Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 5699–5703, July 1991
Biochemistry

Specificity of the Mnt protein determined by binding to randomized operators

(binding affinities/partition function/sequence recognition/information content)

GARY D. STORMO AND MAKOTO YOSHIOKA

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309-0347

Communicated by Peter H. von Hippel, March 22, 1991

ABSTRACT **The relative binding affinities of Mnt protein from bacteriophage P22 are determined for each possible base pair at position 17 of the operator. These are determined from the partitioning of randomized operators into bound and unbound fractions; quantitation is provided by restriction enzyme analysis. Mnt protein is found to have an unusual specificity at this position: a C·G base pair (the wild-type operator) has the highest affinity, a G·C base pair has the lowest affinity, and both orientations of A·T base pairs are intermediate and nearly equivalent. A specific binding constant and specific binding free energy are defined and shown to be directly related to the information content of the operator sequences bound to the protein, taking into account the quantitative differences in binding affinities.**

The Mnt repressor from bacteriophage P22 binds to an operator site on the phage DNA where it blocks transcription from the P_{ant} promoter (1). Sauer and colleagues (2–4) have extensively studied the binding of Mnt protein to its operator DNA. In addition, a large number of Mnt mutants have been isolated that affect the binding specificity (5–8). The specificity of the protein is determined by its amino-terminal domain. It does not contain a helix–turn–helix or zinc-finger DNA binding motif, although sequence homology exists with some other proteins (4, 9). We have devised a random mutagenesis approach combined with a quantitative binding assay that allows us to determine the relative binding affinity of the protein for any base at a particular position in the operator. This also allows us to calculate the information content (10–12) for that position of the operator directly from the thermodynamics of binding.

Information content was originally derived from principles of information theory as a quantitative description of the sequence bias that exists in the DNA binding sites of proteins (10). It can also be derived from principles of likelihood statistics and from statistical mechanics (11, 12). In fact, Eq. 7 of this paper is analogous to equation **2.5** from chapter 1 of Kullback (13), termed the "mean information of discrimination" between two hypotheses, given a data set. Eq. 7 is also analogous to equation **2.3.12** of Hobson (14), where it is shown to be the unique expression that has a set of intuitively reasonable properties for a measure of the information that distinguishes two probability distributions. We first used information content to examine collections of binding sites that were genetically or biochemically defined and known to be important for the proper regulation of gene expression, but for which distinctions based on binding affinities for different sites, or between sites and nonsites, was not known (10). Berg and von Hippel (15) showed how this could be related to differences in binding energies for different binding site sequences, based on some assumptions about the selectivity of the protein and the representativeness of the sample of sites. In this work those assumptions are unnecessary because we obtain complete information about the two site distributions, those bound by the protein and those left unbound. Our analysis is completely analogous to the statistical mechanical derivation of information by Hobson (14), with which we can assign an information content as a property of the DNA binding protein because it provides the force that partitions the collection of binding sites into the two distinguishable distributions. Information content is also shown to be the mean value of a "specific free energy of binding" as defined in the text and shown to have additional reasonable properties.

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.

MATERIALS AND METHODS

Protein and DNA. Wild-type Mnt repressor was kindly provided by Knight and Sauer (8). DNA was synthesized on an Applied Biosystems model 380A DNA synthesizer. The sequences of the template strand, containing the operator sequence with the randomized position, and of the primer strand are shown in Fig. 1. The randomized position was synthesized with the "N" mixture supplied by Applied Biosystems. Detritylated, deprotected DNA was gel-purified. The primer was end-labeled with ^{32}P by using T4 polynucleotide kinase (Promega). Labeled, double-stranded DNA was synthesized by annealing the ^{32}P-labeled primer to the template strand and extending with Sequenase (United States Biochemical). The double-stranded DNA was eluted from a 5% NuSieve (FMC) agarose gel and purified through steps of phenol, phenol/chloroform/isoamyl alcohol, and chloroform/isoamyl alcohol extractions followed by ethanol precipitation.

