Quantitative model for gene regulation by λ phage repressor

gene regulation/repressors and operators/cooperative interaction/thermodynamic model

GARY K. ACKERS*, ALEXANDER D. JOHNSON†‡, and MADELINE A. SHEA*

*Department of Biology and McCollum-Pratt Institute, The Johns Hopkins University, Baltimore, Maryland 21218; and †The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Communicated by Mark Ptashne, November 6, 1981

ABSTRACT A statistical thermodynamic model has been developed to account for the cooperative interactions of the bacteriophage λ repressor with the λ right operator. The model incorporates a general theory for quantitatively interpreting cooperative site-specific equilibrium binding data. Values for all interaction parameters of the model have been evaluated at 37°C, 0.2 M KCl, from results of DNase protection experiments in vitro [A. D. Johnson, B. J. Meyer, & M. Ptashne, Proc. Natl. Acad. Sci. USA (1979) 76, 5061-5065]. With these values, the model predicts repression curves at the divergent promoters P1 and P2, that control transcription of genes coding for the regulatory proteins cro and repressor, respectively. At physiological repressor concentrations, repression at P1 is predicted to be nearly complete whereas P2 is predicted to remain highly active. The results demonstrate the importance of cooperative interactions between repressor dimers bound to the adjacent operator sites O1 and O2 in maintaining a stable lyogenic state and in allowing efficient switchover to the lytic state during induction.

In prokaryotes, genes are commonly switched on and off by the interactions of regulatory proteins with specific DNA sequences. A particularly complex example of such a switch is found at the right operator (O_R) of bacteriophage λ: this operator consists of three tandem DNA sites that are recognized by two phase-encoded regulatory proteins (the λ repressor and cro protein). When phage λ is in the "lysogenic state," the λ repressor (cf. gene product) is synthesized and occupies sites O1 and O2. In this configuration of O_R, the cro gene is repressed and the cl gene is transcribed.

The phage switches into the "lytic state" when the repressor protein is cleaved in half by the recA protein, an action initiated by DNA damage. As repressor is destroyed, the cro gene is derepressed and the cro protein is made and occupies site O2, turning off transcription of the cl gene, yet allowing its own synthesis. In this fashion, the phage can switch from one state (lysogeny) to another (lytic growth) in response to an external signal (for review, see ref. 1).

In this paper, we consider the lysogenic state (repressor on, cro off) in an attempt to understand the physical principles that govern the ways in which the protein–DNA and protein–protein interactions operate in concert to produce the known physiological behavior. We show that a model based on statistical thermodynamic assumptions is sufficient to account for some of the known physiological properties of the repressor–operator regulatory system. A brief summary of certain aspects of this work has been presented elsewhere (1).

Models for interactions at the λ operon have been developed to include effects of inducer and nonspecific DNA on the binding of lac repressor (2–4). The systems of λ and other inducible phages differ from lac by using multiple operator binding sites that have cooperative interactions between bound repressors (1). This requires a more elaborate theoretical approach to the protein–DNA binding problem—one that has not previously been developed.

The mathematical model we present here incorporates a set of rules and assumptions derived from previous genetic, biochemical, and structural studies. Combination of this information with statistical thermodynamic assumptions generates a quantitative formulation that incorporates salient features of the qualitative description developed over the last several years (1, 5–9). This quantitative model has the following significance. (i) It provides a way to test assumptions regarding the physical bases for operation of the system. Thus, it predicts quantitatively the activities of the two O_R-controlled promoters (P1 and P2) as a function of repressor concentration, and these predictions can then be compared with known physiological properties of the system. (ii) It points out features of the λ operator system that were not obvious (or at least fully appreciated) from previous experiments. (iii) It incorporates a theory for interpreting cooperative site-specific equilibrium binding data (e.g., from protection-method experiments) that should be applicable to other regulatory systems. Based primarily on the results of in vitro DNase I protection experiments, we have generated a complete and unique evaluation of the free energies for the system at 37°C, 0.2 M KCl, conditions that may resemble "physiological" (10).

