THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1991 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 266, No. 21, Issue of July 25, pp. 13661-13671, 1991
Printed in U.S.A.

Determination of Binding Constants for Cooperative Site-specific Protein-DNA Interactions Using the Gel Mobility-shift Assay*

(Received for publication, December 14, 1990)

Donald F. Senear‡§ and Michael Brenowitz¶

From the ‡Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92717 and the ¶Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461

We have investigated the question of whether the gel mobility-shift assay can provide data that are useful to the demonstration of cooperativity in the site-specific binding of proteins to DNA. Three common patterns of protein-DNA interaction were considered: (i) the cooperative binding of a protein to two sites (illustrated by the *Escherichia coli* Gal repressor); (ii) the cooperative binding of a bidentate protein to two sites (illustrated by the *E. coli* Lac repressor); and (iii) the cooperative binding of a protein to three sites (illustrated by the λcI repressor). A simple, rigorous, and easily extendable statistical mechanical approach to the derivation of the binding equations for the different patterns is presented. Both simulated and experimental data for each case are analyzed. The mobility-shift assay provides estimates of the macroscopic binding constants for each step of ligation based on its separation of liganded species by the number of ligands bound. Resolution of the binding constants depends on the precision with which the equilibrium distribution of liganded species is determined over the entire range of titration of each of the sites. However, the evaluation of cooperativity from the macroscopic binding constants is meaningful only for data that are also accurate. Some criteria that are useful in evaluating accuracy are introduced and illustrated. Resolution of cooperative effects is robust only for the simplest case, in which there are two identical protein binding sites. In this case, cooperative effects of up to 1,000-fold are precisely determined. For heterogeneous sites, cooperative effects of greater than 1,000-fold are resolvable, but weak cooperativity is masked by the heterogeneity. For three-site systems, only averaged pairwise cooperative effects are resolvable.

The initiation of transcription is a primary control point in the regulation of gene expression in both prokaryotes and eukaryotes. Mechanisms for the regulation of transcription initiation range from a single repressor that binds to a single operator site to occlude the promoter (see Jacob and Monod (1961) for the original operon model) to complex regulatory assemblies. The latter can contain both negative and positive transcription factors that interact with one another and with arrays of DNA sites which are often separated by considerable distances (Tjian and Mitchell, 1989). A nearly ubiquitous feature of these regulatory assemblies is the cooperative binding of regulatory proteins to DNA. In some systems transcriptional regulation is thought to be conferred solely through the control of cooperative interactions between proteins that remain stably bound to the DNA (Sorger and Pelham, 1988). Thus, the separate resolution of the equilibrium binding constants that describe the protein-DNA and cooperative interactions within gene regulatory complexes is a crucial step in understanding the molecular mechanism of gene regulation.

The gel mobility-shift assay (Garner and Revzin, 1981; Fried and Crothers, 1981) has developed into the most widely used method to analyze the binding of proteins to specific sites on DNA. The value of this assay as a qualitative probe of protein-DNA interactions is clearly established. Its popularity is justified by its technical simplicity and by its wide applicability. It is also a valuable tool for the study of protein factors not purified past the cell extract stage. Several investigators have assessed the quantitative value of this method (Carey, 1988; Fried, 1989; Revzin, 1989; Cann, 1989, 1990). The results of these studies indicate that accurate equilibrium binding and kinetic rate constants are obtainable in some cases in which experiments are conducted under carefully controlled conditions.

The question of whether this technique is useful to the quantitative determination of cooperativity in the case of proteins that bind to multiple specific sites on DNA has not been examined as extensively. Only a few experimental studies and theoretical formulations have been published (Senear *et al.*, 1986; Fried, 1989; Tsai *et al.*, 1989; Brenowitz *et al.*, 1990; Cann, 1989; Brenowitz *et al.*, 1991a, 1991b; Cann, 1990). Since the mobility-shift assay does not distinguish between interactions at the different specific sites on DNA, only macroscopic binding parameters are obtainable (Ackers *et al.*, 1983). The macroscopic constants are composite averages of the various local site and cooperativity constants. However, the mobility-shift assay is able to estimate these macroscopic constants precisely through its ability to resolve each step of ligation of the DNA with protein. This technique is exceptional in its resolution of intermediate ligation states (*i.e.* states with some but not all sites liganded). The relative amounts of these intermediate ligation states are extremely sensitive to cooperative interactions between regulatory proteins bound at the different sites. Thus, the question to be addressed is whether there are situations for which it is possible to partition the macroscopic binding constants that are obtained into the microscopic binding constants that describe the *intrinsic* affinity of the protein ligand for a particular site and the *cooperative* interactions between liganded sites.

Because of our continuing interest in quantitative methods to study the protein-DNA interactions involved in transcriptional regulation we have conducted a systematic evaluation

* This work was supported by National Institutes of Health Grants GM41465 (to D. F. S.) and GM39929 (to M. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be sent. Tel.: 714-856-8014.

of the resolvability of cooperativity using mobility-shift assays. As illustrations, we consider three common patterns of protein-DNA interaction: (i) the cooperative binding of a protein to two specific sites (illustrated by the *Escherichia coli* Gal repressor); (ii) the cooperative binding of a bidentate protein to two specific sites (illustrated by the *E. coli* Lac repressor); and (iii) the cooperative binding of a protein to three specific sites (illustrated by the λcI repressor). The analysis of the simulated data is tested against experimental data from our laboratories for each pattern of interactions.

We assume that the equilibrium distribution of liganded species is reported accurately by the mobility-shift assay. This assumption is supported by the experimental data presented. Some limitations of this assumption have been discussed (Carey, 1988; Fried, 1989; Revzin, 1989; Cann, 1989, 1990), and investigators are cautioned that the assumption must be evaluated critically for any system. We illustrate the application of several criteria that are useful to this evaluation. A straightforward statistical-mechanical formalism is used for the construction of binding polynomials (Ackers *et al.*, 1982) and is easily adaptable to any system. We find that it is frequently possible to demonstrate that protein binding is cooperative. However, it is only possible to determine unique values for the cooperativity constants when the interacting sites are identical. The results obtained appear to be generally applicable.

MATERIALS AND METHODS

Protein and DNA Preparation—The repressor protein preparations used in this study are > 95% pure as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Gal and Lac repressor proteins were purified as described (Majumdar *et al.*, 1987; Brenowitz *et al.*, 1991a). The specific DNA binding activity of these preparations was assumed to be 100%. The dimer dissociation equilibrium constant of Gal repressor is estimated to be 2.2 nM (Brenowitz *et al.*, 1990). Lac repressor was assumed to be a stable tetramer (Brenowitz *et al.*, 1991b). λcI repressor protein was purified as described (Brenowitz *et al.*, 1986). The specific DNA binding activity is 55%, based on stoichiometry experiments of the type described by Sauer (1979) and by Johnson (1980) and on $\epsilon^{280\ nm}_{mg/ml} = 1.18$. Total active dimer concentrations were based on the stoichiometry measurements and on a dimer dissociation equilibrium constant of 20 nM (Sauer, 1979).

A 635-bp[1] linear DNA restriction fragment containing the *E. coli gal* operator sequences was isolated as described (Brenowitz *et al.*, 1990). For the Lac repressor binding experiments, the 16-bp gal O_I and O_E recognition sequences were converted to *lac* recognition sequences (designated O_I^{lac} and O_E^{lac}) by making 7- and 6-bp substitutions, respectively (Haber and Adhya, 1988). A 648-bp fragment containing the λO_R sequence was isolated as described (Senear and Ackers, 1990). All DNA fragments were agarose gel purified from CsCl gradient banded plasmid preparations and were essentially free of protein, as determined from the A_{260}/A_{280} ratio (Maniatis *et al.*, 1982).

Mobility-shift Titrations—Mobility-shift titrations of Gal and Lac repressor-operator interactions were conducted as described (Brenowitz *et al.*, 1990, 1991b) using 6% and 4% acrylamide gels (37:1 acrylamide:bisacrylamide), respectively, and TAE electrophoresis buffer (Maniatis *et al.*, 1982). The protein-DNA equilibrium mixtures were incubated at either 0 or at 20 °C (±0.1 °C) as indicated, in a regulated water bath. Aliquots (75 μl) of equilibrated binding reaction mixtures were loaded onto gels that had been preelectrophoresed for 15 min. The gels were run at a constant current of 19 mA/cm^2. The samples contained 3% glycerol to facilitate loading the gel. The samples were loaded with the current on to ensure constant resident time in the well for every sample. The experiments were conducted in Hoeffer SE 600 gel boxes either at room temperature or, for the low temperature experiment, with the temperature of the buffer surrounding the gel maintained at 2 ± 1 °C. For titrations of the λcI repressor-O_R interactions, 5% gels (36:1 acrylamide:bisacrylamide) and TGE buffer (Codish, 1989) were used. Samples (30 or 45 μl) containing 2% Ficoll were loaded onto gels preelectrophoresed for 90 min and maintained at 20 ± 0.5 °C.

Autoradiograms of the dried gels were analyzed densitometrically as described (Brenowitz *et al.*, 1986; Brenowitz and Senear, 1989). Digital representations of the autoradiograms consist of optical density (OD) values for contiguous squares, 210 × 210 μm or less in size. Gal and Lac titration autoradiograms were digitized using a microcomputer-based video densitometer (Reiner and Brenowitz, 1991). Analysis of the digital images was conducted using a microcomputer implementation[2] of the computer software originally described in Brenowitz *et al.* (1986). The two-dimensional imaging and analysis of the cI-repressor titration autoradiograms was conducted using a DEC microVAX-based Eikonix 78/99 photodiode array camera and software developed for this purpose.[3] To quantitate the DNA represented by the different bands, a box surrounding each band was defined, and the total or integrated optical density (OD_{Tot}) of the band was calculated as the sum of OD values for the squares contained within the box. The local film background OD was first subtracted from the OD value for each square.