Binding Reaction and Quantitation. The binding reaction conditions were similar to those of Vershon *et al.* (3), with the addition of a nonionic detergent. The binding buffer was 10 mM Tris·HCl at pH 7.5, 10 mM $MgCl_2$, 0.1 mM EDTA, 200 mM KCl, bovine serum albumin at 100 μg/ml, and 0.1% Nonidet P-40. The detergent was added because it was found to eliminate additional low-mobility bands that otherwise arise in reactions with high concentrations of protein, but it did not affect the binding reaction as monitored by the highest mobility bound band. The binding reactions were done at either room temperature for 30 min or at 37°C for 1 hr, without significant differences in the results. Each reaction contained $\approx 10^{-7}$ M DNA and $\approx 5 \times 10^{-8}$ M Mnt protein. The binding reactions were run on a 4% polyacrylamide gel to separate the bound and unbound fractions (3). In different experiments, the fraction of DNA bound to the protein varied from ≈50% to 80% (data not shown). The bands for each fraction were cut out of the gel and eluted overnight at 37°C.

As shown in Fig. 1, depending on which base is present at the variable position, each operator DNA can be cut with one of four different restriction enzymes. For each enzyme, there is also a control site on the DNA to monitor the efficiency of cutting. The enzymes used were *Spe* I (Boehringer Mannheim or Stratagene), *Sau*96I (Boehringer Mannheim or Strat-

```
gtcgacactagtgagctcagatct aggtccacggtggaNct agtactcctatagtgagtcgtattaattt -3'
                                                gaggatatcactcagcataattaaa -5'

actagt                                act agt      Spe I

                          ggtcc     ggacc          Sau96 I

agct                                  agct         Alu I

      gatc                          gatc           Mbo I
```

FIG. 1. The sequences of the synthetic oligonucleotides are shown. The top strand is the template oligonucleotide with the Mnt operator separated from the rest of the oligonucleotide by spaces. This 17-mer operator sequence corresponds to positions 3–19 of the 21-mer operator defined by Sauer and colleagues (4); the variable position is number 17. The variable (N) position contains all four bases in different individual oligonucleotides. The lower strand is the primer used to create the double-stranded DNA for the binding reactions. The locations of the restriction sites are also shown. For each different base that can occur at the variable position, there is a restriction enzyme that will cut the DNA at that site if and only if the correct base is present. For each such site there is also a control site that should be cut on every DNA.

agene), *Alu* I (Promega), and *Mbo* I (Stratagene). Digestion was performed as prescribed by the enzyme suppliers. Each of the two eluted bands, corresponding to the bound and unbound fractions of the labeled DNA, were divided into five aliquots, to be cut with the four restriction enzymes or left uncut. Quantitation of the labeled restriction fragments was performed on an AMBIS Systems (San Diego) radioanalytic imaging system, using either a 0.8 × 3.2 mm or a 0.8 × 1.6 mm resolution plate and analyzed by using AMBIS software version 2.0.

RESULTS

Relative Binding Affinities. The binding reaction is a coupled equilibrium with the protein and the four different operator DNAs. The binding reaction can be diagrammed as

$$P + D_b \rightleftharpoons P{\cdot}D_b,$$

where P is the Mnt protein and D_b represents the mixture of DNAs, with the subscript b having four values ($b \in \{A, C, G, T\}$) for the four different operator sequences. The association constant for each different DNA can be measured:

$$K_a(b) = \frac{[P{\cdot}D_b]}{[P][D_b]}. \qquad [1]$$

In our experiment the free protein concentration $[P]$ is not known and cannot be accurately estimated because the binding reactions are done with excess DNA. However, the free protein concentration need not be known to determine the relative association constants for each different operator sequence. For example, the ratio of the association constants for the A- and C-containing operators can be determined by the relative partitioning of the A- and C-containing operators into the bound and unbound fractions:

$$\frac{K_a(A)}{K_a(C)} = \frac{[P{\cdot}D_A]}{[D_A]}\,\frac{[D_C]}{[P{\cdot}D_C]}. \qquad [2]$$

The values needed to determine all of the relative association constants can be obtained by quantitating the restriction fragments from the bound and unbound fractions.

Fig. 2 shows the results from one experiment. After the bound and unbound DNAs are separated in the "band-shift" gel (data not shown), the DNA from each fraction is subjected to quantitation with restriction enzymes. Fig. 2*a* shows the results from one experiment of the restriction analysis for the bound DNA fraction, and Fig. 2*b* shows the results from the same experiment for the unbound DNA fraction. In the lane for each restriction enzyme there are three bands. The lower band corresponds to the variable position in the operator; its quantitation is a measure of the partitioning of each different operator into each fraction. The middle band corresponds to the control restriction site, and the upper band corresponds to uncut DNA. The efficiency of cutting is the ratio of the control band to the sum of the control and uncut bands. The proportion of counts in each variable band divided by the cutting efficiency determines the proportion of each base in the bound and unbound fractions, the $[P{\cdot}D_b]$ and $[D_b]$ numbers needed for Eq. 2. Five separate binding reactions were done, and the values of $[P{\cdot}D_b]/[D_b]$ were determined for each. The means and standard deviations for the association constants, relative to the C-containing (wild-type) operator, are given in Table 1. From the relative association constants we can determine the change in free energy of binding for each change from the wild-type base, $\Delta\Delta G$ (16, 17). Table 1 also shows those values. As expected, the C-containing operator has the highest affinity for the protein. The G-containing operator has the lowest affinity, ≈8-fold lower affinity. A- and T-containing operators are each 3- to 4-fold lower in affinity.