THE SYSTEM

We consider the following interacting components (Fig. 1): (i) the λ right operator containing the 17-base pair binding sites O1, O2, and O3, (ii) active repressor dimers capable of binding tightly to the operator sites, and (iii) monomers that cannot bind to DNA but can associate to form active dimers. We wish to calculate (i) the probability that an operator template will exist in a given microscopic configuration (e.g., with O1 occupied and O3 occupied and O2 vacant) and (ii) the probabilities of certain configurations of repressor molecules bound to the operator that have biological significance (e.g., the total probability that O2 is occupied by repressor, obtained by summing all the ways this can occur). The following assumptions and definitions provide a basis for these calculations.

Assumptions and Definitions. 1. We assume that occupancy of operator sites by repressor dimers and the resulting effects

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

‡ Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143.
§ Several of these premises (notably 2, 3, 5, and 6) must be regarded as "facts." They are designated here as "assumptions" solely to indicate their formal role in the mathematical model.
Fig. 1. Interactions of repressor molecules (○—○) at the right operator, O9, of the λ phage genome. The DNA binding sites for repressor (designated O9, O10, and O11) are each composed of 17 base pairs and are separated by spacers of 6 and 7 base pairs. Binding of repressor molecules controls transcription by RNA polymerase from the two divergent promoters, P6 and P4M. Transcription of the cro gene from P6 and consequent synthesis of cro protein leads to lytic growth of the phage genome and cell lysis. Transcription of the cl gene from the promoter P4M and consequent synthesis of repressor is required to maintain the lysogenic state by preventing cro synthesis. (a) Active repressor dimers in equilibrium with inactive monomers. The dimers bind reversibly to each of the three operator sites (arrows). Binding can result in any of the seven combinations shown in Table 1. (b and c) Types of cooperative interaction believed to occur between adjacent bound repressors (configurations 7 and 8 of Table 1).

Microscopic Configurations and Energy States. Assumptions 2, 3, and 4 define a set of eight possible structures (i.e., microscopic configurations of operators) that are assumed to have sufficiently high probability to warrant consideration. Table 1 lists the free energy terms contributing to the various configurations of the operators. The magnitudes and relative values of these interaction energies were determined from experimental results (see below).

Mathematical Relationships of the Model. For each of the eight microscopic configurations of Table 1, we can formulate an exact expression of its probability as a function of repressor concentration according to the principles of statistical thermodynamics. The probability of an operator in the s configuration can be written (cf. ref. 11) as

$$f_s = \frac{\exp(-\Delta G_s/RT)}{\sum_{s} \exp(-\Delta G_s/RT)} [R_{s}]^j,$$

where $\Delta G_s$ is the relative free energy of the s configuration (Table 1), $R_s$ is the gas constant, $T$ is the absolute temperature, $[R_s]$ is the concentration of unbound repressor dimers, and $j$ is the number of repressor dimers bound to an operator in the s configuration. The summation of s is taken from 0 through 8 and $j$ has the values (0, 1, 2, 3) appropriate to the species s. The probability $f_s$ represents the fraction of operators that, at a given concentration of unbound repressor dimers $[R_s]$, will have the configuration whose free energy is $\Delta G_s$.

Once the probabilities for each of the eight operator states are known, we can describe the binding of repressor to each operator site in the full cooperative system. For example, the total occupancy of O9 is given by

$$f_{O9} = f_{09} + f_{09} + f_1 + f_5 = f_{P9M}. \quad [2]$$

Note that, by assumption 5, this is also the fraction $f_{P9M}$ of templates in which P9M is repressed. Likewise (by assumption 4) the fraction of P9-repressed templates is

$$f_{P9} = f_2 + f_3 + f_5 + f_6 + f_7 + f_9. \quad [3]$$

Eqs. 1–3 provide a statistical thermodynamic translation of the model assumptions into the biologically significant quantities $f_s$ and $f_{P9M}$. Thus far, our equations are expressed in terms of the interaction free energies of Table 1 and the concentration

Table 1. Microscopic configurations and free energies for the operator–repressor system O9