The quantities of interest in the gel mobility-shift assay are the fractions of DNA molecules with exactly i ligands bound, Θ_i, where $i = 0$, 1, or 2 for two-site and $i = 0$, 1, 2, or 3 for three-site systems. Θ_i for each band was calculated from $\Theta_i = OD_{Tot,i}/\sum_i OD_{Tot,i}$, where the summation is over all of the bands in a given lane. Since $\sum_i \Theta_i \equiv 1$, there is one less independent experimental measurement than the number of bands observed. To derive expressions for Θ_i, we follow a rigorous and perfectly general statistical mechanical approach (Hill, 1960). This simple approach is applicable to any number and arrangement of binding sites and ligands provided that the microscopic configurations can be explicitly written out. The model for a system composed of two interacting DNA binding sites and a single protein ligand is unambiguous, as shown in Table I. The relative probability of each of the configurations in the table is given by

$$f_s = \frac{e^{-\Delta G_s/RT} L_j}{\sum_{sj} e^{-\Delta G_s/RT} L_j} \tag{1}$$

where ΔG_s is the sum of free energy contributions for configuration s, (Table I), R is the gas constant, T is the absolute temperature, L is the concentration of free protein ligand, and j is the number of ligands bound to configuration s.

The (binding) equations for the Θ_i are obtained by summing the relative probabilities of the appropriate configurations. For a two-site system (Table IA), this yields

$$\Theta_0 = f_1 \tag{2a}$$

$$\Theta_1 = f_2 + f_3 \tag{2b}$$

$$\Theta_2 = f_4 \tag{2c}$$

These equations are greatly simplified by making the familiar substitution $k = e^{-\Delta G/RT}$ to give

$$\Theta_0 = 1/Z \tag{3a}$$

$$\Theta_1 = (k_1 + k_2)\cdot L/Z \tag{3b}$$

$$\Theta_2 = (k_1 k_2 k_{12})\cdot L^2/Z \tag{3c}$$

Z is the binding polynomial (Wyman, 1964) equal to $1 + (k_1 + k_2)\cdot L + (k_1 k_2 k_{12})\cdot L^2$. k_1 and k_2 are microscopic equilibrium association constants for intrinsic binding to sites 1 and 2; k_{12} is the constant describing cooperative interactions when both sites are liganded.

Cooperativity and its potential involvement in a regulatory mechanism are assessed based on k_{12} (or the corresponding ΔG). However, since the three *microscopic* equilibrium constants, k_1, k_2, and k_{12}, appear in only two combinations they can be replaced by two *macroscopic* equilibrium constants defined by $K_1 = k_1 + k_2$ and $K_2 = k_1 k_2 k_{12}$. Therefore, only K_1 and K_2 can be determined from a single mobility-shift experiment. It is evident that it is not possible to determine all three microscopic equilibrium constants uniquely from the three equations described above.

This statistical mechanical approach to the derivation of binding

[1] The abbreviations used are: bp, base pair; Bistris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)-propane-1,3-diol.

[2] M. Brenowitz, P. Reiner, and B. Turner, unpublished computer program, available upon request.

[3] D. F. Senear, unpublished computer program, available upon request.

TABLE I

Configurations and associated free energy states for protein binding to DNA

Binding sites are denoted by L if liganded. Cooperative interactions are denoted by ↔. The total Gibbs free energy of each configuration (ΔG_s) relative to the unliganded reference state is given as the sum of contributions from six energy changes (column 3). ΔG_i (i = 1, 2, or 3) are the intrinsic free energy changes for binding to the individual operator sites; ΔG_{ij} are the free energies of cooperative interaction between liganded sites, defined as the difference in free energy to fill the sites simultaneously (ΔG_T) and the free energy to fill them individually ($\sum_i \Delta G_i$).

A. Two sites (general case)

Species	Binding configurations			Free energy contributions	Total free energy
	Site 1		Site 2		
1				Reference state	ΔG_{s1}
2			L	ΔG_1	ΔG_{s2}
3	L			ΔG_2	ΔG_{s3}
4	L	↔	L	$\Delta G_1 + \Delta G_2 + \Delta G_{12}$	ΔG_{s4}

B. Three sites (specific case)

Species	Binding configurations					Free energy contributions	Total free energy
	Site 1		Site 2		Site 3		
1						Reference state	ΔG_{s1}
2	L					ΔG_1	ΔG_{s2}
3			L			ΔG_2	ΔG_{s3}
4			L			ΔG_3	ΔG_{s4}
5	L	↔	L			$\Delta G_1 + \Delta G_2 + \Delta G_{12}$	ΔG_{s5}
6			L	↔	L	$\Delta G_2 + \Delta G_3 + \Delta G_{23}$	ΔG_{s6}
7	L	⟷			L	$\Delta G_1 + \Delta G_3 + \Delta G_{13}$	ΔG_{s7}
8	L	↔	L		L	$\Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_{12}$	ΔG_{s8}
9	L		L	↔	L	$\Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_{23}$	ΔG_{s9}
10	L	⟵	L	⟶	L	$\Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_{13}$	ΔG_{s10}

equations is easily extended to three or more sites. However, there are additional pairwise cooperativity terms (free energies) since there are multiple configurations with exactly two ligands bound. Also, there are multiple possible patterns of cooperative interactions to be considered in configurations with greater than two ligands bound. For example, the most general model for three sites calls for a free energy term ΔG_{123} (Senear and Ackers, 1990), which simply defines the free energy of cooperative interaction for the configuration in which all three sites are liganded, as the difference between the free energy change to ligate the sites separately, and the free energy change to ligate them together. This is analogous to the definition of ΔG_{12} given in Table 1. However, in most well studied gene regulatory systems, cooperative interactions are thought to be primarily pairwise in nature (Johnson *et al.*, 1979; Dandanell *et al.*, 1987; Haber and Adhya, 1988; Tsai *et al.*, 1989). The salient features of a model for a three-site system which considers only pairwise cooperative interactions are represented by the operator configurations shown in Table 1B. The equations for the fractions of DNA molecules with i ligands bound are given by

$$\Theta_0 = 1/Z \tag{4a}$$

$$\Theta_1 = (k_1 + k_2 + k_3)\cdot L/Z = K_1\cdot L/Z \tag{4b}$$

$$\Theta_2 = (k_1k_2k_{12} + k_1k_3k_{13} + k_2k_3k_{23})\cdot L^2/Z = K_2\cdot L^2/Z \tag{4c}$$

$$\Theta_2 = (k_1k_2k_3(k_{12} + k_{13} + k_{23}))\cdot L^3/Z = K_3\cdot L^3/Z \tag{4d}$$

where $Z = 1 + K_1\cdot L + K_2\cdot L^2 + K_3\cdot L^3$.

Numerical Techniques—Titration experiments for two-site and for three-site operators were simulated by calculating values of Θ_i. For two-site operators, Θ_i (i = 0, 1, and 2) data were calculated by using Equations 3a–3c, at each of 15 concentrations of free repressor. Input values of the equilibrium constants assumed are shown in Table II for representative cases. Repressor concentrations for which Θ_i were calculated were spaced evenly on a log scale and were chosen to cover the range from 1 to 99% fractional saturation of the operators. For three-site operators, Θ_i (i = 0, 1, 2, and 3) data were calculated in a similar manner, except that Equations 4a–4d were used, and Θ_i values were calculated at each of 20 repressor concentrations. The larger number of repressor concentrations allowed us to use increments between data points which were similar to those used in the two-site simulations but still to cover the concentration range from 1% fractional saturation of the lowest affinity site to 99% fractional saturation of the highest affinity site. Input values of the equilibrium constants assumed are shown in Table V for representative cases.

The simulated experiments were designed to reflect the quality of the experimental titration data shown in Figs. 1, 4, and 5. Experimental variation was included by individually adding a small error increment to each calculated Θ_i value. The error increment for each data point was chosen from a Gaussian distribution of numbers with a mean of zero and a standard deviation of either 0.05, 0.08, or 0.10 in different simulated experiments. To approximate closely the actual analysis of experimental data, Θ_i values greater than 1 or less than 0 were set equal to 1 and 0, respectively, and the sum of the Θ_i values for each lane (repressor concentration) was then normalized to 1.

Both simulated and experimentally observed data were analyzed according to the appropriate equations (*e.g.* Equations 3a–3c, where the substitutions $K_1 = k_1 + k_2$ and $K_2 = k_1k_2k_{12}$ are made in the case of two-site operators; Equations 4a–4d where the substitutions $K_1 = k_1 + k_2 + k_3$, $K_2 = k_1k_2k_{12} + k_1k_3k_{23} + k_2k_3k_{13}$, and $K_3 = k_1k_2k_3(k_{12} + k_{23} + k_{13})$ are made in the case of the three-site operators). Nonlinear least squares methods of parameter estimation were used. The analysis program (Johnson and Frasier, 1985) uses a variation of the Gauss-Newton procedure (Hildebrand, 1956) to determine the best fit, model-dependent parameter values corresponding to a minimum in the variance. The N-dimensional parameter space is searched for the variance ratio predicted by an F statistic (Box, 1960) to determine the worst case joint confidence limits for the parameters. The confidence limits reported correspond to approximately 1 standard deviation.

RESULTS AND DISCUSSION

Accuracy of Gel Mobility-shift Experiments—The assumption that the gel mobility-shift assay reports accurately the distribution of liganded states is crucial to the analysis of cooperativity which follows. Inaccuracy is presumed to result primarily from dissociation of the liganded complexes during extended electrophoresis. Cann (1989) argued that this might lead to significant underestimation of the liganded complexes in studies conducted under equilibrium conditions, even for very tight binding protein ligands. The problem is expected to be the greatest for intermediate ligation states, which are those most sensitive to cooperativity. Some (Carey, 1988; Revzin, 1989) suggested that accurate estimates might be possible only for Θ_0, the unliganded DNA.

