Coupling of Local Folding to Site-Specific Binding of Proteins to DNA

Ruth S. Spolar and M. Thomas Record Jr.

Thermodynamic studies have demonstrated the central importance of a large negative heat capacity change (ΔC°_{assoc}) in site-specific protein-DNA recognition. Dissection of the large negative ΔC°_{assoc} and the entropy change of protein-ligand and protein-DNA complexation provide a thermodynamic signature identifying processes in which local folding is coupled to binding. Estimates of the number of residues that fold on binding obtained from this analysis agree with structural data. Structural comparisons indicate that these local folding transitions create key parts of the protein-DNA interface. The energetic implications of this "induced fit" model for DNA site recognition are considered.

Noncovalent interactions between proteins and nucleic acids are fundamental to all steps of expression, replication, and recombination of the genetic information embedded in a DNA sequence. Notable characteristics of most protein-DNA complexes include their stability and specificity: binding constants under typical ionic conditions range from 10^9 to 10^{12} M^{-1} or higher and ratios of specific to nonspecific binding constants range from 10^3 to 10^7. A central goal of many structural and thermodynamic studies is to elucidate the relation between sequences (structures) of the protein and DNA sites and the specificity and stability of the complex in solution.

High-resolution structures of various uncomplexed DNA-binding proteins and their complexes with DNA have been obtained in the crystalline state by x-ray crystallography or in solution by nuclear magnetic resonance (NMR). In many of these systems, structural differences are observed between B-form DNA and DNA in the complex and between the free and bound states of the protein. Conformational changes from linear B-DNA range from relatively smooth bending deformations seen in specific complexes with phage repressor proteins (*1*) to more drastic changes involving sharp bends in the helical axis and disruption of base pair stacking interactions (*2–4*). Conformational changes observed in various proteins include quaternary rearrangements of domains (*5*) or subunits (*6*), ordering of disordered loops or NH_2-terminal residues (*4, 7, 8*), and formation of α helices from unfolded regions in the free state (*9–14*). Most of the folding or ordering transitions of protein residues occur on complementary surfaces in the DNA grooves that comprise the binding site. Frankel and Kim (*15*) and Alber (*16*) have proposed that the folding transitions of GCN4 are an example of the concept of induced fit described by Koshland (*17*) in the context of enzyme-substrate and other specific protein-ligand interactions. However, it is not known whether induced fit is a general feature of site-specific binding of proteins to DNA.

Thermodynamic studies of site-specific protein-DNA associations demonstrate the central importance of a large negative heat capacity change ΔC°_{assoc} (*18–20*). As a consequence, the entropic ($T\Delta S^\circ_{assoc}$) and enthalpic (ΔH°_{assoc}) contributions to ΔG°_{assoc} of site-specific complexation vary with temperature in a nearly parallel manner, typically changing sign in the physiological temperature range. We proposed that a large negative ΔC°_{assoc} is a distinctive feature of site-specific binding, and that the sign of ΔC°_{assoc} is determined by the removal of large amounts of nonpolar surface from water on complex formation (*18*). Because the amount of nonpolar surface involved appeared too large to be accounted for in a "rigid body" association, we proposed that conformational changes in the protein that buried large amounts of nonpolar surface were coupled to binding (*18, 19*). Although this prediction appears consistent with the emerging picture of site-specific binding from recent structural studies, most systems that have been characterized thermodynamically lacked complete structural data, especially data from structures in solution. Hence we turned to a thermodynamic analysis to predict whether conformational changes are coupled to binding.

In principle, analysis of the entropy change ΔS°_{assoc} should allow discrimination between "rigid body" associations and ones involving coupled conformational changes, especially folding transitions (*21*). We used thermodynamic data from model compound transfers and protein folding studies to dissect ΔS°_{assoc} for a protein-ligand association into contributions from the burial of nonpolar surface area, changes in chain conformation, loss of translational and rotational entropy, and other effects. Application of this analysis to site-specific protein-DNA interactions leads us to propose that local (and, in some cases, global) protein folding transitions are coupled to DNA binding. Where binding free energy drives folding of residues to create key parts of the binding interface, this adaptation may in principle be DNA sequence–dependent and lead to different ordered or folded structures on different DNA sequences (*16*). Such a range of induced fits may affect both specificity and stability; accordingly, the energetics of specific recognition should be intrinsically nonadditive (*22, 23*).

