

## Removing Water From an EcoRI-Noncognate DNA Complex With Osmotic Stress

<http://www.adeninepress.com>

**Nina Y. Sidorova\*,  
and Donald C. Rau**

Laboratory of Physical  
and Structural Biology,  
NICHD,

National Institutes of Health,  
Bethesda, MD 20892-0924

### **Abstract**

We recently showed that a nonspecific complex of the restriction nuclease EcoRI with poly (dI-dC) sequesters significantly more water at the protein-DNA interface than the complex with the specific recognition sequence. The nonspecific complex seems to retain almost a full hydration layer at the interface. We now find that at low osmotic pressures a complex of the restriction nuclease EcoRI with a DNA sequence that differs by only one base pair from the recognition site (a 'star' sequence) sequesters about 70 waters more than the specific one, a value virtually indistinguishable from nonspecific DNA. Unlike complexes with oligo (dI-dC) or with a sequence that differs by two base pairs from the recognition sequence, however, much of the water in the 'star' sequence complex is removed at high osmotic pressures. The energy of removing this water can be calculated simply from the osmotic pressure work done on the complex. The ability to measure not only the changes in water sequestered by DNA-protein complexes for different sequences, but also the work necessary to remove this water is a potentially powerful new tool for coupling inferred structural changes and thermodynamics.

### **Introduction**

It is becoming increasingly apparent that hydration properties and energetics play a key role in determining the specificity and strength of protein-DNA recognition, in particular, and in defining the structure and interactions of macromolecules, in general (1-5). Crystal structures of many specific DNA-protein complexes (reviewed in 6 and 7) show that the DNA-water and protein-water interactions in dilute solution are mainly replaced by direct DNA-protein contacts at the interface. Increasingly, however, crystal or NMR structures of DNA-protein complexes are showing water molecules structured at the interface mediating contacts between the two surfaces (8-11). Additionally, Gewirth and Sigler (12) found that the complex of a steroid receptor-like domain with a noncognate DNA sequence had several more waters incorporated at the interface than the specific complex, suggesting a relationship between the number of mediating waters and the binding strength. Thermodynamically, the role of hydration is inferred through the large heat capacity changes accompanying the formation of specific DNA-protein complexes that have been attributed to the release of structured water from the surfaces (13-16).

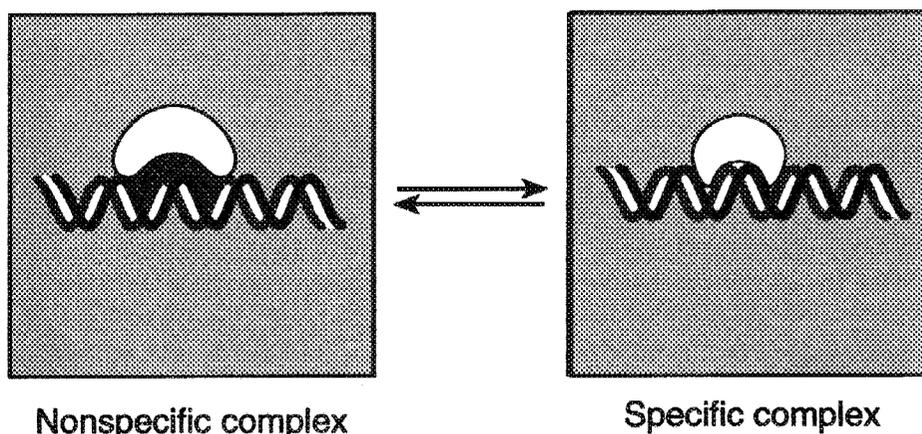
In spite of this generally accepted importance of water, there are few techniques for measuring changes in the numbers of 'bound' waters accompanying DNA-protein binding reactions. A difference in the numbers of water molecules released in the binding of a protein to two different DNA sequences can be measured through the dependence of the relative binding constant on water activity (or, equivalently, on osmotic pressure). Water activity can be varied by adding neutral solutes that do not themselves directly affect the DNA-protein binding. Defined in this thermodynamic sense, 'bound' water is the protein, DNA, or complex associated water that excludes solute (17). This inference of a change in bound water through the dependence of

\*Phone: 301-402-4698;  
Fax: 301-496-9462;  
E-mail: [sidorova@helix.nih.gov](mailto:sidorova@helix.nih.gov)

binding free energy on water chemical potential is strictly analogous to measuring ion release through the dependence of binding constants on salt activity, or protonation through a pH sensitivity, or even entropy through a temperature dependence. The osmotic stress technique (17) has been used to measure the changes in water binding accompanying the DNA binding of several proteins: *E. coli gal* repressor (18), *E. coli* CAP protein (19), Hin recombinase (20), Ultrathorax and Deformed homeodomains (21), *E. coli tyr* repressor (22), EcoRI (23,24), and Sso7d protein (25).

The binding of the restriction nuclease EcoRI to DNA is a paradigm for one extreme of specificity in recognition. As extensively studied by Jen-Jacobson and coworkers (26-28), even small perturbations in DNA structure and sequence can have large consequences for binding energy. The ratio of binding constants of the enzyme to the specific recognition sequence and to sequences that differ by two or more base pairs is  $\sim 10^4$ . Even changing a single base pair at the end of the recognition sequence (a 'star' sequence) decreases the binding constant by a factor of  $\sim 10^3$ . Most of the specific binding energy is lost with the change of one in six base pairs. In spite of the loss of binding energy 'star' sites can still be inefficiently cleaved by the enzyme. This cleavage at 'star' sites is intriguingly enhanced by adding osmolytes.

We recently showed that a striking difference in the binding of the restriction nuclease EcoRI to its specific recognition sequence and to the nonspecific DNA poly (dI-dC) is the amount of water sequestered at the protein-DNA interface (23). The complex of protein with poly(dI-dC) retains about 110 more waters at 25°C than the specific complex of the protein with a DNA fragment containing its recognition sequence GAATTC. Unlike the binding of free *gal* repressor to its operator sequences (18), the difference in excluded water between specific and nonspecific EcoRI complexes is not dependent on solute size or chemical nature. This independence implies that the water retained by the nonspecific complex is sequestered in a cavity probably at the DNA-protein interface that is sterically inaccessible to solute. Figure 1 schematically illustrates the difference in sequestered water between specific and nonspecific EcoRI binding.



**Figure 1:** A schematic representation of the difference in sequestered water between EcoRI specifically bound to its recognition sequence and the nonspecific protein-DNA complex. The crystal structure (29) shows that the specific complex is characterized by direct protein-DNA contacts with no intervening water. From the osmotic dependence of the difference between specific and nonspecific binding of EcoRI (23), the nonspecific complex retains practically a full layer of water at the interface between DNA and protein surfaces (shown as the crosshatched area). From the insensitivity of the number of sequestered waters to solute size and chemical nature, it is probable this water sterically excludes solutes.