To address the question of accuracy we introduce three criteria and demonstrate their application to the experimental data used in our further analysis of cooperativity. The arguments for accuracy are based on (i) comparison of gel mobility-shift data to isotherms predicted by an independent experimental method (DNase I footprinting); (ii) comparison of the isotherm that represents the disappearance of unliganded DNA with the isotherm that represents the appearance of liganded DNA, for DNA with a single binding site; and (iii) a conservation of mass argument, in which only the DNA in tightly defined bands is considered.

Comparison between Gel Mobility-shift and Footprint Analysis of λcI Repressor Binding to O_R—The binding of the λcI repressor to the three-site operator O_R was studied using the mobility-shift method (Fig. 1). In our experiments on O_R only three bands are observed rather than the four expected (*i.e.* corresponding to unliganded and three ligation states). We interpret the first retarded band as being composed of both singly and doubly liganded DNA, based on experiments conducted using mutant operators (reduced valence) in which

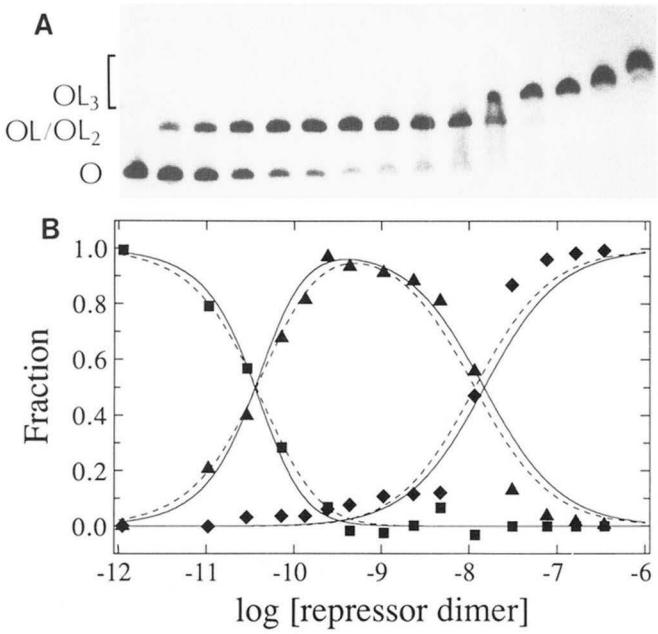

FIG. 1. *A*, gel mobility-shift titration of λcI repressor binding to O_R. Binding conditions are 10 mM Bistris, pH 7.00, 2.5 mM $MgCl_2$, 1 mM $CaCl_2$, 100 μg/ml bovine serum albumin, 2 μg/ml calf thymus DNA, and 200 mM KCl at 20 °C. Bands corresponding to unliganded operator DNA (O) complexes with 1 or 2 (OL, OL_2) and complexes with 3 (OL_3) or more repressors bound are marked. *B*, fractions of DNA molecules with i ligands bound (Θ_i) were calculated from the bands in *A* as described under "Materials and Methods." *Squares*, unliganded operator; *triangles*, operator with one or two repressors bound; *diamonds*, operator with three repressors bound. The *solid curves* result from fitting the data according to Equations 4*a*–4*d*, to give $K_1 = 2.24$ (+1.08, −0.73) × 10^{10} M^{-1}, $K_2 = 1.36$ (+3.75, −1.00) × 10^{20} M^{-2}, $K_3 = 1.20$ (+3.26, −0.92) × 10^{28} M^{-3}. The square root of the variance is 0.056. *Broken curves* are predicted from $K_1 = 1.43 \times 10^{10}$ M^{-1}, $K_2 = 3.94 \times 10^{20}$ M^{-2}, $K_3 = 2.70 \times 10^{28}$ M^{-3}, calculated from values of the interaction energies obtained from DNase I footprint experiments.

only two sites are active in specific repressor binding.[4] At the highest repressor concentrations some additional retardation of the third band is evident. The interpretation that this is caused by the nonspecific binding of repressor to the DNA outside of the operator region is consistent with our separate studies of the nonspecific binding of repressor to DNA (Senear and Batey, 1991).

As predicted by Cann (1989) slight dissociation of the liganded complexes during electrophoresis is evidenced by a slight smear forward of the liganded bands (clearly visible in the original autoradiogram) except at the highest repressor concentrations. Therefore, to analyze these data each band was defined as extending from its well defined trailing edge to the trailing edge of the next band forward. The data plotted are the fractions determined in this manner. The *solid curves* represent the best fit to the data, using Equations 4a–4d. The *dashed curves* are the isotherms predicted by the macroscopic equilibrium constants (K_1, K_2, K_3) calculated from values of the interaction free energies obtained from DNase I footprint experiments (Senear and Ackers, 1990). The correspondence between the two sets of isotherms is excellent. In addition, the values of the calculated macroscopic constants are within the confidence limits of the fitted values. This correspondence supports our analysis of the mobility-shift data and argues that accurate results are obtainable even when slight dissociation of liganded complexes is observed. Similar correspondence has also been observed at another reaction condition[4] and in the case of at least one other repressor-operator interaction (Carey, 1988).

Separate Analysis of Liganded and Unliganded Bands—Another means to assess the accuracy of the titration data is to compare isotherms determined separately from the densities of the liganded and unliganded bands in mobility-shift assays of binding to single-site DNAs. A significant difference between this analysis and the general procedure described under "Materials and Methods" is that the absolute band densities representing complexed and free DNA are considered as independent experimental determinations, free from the constraint, $\Sigma\ \Theta_i \equiv 1$. The resulting curves are treated as transition curves, with upper and lower limits determined numerically (Brenowitz *et al.*, 1986). These curves will be coincident only if complex dissociation does not occur. Fig. 2 presents the results of one such analysis of Gal repressor binding to a single operator. The titrations for liganded and unliganded bands are described by the Langmuir isotherm, $\bar{Y} = k \cdot L/(1 + k \cdot L)$, and by $1 - \bar{Y}$, respectively, where $\bar{Y}$ is the fractional saturation of the site.

Quantitative analysis appears to indicate slightly tighter binding defined by the curve for the free DNA than by the curve for the liganded DNA. The difference accounts for a factor of 1.5-fold in k, or 0.2 kcal/mol difference in the free energy change. This result again supports the prediction of Cann (1989) that some dissociation occurs during electrophoresis of even tightly bound protein-DNA complexes. It should be noted, however, that the fitted k values are distinguishable (*i.e.* the confidence limits do not overlap), and therefore the slight dissociation is discernible, only because of the remarkable precision of these data. The variances of the fitted curves are approximately half of those of the other experimental data presented herein (compare with Figs. 1, 4, and 5). At the more usual precision obtained, differences of this magnitude are not discernible.

Conservation of Mass—A third potential criterion for accuracy was also evaluated. This criterion relies on a conser-

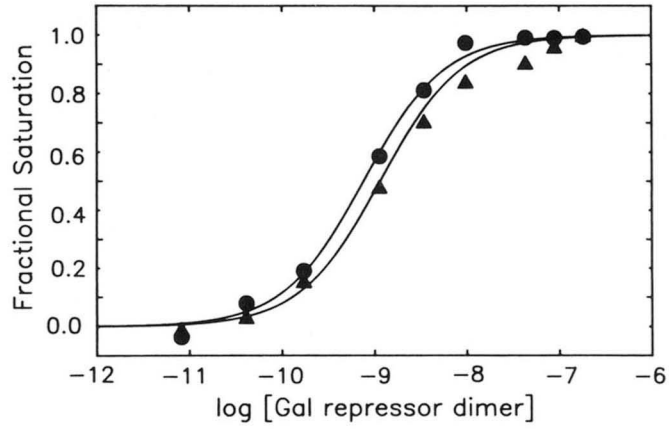

FIG. 2. **Separate analysis of the protein complexed and free DNA bands from a mobility-shift titration of Gal repressor binding to DNA in which only site O_E binds repressor.** Site O_I was inactivated by base pair substitutions. Binding conditions are 25 mM Bistris, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 2 mM dithiothreitol, 50 μg/ml bovine serum albumin, 2 μg/ml calf thymus DNA, and 100 mM KCl, titrated to pH 7.0 at 20 °C. *Triangles* and *circles* represent the protein-DNA complex and free DNA, respectively. The *solid lines* indicate the best fit curves obtained as described under "Materials and Methods." $k_1 = 8.71$ (±1.31) × 10^8 M^{-1} ($\sigma = 0.027$) and $k_1 = 1.32$ (±0.20) × 10^9 M^{-1} ($\sigma = 0.027$) were obtained for the complexed and free DNA bands, respectively. These values compare with $k_1 = 1.12$ (±0.18) × 10^9 M^{-1} ($\sigma = 0.018$) resolved by using the standard analysis described under "Materials and Methods."

[4] D. F. Senear, unpublished observations.

vation of mass (DNA) argument and would be useful when there are multiple shifted bands (*i.e.* multiple ligation states). The question is whether the total DNA (measured as optical density on an autoradiogram) observed in the bands in each lane of the gel is equal to the DNA that was originally loaded. The latter is experimentally determined as the DNA in the (unliganded) band in a reference lane that contains no ligand. When each band is conservatively defined by a box that just encloses its contours (see "Materials and Methods"), any DNA that smears forward from a liganded band because of dissociation during electrophoresis (Cann, 1989) will not be counted. If there is no evidence for dissociation during the electrophoresis then it is reasonable to infer that the original equilibrium distribution of species is observed accurately.