Interpretation of the heat capacity and entropy changes of protein folding. Model compound transfer data have been used to interpret the massive reduction in partial molar heat capacity (of the order of 1 to 6 kcal mol^{-1} K^{-1}) (*24, 25*) of a protein on folding in aqueous solution (ΔC°_{fold}) (*26–30*). We observed that ΔC°_{tr} of transfer of hydrocarbons and amides from water to the pure liquid phase and ΔC°_{fold} of proteins depend in the same way on the accompanying changes in water-accessible nonpolar (ΔA_{np}) and polar (ΔA_p) surface area (Å^2) (*26*)

$$\Delta C^\circ = (0.32 \pm 0.04)\Delta A_{np} - (0.14 \pm 0.04)\Delta A_p \text{ cal mol}^{-1}\text{ K}^{-1} \quad (1)$$

In applications of Eq. 1 to processes involving proteins, the contribution from the burial of nonpolar surface, $(0.32 \pm 0.04)\Delta A_{np}$, is always the dominant term in ΔC°. Although the choices of nonaqueous phase and of compounds used to model thermodynamic contributions to folding continue to be debated, models of ΔC°_{fold} that are based only on changes in water-accessible surface area account for all of the observed heat capacity change within experimental uncertainty. Changes in vibrational modes (or other effects of folding) apparently do not contribute significantly to ΔC°_{fold} although a vibrational contribution was anticipated (*31*).

Baldwin (*32*) observed that the entropy of transfer of all liquid hydrocarbons from water to the pure liquid phase extrapolates to zero at 386 K, on the assumption that ΔC°_{tr} was temperature-independent. If the thermodynamics of the hydrophobic effect (HE) are defined in terms of the thermodynamics of this transfer process for liquid hydrocarbons (*18, 26–28, 32–34*), then it follows that the entropic contribution from the hydrophobic effect vanishes at 386 K (*35*). Because changes in the exposure of nonpolar surface to water are thought to be

R. S. Spolar is in the Department of Chemistry and M. T. Record Jr. is in the Departments of Chemistry and Biochemistry, University of Wisconsin–Madison, Madison, WI 53706, USA.

the major contributor to the heat capacity change of processes involving proteins (*31*), Baldwin (*32*) proposed that the contribution to the observed entropy change of protein folding from the hydrophobic effect (ΔS°_{HE}) could be obtained from the relation $\Delta S^{\circ}_{HE}(T) = \Delta C^{\circ}_{fold} \ln(T/386)$, where T is temperature in degrees kelvin.

At a higher level of resolution, because ΔC°_{fold} appears to contain contributions from changes in both nonpolar and polar water-accessible surface area (*26*), the following equation may be more appropriate

$$\Delta S^{\circ}_{HE}(T) = 0.32\ \Delta A_{np} \ln (T/386) \quad (2)$$

where ΔA_{np} is in Å^2, T in degrees K, and $\Delta S^{\circ}_{HE}(T)$ in entropy units (e.u., or cal K^{-1} mol^{-1}). We propose that Eq. 2 is generally applicable to all protein processes to quantify the entropic contribution of the hydrophobic effect (ΔS°_{HE}) from structural information. Where ΔA_{np} is not known but where it is reasonable to assume that the ratio of $\Delta A_p/\Delta A_{np} = 0.59$ as in the folding of globular proteins (*26, 36*), then

$$\Delta S^{\circ}_{HE}(T) = 1.35\ \Delta C^{\circ} \ln (T/386) \quad (3)$$

Because $|\Delta C^{\circ}_{fold}|$ is large, the entropy change of folding (ΔS°_{fold}) is temperature-dependent. Hence a temperature T_S exists where ΔS°_{fold} is zero. At T_S, the contributions to ΔS°_{fold} from the hydrophobic effect [$\Delta S^{\circ}_{HE}(T_S)$] and from other sources (ΔS°_{other}; assumed as a first approximation to be temperature-independent) must sum to zero

$$\Delta S^{\circ}_{fold}(T_S) = 0 = \Delta S^{\circ}_{HE}(T_S) + \Delta S^{\circ}_{other} \quad (4)$$

Experimental values of T_S for folding of globular proteins, the number of residues folded ($\Re$), and values of $\Delta S^{\circ}_{HE}(T_S)$ calculated from the amount of nonpolar surface buried in folding (Eq. 2) are given in Table 1. Application of Eq. 4 to the data of Table 1 indicates the existence of a large unfavorable entropy contribution (ΔS°_{other}). Expressed per residue ($\Delta s^{\circ}_{other} \equiv \Delta S^{\circ}_{other}/\Re$), this unfavorable entropic contribution is relatively constant for the entire set of globular proteins ($\Delta s^{\circ}_{other} = -5.6 \pm 0.5$ e.u.) (*37*).

The analysis in Table 1 is consistent with previous proposals (*24, 32, 34*) of two dominant and opposing contributions to the observed entropy of protein folding: one from the hydrophobic effect, or the "release" of water on burial of nonpolar surface, and the other from the reduction in conformational entropy. We can then ask whether the magnitude of Δs°_{other} in Table 1 is entirely due to the reduction in conformational entropy upon folding. Summation of theoretical values for the change in conformational entropy of the backbone (*38*) and side chains (*39*) on folding yields an average change in conformational entro-

Table 1. Entropic contribution to protein folding from the hydrophobic effect.