We now report that the abrupt decrease in binding energy with even a single base pair change ('star' site) from the recognition sequence is accompanied by an abrupt increase in the water sequestered by the EcoRI-DNA complex, supporting a connection between hydration and binding energy. At low applied pressures, the osmotic stress dependence of the relative binding energies of DNA sequences that differ by one, two, or even six base pairs from the recognition sequence indicate that all these noncognate complexes sequester  $\sim 70$  more waters at  $\sim 0^\circ\text{C}$  than the specific complex (Figure 1). At higher pressures, however, we now see that the behavior of a 'star' sequence complex is strikingly different from complexes with the sequences having two or six wrong base pairs. Although the two or six wrong base pairs containing complexes still sequester 70 waters at the highest osmotic pressures applied, a 'star' sequence complex loses most of its water. This novel result can be straightforwardly

## Removing Water From an EcoRI-Noncognate DNA Complex

understood if it is realized that water sequestered by these complexes should not be considered fixed and invariant. In principle, any sequestered water can be removed by applying high enough osmotic stress (or, equivalently, low enough water activity or relative humidity). The work necessary to dehydrate complexes will naturally depend on the DNA-protein contacts resulting from the removal of water. This reasoning is consistent with the observed loss of water from 'star' sequence complex with one wrong base pair under osmotic conditions at which the complexes with sequences having two or six wrong base pairs show no loss of water.

### *Materials and Methods*

#### *Materials.*

The EcoRI restriction nuclease used in most of the experiments reported here was purchased from New England Biolabs and used without further purification. Active protein concentrations were determined by direct titration with a 322 bp fragment containing the recognition sequence under conditions of stoichiometric binding. Additional control experiments were performed with highly purified EcoRI (a generous gift of Dr. L. Jen-Jacobson) to ensure that the results obtained with the commercial enzyme were not artifactual. In agreement with several others (24,30), we found no significant differences in the binding properties of the two enzyme preparations. Both the competition for EcoRI binding between a DNA fragment containing the specific recognition site and nonspecific oligo(dI-dC)<sub>12</sub> and the osmotic stress dependence of the competitive binding constant are within experimental error for the two enzyme preparations.

A 322 bp DNA fragment carrying one EcoRI binding site, GAATTC, was isolated from the PvuII digestion products of pUC19 using standard techniques. Both pUC19 and the restriction nuclease PvuII were purchased from New England Biolabs and used without further purification. The self-complementary oligonucleotides used as competitor DNA were (dI-dC)<sub>12</sub>, 5'-ggcgatcgaGAATTCtgcgccc-3' carrying one EcoRI specific site (shown in capital letters), 5'-ggcccaTAATTCaccggtGAATTA<sup>g</sup>ggcgc-3' that has two 'star' sites (shown in capital letters) differing by one base pair (underlined) from the recognition site, and 5'-ggcgacGATATCGATATCg<sup>g</sup>tcgcc-3' that has two potential EcoRI binding sites differing by two base pairs from the specific sequence. Several experiments were performed with a double helical oligonucleotide containing a single 'star' sequence (5'-ggcgccatcTAATTCatccg<sup>g</sup>gg-3' and its complement) to ensure that the binding behavior of the two 'star' sequence oligonucleotide with increasing osmotic stress was not a consequence of the two sites.

All oligonucleotides were purchased from Gibco BRL, dissolved in 10-1 TE buffer, separated from small molecular weight impurities using P6 Bio-Spin columns at room temperature, and annealed. The double-stranded character of the oligonucleotides was confirmed by polyacrylamide gel electrophoresis. The concentrations of the DNA fragment and the oligonucleotides were determined spectrophotometrically, using an extinction coefficient of 0.0148 (μM base pairs)<sup>-1</sup> cm<sup>-1</sup> at 250 nm for (dI-dC)<sub>12</sub> and 0.013 (μM base pairs)<sup>-1</sup> cm<sup>-1</sup> at 260 nm for the other three oligonucleotides and the DNA fragment.

Betaine glycine was purchased from United States Biochemical and was used without further purification. Osmolal concentrations of betaine were determined by direct measurement using a vapor pressure osmometer (Wescor, Logan, UT; model 5100) at 20°C. Betaine osmotic pressures measured by freezing point depression are consistent with the 20°C data. Water chemical potentials are linearly proportional to solute osmolal concentrations, i.e.,  $\Delta\mu_w = \mu_w - \mu_w^{ref} = -RT [\text{osmolal}]/55.6$ , where  $\mu_w$  and  $\mu_w^{ref}$  are the water chemical potentials of the solution with and without added osmolyte, respectively.

Sidorova and Rau

Since we are particularly concerned with the differences in the osmotic stress dependence for the binding of EcoRI to various DNA sequences and not with the absolute dependence, we use a competition assay to measure directly the ratio of binding constants to different sequences. The decrease in EcoRI binding to a DNA fragment containing the specific recognition sequence is measured by the gel mobility shift assay as the concentration of oligonucleotide competitor is increased. The experimental procedure was described in detail previously (23). Briefly, EcoRI was added to mixtures of the 322 bp fragment (2.4 nM in specific EcoRI binding sites) and varying concentrations of oligonucleotide in 25 mM TrisCl (pH 7.5), 0.1 M NaCl, 2.5 mM EDTA, 1mM DTT, 0.1 mg/ml BSA, and 2.5% ficoll (70,000 m.w.), incubated on ice. The total reaction volume was 25  $\mu$ l. Sufficient EcoRI was added to give 40-60% stoichiometrically bound fragment without added competitor. Titration both of the protein with the specific DNA fragment and of the specific fragment with protein verified that the binding of active protein to the specific sequence was stoichiometric under the experimental conditions and DNA concentrations used. We observed no cleavage of the DNA in the absence of  $Mg^{2+}$ . At betaine concentrations < 2 osmolal, 30 minutes incubation was sufficient to reach equilibrium. At higher concentrations, the reaction rates are sufficiently slow such that longer incubation times are required. Separate kinetic measurements were performed at these higher betaine concentrations to determine the time necessary to reach equilibrium. The reaction mixtures were then electrophoresed in a 1.5% agarose gel, 0.5x TBE buffer, at 120 V, in the cold room, for 2 hrs to separate free DNA fragment and EcoRI-bound complex. As shown previously (31), the specific EcoRI-DNA complex is remarkably stable in the agarose gel over the course of the experiment. We observed no dependence of the fraction of DNA in complex on the time of gel electrophoresis between 30 min. and 2 hours.