As a test, a calculation was performed on a set of 20 representative mobility-shift titrations of Gal repressor binding to either one-site or two-site operators. The ratio of DNA in a lane to the DNA in the reference lane (no Gal repressor) was calculated for each lane with a non-zero Gal repressor concentration. The ratio was then plotted as a function of repressor concentration. A systematic decrease in this ratio at subsaturating protein concentrations would indicate significant dissociation. This was not observed. However, substantial random variation in the ratio (mean = 0.97 ± 0.23) was observed. We note that a loss of even 10–15% of the complex DNA caused by dissociation could perturb significantly the apparent equilibrium binding constants yet is below the level of the precision of these data. Therefore, we conclude that this criteria is not useful, at least as the experiment is conducted currently in our laboratories.

Accuracy of Densitometric Detection of DNA—Quantitation of the DNA in different bands on mobility-shift gels was accomplished by the use of computer-aided two-dimensional densitometric analysis of autoradiograms. The autoradiograms were exposed and developed under strictly controlled conditions. Concerns about the linearity and sensitivity of film response have led some investigators to question the accuracy of binding curves obtained in this manner (Fried, 1989; Letovsky and Dynan, 1989). The development of spatially restricted, direct radioactivity detectors (*i.e.* Ambis® and Betagen® scanners) offers an alternative to autoradiography for the quantitation of gel electrophoresis experiments. The main advantage of these detectors is that direct quantitation of radioactivity obviates potential problems with the range and linearity of the photographic film response.

To address the concerns that have been raised, we have analyzed mobility-shift titrations both by using autoradiography as described above and by using a direct detector. Densitometric analysis was conducted using the video capture and display system described under "Materials and Methods." Direct detection was accomplished by using an Ambis® scanner equipped with a 0.8 × 3.2-mm mask. Sixteen-hour and 90-min scans with the Ambis® scanner yielded essentially identical results. Representative results shown in Fig. 4*A* compare the two methods for analyzing the same experiment. The direct detector offers no advantage in the precision of the titration data. The precision appears to be limited by the manipulations involved in conducting the titrations rather than by the detection technique employed. The two methods of detection appear equally sensitive and accurate since the two sets of points describe the same isotherms. We conclude that carefully conducted autoradiography can provide high quality data for quantitative analysis of mobility-shift titrations. It should be noted, however, that two-dimensional analysis of gel electrophoretograms does offer critical advantages in both accuracy and precision over one-dimensional techniques (Bossinger *et al.*, 1979; Brenowitz *et al.*, 1986).

Resolution of Cooperativity by Mobility-shift Assays—Cooperatively interacting binding systems are characterized by decreased populations of intermediate ligation states. In classical binding experiments in which the quantity determined is the fraction of sites liganded, this fact generates steeper binding isotherms (sigmoidal curves on a linear scale). The resolution of macroscopic equilibrium constants depends on fitting the detailed shape of the binding curve. However, the sensitivity of the curve shape to cooperativity is low so that even reasonably precise data usually fail to resolve uniquely the equilibrium constants for the intermediate ligation states. It is not possible to address the issue of cooperativity with such data.

A particular advantage of the mobility-shift assay is its direct detection of the intermediate ligation states. Fig. 3 illustrates the dramatic effect of even moderate cooperativity on the population of singly liganded DNA, for DNA with two identical sites. When binding is noncooperative, the populations of unliganded, singly liganded, and doubly liganded DNA are binomially distributed. Θ_1 reaches a maximum (Θ_1^{max}) of 0.5 at half-saturation. A very moderate 5.6-fold cooperative effect ($\Delta G_{12} = -1$ kcal/mol) reduces Θ_1^{max} to 0.3. Even with relatively noisy data, of the type commonly obtained in these experiments, Θ_1^{max} is usually able to be estimated to a precision of about 0.02–0.03 from nonlinear least squares analysis. As demonstrated below this is sufficient to resolve the cooperativity to within severalfold, provided that k_{12} is less than about 1,000.

Of course as discussed above the combinations of parameters used to simulate the curves 2–5 in Fig. 3 are not unique. There are always alternative combinations of k_1, k_2, and k_{12} which give the same values of K_1 and K_2. For example, the curves represented by the *broken lines* in Fig. 3 are generated by $k_1 = k_2 = 9.0 \times 10^8$ M^{-1} and $k_{12} = 31$ (equivalent to −2 kcal/mol at 20 °C) and also by $k_1 = 1.7 \times 10^9$ M^{-1}, $k_2 = 9.0 \times 10^7$ M^{-1}, and $k_{12} = 163$. Therefore, these curves do not address the issue of site heterogeneity. However, as will be shown

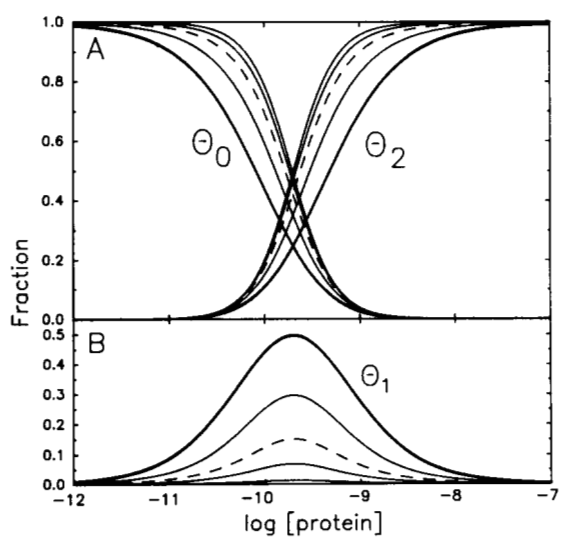

FIG. 3. **Simulated mobility-shift titration curves for a protein ligand binding to two identical sites *A*, Θ_0 and Θ_2; *B*, Θ_1.** The *bold curves* reflect $k_i = 5.0 \times 10^9$ M^{-1} ($\Delta G_i = -13.0$ kcal/mol) and $k_{12} = 1.0$ ($\Delta G_{12} = 0.0$ kcal/mol). In successive curves, the total energy of binding is held constant while the magnitude of cooperative free energy, ΔG_{12}, is increased from 0 to −5 kcal/mol ($k_{12} = 5.4 \times 10^3$) by increments of −1 kcal/mol. The curve depicting $\Delta G_{12} = -4.0$ kcal/mol was omitted for clarity. The *dashed curves* depict $\Delta G_{12} = -2.0$ kcal/mol.

TABLE II

Analysis of simulated titration experiments for binding of protein to two DNA binding sites

Mobility-shift data consisting of values of Θ_0, Θ_1 and Θ_2 were simulated as described under "Materials and Methods," using the microscopic binding constants listed. The product of the microscopic constants ($k_1 \cdot k_2 \cdot k_{12}$) was held constant at 2.5×10^{19} M^{-2} ($\Delta G = -26$ kcal/mol) as increasing extent of cooperativity (k_{12}) was introduced. The site heterogeneity, defined by k_2/k_1, was also held constant, equal to 1 ($\Delta G = 0$ kcal/mol), 31 ($\Delta G = -2.0$ kcal/mol), or 173 ($\Delta G = -3.0$ kcal/mol). The random error level was 0.05. The simulated data were analyzed by using Equations 3a–3c, which were reformulated with macroscopic product binding constants $K_1^2/4$ and K_2.

Microscopic binding constants used for simulations				Macroscopic binding constants obtained from curve fitting ($\times 10^{-19}$)		
k_{12}	ΔG_{12}[a]	$k_1(\times 10^{-9})$	$k_2(\times 10^{-9})$	K_2	$K_1^2/4$	$K_2?K_1^2/4$
$k_2/k_1 = 1$						
1	0	5.02	5.02	1.81 ± 0.63	1.46 ± 0.77	=
6	−1	2.13	2.13	2.54 ± 0.93	0.56 ± 0.37	>
31	−2	0.901	0.901	2.63 ± 1.11	0.14 ± 0.13	>
$k_2/k_1 = 31$						
1	0	0.901	28.0	1.98 ± 0.63	16.5 ± 7.3	<
6	−1	0.381	11.9	2.54 ± 0.84	5.19 ± 2.45	=
31	−2	0.162	5.02	2.92 ± 1.16	0.62 ± 0.45	>
173	−3	0.0684	2.13	2.11 ± 0.76	0.088 ± 0.057	>
$k_2/k_1 = 173$						
6	−1	0.162	28.0	2.62 ± 0.84	19.8 ± 8.8	<
31	−2	0.0684	11.9	2.67 ± 0.83	4.71 ± 1.61	=
173	−3	0.0290	5.02	2.00 ± 0.99	0.51 ± 0.43	>
966	−4	0.0123	2.13	3.12 ± 1.15	0.19 ± 0.11	>

[a] In kcal/mol at 20 °C.

below, there is no combination of k_1 and k_2 and $k_{12} < 31$ which generates these curves. Therefore, cooperativity is still inferred in this case.

A general rule for this behavior is obtained by comparing the macroscopic product constants, K_1 and K_2. Consider two noncooperative ($k_{12} = 1$) and identical sites ($k_1 = k_2$ defined as equal to an intrinsic binding constant, k_i). This leads to, $K_1 = 2 \cdot k_i$ and $K_2 = k_i^2 = K_1^2/4$. For heterogeneous ($k_1 \neq k_2$) noncooperative sites, $K_2 < K_1^2/4$. Whenever $K_2 < K_1^2/4$, cooperativity can not be inferred because identical isotherms are obtained for cooperative binding to heterogeneous sites and for noncooperative binding to homogeneous sites. However, if $K_2 > K_1^2/4$ then $k_{12} > 1$, and binding is cooperative. Notice that all cases of negative cooperativity ($k_{12} < 1$) lead to $K_2 < K_1^2/4$, and so are indistinguishable from noncooperative binding to heterogeneous sites. Although critically dependent on the uncertainty of the estimates of K_1 and K_2 and therefore on the precision of the data we might expect to demonstrate positive cooperativity whenever its magnitude equals or exceeds the site heterogeneity.