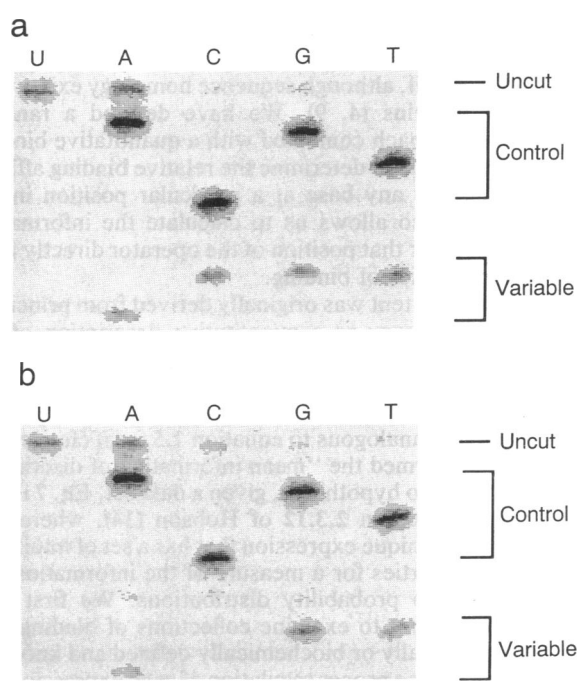

FIG. 2. Example of the quantitation of the bound (*a*) and unbound (*b*) fractions by use of the restriction enzymes. The lanes are labeled according to which base is present in order for it to be cut by the restriction enzyme: U, uncut; A, *Spe* I cut; C, *Sau*96I cut; G, *Alu* I cut; T, *Mbo* I cut. The highest band in each lane is that left uncut. The middle band is the control band that is present on each oligonucleotide. The lower band is the variable band and corresponds to the amount of each different operator sequence in each of the bound and unbound fractions.

Table 1. Relative and specific binding constants and free energies

Parameter	Base (*b*)			
	A	C	G	T
$K_a(b)/K_a(C)$	0.31 ± 0.06	1	0.12 ± 0.04	0.28 ± 0.11
$\Delta\Delta G(b)$*	0.69	0	1.25	0.75
$K_s(b)$	0.71 ± 0.08	2.36 ± 0.26	0.28 ± 0.05	0.64 ± 0.18
$\Delta G_s(b)$*	0.20	−0.51	0.75	0.26

*ΔG values are in kcal/mol, using $RT = 0.59$.

Specific Binding Constant. The data presented allow us to calculate the relative binding constants only, not the absolute binding constants. However, the absolute association constant to the C-containing operator has been measured as 4.5×10^{10} M^{-1} [under our binding conditions (3)], allowing us to also know the absolute binding constants to all of the mutants. We are primarily interested in understanding the mechanism of specific binding, that portion of the total binding that changes with changes in the operator sequence.

We define a "specific binding constant" for each base as

$$K_s(b) = \frac{K_a(b)}{K_n}, \quad [3]$$

where

$$K_n = \frac{\Sigma_b[P{\cdot}D_b]}{[P]\Sigma_b[D_b]} \quad [4]$$

is the overall binding constant to the complete mixture of DNA. This is to say that we divide the association constant for the operator *b* into the product of a specific component that depends on *b* and a nonspecific component that does not. This is completely analogous to typical methods of determining nonspecific binding constants by measuring the affinity to some heterologous bulk DNA, such as salmon sperm DNA. The main difference is that we have complete information about the proportions of each potential binding site in the DNA mixture and can easily identify sequences that bind both better and worse than the average. Note that, in general, K_n (and therefore also K_s) is not a constant, but depends on the mixture of the different operators in the reaction. For example, given the data in Table 1 it is clear that if the mixture of binding sites is mostly C-containing the total binding will be higher than if the mixture is mostly G-containing. However, we can define a "standard condition" under which K_n and K_s will be true constants. An appropriate standard condition is for the unbound operators to be present in equal concentrations. This is a simple constraint to apply experimentally: have the total DNA mixture contain each operator in equal concentrations and do the binding with a vast excess of DNA. However, it is not necessary to perform the reactions under the standard conditions, just as those reactions described above were not, because the relative binding affinities are sufficient to calculate the results that would be obtained under any other conditions. Therefore the reactions can be performed under whatever conditions are convenient, and then specific binding constants can be calculated for the standard condition, as described below. (See ref. 18 for another derivation.)