*Quantitation and Data Analysis*

Electrophoretic bands containing free DNA fragment and DNA-protein complex were stained with SYBR Green I (Molecular Probes) on agarose gels and quantitated using fluorescent intensities as described previously (23). The linearity of fluorescent intensity versus DNA amount per band over the range of concentrations studied was confirmed using pBR322 DNA fragments generated by MspI digestion.

The ability of an oligonucleotide to compete with the specific site fragment for the EcoRI binding depends on the ratio between specific and oligonucleotide DNA association constants ( $K_{rel} = K_{sp}/K_{oligo}$ ). At equilibrium,

$$K_{rel} = \frac{K_{sp}}{K_{oligo}} = \frac{[(DNA_{sp} \cdot EcoRI)][DNA_{oligo}]_{free}}{[(DNA_{oligo} \cdot EcoRI)][DNA_{sp}]_{free}} \quad [1]$$

The fraction of specific fragment bound with EcoRI,  $f_b = [(DNA_{sp} \cdot EcoRI)]/[DNA_{sp}]_{total}$ , is experimentally determined from the gel mobility-shift assay. As developed previously (18,23), if the binding of EcoRI to the competitor oligonucleotides is much weaker than specific fragment binding ( $K_{oligo} \ll K_{sp}$ ), then, under conditions of virtually stoichiometric protein binding, the decrease in specific fragment binding with increasing competitor DNA concentration is given by,

$$f_b = - \frac{1}{K_{rel}} \frac{f_b}{1 - f_b} \frac{[DNA_{oligo}]_{total}}{[DNA_{sp}]_{total}} + f_b^0, \quad [2]$$

where  $f_b^0$  is the fraction of specifically bound DNA fragment in the absence of com-

## Removing Water From an EcoRI-Noncognate DNA Complex

petitor oligonucleotide. This expression is a rearrangement of the equation used previously. The relative binding constant,  $K_{rel}$ , can be straightforwardly calculated from the slope of  $f_b$  versus  $\frac{f_b}{1-f_b} [DNA_{oligo}]_{total}$ , while holding specific site and protein concentrations constant.

In order to compare our relative binding constants of EcoRI to 'star' and to non-specific sequences and subsequently with the results of Lesser *et al.* (26), we must correct for the two 'star' sites contained in the 'star' sequence oligonucleotide. Even though there is little difference between binding to oligo (dI-dC) and to the GATATC oligonucleotide or, by extension, to any other nonspecific DNA sequence the relative binding constants are simply calculated assuming that each nonspecific oligonucleotide also has two binding sites. The difference in the number of waters sequestered from solute between the specific and nonspecific complexes, calculated from the dependence of  $K_{rel}$  on the solution osmotic pressure, is, of course, independent of the assumed number of binding sites/oligonucleotide.

The data in Figure 3 for the dependence of  $\Delta G_{rel}$  ( $= -RT \ln(K_{rel})$ ) on the osmotic stress  $\Pi$  for the two 'star' sequence oligonucleotide competition was fit to a model assuming an equilibrium between two states of the 'star' sequence complex with different numbers of sequestered waters (relative to the specific site complex). The particular equation used for fitting is,

$$RT \ln(K_{rel,app}) = RT \ln(K_{rel,1}^0) + \Delta N_{w1} \bar{v}_w \Pi - RT \ln(1 + e^{-(\Delta G_{12}^0 + (\Delta N_{w2} - \Delta N_{w1}) \bar{v}_w \Pi)/RT}) \quad [3]$$

where  $\bar{v}_w$  is the average partial molar volume of water sequestered in the complex ( $\sim 18$  ml/mole),  $RT$  is the thermal energy,  $K_{rel,app}$  is the apparent binding constant of the 'star' sequence complex relative to the specific site,  $K_{rel,1}^0$  is the relative constant for state 1 at  $\Pi = 0$ ,  $\Delta N_{w1}$  and  $\Delta N_{w2}$  are the numbers of waters sequestered (relative to the specific site complex) by states 1 and 2, respectively, and  $\Delta G_{12}^0$  is the free energy difference between the two states at  $\Pi = 0$ . This nonlinear equation was fit to the data using the software TableCurve 2D (Jandel Scientific) that employs a standard Levenburg-Marquardt fitting algorithm.

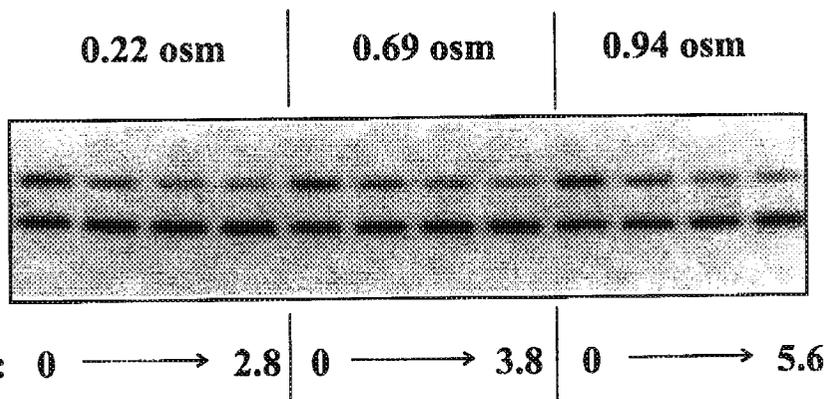
### Results

The general strategy is the same as used previously (23) and is outlined in Figure 2. Relative binding constants of EcoRI to different DNA sequences are measured by a competition assay. Competitive binding reactions allow a more direct and sensitive determination of the thermodynamic differences between DNA sequences than traditional measurements of the binding of protein free in solution to DNA. The loss of binding of EcoRI to a 322 bp DNA fragment containing its specific recognition sequence is measured using the gel mobility shift assay (32,33) as the concentration of a competitor oligonucleotide containing the sequence of interest is increased. The ratio of binding constants to fragment and to oligonucleotide can be extracted from the decrease in specific binding since the free protein concentration is much less than the protein-DNA complex concentration, i.e., binding is essentially stoichiometric. Both the reported dissociation constant of  $\sim 0.005$  nM (26) for the EcoRI-specific site complex under similar salt conditions and our previously estimated upper bound of  $< 0.03$  nM from protein-DNA titration experiments (23) are indeed much smaller than the DNA concentrations used,  $\sim 2.4$  nM sites. A typical gel mobility shift experiment at three water activities set by betaine glycine concentration and the relative binding constant analysis are shown in Figure 2.