To test this expectation several series of mobility-shift titration experiments were simulated, as described under "Materials and Methods." In each series the product of the microscopic constants ($k_1 \cdot k_2 \cdot k_{12}$) was held constant, equal to 2.5×10^{19} M^{-2} (describing a constant total Gibbs free energy change to saturate the two sites, $\Delta G = -26$ kcal/mol). The site heterogeneity, defined by k_2/k_1, was also held constant whereas the extent of cooperativity (k_{12}) was varied systematically. The data were then fit to Equations 3a–3c, which were formulated in terms of $K_1^2/4$ and K_2. The results of the analysis of representative simulated titrations are shown in Table II. Assumed values of k_1, k_2, and k_{12} are indicated in the table. The random error level in these simulated data was 0.05 (standard deviation).

For virtually every simulated titration the fitted values of $K_1^2/4$ and K_2 are well within experimental error of the values of $K_1^2/4$ and K_2 calculated from k_1, k_2, and k_{12} used to simulate the data. This supports our expectation that the mobility-shift data are capable of resolving these binding constants. As expected, K_2 is distinguishably greater than $K_1^2/4$ whenever the cooperativity equals or exceeds the site heterogeneity. For example, if the site hetereogeneity is defined by $k_2/k_1 = 31$, then K_2 is distinguishably greater than $K_1^2/4$ (indicating positive cooperativity) when $k_{12} \geq 31$. Higher levels of random absolute error affect the precision of the estimates of K_1 and K_2, and correspondingly more cooperativity is required to stand out above the experimental uncertainty.

Identical Binding Sites—Cooperativity is exactly determined for the special case in which the two sites have identical affinity for ligand. In protein-DNA systems such equivalence of binding sites might be inferred from the identify of the DNA sequences of the sites. However, numerous examples of context effects of surrounding sequence on binding site affinity suggest caution when sequence identity is the only criterion. Equivalence of the protein binding sites on DNA should be confirmed experimentally by analysis of mutants in which one or more of the binding sites are inactivated by base pair substitutions (for discussion see Brenowitz *et al.*, 1986).

When the binding sites are equivalent the data can be analyzed directly with the microscopic binding and cooperativity constants, k_i and k_{12}, since $K_1 = 2k_i$ and $K_2 = k_i^2 k_{12}$. When simulated titrations for which $k_1 = k_2$ was assumed (*e.g.* as in lines 1–3 of Table II) were analyzed in this manner, the fitted values of k_i were identical (to within experimental error) to the assumed values of k_1 and k_2 in every case. For simulated titrations for which assumed values of k_{12} were less than 966 ($\Delta G_{12} = -4$ kcal/mol) k_{12} was also resolved accurately and precisely. However, as k_{12} is increased above 966 M^{-1}, its resolution becomes increasingly less precise, as indicated by increasingly large and asymmetric confidence limits. This result reflects an asymptotic limit to the mobility-shift curves (Fig. 3) at which the fraction of singly liganded intermediate approaches zero and becomes insensitive to further cooperativity. Thus, one significant advantage of fitting k_i and k_{12} directly is that confidence limits reflecting the precision of the estimates are obtained.

Mobility-shift Titrations of Gal and Lac Repressor-Operator Interactions—The principles outlined above have been applied to the analysis of the site-specific binding of several proteins to DNA. An illustrative example is the binding of *E. coli* Gal repressor to its operators, O_E and O_I. This system is of interest because Gal repressor dimers are proposed to bind

cooperatively to the two operators (separated by 11 helical turns) to form a DNA loop (Adhya, 1989). Identical binding affinities of O_E and O_I for Gal repressor have been demonstrated (Brenowitz *et al.*, 1990). Mobility-shift titrations of this binding under two conditions (20 and 0 °C) are shown in Fig. 4. Three electrophoretic bands are observed. These correspond to free DNA and to DNA with one and with two proteins bound. This pattern of bands is that expected for a system in which a protein binds to two sites on the DNA and the protein-DNA complexes do not assume unusual conformations. Analysis of these data using Equations 3a–3c, subject to the constraint $k_1 = k_2 = k_i$ yielded unique values of k_i and k_{12} (*A*, $k_i = 2.14\ (\pm 0.32) \times 10^9\ M^{-1}$, $k_{12} = 1.10\ (\pm 0.25)$, $\sigma = 0.039$ (σ is the square root of the variance of the fitted curves); *B*, $k_i = 1.98\ (\pm 0.26) \times 10^7\ M^{-1}$, $k_{12} = 3.56\ (\pm 0.83)$, $\sigma = 0.045$). These are in good agreement with values determined by DNase I footprint titration studies (Brenowitz *et al.*, 1990). Thus, for two-site systems, the mobility-shift experiment is quite capable of precise and accurate resolution of noncooperative and of mildly cooperative binding effects when analyzed in this manner.

Mobility-shift titrations of the Lac repressor binding to DNA containing two operator sites have allowed us to assess the importance of separate resolution of each intermediate ligation state to the estimation of binding and cooperativity constants. Representative titrations of Lac repressor binding to *gal* operator DNA in which the operator sites O_E and O_I were both converted to the *lac* recognition sequence (O_E^L and O_I^L) are shown in Fig. 5*A*. These titrations differ strikingly from Gal repressor titrations. The differences are caused by the fact that the bidentate Lac repressor tetramers can bind simultaneously to both operator sites to mediate the formation of a protein-DNA looped complex (Gralla, 1989). Two mobility-shifted bands are observed: a strongly retarded, low repressor concentration band, and an intermediate mobility band at higher repressor concentration. These bands have been identified previously as corresponding to operator DNA with a single repressor bound (including both the species in which Lac repressor is bound to only one site, and the looped complex with repressor bridging the two sites) and to DNA with repressor bound independently to both sites. A detailed discussion of these band assignments can be found in Brenowitz *et al.* (1991b). A phenomenological model for these interactions which accounts for the bidentate nature of the repressor is in Table III.

To assess the necessity for separate resolution of each of the intermediate states these data were analyzed in two ways. First, the three bands, corresponding to Θ_0, Θ_1, and Θ_2, were quantitated separately (Fig. 5*A*) and analyzed according to the model in Table III. The results (Table IV) indicate precise estimation of the cooperative binding constant, k_{12}. Note that substantial asymmetry in the fitted curve for Θ_2 reflects the data correctly. Second, the bands corresponding to Θ_1 and Θ_2 were combined (Fig. 5*B*) to mimic a situation in which separate resolution is not obtained. The data were analyzed as before except that the equations for Θ_1 and Θ_2 were summed. The least squares program could not converge to a unique solution when k_{12} was a fitted parameter. Table IV shows the result when k_{12} was fixed at different values. In this case the goodness of the fit, reflected in the variance of the fitted curves, is insensitive to k_{12} over a very broad range, indicating that k_{12} is not resolved. This highlights the strong dependence of the analytical resolution of cooperativity on the experimental resolution of each of the intermediate ligation states.

Resolution of Cooperative Effects in Three Site Systems—To assess the generality of the conclusion drawm from the analysis of two-site protein-DNA systems the issue of cooperativity in three-site systems was also addressed. As with two-site systems, there are simple relationships among the macroscopic equilibrium constants, K_1, K_2 and K_3, which are useful to the analysis of cooperativity. Thus, if $K_2 > K_1^2/3$, then the second ligand binding event is cooperative. Of course, this alone provides no information as to which of the three microscopic pairwise interactions (Table IB) is/are cooperative. Similarly, if $K_3 > K_2K_1/3$, then the third ligand binding event is cooperative.

To test the utility of these relationships a series of mobility-shift experiments was again simulated, as described earlier. There are significant complications inherent to three (or more) site systems. First, all pairs of sites need not interact with the same cooperativity. Second, when more than two ligands are bound the cooperative interaction may be different from a linear combination of the pairwise interactions. To simplify the presentation of three site systems we consider first the situation in which all pairwise cooperative interactions between liganded sites, denoted by k_c, are the same, and

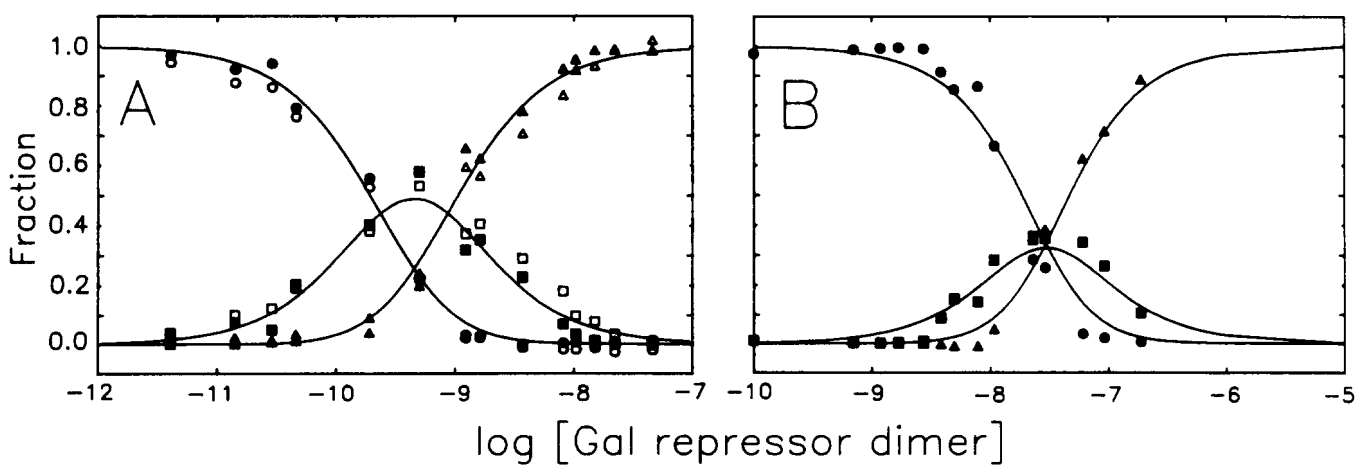

FIG. 4. **Mobility-shift titration of Gal repressor binding to double operator (O_E/O_I) DNA.** *A*, data at 20 °C. *B*, data at 0 °C. Binding conditions are as in Fig. 2 except that 50 mM KCl was used at 20 °C. *Circles*, Θ_0; *squares*, Θ_1; *triangles*, Θ_2. The *solid lines* represent the best fits of Equations 3a–3c to the data. *A* also compares quantitation by autoradiography and densitometry and by direct radioactivity detection. *Solid* and *open symbols* indicate the densitometrically and directly scanned data, respectively.