$K_s(b)$ is a constant that represents the partitioning of the *b*-containing operator between the bound and unbound fractions. From Eqs. **1**, **3**, and **4**, for an experiment done under standard conditions,

$$K_s(b) = \frac{[P{\cdot}D_b]}{\Sigma_b[P{\cdot}D_b]} \frac{\Sigma_b[D_b]}{[D_b]} = 4f_b, \quad [5]$$

where $f_b = [P{\cdot}D_b]/\Sigma_b[P{\cdot}D_b]$ is the proportion of the *b*-containing operator in the bound fraction, and $[D_b]/\Sigma_b[D_b] = 1/4$ is the definition of the standard condition. Clearly, $\Sigma_b K_s(b) = 4$. Since the ratios of the binding constants do not depend on the mixture of operator sequences, relative binding constants from any experiment may be converted into specific binding constants. For example, the sum of the relative binding constants shown in the first row of Table 1 is 1.71, so the $K_s(b)$ values can be calculated by multiplying each of those relative values by 4/1.71. [The actual values for $K_s(b)$ in Table 1 differ from this slightly because they were calculated for each of the five binding reactions separately and then averaged. This has the advantage of providing standard deviations for each of the different operators.] Note that $K_s(b) = 1$ when the base *b* is 1/4th of both the bound and unbound fractions. For any DNA binding protein, either $K_s(b) = 1$ for all *b*, or it will be >1 for some *b* and <1 for other *b*.

We can also derive a "specific binding free energy" as $\Delta G_s(b) = -RT\ln K_s(b)$, and this is also given in Table 1. Note that by this definition, $\Delta G_s(b) = 0$ when $K_a(b) = K_n$. Negative values indicate bases that are preferred by the protein, and positive values indicate those bases that are discriminated against by the protein. These seem to us to be reasonable properties for specific binding constants and specific binding energies to have. Furthermore, the definitions presented in this paper show directly the relationship between "information" involved in specificity and the energetics of partitioning as derived from statistical mechanics (ref. 14, especially section 2.3).

Information Content. Information content is a measure of the amount of specificity in a protein's binding sites, independent of the mechanism by which the specificity is obtained (10–12). Given a collection of binding sites, the information content based on those sites is

$$I_{Seq} = \sum_{i=1}^{L} \sum_{b=A}^{T} f_{i,b} \log_2 \frac{f_{i,b}}{p_b}, \quad [6]$$

where *i* corresponds to the positions in the binding sites, $f_{i,b}$ is the proportion of each base at each position in the binding sites, and p_b is the proportion of each base in the genome. If a particular base is absolutely required at some position for binding, then that position contains 2 bits of information (or possibly more if $p_b < 0.25$), whereas a position that does not influence the binding, for which $f_{i,b} = p_b$, has 0 bits. All numbers between those two extremes are allowed. Information content has usually been determined from known examples of the binding sites for some protein (10, 18). Berg and von Hippel (15, 19, 20) have shown how the information content can be related to the binding energy of the protein to different sequences, given some assumptions about the randomness of the sample and the selection of the sites.

Given the data presented in this paper, we can determine the information content for one position of the binding site based directly on the thermodynamics of the binding reaction. We have challenged the protein with operators containing all possible bases at one position and determined the partitioning of each different operator sequence into the bound and unbound fractions. The bound fraction constitutes

the collection of sites; each different site is represented in proportion to its binding affinity. The sample size is large, approximately a picomole of each different operator in these experiments, and we repeated the experiment several times to determine the reliability of the measurements. The information content for the variable position, based on the specificity, is

$$I_{\text{spec}} = \sum_b f_b \log_2 \frac{f_b}{p_b}, \quad [7]$$

where f_b is as defined in Eq. 5 and $p_b = 0.25$ for the standard condition. In terms of the specific binding constants and energies,

$$I_{\text{spec}} = \frac{1}{4} \sum_b K_s(b) \log_2 K_s(b) = \frac{-1}{RT \ln 2} \sum_b f_b \Delta G_s(b). \quad [8]$$

The last equation shows that information content, derived from quantitation of an *in vitro* binding reaction, is directly proportional to the average specific binding energy, where the average is taken over all of the DNA–protein interactions. For the interaction of the Mnt protein with the operators studied in these experiments, the information content of the variable position is 0.41 bits.