The difference in binding free energy of EcoRI to the specific sequence and to a DNA sequence on an oligonucleotide is related to the ratio of association binding constants by  $\Delta G_{rel} = -RT \ln(K_{rel}) = -RT \ln(K_{sp}/K_{oligo})$ , where  $RT$  is the thermal energy. Figure 3 shows the dependence of the relative binding free energies of EcoRI to three different oligonucleotide DNA sequences on the osmolal concen-

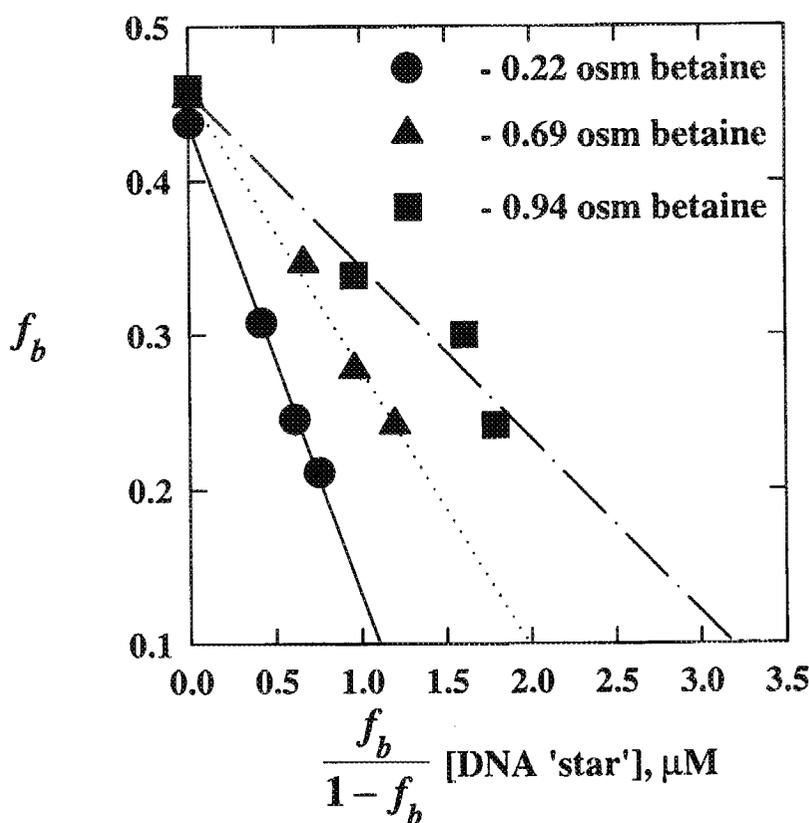
bound DNA  
free DNA

[DNA 'star'],  $\mu\text{M}$ : 0  $\longrightarrow$  2.8 | 0  $\longrightarrow$  3.8 | 0  $\longrightarrow$  5.6



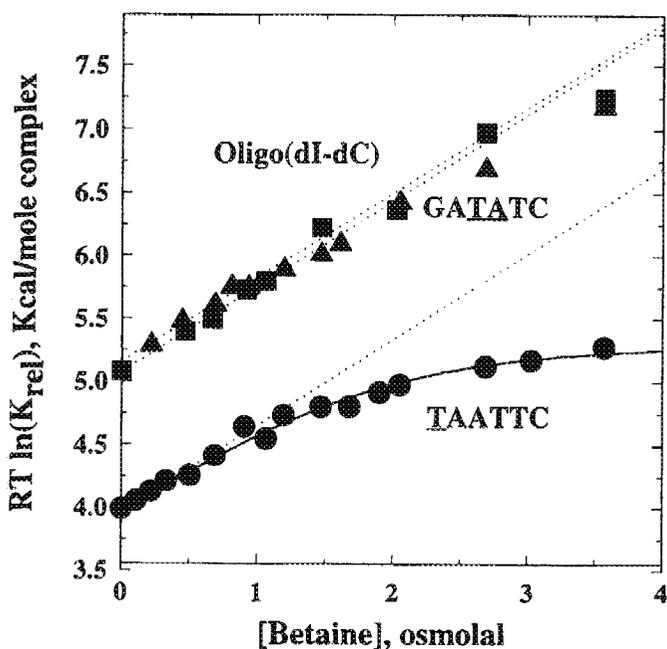
**Figure 2:** Measurement of relative binding constants. (A) The gel mobility-shift assay is used to monitor the loss of EcoRI binding to a 322 bp DNA fragment containing the enzyme's specific recognition sequence, GAATTC, as the concentration of a 30 bp oligonucleotide competitor containing two copies of the 'star' sequence TAAATC is increased. Three series of experiments for different osmolar concentrations of betaine are shown.

(B) The ratio of specific and 'star' sequence EcoRI association binding constants ( $K_{\text{rel}} = K_{\text{sp}}/K_{\text{star}}$ ) is extracted from the competition between the 322 bp fragment and the 'star' sequence oligonucleotide for EcoRI binding. The fraction of fragment with specifically bound EcoRI,  $f_b$ , is quantitated from the fluorescent intensities of the DNA bands in the gel. The linear slope of the plot  $f_b$  vs  $(f_b/(1-f_b))$  [DNA<sub>star</sub>] is  $-1/(K_{\text{rel}}[\text{DNA}_{\text{sp}}])$ , where [DNA<sub>sp</sub>] and [DNA<sub>star</sub>] are the concentrations of DNA fragment specific recognition sites and oligonucleotide 'star' sequences, respectively. For 0.22 (●), 0.69 (▲) and 0.94 (■) osmolar betaine,  $K_{\text{rel}} = (2.0 \pm 0.2) \times 10^3$ ,  $(3.3 \pm 0.28) \times 10^3$ ,  $(5.3 \pm 0.52) \times 10^3$ , respectively.



The relative binding energies of EcoRI to a 24 bp oligonucleotide containing a site that differs from the recognition site by two base pairs (GATATC) and to  $(\text{dI-dC})_{12}$  are closely similar. In the absence of added neutral solutes, the binding constants to these sequences are a factor of  $1.2 \times 10^4$  smaller than the binding constant to the specific site. Comparing the specific sequence with three doubly substituted sequences and the inverted site, Lesser *et al.* (26) reported a range of values,  $0.4 - 1.6 \times 10^4$ , for the ratio of specific and nonspecific binding constants under similar salt conditions but at 25°C. With added solutes, the difference in free energy varies linearly with osmotic stress at least through  $\sim 3$  osmolar (corresponding to  $\sim 70$  atm osmotic pressure or to  $\sim 95\%$  relative humidity). The linearity indicates that the difference between the numbers of excess waters associated with specific and nonspecific complexes of DNA with EcoRI,  $N_{\text{w,sp}}$  and  $N_{\text{w,nonsp}}$ , respectively, is constant over this range of betaine glycerine concentrations. The difference in the excess numbers of sequestered waters between the specific and these two nonspecific complexes can be extracted from the slopes through  $d[\text{RT} \ln(K_{\text{rel}})]/d[\text{osmolar}] = -\text{RT}$

### Removing Water From an EcoRI-Noncognate DNA Complex



**Figure 3:** The binding free energies of EcoRI to three oligonucleotides each relative to the specific sequence,  $-\Delta G_{rel} = RT \ln(K_{rel})$ , are shown as a function of the osmolal concentration of betaine. The competitor sequences shown are: (dI-dC)<sub>12</sub> (■), GATATC (two wrong base pairs) (▲), and TAATTC ('star' sequence) (●). Each point is the average of 2-4 separate experiments, with an error of at most ~ 15%. The dashed lines are linear regression fits to the initial, low pressure data, from 0 to 1 osmolal, for each oligonucleotide. The solid line is the best fit of the TAATTC oligonucleotide data to the two state model described in the text (Equation [3] of Materials and Methods).