<table>
<thead>
<tr>
<th>Species</th>
<th>Configuration</th>
<th>Free energy contributions</th>
<th>Total free energy ($\Delta G_s$, kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O9 O O</td>
<td>Reference state</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>O9 O O</td>
<td>$\Delta G_1$</td>
<td>-11.7</td>
<td></td>
</tr>
<tr>
<td>O9 O O</td>
<td>$\Delta G_2$</td>
<td>-10.1</td>
<td></td>
</tr>
<tr>
<td>O9 O O</td>
<td>$\Delta G_3$</td>
<td>-10.1</td>
<td></td>
</tr>
<tr>
<td>O9 O O</td>
<td>$\Delta G_4 + \Delta G_5 + \Delta G_7$</td>
<td>-23.8</td>
<td></td>
</tr>
<tr>
<td>O9 O O</td>
<td>$\Delta G_6$</td>
<td>-21.8</td>
<td></td>
</tr>
<tr>
<td>O9 O O</td>
<td>$\Delta G_5 + \Delta G_6 + \Delta G_7 + \Delta G_12$</td>
<td>-32.2</td>
<td></td>
</tr>
<tr>
<td>O9 O O</td>
<td>$\Delta G_5 + \Delta G_6 + \Delta G_7 + \Delta G_12$</td>
<td>-33.9</td>
<td></td>
</tr>
</tbody>
</table>

Individual operator sites are denoted by O if vacant or R if occupied by a bound repressor dimer. SS Pairwise interaction between adjacent bound repressors. $\Delta G_s$ represents the standard free energy of formation relative to the reference template ($s = 1$) for each operator species $s$. $\Delta G_1$, $\Delta G_2$, and $\Delta G_3$ are intrinsic free energies for binding repressor dimers to the various sites and are each related to a corresponding equilibrium constant $K_s$ by the standard relationship $\Delta G_s = -RT\ln K_s \quad (i = 1, 2, 3, or 3).$

* 1 cal = 4.18 J.
of unbound repressor dimers \([R_D]\). To relate these quantities to the total concentration of repressors \([R_i]\), we note that

\[
[R_i] = [R_j] + 2[R_d] + 2[O_j] \Sigma \Delta f_i
\]  

where \([R_i]\) is the concentration of repressor monomers (which cannot bind DNA, see above) and \([O_j]\) is the total concentration of operators. Eqs. 1–4 plus the equation for dimer formation \([\Delta f_d = K(R_i)^2] (12)\) comprise the mathematical formulation of the model. This provides us with the means to predict the \(P_R\) and \(P_{RM}\) repression curves, \(f_P\) and \(f_{P_{RM}}\) as a function of repressor concentration, once the interaction free energies are known.

## EVALUATION OF THE INTERACTION PARAMETERS

Having set up the formal model, we now consider the values for the five free energy terms of Table 1 \((\Delta G_1, \Delta G_2, \Delta G_3, \Delta G_{12}, \text{and} \Delta G_{23})\). From the DNase protection experiments of Johnson et al. (6), the concentration of repressor dimers required for half saturation of the individual operator sites has been determined for the wild-type operator and for mutant operators in which one or two of the three repressor binding sites have been destroyed by point mutation and deletion. To calculate the five free energy terms from these data, we use the following expressions for occupancy of the three sites in wild-type operators:

\[
f_{O_1} = f_s + f_s + f_s + f_s
\]  

\[
f_{O_2} = f_s + f_s + f_s + f_s
\]  

\[
f_{O_3} = f_s + f_s + f_s + f_s
\]  

For the mutant operators, the probabilities have similar form, each containing the appropriate combinations of terms. Each of the 11 data points (Table 2) gives a value of \(R_i\) at which the individual binding site is half occupied; thus, each of these data points defines an equation of the form shown above for which \(f_{O_{10}} = 0.5\). The system of 11 simultaneous equations was then solved by a nonlinear least-squares procedure (13) that estimates the best values for the five energies \((\Delta G_1, \Delta G_2, \Delta G_3, \Delta G_{12}, \Delta G_{23})\) and the confidence limits associated with each value. Results of these calculations (performed on a Hewlett-Packard 1000 system) are given in Table 3.

The effects of experimental uncertainty in repressor concentration (as high as 30% random error) were explored. Such errors were found to cause changes of only a few tenths of a kilocalorie in the energies resolved from the 11 simultaneous equations. A full treatment of the numerical analysis of these and similar data will be presented elsewhere.