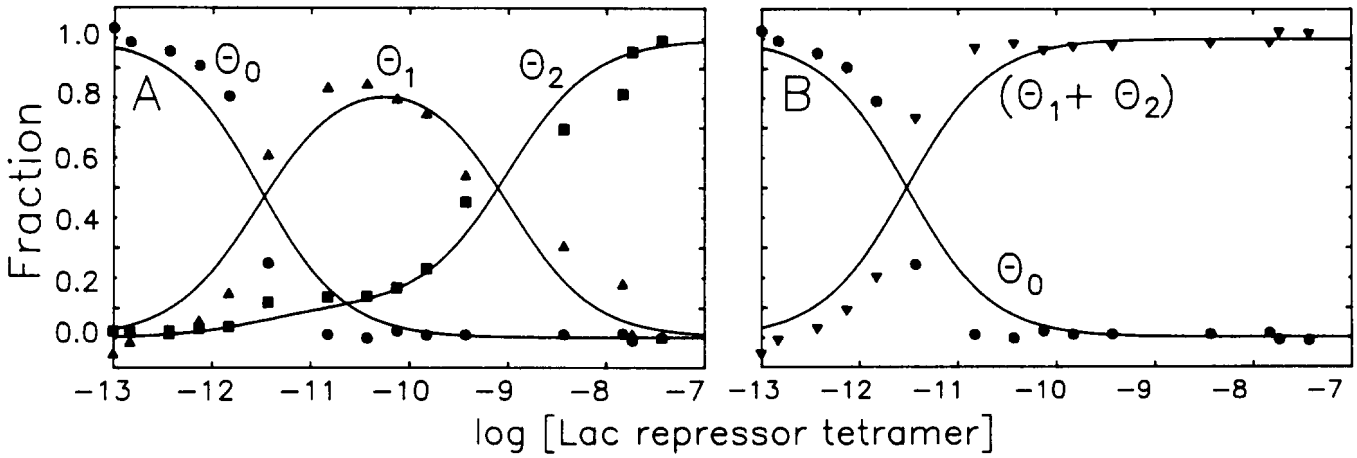

FIG. 5. **Mobility-shift titrations of Lac repressor binding to double operator (O_1^L/O_E^L) DNA.** Experimental conditions are as in Fig. 2. In *A* the bands representing the singly and doubly liganded complexes were analyzed separately. In *B* they were analyzed together. *Solid lines*, best fit curves from analysis of the data according to the model in Table III. Each of the pairs of parameter values (k_i and k_{12}) listed in the last four lines of Table IV predicts the same titration curves, which are those depicted as the *solid lines* in *panel B*. *Circles*, Θ_0; *triangles*, Θ_1 (including complexes looped by bidentate tetramer); *squares*, Θ_2; *inverted triangles*, the sum ($\Omega_1 + \Theta_2$).

TABLE III

Binding configurations and associated free energy states for bidentate Lac repressor binding to two sites

Simplified model for the interaction of Lac repressor tetramer with two operator binding sites on DNA. The model assumes equivalent binding sites ($\Delta G_1 = \Delta G_2 = \Delta G_i$) and that Lac repressor tetramer does not dissociate to dimers in solution (see Brenowitz *et al.*, 1991b for discussion).

Species	Binding configurations		Free energy contributions	Total free energy
	Site 1	Site 2		
1			Reference state	ΔG_{s1}
2		L	ΔG_i	ΔG_{s2}
3	L		ΔG_i	ΔG_{s3}
4[a]	←L→		$\Delta G_i + \Delta G_{12}$	ΔG_{s4}
5	L	L	$2 \cdot \Delta G_i$	ΔG_{s5}

[a] Species 4 represents a looped complex, cross-linked by the bidentate repressor tetramer. ΔG_{12} is defined phenomenologically, as in Table I. No mechanistic interpretation is intended.

TABLE IV

Equilibrium constants for the binding of bidentate Lac repressor to two sites

The titration data in Fig. 5 were analyzed according to the statistical mechanical model in Table III.

k_i (M^{-1})	k_{12}	σ[a]
Fig. 5*A*		
$9.25 (\pm 2.68) \times 10^9$	19.1 (±5.45)	0.081
Fig. 5*B*		
$4.31 (\pm 1.07) \times 10^{10}$	(0.01)	0.074
$3.69 (\pm 0.90) \times 10^{10}$	(1)	0.075
$9.61 (\pm 2.52) \times 10^9$	(20)	0.077
$2.43 (\pm 0.63) \times 10^8$	(1,000)	0.077

[a] Square root of the variance of the fitted curves.

the fully saturated operator is represented by three states, each exhibiting one pairwise cooperative interaction (species 8–10, Table IB). Results of the analysis of representative simulated titrations are in Table V.

The conclusions to be drawn from these results echo the conclusions drawn from the analysis of two-site systems. Cooperativity at the second ligation step is discernible, that is, stands above the uncertainty in the estimation of the macroscopic constants, when the pairwise cooperativity equilibrium constant, k_{ij}, equals or exceeds the site heterogeneity, defined by the ratio, k_i/k_j. Three different patterns of site heterogeneity are presented: (i) two sites of equal affinity, plus a higher affinity site; (ii) two sites of equal affinity, plus a lower affinity site; (iii) three sites of differing affinity. The most important feature to the resolution of cooperativity seems to be the relationship $k_{ij}/(k_i/k_j)$ where i and j represent the two highest affinity sites. However, the data also indicate that it requires slightly greater cooperativity to stand above the uncertainty in the estimates of the macroscopic equilibrium constants when all three sites have different affinities than when two sites are of the same affinity (compare lines 5 and 9 in Table V).

If it is known independently that two of the sites have equal affinities then it is feasible to formulate Equations 4a–4d with parameters k_i, k_j, and k_c. k_i and k_j are the intrinsic binding constants. k_c is interpreted as an averaged pairwise cooperativity parameter. Unique estimates for the three parameters are obtainable by nonlinear least squares analysis. A similar approach is applicable when the three sites might all be different. In this case, one must formulate Equations 4a–4d with parameters k_1, k_2, k_3, and $k_c \cdot k_c$ is fixed at different values and the data fitted to determine the k_i values and the variance of the fit. A minimum in the variance indicates the most probable value of k_c, as demonstrated in Fig. 6.

The three-site model presented in Table IB provides for no additional cooperativity at the third ligand binding event. This is reflected in the comparison between $K_2K_1/3$ and K_3 in the first three cases shown in Table V. To assess the possibility of discerning additional cooperativity in the third ligand binding event, a model that incorporates additional cooperativity at the third ligand binding event was considered. In this model there is a single species with all three sites liganded, in which all pairwise cooperative interactions are made. The cooperativity constant for this species is the product, $k_{12}k_{13}k_{23}$. Again, only the simplest case in which $k_{12} = k_{13} = k_{23} = k_c$ was considered, so that $k_{12}k_{13}k_{23} = k_c^3$. The results of analysis of the simulated data (last case, Table V) indicate that it is possible to assess separately cooperative interactions in the second and third ligand events. However, this model for the interactions in the fully liganded system presents an extreme case. Results with other patterns of cooperative interaction

TABLE V

Analysis of simulated titration experiments for binding of protein to three DNA binding sites

Mobility-shift data were simulated as described under "Materials and Methods," using the microscopic binding constants listed and the interaction model shown in Table I. The product of the microscopic constants ($k_1 \cdot k_2 \cdot k_3 \cdot (3k_c)$) was held constant at 1.25×10^{29} M^{-3} (corresponding to a total free energy change to saturate the sites, $\Delta G = -39$ kcal/mol) while pairwise cooperativity ($k_c = k_{12} = k_{13} = k_{23}$) was introduced. The random error level was 0.05. The macroscopic product binding constants K_1, K_2, and K_3 were estimated by using nonlinear least squares analysis. Confidence limits for the derived constants $K_1^3/3$ and $K_1K_2/3$ were estimated by propagating the confidence limits for the fitted constants.