Protein	$\Re$	$-\Delta C^{\circ}_{fold}$ (cal mol^{-1} K^{-1})	T_S* (K)	$-\Delta A_{np}$† (Å^2)	$\Delta S^{\circ}_{HE}(T_S)$‡ (e.u.)	$-\Delta s^{\circ}_{other}$§ (e.u.)
Streptococcal protein G, domain B1	56	620 (*68*)	272	2900	325	5.8
BPTI	58	720 (*24*)	306	2640	196	3.4
		400 (*69*)	221		471	8.1
Parvalbumin b	108	1100 (*70*)	268	5485	640	5.9
Ribonuclease A	124	1230 (*25*)	255	5815	771	6.2
Lysozyme (hen egg white)	129	1540 (*25*)	270	6870	786	6.1
Ferricytochrome c	104	1730 (*25*)	294	5540	483	4.6
Staphylococcal nuclease	141	1820 (*25*)	288	7880	738	5.2
Holo‖ myoglobin	153	2770 (*25*)	301	9710	773	5.1
β trypsin	223	2850 (*25*)	281	11830	1200	5.4
Papain	212	2920 (*25*)	290	12755	1167	5.5
α chymotrypsin	245	3020 (*25*)	280	14770	1517	6.2
Carbonic anhydrase	256	3820 (*25*)	290	15760	1442	5.6
Pepsinogen	370	6090 (*25*)	297	23730	1990	5.4
					Average¶	5.6 ± 0.5

*Values of T_S were calculated from values of ΔC°_{fold} and ΔS°_{fold} cited in the reference indicated in column 3. Reported uncertainties in ΔC°_{fold} range from 5 to 20 percent. Corresponding uncertainties in T_S range from 1 to 7 K degrees. †Calculations of ΔA_{np} model the denatured state as an extended β chain (*26, 27*). The value of ΔA_{np} for folding the B1 domain of streptococcal protein G was calculated as described in (*27*) from Brookhaven Protein Database (*67*) file 2GB1. All other values of ΔA_{np} are from (*26*). ‡Eq. 2. §$\Delta s^{\circ}_{other} = \Delta S^{\circ}_{other}/\Re$, calculated from Eq. 4. ‖In this and subsequent tables, holo refers to the protein associated with its cofactor. ¶Not including BPTI.

Table 2. Comparison of predicted and observed heat capacity changes for protein-ligand association. For entries where folding is coupled to the association process (as noted in last column), the unfolded region was modeled as an extended chain of the appropriate amino acid sequence by means of Insight II (Biosym Technologies). Structures used to calculate surface area changes on association are referenced below.

Process	$-\Delta A_{np}$ (interface) (Å^2)	$-\Delta A_{np}$ (fold) (Å^2)	$-\Delta A_p$ (interface) (Å^2)	$-\Delta A_p$ (fold) (Å^2)	$-\Delta C^{\circ}_{calc}$* (cal mol^{-1} K^{-1})	$-\Delta C^{\circ}_{assoc}$ (cal mol^{-1} K^{-1})	Coupled process
Subtilisin + inhibitor → complex	985†	0	690	0	220 ± 50	240 ± 25 (*71*)	None
Yeast hexokinase + glucose → complex	175‡	0	22	0	50 ± 10	50 ± 110 (*72*)	Hinge bend
FK506 + FKBP-12 → complex	650 (*73*)	0	−7	0	210 ± 25	260 ± 80 (*73*)	None
2 α chymotrypsin → dimer	1260§	0	945	0	270 ± 60	700 (*74*)	None
	1283‖	0	845	0	290 ± 60		
S-protein + S-peptide → ribonuclease S	758¶	320	672	365	200 ± 60	425 ± 250 (at 273 K) (*47*)	Folding
L-tryptophan + apo# TrpR monomer → complex	248 (*55*)	785	17	444	270 ± 45	230 ± 40 (*55*)	Folding
Avidin + biotin → complex	302 (*75*)	635	278	322	220 ± 45	237 ± 12 (*76*)	Folding
Angiotensin II + antibody Fab 131 → complex	993 (*48*)		745		215 ± 50	240 ± 20 (*48*)	Folding
Holo TrpR dimer + *trp* operator DNA → complex	1471 (*55*)	945	1884	608	420 ± 140	540 ± 100 (*55*)	Folding
2 arc repressor → dimer	2356 (*77*)	3312	1252	3010	1220 ± 280	1400 ± 200 (*78*)	Folding
2 λ cro repressor → dimer	922 (*79*)	5617	490	2965	1610 ± 300	1530 ± 170 (*80*)	Folding
2 GR DBD + DNA site → complex	1885**	2060	2213	1800	700 ± 225	1000 ± 200 (*56*)	Folding
3 glucagon → trimer	1044††	2169	418	951	840 ± 140	630 ± 100‡‡	Folding
4 melittin → tetramer	3775§§	3332	660	1794	1930 ± 300	1100 ± 240 (*82*)	Folding