DISCUSSION

All of the G·C base pairs in the Mnt operator are important for binding by the Mnt protein, as mutations at any of them create operator-constitutive mutants (M. Susskind, personal communication). Binding of the operator by Mnt strongly protects all of the guanine bases from methylation by dimethyl sulfate, and prior methylation interferes with binding (3). At the variable position in our experiments [position 17 in the numbering of Sauer and colleagues (4)], changing the C to either G or T was found to result in operator-constitutive mutants (M. Susskind, personal communication). If the *in vivo* binding is consistent with our *in vitro* measurements, then a reduction of only 8-fold is sufficient to create an operator-constitutive mutant. The C to T change in our binding studies results in only a 3- to 4-fold reduction, but it creates a site that will be methylated by the *dam*-encoded methylase *in vivo*, thereby creating a double mutant: the wild-type C·G base pair is changed to T·N^6-methyl-A and the adjacent A·T base pair becomes N^6-methyl-A·T. A symmetric double-mutant operator, which created two *dam* methylase sites, was used to select an Mnt protein with altered specificity, which had acquired a histidine to proline change at amino acid 6 (5). The high-affinity binding of the mutant protein required the methylation of the operator (6). Another Mnt mutant of amino acid 6, containing a histidine to alanine change, has also been isolated (4, 8).

The interaction of the λ repressor and Cro proteins with their operators has been extensively studied, both biochemically and genetically. All single base-pair mutations to the wild-type O_R1 operator have been synthesized, and the binding affinities for each protein have been determined (16, 17). The information content of 0.41 bits that we determine for Mnt protein at operator position 17 is somewhat lower than the average per position for either of the λ proteins, but each contains several positions with less. Recently, all double-symmetric base-pair changes to two versions of the Trp operator have been tested for their ability to be repressed by the Trp repressor *in vivo* (21). Surprisingly, none of the operator positions for either λ protein or for Trp repressor matches the qualitative specificity seen for the Mnt protein, where a C·G base pair has the highest affinity in one orientation and the lowest affinity in the other, whereas both orientations of an A·T base pair are intermediate and nearly equivalent. Perhaps this is a characteristic of recognition by a histidine, which none of the other proteins contain in their DNA recognition helices.

Our information content analysis of the binding results gives an overview of the Mnt protein's specificity for operator position 17. This is analogous to, but not identical with, information content analyses based on collections of known binding sites. In particular, our measurements depend only on the interaction of the protein and the DNA and are independent of other constraints on the sequences that might contribute to their nonrandomness (15, 19). Some of the extraneous constraints may be removed with *in vivo* assays that focus on particular functions (18), but other constraints may be unavoidable. For example, *in vivo* there may be threshold effects that limit the range of binding variation that can be detected. The C to G change at the variable position is reduced only ≈8-fold in our *in vitro* assays but is an operator-constitutive mutant *in vivo*. This is probably because a small decrease in the occupancy of the operator is sufficient to allow enough antirepressor protein to be made that the infected cell cannot become lysogenic. A mutation that overproduced the Mnt protein should suppress such an operator mutant, but in the absence of such suppression the collection of all viable operator sequences would not include any with a G in that position (assuming there are not other mutations that could compensate for the decreased binding observed for the G). Such threshold effects will always lead to information content measurements based on example binding sites being larger than measurements based on binding affinities.

In our assays the distinction between a site and a nonsite is intentionally blurred. Rather, we assume that any sequence will be bound by the protein if it is present at a high enough concentration. We are more interested in the continuous distribution of affinities from the best binding sequence to the worst. The information content measurement presented in this paper is based on the relative partitioning of each potential binding site into bound and unbound fractions and is equivalent to an average specific binding free energy, as defined above. The analysis would be equally appropriate for a complete randomization of the entire operator, but it would be impossible to obtain the quantitative data for all 4^{21} different sequences. It is possible to obtain quantitative data for each position of the operator independently. The reliability of using such data to determine the information content of the entire operator depends on how independent the binding contributions are from the different positions (i.e., whether the $\Delta\Delta G$ values from the separate positions can be added to give accurate values for operators with multiple changes). For both the λ repressor and Cro proteins, measured binding affinities to several operators, each with multiple changes, are remarkably well correlated (>95%) with predictions based on the single-position $\Delta\Delta G$ values (16, 17). Whether this will be true for other proteins, or even for the λ proteins for sequences that are more different from the consensus binding site, remains to be determined.