Microscopic binding constants used for simulations					Macroscopic binding constants obtained from curve fitting					
k_c	ΔG_c	$k_1(\times 10^{-9})$	$k_2(\times 10^{-9})$	$k_3(\times 10^{-9})$	$K_2(\times 10^{-20})$	$K_1^2/3(\times 10^{-20})$	$K_2?K_1^2/3$	$K_3(\times 10^{-30})$	$K_1K_2/3(\times 10^{-30})$	$K_3?K_1K_2/3$
$k_3 = k_2 = k_1$										
1	0	5.00	5.00	5.00	0.88 ± 0.27	0.95 ± 0.24	=	0.13 ± .04	0.49 ± 0.18	<
6	−1	1.91	1.91	1.91	0.65 ± 0.15	0.10 ± 0.04	>	0.13 ± .04	0.36 ± 0.12	<
31	−2	1.10	1.10	1.10	1.45 ± 0.43	0.05 ± 0.05	>	0.14 ± .05	0.19 ± 0.13	<
$k_3 = k_1$, $k_2/k_1 = 31$[a]										
1	0	1.59	49.3	1.59	1.41 ± 0.07	6.71 ± 1.54	<	0.10 ± .04	2.12 ± 0.72	<
6	−1	0.607	18.8	0.607	1.33 ± 0.45	1.56 ± 0.42	=	0.11 ± .04	0.96 ± 0.47	<
31	−2	0.351	10.9	0.351	3.04 ± 0.83	0.57 ± 0.19	>	0.18 ± .06	1.33 ± 0.48	<
73	−2.5	0.264	8.19	0.264	3.55 ± 0.96	0.40 ± 0.18	>	0.15 ± .05	1.29 ± 0.56	<
$k_3/k_2 = k_2/k_1 = 31$[a]										
1	0	0.161	5.00	155.	9.14 ± 2.09	86.80 ± 16.2	<	0.14 ± .04	49.2 ± 12.3	<
6	−1	0.0615	1.91	59.1	6.37 ± 1.73	12.70 ± 2.90	<	0.12 ± .04	13.3 ± 4.2	<
31	−2	0.0356	1.10	34.2	11.40 ± 3.87	4.44 ± 1.61	>	0.10 ± .04	14.0 ± 6.1	<
73	−2.5	0.0268	0.830	25.7	16.40 ± 4.45	1.56 ± 0.80	>	0.12 ± .04	12.0 ± 5.5	<
$k_3/k_2 = k_2/k_1 = 31$[b]; $k_{ijk} = k_c^3$										
1	0	0.161	5.00	155.	8.11 ± 2.20	99.60 ± 22.8	<	0.12 ± .04	46.7 ± 14.8	<
6	−1	0.0269	0.833	25.8	1.38 ± 0.44	2.44 ± 0.61	<	0.12 ± .04	1.24 ± 0.45	<
31	−2	0.00520	0.161	5.00	0.22 ± 0.09	0.080 ± 0.037	>	0.10 ± .04	0.036 ± 0.019	>
73	−2.5	0.00221	0.0685	2.12	0.081 ± 0.038	0.020 ± 0.014	>	0.12 ± .04	0.0065 ± 0.033	>

[a] In these simulations, site heterogeneity (defined by k_3/k_2 and k_2/k_1) was introduced for one or all sites and was also held constant, equal to 31 ($\Delta G = -2.0$ kcal/mol).

[b] In this simulation, cooperativity was introduced to the third ligand binding step by employing a model with only one triply liganded state, in which all pairwise cooperative interactions are made, *i.e.* ($k_1 \cdot k_2 \cdot k_3 \cdot k_c^3$) was held constant at 1.25×10^{29} M^{-3}.

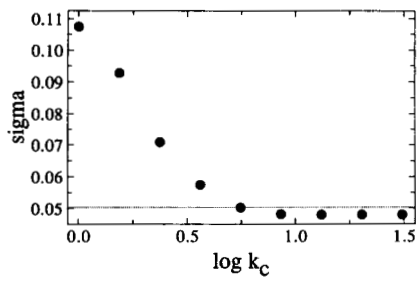

FIG. 6. **Square root of the variance (σ) *versus* averaged pairwise cooperativity constant, k_c.** Mobility-shift data for the binding of the λcI repressor to O_R under the conditions used for the experiment in Fig. 1 were simulated, including a separate band for Θ_1. Values of the microscopic binding and cooperativity constants, $k_1 = 1.41 \times 10^{10}$ M^{-1}, $k_2 = 1.15 \times 10^8$ M^{-1}, $k_3 = 1.73 \times 10^7$ M^{-1}, $k_{12} = 244$, $k_{13} = 1$, $k_{23} = 813$, and $k_{123} = 967$ are from Senear and Ackers (1990). The data were analyzed using the statistical mechanical model for the interactions represented by species 1–9 in the bottom half of Table I, subject to the constraint, $k_{12} = k_{23} = k_c$, and using different fixed values of k_c. The *horizontal dashed line* defines the variance at the 65% confidence interval (see "Materials and Methods").

(not shown) indicate that resolution of cooperativity in the third ligand binding event is much more difficult than at the second ligand binding event. Even moderate site heterogeneity will frequently mask any additional cooperativity at the third ligand binding event.

Binding of the λcI Repressor to O_R—This system is considered as an example of a three-site, cooperative system. Senear and Ackers (1990) showed that cooperativity with all three sites liganded is given by k_{123} (equal to $e^{-\Delta G_{123}/RT}$, see "Materials and Methods") which is unrelated to the pairwise terms k_{12}, k_{13}, and k_{23}. They also obtained values $k_1 = 1.41 \times 10^{10}$ M^{-1}, $k_2 = 1.15 \times 10^8$ M^{-1}, $k_3 = 1.73 \times 10^7$ M^{-1}, $k_{12} = 244$, $k_{13} = 1$, $k_{23} = 813$, and $k_{123} = 967$ for the experimental conditions in Fig. 1. This represents a fairly extreme case of site heterogeneity with relative intrinsic affinities of operator sites, O_R1, O_R2 and O_R3, given by $k_1/k_3 = 813$ and $k_2/k_3 = 66$. O_R also exhibits strong pairwise cooperativity between adjacent sites, and cooperativity in the species with all three sites liganded exceeds the pairwise terms. Values of the macroscopic constants, $K_1 = 1.43 \times 10^{10}$ M^{-1}, $K_2 = 3.94 \times 10^{20}$ M^{-2}, and $K_3 = 2.70 \times 10^{28}$ M^{-3} are calculated using the relationships for K_1 and K_2 in Equations 4a and 4b, and $K_3 = k_1k_2k_3k_{123}$. Based on these values, $K_2/(K_1^2/3) = 5.8$, so that pairwise cooperativity at the second ligation step should be discernible, given high precision estimates of the macroscopic constants. However, $K_3/(K_1K_2/3) = 0.014$, so that cooperativity at the third ligation step should not be discernible.

The results of the analysis of the mobility-shift data shown in Fig. 1 are given in the figure legend. K_2 (1.36 (+3.75, −1.00) $\times 10^{20}$ M^{-2}) and $K_1^2/3$ (1.67 (+1.61, −1.09) $\times 10^{20}$ M^{-2}) are the same to within the limits of experimental uncertainty. When analyzed in this manner these data are not sensitive to the pairwise cooperativity in the second binding step. This appears to result primarily from the fact that a separate band corresponding to Θ_1 is not resolved by the mobility-shift assay in this case. The loss of this information causes resolution of the macroscopic binding constants to depend primarily on the shapes of the curves. In particular, this leads to high correlation between K_2 and K_3 (correlation coefficient = 0.986) and very low precision in their estimation. The latter precludes a very meaningful comparison to $K_1^2/3$. Analysis of these data using the microscopic binding constants k_1, k_2, k_3, and k_c ($k_c = k_{12} = k_{23} = k_{123}$, $k_{13} = 1$) as described above (Fig. 6) did yield

a minimum in the variance for $k_c > 1$; however, the improvement in the variance over that for $k_c = 1$ was not statistically meaningful.

The question of the resolution of cooperativity for this system from an experiment with separate bands for Θ_1 and Θ_2 was addressed by analysis of simulated experimental data. The simulated data were based on the interaction model and on values of the macroscopic constants, K_1, K_2, and K_3, calculated from the data of Senear and Ackers (1990). From the analysis of eight different simulated experiments with random noise levels of 0.05 and 0.08, the average of the ratios, $K_2/(K_1^2/3)$ was found to be 3.6 (±1.0). Analysis of these data using the microscopic binding constants k_1, k_2, k_3, and k_c was particularly valuable. This yielded a minimum estimate of the average pairwise cooperativity, $k_c \geq 6$, exactly as expected. This reinforces the conclusion, drawn from the analysis of the Lac repressor experiment, that resolution of cooperativity is critically dependent on separate resolution of all of the intermediate ligation states.

Hill Analysis of Mobility-shift Titrations—Recently, the Hill binding equation has been used to analyze mobility-shift titrations, both to assess cooperativity (Ingraham *et al.*, 1990) and to estimate binding stoichiometry (Hudson *et al.*, 1990). The Hill equation describes the concerted binding of n_H ligands to a macromolecule,

$$\bar{Y} = \frac{(L \cdot k)^{n_H}}{1 + (L \cdot k)^{n_H}} \quad (5)$$

According to the chemical model, the Hill coefficient (n_H) is the stoichiometry of ligand binding. In practice, nonintegral Hill coefficients less than the ligand stoichiometry are observed routinely. These are usually interpreted to indicate the degree of cooperativity; an n_H equal to 1 indicates noncooperative ligand binding whereas an n_H equal to the stoichiometry indicates infinite cooperativity. Notice that an n_H greater than the stoichiometry has no meaningful physical or chemical interpretation.

Equation 5 is inappropriate for the analysis of mobility-shift titrations for three reasons. First, the mobility-shift assay measures Θ_i, the fraction of macromolecules with i ligands bound, not $\bar{Y}$, the fractional saturation. Consequently, Hill analysis of Θ_0 for DNA containing two or more binding sites does not yield $n_H = 1.0$ when ligand binding is noncooperative. This point is clearly illustrated by the analysis presented in Table VI of the Θ_0 curves shown in Fig. 3. Analysis of the noncooperative binding curve (Fig. 3, *bold lines*) by Equation 5 yields $n_H = 1.2$ for the two-site operator. This reflects the statistics of the binding of a single ligand to either of two sites. These statistics are also reflected in k_{app} (10.0×10^9), which is twice the value assumed to simulate the curves. Second, n_H reflects the steepness of the binding curves (Fig. 3*A*) which is less sensitive to cooperativity than is Θ^{max} for the intermediate states. For example, analysis of the curves simulated with $k_{12} = 31$ (Fig. 3, *dashed lines*) yielded a 33% increase in n_H (from 1.2 to 1.6) as compared with the noncooperative case but a 40% decrease in Θ^{max} (from 0.5 to 0.3). This apparently modest difference is compounded greatly by the inherent difficulty in precisely determining the curve shape from real data. Third, Equation 5 defines isotherms that are symmetric with respect to their midpoints, which is not a necessary feature of mobility-shift titration data.