$\Delta N_w/55.6$ , where  $\Delta N_w = N_{w,sp} - N_{w,nonsp}$ . The two nonspecific complexes sequester about 70 more water molecules ( $70 \pm 3$  for the GATATC oligonucleotide and  $69 \pm 7$  for (dI-dC)<sub>12</sub>) from the bulk solution than the specific complex at  $\sim 0^\circ\text{C}$ . This is in contrast to the 110 waters found previously at  $25^\circ\text{C}$  (23). In data not shown, the relative binding constant of EcoRI to the specific sequence on the 322 bp fragment and to the specific sequence on a 24 bp oligonucleotide ( $K_{sp,frag}/K_{sp,oligo} = 0.4$ , at  $\sim 0^\circ\text{C}$ ) does not depend on osmotic stress through the highest betaine concentration used ( $\sim 3$  osmolal). This signifies that there is no difference in sequestered water between the two specific complexes, as expected.

Figure 3 also shows data for the binding of EcoRI to a 'star' sequence (TAATTC) oligonucleotide that differs by one base pair from the recognition sequence. With no added osmolyte, the binding of enzyme is about 8 times stronger to the oligonucleotide containing TAATTC 'star' sequences than to GATATC oligonucleotide or (dI-dC)<sub>12</sub>, but still about a factor of  $1.6 \times 10^3$  weaker than to the recognition sequence. Lesser *et al.* (265) reported a factor of  $1.1 \times 10^3$  for an oligonucleotide also containing this 'star' sequence but at  $25^\circ\text{C}$ . The dependence of the binding free energy for the 'star' sequence oligonucleotide on the osmolal concentration of betaine glycine is very different from the two other noncognate sequences. Only at low stresses is the dependence approximately linear. The initial slope translates into a difference of  $70 \pm 5$  waters between the 'star' and specific sequence complexes. At higher osmotic stresses, however, the dependence of the relative binding free energy on osmolal concentration is clearly nonlinear. Consequently, EcoRI binding to the 'star' sequence oligonucleotide relative to the other two nonspecific sequence oligonucleotides becomes even stronger at higher osmotic stresses. At  $\sim 3$  osmolal, for example, the binding constant of EcoRI to the 'star' sequence oligonucleotide is about 40 times larger than to the nonspecific sequence oligonucleotides rather than the factor of 8 seen at betaine glycine concentrations less than 1 osmolal.

Since betaine glycine is still acting osmotically at these high stresses on the competition between specific EcoRI binding and binding to the nonspecific GATATC and (dI-dC) oligonucleotides, it is probable that this osmolyte is also acting osmotically on the competition reaction between the 'star' and specific sequence complexes. Otherwise there must be a specific binding of betaine glycine to the 'star' sequence complex that does not occur with either the nonspecific GATATC and (dI-dC) oligonucleotide complexes or the specific sequence complex. We consider this unlikely. The downward curvature of the data in Figure 3 would indicate that

the 'star' complex sequesters significantly fewer waters at higher stresses. The approximate slope at the highest pressures measured translates into a difference of only ~ 15 waters between the 'star' and specific sequence complexes.

The simplest model that can adequately fit the 'star' sequence complex data (the solid line in Figure 3) assumes an osmotic stress dependent equilibrium between two discrete states. Within this model (equation [3] of Materials and Methods), one state sequesters  $\Delta N_{w1}$  water molecules and the other  $\Delta N_{w2}$  (both relative to the specific sequence complex). The other variables are the association constant for the first state relative to the specific sequence in the absence of added solute,  $K_{rel,1}^0 (= K_{sp}^0/K_1^0)$ , and the free energy difference between the two states in the absence of added solute,  $\Delta G_{12}^0$ . The best fit of this two-state model to the 'star' sequence data gives  $RT \ln(K_{rel,1}^0) = 4.05 \pm .05$  and  $\Delta N_{w1} = 75 \pm 10$ . The parameters  $\Delta G_{12}^0$  and  $\Delta N_{w2}$  are tightly coupled and difficult to estimate as accurately. The best fitting value for the remaining number of sequestered waters in the second state,  $\Delta N_{w2}$ , is  $5 \pm 5$  waters and the corresponding best fitting value for  $\Delta G_{12}^0$  indicates that, with no applied stress, the state with 75 sequestered waters is more stable by  $1.4 \pm 0.5$  Kcal/ mole complex. The drier state of the 'star' sequence complex has very little sequestered water relative to the specific complex. The effect of osmotic stress is to modulate the free energy difference between the two binding modes of EcoRI to the 'star' sequence. While, with no added osmolyte, a calculated 93% of the star sequence complexes are predicted to be in the state with 75 sequestered waters, ~ 86% of the complexes are in the drier state at 3.5 osmolal.

### Discussion

Understanding the physical basis underlying the strength and specificity that characterizes the interaction of sequence specific DNA binding proteins with their target sequences will require measuring differences in structure and thermodynamics of complexes with varying DNA sequences. We have previously found for EcoRI (23) and for *E. coli* gal repressor (18) that there are large differences in the release of water between binding to the specific recognition sequences and in forming non-specific complexes. The osmotic stress approach we use to measure these differences in water is nothing more than a particular form of the fundamental Gibbs-Duhem equation that focuses on the role of water chemical potential in modulating reactions between states with different numbers of sequestered waters.

The exclusion of solute from water associated with macromolecular complexes can occur for several reasons. Many osmolytes, such as sucrose and betaine glycine, are excluded to varying extents from exposed surfaces. Reactions that result in changes in exposed surface area will necessarily depend on osmolyte concentration. Both crowding (34,35) and preferential hydration (36,37) are formulations that have been developed to describe exclusion from exposed surfaces. In the language of crowding, osmolyte exclusion is simply a matter of solute size. Preferential hydration recognizes an additional physical reality, i.e., exclusion from exposed surfaces should also depend on the chemical nature of the solute. Both crowding and preferential hydration' view the effect of solute on reactions from the standpoint of the excluded osmolyte. The osmotic stress method simply views these same reactions from the standpoint of the water that is necessarily included. The various viewpoints are complementary and not exclusive.