## PREDICTED BEHAVIOR OF THE SYSTEM

Repression Curves at \(P_R\) and \(P_{RM}\). Having resolved the five free energies that define the operator–repressor interactions

### Table 2. Values of \([R_D]\) for half occupation of individual sites

<table>
<thead>
<tr>
<th>DNA Template</th>
<th>(O_{D3})</th>
<th>(O_{D2})</th>
<th>(O_{D1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(O_1) (wild type)</td>
<td>25</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>(O_{12})</td>
<td>5</td>
<td>5</td>
<td>—</td>
</tr>
<tr>
<td>(O_{13})</td>
<td>25</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>(O_{12}), (O_{13})</td>
<td>25</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(O_{13}), (O_{13})</td>
<td>—</td>
<td>25</td>
<td>—</td>
</tr>
<tr>
<td>(O_{13})</td>
<td>—</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Data are from DNase protection titrations in units of 3 nM. All results are from ref. 6 except that for \(O_{D3}\).

and using \(K = 5 \times 10^7 \text{M}^{-1} \) and \([O_1] = 10^{-9} \text{M} \) (11), we calculated the repression curves from Eqs. 2–4 for the wild-type operon. The predicted curves are presented in Fig. 2. We note two characteristics of these curves. First, they agree with experimental findings in vivo and in vitro that repressor turns off \(P_R\) at lower concentrations than those required to turn off \(P_{RM}\). The curves predict that a 25-fold more repressor is needed to half-repress \(P_{RM}\) than to half-repress \(P_R\). This value agrees with that determined experimentally in vitro by using the abortive initiation assay (14) to measure the activities of the two promoters as a function of repressor concentration (D. Hawley and W. R. McClure, personal communication). In addition, Maurer et al. (7) and Meyer and Ptashne (8, 9) have shown in vivo that at least 10- to 15-fold more repressor is required to half-repress \(P_{RM}\) than to half-repress \(P_R\). This value was difficult to determine accurately due to the difficulty of measuring low concentrations of repressor in vivo.

Second, we note that the predicted \(P_R\) and \(P_{RM}\) repression curves differ in shape; for example, the \(P_R\) curve is steeper. This difference exists because \(P_R\) is controlled by two operator sites to which repressor binds cooperatively whereas \(P_{RM}\) is controlled by single site. Hawley and McClure (using the abortive initiation assay) have shown that \(P_R\) turnover is, in fact, a steeper function of repressor concentration than is \(P_{RM}\) turnover.

### Maintenance of the Lysogenic State

In a lysogen, \(P_R\) is very tightly repressed whereas \(P_{RM}\) is turned down only about 20% (8). The predicted curves (Fig. 2) are consistent with this finding. For example, at a total repressor concentration of \(10^{-7} \text{M}\), the calculated values of \(f_{P_R}\) and \(f_{P_{RM}}\) are 0.99 and 0.35, respectively. Although the values calculated cannot be taken as an exact rep-

### Table 3. Resolved interaction free energies for \(O_R\)

<table>
<thead>
<tr>
<th>Energy, kcal</th>
<th>Individual site binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta G_1)</td>
<td>-11.69 ± 0.03</td>
</tr>
<tr>
<td>(\Delta G_2)</td>
<td>-10.10 ± 0.05</td>
</tr>
<tr>
<td>(\Delta G_3)</td>
<td>-10.09 ± 0.02</td>
</tr>
<tr>
<td>Cooperative interaction</td>
<td></td>
</tr>
<tr>
<td>(\Delta G_{12})</td>
<td>-1.99 ± 0.06</td>
</tr>
<tr>
<td>(\Delta G_{23})</td>
<td>-1.94 ± 0.06</td>
</tr>
</tbody>
</table>

Results are ±65% confidence limits for estimated values (13).
representation of in vivo repression, these curves correctly predict that \( P_R \) will be nearly completely repressed while \( P_{RM} \) will be highly active. Although the concentration of repressor in a lysogen is not known accurately, it is at least \( 1-2 \times 10^{-7} \) M (100–200 monomers per cell) (15, 16).