TABLE VI

Hill coefficient and apparent binding affinity as a function of the cooperativity for DNA containing two identical binding sites

Mobility-shift curves shown in Fig. 3 for the binding of a ligand to two identical sites ($k_1 = k_2$) with increasing amounts of cooperativity, were fit to Equation 5.

Simulated cooperativity		Fitted parameters		
			Binding affinity	
k_{12}	ΔG_{12}	n_H	$k_{app} \times 10^9$	ΔG_{app}
			M^{-1}	*kcal/mol*
1	0	1.2	10.0	−13.4
6	−1	1.4	8.4	−13.3
31	−2	1.6	7.1	−13.2
173	−3	1.8	6.0	−13.1
966	−4	1.9	5.0	−13.0
5,383	−5	2.0	5.0	−13.0

These conclusions are further borne out by analysis of Θ_0 from the Gal and Lac titrations (Figs. 4 and 5) using Equation 5. For the noncooperative and cooperative Gal titrations (Fig. 4, *A* and *B*) this analysis yielded $n_H = 1.5 \pm 0.3$ ($k_{app} = 4.51$ (±0.61) $\times 10^9$ M^{-1}, $\sigma = 0.033$) and $n_H = 2.0 \pm 0.3$ ($k_{app} = 6.41$ (±0.46) $\times 10^7$ M^{-1}, $\sigma = 0.035$). As expected, n_H exceeds 1.0 in the absence of cooperativity, and the confidence limits indicate poor resolution of the cooperative effect (compare with Table VI). The Hill coefficient for the data in Fig. 4*B* does not encompass the true value, which should correspond to $k_{12} = 3.6$. Even at its lower confidence limit of 1.7, the value obtained for n_H implies $k_{12} \geq 75$. Hill analysis of the Lac repressor titration (Fig. 5*A*) yielded $n_H = 2.5 \pm 0.3$ ($k_{app} = 4.10 \pm 0.24 \times 10^{11}$ M^{-1}, $\sigma = 0.022$), a physically meaningless value. A similar meaningless value of n_H from a Hill analysis of a very precise mobility-shift titration was reported recently (Ingraham *et al.*, 1990). Of course, these effects preclude meaningful use of n_H to estimate binding stoichiometry.

CONCLUSIONS

One strategy that can provide *unique* estimates of all of the intrinsic and cooperative microscopic binding constants in a complex system involves the comparative analysis of protein binding to wild-type DNA and to reduced-valence mutant DNAs in which one or more protein binding sites are inactivated. The application of individual site binding curves obtained by quantitative footprint assays to such an analysis has been extensively described (Johnson *et al.*, 1979; Ackers *et al.*, 1982; Brenowitz *et al.*, 1986, 1990; Senear *et al.*, 1986; Senear and Ackers, 1990). The use of mobility-shift titrations provides a plausible alternative to the use of technically more difficult footprint assays in such a strategy. Mobility-shift titrations of binding to wild-type DNA alone, as described herein, offer an attractive first step in the analysis of potentially cooperative systems. Demonstration of cooperativity and/or site heterogeneity provides strong direction to the formulation of models for the interactions and mechanism of biological regulation. Mobility-shift data can provide a basis for such formulations. However, it is clear that the resolution of cooperative effects is robust only for thermodynamically identical binding sites. Also, there is high potential to be misled if the data analysis is not done correctly.

In a few instances slight differences between the abilities of the actual and simulated experiments to discern cooperativity were noted. This leads us to the caution that the simulated data contain only randomly distributed noise. Actual experiments always suffer from sources of systematic error for which the mathematical formulation of the interaction model does not account. However, in most of the cases we have presented, there is general agreement between the simulated and experimental data in their ability to resolve cooperative effects. This indicates that the resolution of cooperativity is governed by the random noise in these experi-

ments, not by experimental artifacts, and that the results of the simulations are applicable to real experiments.

The quantitative analysis of gel mobility-shift titration data which we have followed is one of several theoretically correct formalisms that have been described recently (Hudson and Fried, 1990; Tsai *et al.*, 1989). In these other approaches cooperativity is evaluated separately for individual lanes of the gel by taking ratios of the DNA present in the bands representing the intermediate ligation states to the unliganded DNA. The approach presented in this paper provides several significant advantages over these other methods. First, the need to make assumptions concerning the proper macromolecule concentrations at which to measure ratios of bands (Hudson and Fried, 1990) is obviated. Second, least squares minimization of the binding expressions to all of the titration data provides sensitivity to systematic deviation. Deviation caused by either an incorrect model or systematic experimental error is often detectable in residual plots. This is particularly valuable in cases of substantially asymmetric binding curves, as in Fig. 5. Third, correct calculation of joint confidence limits for the fitted parameters provides realistic estimates of the precision of the estimates (Johnson and Frasier, 1985). It is usually simple to formulate the binding equations with the cooperativity as one parameter (see Table II for which $K_1^2/4$ was fitted directly). Finally, the analysis provides for simultaneous analysis of binding data for wild-type and reduced valence mutant DNA templates (Senear and Bolen, 1991) as necessary to the exact resolution of cooperativity whenever all binding sites are not identical.

Acknowledgments—We thank Sankar Adhya for the generous gifts of the Gal and Lac repressors and for the strains containing plasmids used in this study. The Gal and Lac mobility-shift titration experiments were conducted by Elizabeth Jamison and Amy Pickar.

REFERENCES

Ackers, G. K., Johnson, A. D., and Shea, M. A. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 1129–1133

Ackers, G. K., Shea, M. A., and Smith, F. R. (1983) *J. Mol. Biol.* **170,** 223–242

Adhya, S. (1989) *Annu. Rev. Genet.* **23,** 207–250

Bossinger, J., Miller, M. J., Vo, K. P., Geiduschek, E. P., and Xuong, N. (1979) *J. Biol. Chem.* **254,** 7986–7998

Box, G. E. P. (1960) *Ann. N. Y. Acad. Sci.* **86,** 792

Brenowitz, M., and Senear, D. F. (1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) Vol. 2, Supplement 7, Unit 12.4, Wiley-Interscience Associates, New York

Brenowitz, M., Senear, D. F., Shea, M. A., and Ackers, G. K. (1986) *Methods Enzymol.* **130,** 132–181

Brenowitz, M., Jamison, E., Majumdar, A., and Adhya, S. (1990) *Biochemistry* **29,** 3374–3383

Brenowitz, M., Mandal, N., Pickar, A., Jamison, E., and Adhya, S. (1991a) *J. Biol. Chem.* **266,** 1281–1288

Brenowitz, M., Pickar, A., and Jamison, E. (1991b) *Biochemistry* **30,** in press

Cann, J. R. (1989) *J. Biol. Chem.* **264,** 17032–17040

Cann, J. R. (1991) *J. Mol. Biol.* **216,** 1067–1075

Carey, J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 975–979

Codish, L. A. (1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) Vol. 2, Supplement 7, Unit 12.3, Wiley-Interscience Associates, New York

Dandanell, G., Valentine-Hansen, P., Larsen, J. E. L., and Hammer, K. (1987) *Nature* **325,** 823–827

Fried, M. (1989) *Electrophoresis* **10,** 366–376

Fried, M., and Crothers, D. M. (1981) *Nucleic Acids Res.* **9,** 6505–6525

Garner, M. M., and Revzin, A. (1981) *Nucleic Acids Res.* **9,** 3047–3060

Gralla, J. D. (1989) *Cell* **57,** 193–195

Haber, R., and Adhya, S. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 9683–9687

Hildebrand, F. B. (1956) *Introduction to Numerical Analyis,* McGraw-Hill, New York

Hill, T. L. (1960) *Introduction to Statistical Mechanics,* Addison Wesley, Reading, MA

Hudson, J. M., and Fried, M. G. (1990) *J. Mol. Biol.* **214,** 381–396

Hudson, J. M., Crowe, L. G., and Fried, M. G. (1990) *J. Biol. Chem.* **265,** 3219–3225

Ingraham, H. A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M. S., Wilson, L., and Rosenfeld, M. G. (1990) *Cell* **61,** 1021–1033

Jacob, F., and Monod, J. (1961) *J. Mol. Biol.* **3,** 318–356

Johnson, A. D. (1980) Ph.D. dissertation, Harvard University, Cambridge, MA

Johnson, A. D., Meyer, B. J., and Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 5061–5065

Johnson, M. L., and Frasier, S. G. (1985) *Methods Enzymol.* **117,** 301–342

Letovsky, J., and Dynan, W. S. (1989) *Nucleic Acids Res.* **17,** 2639–2653

Majumdar, A., Rudikoff, S., and Adhya, S. (1987) *J. Biol. Chem.* **262,** 2326–2331

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Reiner, P., and Brenowitz, M. (1991) *CABIOS,* in press

Revzin, A. (1989) *Biotechniques* **7,** 346–355

Sauer, R. T. (1979) Ph.D. dissertation, Harvard University, Cambridge, MA

Senear, D. F., and Ackers, G. K. (1990) *Biochemistry* **29,** 6568–6577

Senear, D. F., and Batey, R. (1991) *Biochemistry* **30,** in press

Senear, D. F., and Bolen, D. W. (1991) *Methods Enzymol.,* in press

Senear, D. F., Benowitz, M., Shea, M. A., and Ackers, G. K. (1986) *Biochemistry* **25,** 7344–7354

Sorger, P. K., and Pelham, H. R. B. (1988) *Cell* **54,** 855–864

Tjian, R., and Mitchell, P. J. (1989) *Science* **245,** 371–378

Tsai, S. Y., Tsai, M.-J., and O'Malley, B. W. (1989) *Cell* **57,** 443–448

Wyman, J. (1964) *Adv. Protein Chem.* **19,** 223–286