In contrast to the effect of solute on reactions with significant changes in exposed surface areas, there are several systems, the opening and closing of membrane channels (38), the oxygenation of hemoglobin (39), and the specific-nonspecific competition binding of EcoRI (23), that are characterized by a  $\Delta N_w$  that is insensitive, within experimental error, to solute size and chemical nature. We have concluded that in these cases  $\Delta N_w$  is due to a difference in water sequestered in cavities, grooves, or pockets that are sterically inaccessible to the solutes.

## Removing Water From an EcoRI-Noncognate DNA Complex

Our previous measurements of  $\Delta N_w$  between specifically and nonspecifically bound EcoRI are in this class. Five different solutes, glycine, methyl glucoside, sucrose, triethylene glycol, and betaine glycine, all gave  $\Delta N_w$  results values within 10% of the average of 110 waters at 20°C (23). We concluded that the extra water sequestered in the nonspecific EcoRI complex, compared with the specific one, is in some pocket that sterically excludes solute. We presumed this water is probably at the DNA-protein interface, of sterically sequestered water. By directly comparing the binding of protein to specific and nonspecific DNA sequences using a competitive assay we avoid the contribution from changes in exposed surface accompanying the binding of free protein.

These results do not preclude binding of solutes to complexes or their exclusion from exposed surfaces, only that the differences in binding or exclusion from surfaces between specific and nonspecific sequence complexes is negligible. Measuring  $\Delta N_w$  between specifically and nonspecifically bound enzyme directly using a competition assay rather than measuring separate absolute binding constants has the additional practical advantage that a single direct measurement, of course, involves less error than for determining the two binding constants separately. The measurement of a binding constant with nonspecific DNA that necessarily requires high protein concentrations can be particularly problematic.

We are now comparing  $\Delta N_w$  between specific and nonspecific complexes for different noncognate sequences. Only differences in the structure of complexes with the noncognate sequences and the consequent differences in their interaction with osmolytes will contribute to differences in  $\Delta N_w$ . The effect of solute on the energetics of the specific sequence complex is simply a common baseline.

The structural counterpart to the thermodynamical strong sequence specificity shown by EcoRI can be inferred from the number of waters sequestered by complexes with the different DNA sequences studied here. The dramatic decrease in binding constant to these noncognate sequences relative to the specific sequence is accompanied by a large increase in the number of waters sequestered by the complexes relative to the specific. Under low osmotic stress conditions, the complexes of EcoRI with the three noncognate sequences, (dI-dC), GATATC, and the 'star' sequence TAATTC oligonucleotides, all sequester ~ 70 more water molecules than the specific sequence complex. Since the binding constant to the 'star' sequence oligonucleotide is about ten fold higher than for the other two sequences, ~ 90% of the protein is bound to the 'star' sequence rather than to the flanking regions and, therefore, the 70 waters does indeed correspond to the 'star' sequence complex. The change of even one base pair from the specific recognition sequence is sufficient to result in an abrupt change between specific and nonspecific binding modes, at least as defined by sequestered water. If this water is indeed at the protein-DNA interface, then this transition is perhaps a consequence of the extensive, highly interconnected network of protein-protein and protein-DNA hydrogen bonds that characterizes the specific EcoRI-DNA complex (29,40,41). The interacting surfaces are perhaps too rigid to adapt to local, non-complementary patches, while maintaining close association of the remaining complementary regions.

Significant differences between complexes with the noncognate sequences are seen in the dependence of binding free energy on water activity at high stress. The application of large osmotic pressures can cause additional complications in binding reactions. Both the protein and DNA conformations, for example, could be sensitive to solute stress. The observation, however, that relative binding free energies of EcoRI to GATATC and oligo (dI-dC) continue to scale linearly with water chemical potential energy even at these high osmotic pressures (up to ~ 3 osmolal betaine) strongly indicates that betaine glycine is still acting simply osmotically. Additionally, at the high solute concentrations, the binding of the enzyme to specific sequence remains stoichiometric and there is no apparent loss of actively bind-

ing protein. Any change in protein or DNA conformation induced by these high solute concentrations is either small or affects specific and nonspecific sequence binding equally.

It is only with the 'star' sequence oligonucleotide that the dependence of  $\Delta G_{rel}$  on  $\Pi$  is distinctly nonlinear. While at low solute concentrations the osmotic pressure dependence of the binding constant ratio for competing specific and 'star' sequences closely resembles the results for the other two nonspecific sequences, at high solute concentrations the insensitivity to osmotic pressure more closely resembles the results for the specific sequence fragment - specific sequence oligonucleotide competition. One explanation is that betaine glycine binds specifically to the 'star' sequence complex, but not to the specific or nonspecific sequence complexes. Though possible, we consider this improbable. Rather, given the continued osmotic action of betaine glycine on the competition for EcoRI binding between the specific sequence and the nonspecific GATATC and (dI-dC) oligonucleotides, we interpret the decrease in the slope of  $\Delta G_{rel}$  vs.  $\Pi$  as indicating that the average number of waters sequestered by the 'star' sequence complex decreases with increasing stress.

The slope of  $\Delta G_{rel}$  vs.  $\Pi$  indicates that only ~15 waters/ 'star' sequence complex are sequestered from osmolyte at high stresses (~3 osmolal), a net loss of ~55 waters relative to the low stress complex. There are alternate possibilities that can account for this loss of water. The most straightforward explanation is that this loss is directly from the 70 initially sequestered waters. Since these waters are likely at the DNA-protein interface, the loss of this water would force closer DNA-protein contact. Alternatively, the net loss of water may be due to other changes in the 'star' sequence complex that are separate from the 70 sequestered waters. Changes that in structure that bury additional surface area, as, for example, due to a protein conformational transition or a protein-protein dimerization reaction, could also explain the data. Such changes, however, must then be particular for the 'star' sequence complex and not occur for the specific or nonspecific GATATC and (dI-dC) oligonucleotide complexes. Although we can not definitively exclude this possibility (this question could, in principle, be resolved using other solutes in addition to betaine), we think it more probable that the water loss is directly from the 70 sequestered waters. The difference between the 'star' sequence and the (dI-dC) and GATATC oligonucleotide complexes would then very naturally be the difference in energy required to remove the 70 waters and force more direct protein-DNA contacts between a sequence with one wrong base pair compared with sequences with having two or more wrong base pairs.