The Importance of Cooperativity. We believe that cooperative interaction between adjacent bound repressor molecules helps to maintain the stability of the lysogenic state and yet allows the phage to be easily induced. First, the cooperativity effectively increases the repressor concentration of repressor. This can be seen by comparing the fully cooperative repressor repression curve (A) with the predicted noncooperative repression curve (C) obtained by setting \( \Delta G_R \) = \( \Delta G_{R*} \) = 0 but not changing \( \Delta G_{2} \), \( \Delta G_{2*} \), and \( \Delta G_{3} \). At lysogenic concentrations of repressor (e.g., when \( P_R \) is \( \approx \)20% repressed), the effect of cooperativity is to significantly tighten the repression of \( P_R \). Second, we believe that cooperative repressor binding aids in effectively switching the phage from the lysogenic to the lytic state. Since \( f_{P} \) is a steep function of repressor concentration (Fig. 2), a modest decrease in repressor concentration can lead to a relatively large decrease in extent of repression and a consequent burst of cro production.

Nonspecific DNA Binding. The probability of any configuration of the operator is predicted by Eq. 1. It is determined by the relative free energy of that configuration and the concentration of free repressor dimers in solution. We have tacitly assumed (Eq. 4) that no substantial fraction of the intracellular repressor is sequestered on nonoperator DNA (or other cellular structures). Although we are not certain that this assumption is valid, we do know that, at 0°C in the presence of low (50 mM) salt, the ratio of nonspecific to specific binding for \( \lambda \) repressor is \( \approx 10^{-8} \) (12). If this ratio remains the same under physiological conditions (37°C, 0.2 M salt; see above), then we estimate that little or none of the repressor in a lysogen is bound to nonoperator DNA (see ref. 1). We emphasize that our theoretical treatment of \( O_R \) occupancy by repressor holds irrespective of the degree of nonspecific DNA binding.

DISCUSSION

Several theoretical treatments have been developed to analyze regulatory systems governed by protein–DNA interactions (19–21); however, none of them deal with the problem of interpreting cooperative binding curves at specific nonoverlapping sites. The formulations developed in this study provide a simple rigorous theory to analyze systems of this type. The method can readily be extended to systems other than bacteriophage \( \lambda \) (and its relatives P22 and 434) in which the numbers of binding sites and the rules of interaction and biological function may be different. For example, control of the Escherichia coli arabinose operon by araC and CAP protein (22, 23), autogenous regulation of the E. coli lexA gene (24), and control of the smirvin virus 40 early genes by the large tumor antigen (25) are all thought to depend on cooperative interactions between adjacent DNA-bound regulatory proteins. The development of DNA sequence analysis and DNAse protection methods (26–28) has made possible the generation of individual-site binding data, such as that analyzed here.

In this study, we have shown that a statistical thermodynamic mechanism for repressor–operator interactions provides a working model that accounts, so far, for certain in vivo characteristics of the lysogenic phase. This demonstration of sufficiency of the physical assumptions does not, of course, prove them to be correct or necessary. It is possible that other theories, based for example on purely kinetic mechanisms, could account equally well for the available body of physiological observations.

We believe the present model to be highly credible, however, because it does not depend upon any “adjustment of parameters.” Evaluation of the model parameters was carried out independently of their application to in vivo situations and was rigorously exact since the interacting components of the in vitro binding experiments were at thermodynamic equilibrium. These parameters were then used to predict the in vivo behavior of the system. Application of the model to an in vivo situation depends on the supposition that the probabilities of interaction will still be governed by their independently determined thermodynamic energies. Evaluation of this assumption can only be obtained by comparing predicted properties with those observed in vivo. We found that, without any adjustment of constants, the repression curves and other features predicted by the model are in very good agreement with the physiological observations. This lends credence to the physical assumptions on which the model is based. Further tests of the approach developed here could be carried out by extending the model to include the mutual effects of cro and RNA polymerase. These effects, along with dynamic aspects of the system, will be presented in detail elsewhere.

We thank Mark Ptashne for his continuous interest and support in this work. We also thank Benjamin W. Turner, Gary Ketner, William McClure, and Diane Hawley for valuable discussions of the ideas and information described in this paper. This work has been supported by grants from the National Institutes of Health and the National Science Foundation.

We thank Mark Ptashne for his continuous interest and support in this work. We also thank Benjamin W. Turner, Gary Ketner, William McClure, and Diane Hawley for valuable discussions of the ideas and information described in this paper. This work has been supported by grants from the National Institutes of Health and the National Science Foundation.