation. *Current Opinion in Structural Biology*. 2007; 17(6):633–640. <https://doi.org/10.1016/j.sbi.2007.09.011> PMID: 18024008
26. Cui Q, Karplus M. Allostery and Cooperativity Revisited. *Protein Science*. 2008; 17(8):1295–1307. <https://doi.org/10.1110/ps.03259908> PMID: 18560010

27. Park PSH, Lodowski DT, Palczewski K. Activation of G Protein–Coupled Receptors: Beyond Two-State Models and Tertiary Conformational Changes. *Annual Review of Pharmacology and Toxicology*. 2008; 48(1):107–141. <https://doi.org/10.1146/annurev.pharmtox.48.113006.094630> PMID: 17848137
28. del Sol A, Tsai CJ, Ma B, Nussinov R. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways. *Structure*. 2009; 17(8):1042–1050. <https://doi.org/10.1016/j.str.2009.06.008> PMID: 19679084
29. Lanfranco MF, Gárate F, Engdahl AJ, Maillard RA. Asymmetric Configurations in a Reengineered Homodimer Reveal Multiple Subunit Communication Pathways in Protein Allostery. *The Journal of Biological Chemistry*. 2017; 292(15):6086–6093. <https://doi.org/10.1074/jbc.M117.776047> PMID: 28188293
30. Lin SH, Lee JC. Communications Between the High-Affinity Cyclic Nucleotide Binding Sites in *E. coli* Cyclic AMP Receptor Protein. *Biochemistry*. 2002; 41(39):11857–11867. <https://doi.org/10.1021/bi026099z> PMID: 12269830
31. Dai J, Lin SH, Kemmis C, Chin AJ, Lee JC. Interplay Between Site-specific Mutations and Cyclic Nucleotides in Modulating DNA Recognition by *Escherichia coli* Cyclic AMP Receptor Protein. *Biochemistry*. 2004; 43(28):8901–8910. <https://doi.org/10.1021/bi0499359> PMID: 15248748
32. Lin SH, Kovac L, Chin AJ, Chin CCQ, Lee JC. Ability of *E. coli* Cyclic AMP Receptor Protein To Differentiate Cyclic Nucleotides: Effects of Single Site Mutations. *Biochemistry*. 2002; 41(9):2946–2955. <https://doi.org/10.1021/bi0119215> PMID: 11863432
33. Heyduk E, Heyduk T, Lee JC. Intersubunit Communications in *Escherichia coli* Cyclic AMP Receptor Protein: Studies of the Ligand Binding Domain. *Biochemistry*. 1992; 31(14):3682–3688. <https://doi.org/10.1021/bi00129a017> PMID: 1314647
34. Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically Driven Protein Allostery. *Nature Structural and Molecular Biology*. 2006; 13(9):831–838. <https://doi.org/10.1038/nsmb1132> PMID: 16906160
35. Takahashi M, Blazy B, Baudras A, Hillen W. Ligand-modulated Binding of a Gene Regulatory Protein to DNA. Quantitative Analysis of Cyclic-AMP Induced Binding of CRP from *Escherichia coli* to Non-specific and Specific DNA Targets. *Journal of Molecular Biology*. 1989; 207(4):783–796. [https://doi.org/10.1016/0022-2836\(89\)90244-1](https://doi.org/10.1016/0022-2836(89)90244-1) PMID: 2547972
36. Yu S, Maillard RA, Gribenko AV, Lee JC. The N-terminal Capping Propensities of the D-helix Modulate the Allosteric Activation of the *Escherichia coli* cAMP Receptor Protein. *The Journal of Biological Chemistry*. 2012; 287(47):39402–39411. <https://doi.org/10.1074/jbc.M112.404806> PMID: 23035121
37. Marzen S, Garcia HG, Phillips R. Statistical Mechanics of Monod-Wyman-Changeux (MWC) Models. *Journal of Molecular Biology*. 2013; 425(9):1433–1460. <https://doi.org/10.1016/j.jmb.2013.03.013> PMID: 23499654
38. Tzeng SR, Kalodimos CG. Protein Activity Regulation by Conformational Entropy. *Nature*. 2012; 488:236–240 <https://doi.org/10.1038/nature11271> PMID: 22801505
39. Heyduk T, Lee JC. Application of Fluorescence Energy Transfer and Polarization to Monitor *Escherichia coli* cAMP Receptor Protein and *lac* Promoter Interaction. *Proceedings of the National Academy of Sciences*. 1990; 87(5):1744–1748.