The data is not sufficiently precise to distinguish unambiguously an osmotic stress dependent equilibrium between two distinct states, i.e., between a 'star' sequence complex with ~ 70 sequestered waters and a second discrete binding mode with much less associated water, or among a continuum of states, i.e., a gradual loss of water from the 'star' complex. The osmotic work done in removing water from the 'star' sequence complex can be calculated independently of the mechanism. This  $\Pi dV$  work is straightforwardly calculated as,

$$W = - \int_{\Delta N_{w0}}^{\Delta N_{wf}} \Pi (N_w) \bar{v}_w dN_w = - \bar{v}_w \Delta N_{wf} \Pi_f + \int_0^{\Pi_f} \bar{v}_w N_w d\Pi, \quad [4]$$

where  $\Delta N_{w0}$  and  $\Delta N_{wf}$  are the numbers of waters sequestered by the complex at initial and final pressures 0 and  $\Pi_f$ , respectively, and  $\bar{v}_w$  is the average molar volume of water sequestered in the complex (~18 ml/mole). Since the excess number of waters sequestered by the 'star' sequence complex relative to the specific complex,  $\Delta N_w$ , is, by definition,  $d\Delta G_{rel}/\bar{v}_w d\Pi$ , the work is,

$$W = + \Delta G_{rel}(\Pi_f) - \bar{v}_w \Delta N_{wf} \Pi_f - \Delta G_{rel}(0). \quad [5]$$

## Removing Water From an EcoRI-Noncognate DNA Complex

This work is simply the difference between the measured relative binding free energy at  $\Pi = 0$  and the free energy change linearly extrapolated from the high stress region ( $\Pi_f$ ) back to  $\Pi = 0$ . If the slope at 3 osmolal for the 'star' sequence in Figure 3 is estimated as indicating that ~ 15 waters are left in the complex, then the work of removing ~ 55 of the 70 waters initially present is ~ 0.8 Kcal/mole complex. An estimate of the energy to remove all the waters is model dependent. The two state model used in calculating the best fitting curve in Figure 3 predicts that ~ 1.4 Kcal/mole complex is necessary to remove essentially all the water (the energy difference between the two states at  $\Pi = 0$ ). This means that the energy difference between the dehydrated 'star' sequence and specific site complexes is ~ 5.4 Kcal/mole complex (the plateau value of  $RT \ln(K_{rel})$  in Figure 3). This would be a direct measure of the interaction of the one wrong base pair with the protein both if the loss of water from the 'star' sequence complex is from the protein-DNA interface and if the structures of the specific sequence and the dehydrated 'star' sequence complexes are comparable. Any difference in structure of the dehydrated 'star' sequence complex from the specific binding mode at the high osmotic stresses would lower the energy difference.

It has long been known that EcoRI will infrequently cleave 'star' sequences and that the enzymatic activity at 'star' sequences increases with increasing concentrations of osmolytes, such as glycerol. Robinson and Sligar (42,43) have investigated the dependence of the cleavage rate at 'star' sites by EcoRI on solute concentration and have reported an osmotic effect that is qualitatively similar to our results for the competitive binding of EcoRI to specific and nonspecific sequences. There is a common curve for cleavage rate at 'star' sites vs. osmotic pressure for a very wide variety of solutes (glycerol, dimethyl sulfoxide, ethanol, ethylene glycol, dextrose, sucrose, 2-propanol, and N-methylformamide) even up to high osmotic pressures (~ 3 osmolal). These solutes are acting osmotically on a sterically sequestered water pocket to increase the cleavage rate at 'star' sequences. We suggest that the loss of water from the 'star' sequence complex inferred from the data in Figure 3 is directly related to the enzymatic activity at these sites.

More recent results of Robinson and Sligar (24), however, are not in agreement with the data presented here. Although Robinson and Sligar find a difference of ~ 70 waters at 25°C between the specific EcoRI complex and the complex with the nonspecific sequence TAGACG, a 'star' sequence complex is seen to release 140 waters more than the nonspecific complex even at low osmotic pressures. The 'star' sequence complex seems to sequester some 70 fewer waters than the specific one, even though the X-ray crystal structure (29,40,41) shows very few or no waters remaining at the protein-DNA interface of the specific complex already. These differences in water were determined from the differences in the osmotic dependence of absolute binding constants. We do not understand why these results are so different from our own. The relative binding constants with no osmotic stress measured by Robinson and Sligar are consistent with those of Lesser *et al.* (26) and with those reported here. We would point out that the protein concentrations necessary for titration of the nonspecific and 'star' sequence DNA are  $10^3 - 10^4$  times higher than with the specific sequence or than used here. It is possible that high osmotic stresses or the particular osmolyte used, ethylene glycol, causes additional complications at these high protein concentrations.

It should be noted that the osmotic stresses needed to remove water from the 'star' sequence complex are not particularly high compared with water activities used to crystallize many DNA-protein complexes. These solutions typically use polyethylene glycol or methyl pentanediol to control water activity in much the same way that betaine glycine was used here. The water seen at DNA-protein interfaces in several crystal structures may only reflect a fraction of the waters present with much lower osmotic pressures.

The osmotic stress dependence of the energetics of DNA-protein complexes offers novel possibilities for using thermodynamics to infer changes in structure. A standard strategy for dissecting the energetics of DNA-protein interactions is to mutate the complementary surfaces and to determine the change in binding free energy, enthalpy, or heat capacity. The changes are often difficult to interpret in terms of interactions between individual groups on the protein and DNA since these thermodynamic measurements contain little direct information about concomitant changes in complex structure. The simultaneous measurement of the change in the number of waters sequestered by DNA-protein complexes can help bridge this gap. Furthermore, one can now use osmotic pressure to remove water from complexes with different sequences and measure comparative energies associated with a structural change.

The release of structured water in forming specific complexes is thought to be responsible for the large heat capacity changes characteristic of specific DNA-protein binding (13-16). This can now be probed for EcoRI binding by measuring the temperature dependence both of the number of sequestered waters for nonspecific complexes in general and of the work to remove these waters from the 'star' sequence complex in particular. The difference of 70 waters between specific and nonspecific sequence complexes of EcoRI reported here at 0°C is significantly smaller than the 110 waters found at 25°C previously (23). This large difference suggests a large difference either in protein or DNA structure or in the energetics of hydration at the two temperatures that is directly pertinent to the inferred heat capacity.

Recent experiments directly measuring forces between many biopolymers in macroscopic arrays suggest that structuring of the intervening water dominates intermolecular interactions at close spacing (1), rather than van der Waals interactions or electrostatics. These 'hydration forces' can be either strongly repulsive or attractive depending sensitively on the complementarity of the apposing surfaces. The correlation between numbers of sequestered waters and binding free energies as surfaces are mutated or as water is removed at high stresses is directly pertinent to uncovering a link between recognition and hydration. Indeed, if the loss of water from the nonspecifically bound star complex is continuous, then the data in Figure 3 can be presented in a way that is strictly analogous to a force curve, i.e., as  $\Delta\Delta G$  or  $\Pi$  vs.  $\Delta N_w$ .