The experimental method presented in this paper depends on the ability to distinguish different operator sequences by their being cut by different restriction enzymes. This does not allow the method to be applied at any position in a binding site, although a surprising number of positions can be analyzed in this manner, particularly if one is willing to analyze one position in the background of a single mutation at another position. However, there is a generalization of the method that will allow it to be applied to any position in a binding site, regardless of whether there exists a complement of appropriate restriction enzymes. This requires the ability to obtain "quantitative sequencing" information from the bound and unbound fractions. While conceptually simple, we have

found this generalization of the method to be technically difficult. Once perfected, the quantitative sequencing method would allow for the analysis presented in this paper to be applied to any position in a binding site and, in fact, to several positions simultaneously. This would greatly enhance our ability to rapidly obtain quantitative specificity measurements for any DNA binding protein. Other types of quantitative data may also be analyzed in the manner presented, such as the relationships between sequences and functional activities *in vivo* (22, 23).

We thank Robert Sauer and Kendall Knight for providing us with the Mnt protein and many useful discussions and Miriam Susskind for providing unpublished data. Tom Schneider was an important contributor to early discussions on the methodology, and Judy Ruckman and Julie Rhode provided technical assistance in developing the assays. This work was supported by Grant GM28755 from the National Institutes of Health.

1. Susskind, M. M. & Youderian, P. (1983) in *Lambda II*, eds. Hendrix, R., Roberts, J. & Weisberg, R. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 347–363.
2. Vershon, A. K., Youderian, P., Susskind, M. M. & Sauer, R. T. (1985) *J. Biol. Chem.* **260**, 12124–12129.
3. Vershon, A. K., Liao, S.-M., McClure, W. R. & Sauer, R. T. (1987) *J. Mol. Biol.* **195**, 311–322.
4. Knight, K. L., Bowie, J. U., Vershon, A. K., Kelley, R. D. & Sauer, R. T. (1989) *J. Biol. Chem.* **264**, 3639–3642.
5. Yoderian, P., Vershon, A., Bouvier, S., Sauer, R. T. & Susskind, M. M. (1983) *Cell* **35**, 777–783.
6. Vershon, A., Youderian, P., Weiss, M., Susskind, M. & Sauer, R. (1985) in *Sequence Specificity in Transcription and Translation*, eds. Calendar, R. & Gold, L. (Liss, New York), pp. 209–218.
7. Knight, K. L. & Sauer, R. T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 797–801.
8. Knight, K. L. & Sauer, R. T. (1989) *J. Biol. Chem.* **264**, 13706–13710.
9. Bowie, J. U. & Sauer, R. T. (1990) *J. Mol. Biol.* **211**, 5–6.
10. Schneider, T. D., Stormo, G. D., Gold, L. & Ehrenfeucht, A. (1986) *J. Mol. Biol.* **188**, 415–431.
11. Stormo, G. D. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 241–263.
12. Stormo, G. D. (1990) *Methods Enzymol.* **183**, 211–221.
13. Kullback, S. (1959) *Information Theory and Statistics* (Wiley, New York).
14. Hobson, A. (1971) *Concepts in Statistical Mechanics* (Gordon & Breach, New York).
15. Berg, O. G. & von Hippel, P. H. (1987) *J. Mol. Biol.* **193**, 723–750.
16. Takeda, Y., Sarai, A. & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 439–443.
17. Sarai, A. & Takeda, Y. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6513–6517.
18. Schneider, T. D. & Stormo, G. D. (1989) *Nucleic Acids Res.* **17**, 659–674.
19. Berg, O. G. & von Hippel, P. H. (1988) *J. Mol. Biol.* **200**, 709–723.
20. Berg, O. G. (1988) *Nucleic Acids Res.* **16**, 5089–5105.
21. Staacke, D., Walter, B., Kisters-Woike, B., v. Wilcken-Bergmann, B. & Muller-Hill, B. (1990) *EMBO J.* **9**, 1963–1967.
22. Stormo, G. D. (1991) *Methods Enzymol.*, in press.
23. Stormo, G. D., Schneider, T. D. & Gold, L. (1986) *Nucleic Acids Res.* **14**, 6661–6679.