40. Heyduk T, Lee JC. *Escherichia coli* cAMP Receptor Protein: Evidence for Three Protein Conformational States with Different Promoter Binding Affinities. *Biochemistry*. 1989; 28(17):6914–6924. <https://doi.org/10.1021/bi00443a021> PMID: 2554959
41. Pyles EA, Lee JC. Mode of Selectivity in Cyclic AMP Receptor Protein-dependent Promoters in *Escherichia coli*. *Biochemistry*. 1996; 35(4):1162–1172. <https://doi.org/10.1021/bi952187q> PMID: 8573570
42. Harman JG. Allosteric Regulation of the cAMP Receptor Protein. *Biochimica et Biophysica Acta*. 2001; 1547(1):1–17. [https://doi.org/10.1016/S0167-4838\(01\)00187-X](https://doi.org/10.1016/S0167-4838(01)00187-X) PMID: 11343786
43. Li G, Young KD. A cAMP-independent Carbohydrate-driven Mechanism Inhibits *tnaA* Expression and TnaA Enzyme Activity in *Escherichia coli*. *Microbiology (United Kingdom)*. 2014; 160(PART 9):2079–2088.
44. Goldenbaum PE, Hall GA. Transport of Cyclic Adenosine 3', 5'-monophosphate Across *Escherichia coli* Vesicle Membranes. *Journal of Bacteriology*. 1979; 140(2):459–467. PMID: 227841
45. Cossart P, Gicquel-Sanzey B. Regulation of Expression of the *crp* Gene of *Escherichia coli* K-12: *In Vivo* Study. *Journal of Bacteriology*. 1985; 161(1):454–457. PMID: 2981809
46. Daber R, Sochor MA, Lewis M. Thermodynamic Analysis of Mutant Lac Repressors. *Journal of Molecular Biology*. 2011; 409(1):76–87. <https://doi.org/10.1016/j.jmb.2011.03.057> PMID: 21459098
47. Hines KE, Middendorf TR, Aldrich RW. Determination of Parameter Identifiability in Nonlinear Biophysical Models: A Bayesian Approach. *The Journal of General Physiology*. 2014; 143(3):401–416. <https://doi.org/10.1085/jgp.201311116> PMID: 24516188

48. Transtrum MK, Machta BB, Brown KS, Daniels BC, Myers CR, Sethna JP. Perspective: Sloppiness and Emergent Theories in Physics, Biology, and Beyond. *Journal of Chemical Physics*. 2015; 143(1):010901. <https://doi.org/10.1063/1.4923066> PMID: 26156455
49. Hilser VJ, Garcia-Moreno EB, Oas TG, Kapp G, Whitten ST. A Statistical Thermodynamic Model of the Protein Ensemble. *Chemical Reviews*. 2006; 106(5):1545–1558. <https://doi.org/10.1021/cr040423+> PMID: 16683744
50. Gunawardena J. Models in Biology: Accurate Descriptions of Our Pathetic Thinking. *BMC Biology*. 2014; 12(1):29. <https://doi.org/10.1186/1741-7007-12-29> PMID: 24886484
51. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, et al. Transcriptional Regulation by the Numbers: Models. *Current Opinion in Genetics & Development*. 2005; 15(2):116–124. <https://doi.org/10.1016/j.gde.2005.02.007>
52. Youn H, Kerby RL, Conrad M, Roberts GP. Study of Highly Constitutively Active Mutants Suggests How cAMP Activates cAMP Receptor Protein. *Journal of Biological Chemistry*. 2006; 281(2):1119–1127. <https://doi.org/10.1074/jbc.M509421200> PMID: 16260780
53. Razo-Mejia M, Barnes SL, Belliveau NM, Chure G, Einav T, Lewis M, et al. Tuning Transcriptional Regulation Through Signaling: A Predictive Theory of Allosteric Induction. *Cell Systems*. 2018; 6(4):456–469. <https://doi.org/10.1016/j.cels.2018.02.004> PMID: 29574055
54. Caprio MA. LevelScheme: A Level Scheme Drawing and Scientific Figure Preparation System for Mathematica. *Computer Physics Communications*. 2005; 171(2):107–118. <https://doi.org/10.1016/j.cpc.2005.04.010>
55. Rohatgi A. WebPlotDigitizer; 2017. Available from: <https://automeris.io/WebPlotDigitizer>.