Alternatively, many other specific sequence DNA binding proteins, as, for example,  $\lambda$  Cro protein (13) and the *E. coli lac* repressor (44) are not as stringent as EcoRI and show a more gradual change in binding energy as the consensus recognition sequence is changed. The correlation between binding free energy and sequestered water provides complementary information to the osmotic work in removing water and is a further way to link hydration and binding energy. This newly realized possibility to measure not only differences in water sequestered by DNA-protein complexes but also changes in this water with osmotic stress can be a powerful tool for understanding sequence specific recognition.

#### *Acknowledgments*

We are deeply grateful to V.A. Parsegian, M.M. Garner, and M.G. Fried for the fruitful discussions and to L. Jen-Jacobson for the kind gift of highly purified EcoRI.

#### *References and Footnotes*

1. S. Leikin, V.A. Parsegian and D.C. Rau, *Annu. Rev. Phys. Chem.* 44, 369-395 (1993).
2. H.M. Berman, *Curr. Opin. Struct. Biol.* 4, 345-350 H.M. (1994).
3. N.D. Arbuckle and B. Luisi, *Nature Struct. Biol.* 2, 341-346 B. (1995).
4. T. Hard and T. Lundback, *Biophys. Chem.* 62, 121-139 (1996).
5. J.W.R. Schwabe, *Curr. Opin. Struct. Biol.* 7, 126-134 (1997).
6. T.A. Steitz, *Quart. Rev. Biophys.* 23, 205-280 (1990).
7. C.O. Pabo, *Annu. Rev. Biochem.* 61, 1053-1095 (1992).

## Removing Water From an EcoRI-Noncognate DNA Complex

8. Z. Otwinowski, R.W. Schevitz, R.-G. Zhang, C.L. Lawson, A. Joachimiak, R.Q. Marmorstein, B. Luisi and P.B. Sigler, *Nature* 335, 321-329 (1988).
9. D.S. Wilson, B. Guenther, C. Desplan and J. Kuriyan, *Cell* 82, 709-719 (1995).
10. T. Li, M.R. Stark, A.D. Johnson and C. Wolberger, *Science* 270, 262-269 (1995).
11. J.A. Hirsh and A.K. Aggarwal, *EMBO J.* 14, 6280-6291 (1995).
12. D.T. Gewirth and P.B. Sigler, *Nature Struct. Biol.* 2, 386-394 (1995).
13. Y. Takeda, P.D. Ross and C.P. Mudd, *Proc. Natl. Acad. Sci. USA* 89, 8180-8184 (1992).
14. J.E. Ladbury, J.G. Wright, J.M. Sturtevant and P.B. Sigler, *J. Mol. Biol.* 238, 669-681 (1994).
15. R.S. Spolar and T.M. Record, Jr., *Science* 263, 777-784 (1994).
16. C.J. Morton and J.E. Ladbury, *Prot. Sci.* 5, 2115-2118 (1996).
17. V.A. Parsegian, R.P. Rand and D.C. Rau, *Meth. Enzym.* 259, 43-94 (1995).
18. M.M. Garner and D.C. Rau, *EMBO J.* 14, 1257-1263 (1995).
19. K.M. Vossen, R. Wolz, M.A. Daugherty and M.G. Fried, *Biochemistry* 36, 11640-11647 (1997).
20. C.R. Robinson and S.G. Sligar, *Protein Science* 5, 2119-2124 (1996).
21. L. Li and K.S. Matthews, *Biochemistry* 36, 7003-7011 (1997).
22. J. Poon, M. Bailey, D.J. Winzor, B.E. Davidson, and W.H. Sawyer, *Biophys. J.* 73, 3257-3264 (1997).
23. N.Y. Sidorova and D.C. Rau, *Proc. Natl. Acad. Sci. USA* 93, 12272-12277 (1996).
24. C.R. Robinson and S.G. Sligar, *Proc. Natl. Acad. Sci. USA* 95, 2186-2191 (1997).
25. T. Lundback, H. Hansson, S. Knapp, R. Ladenstein and T. Hard, *J. Mol. Biol.* 276, 775-786 (1998).
26. D.R. Lesser, M.R. Kurpiewski and L. Jen-Jacobsen, *Science* 250, 776-786 (1990).
27. L. Jen-Jacobsen, *Meth. Enzym.* 259, 305-344 (1995).
28. L. Jen-Jacobsen, L.E. Engler, D.R. Lesser, M.R. Kurpiewski, C. Yee, C. and B. McVerry, *EMBO J.* 15, 2870-2882 (1996).
29. J.A. McClarin, C.A. Frederick, B.-C. Wang, P. Greene, H.W. Boyer, J. Grable and J.M. Rosenberg, *Science* 234, 1526-1541 (1986).
30. S. Venditti and R.D. Wells, *J. Biol. Chem.*, 266, 16786-16790 (1991).
31. N.Y. Sidorova, P. Gazoni and D.C. Rau, *J. Biomol. Struct. and Dynam.* 13, 367-385 (1995).
32. M.G. Fried and D.M. Crothers, *Nucl. Acids Res.* 9, 6505-6525 (1981).
33. M.M. Garner and A. Revzin, *Nucl. Acids Res.* 9, 3047-3060 (1981).
34. S.B. Zimmerman and A.P. Minton, *Annu. Rev. Biomol. Struct.* 22, 27-65 (1993).
35. A.P. Minton, *Meth. Enzym.* 295, 127-149 (1998).
36. S.N. Timasheff, *Annu. Rev. Biophys. Biomol. Struct.* 22, 67-97 (1993).
37. S.N. Timasheff, *Adv. Protein Chem.* 51, 355-432 (1998).
38. S.M. Bezrukov, I. Vodyanoy and V.A. Parsegian, *Nature (London)* 370, 279-281 (1994).
39. M.F. Colombo, D.C. Rau and V.A. Parsegian, *Science* 256, 655-659 (1994).
40. J.M. Rosenberg, *Curr. Opin. Struct. Biol.* 1, 104-113 (1991).
41. Y. Kim, J.C. Grable, R. Love, P.J. Greene and J.M. Rosenberg, *Science*, 249, 1307-1309 (1990).
42. C.R. Robinson and S.G. Sligar, *J. Mol. Biol.* 24, 302-306 (1993).
43. C.R. Robinson and S.G. Sligar *Biochemistry* 33, 3787-3793 (1994).
44. D.E. Frank, R.M. Saecker, J.P. Bond, M.W. Capp, O.V. Tsodikov, S.E. Melcher, M.M. Levandoski and T. M. Record, Jr., *J. Mol. Biol.* 267, 1186-1206 (1997).

*Date Received: March 10, 1999*

**Communicated by the Editor Ramaswamy H. Sarma**