*Eq. 1. †Brookhaven Protein Database (PDB) (*67*) file code 2SIC. ‡PDB files 2YHX, 1HKG. §PDB file 4CHA. ‖PDB file 5CHA. ¶PDB file 1RNV. #In this and subsequent tables, apo refers to the protein in the absence of its cofactor. **Surface area buried upon dimerization calculated from PDB file 1GRD. Surface area buried from folding of the second Zn loop: residues (476 to 495) were modeled as an extended chain in each monomer in 1GRD (*13, 14*). To obtain ΔA_{np} and ΔA_{np} for the protein-DNA interface, we used the estimate of the total surface buried in the interface from (*14*), and assumed that the ratio $\Delta A_{np}/\Delta A_p$ was 45:55, similar to that buried in the TrpR-DNA interface. ††PDB file 1GCN. ‡‡Data of (*81*) analyzed according to (*18*). §§PDB file 2MLT.

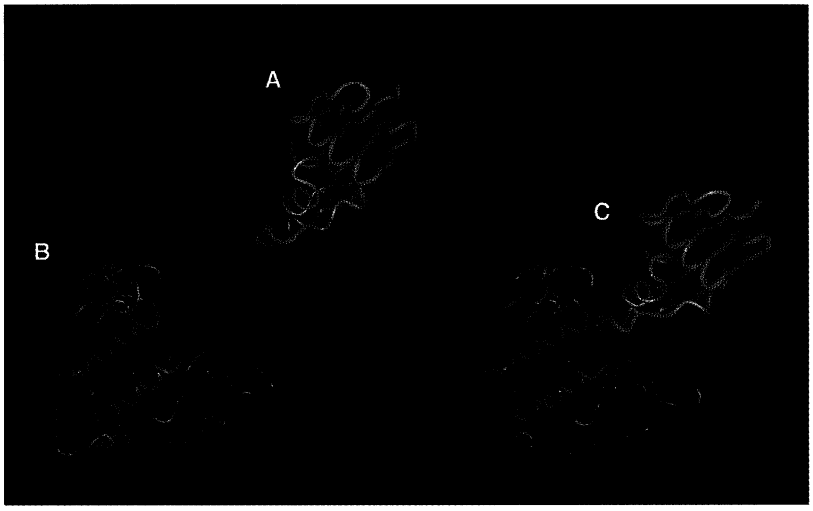

Fig. 1. Ribbon model of a "rigid-body" association. X-ray crystallographic structures of (**A**) subtilisin inhibitor monomer (purple, PDB file 2SSI) and (**B**) uncomplexed subtilisin (green, PDB file 2ST1), shown in the same orientation as in the complex. (**C**) Enzyme-inhibitor complex (PDB file 2SIC), same colors as in (A).

py per residue (Δs°_c) on folding of $-(6 \pm 2)$ e.u. Because $\Delta s^\circ_{other} = \Delta s^\circ_c$, we conclude that possible entropic contributions from dehydration of polar surfaces (*40*), changes in hydrogen bonding, and changes in vibrational entropy (*41*) must collectively make no large contribution to the overall ΔS°_{fold}.

Analysis of heat capacity and entropy changes in protein-ligand associations. Protein-ligand and protein-protein associations are also accompanied by large negative heat capacity changes (ΔC°_{assoc}). To examine whether ΔC°_{assoc} can be interpreted solely in terms of changes in polar and nonpolar surface, we compared experimental values of ΔC°_{assoc} with those calculated from Eq. 1 for all protein-ligand and protein-protein interactions for which both thermodynamic and high resolution structural data are available (Table 2). Crystal structures were used, except for glucocorticoid receptor and trp repressor, where crystal structures of the complexes have been supplemented by NMR solution structures of the uncomplexed proteins. The associations in Table 2 represent diverse interactions including enzyme-substrate, antibody-antigen, protein-drug, protein-DNA, protease-protease inhibitor, and subunit-subunit assemblies. These processes bury widely different amounts and types of surfaces, both in and distant from the binding site. The first four entries in Table 2 are examples of rigid-body associations where crystal structural analyses provide no evidence for disorder-order transitions on binding; in these cases, ΔA_{np} and ΔA_p result entirely from burial of preexisting complementary surfaces (Fig. 1). The last ten entries are for systems where biophysical data indicate that folding is coupled to association (Fig. 2). In these systems, surface area is removed from water both in the contact between any preexisting components of the interface and in coupled conformational changes (driven by binding free energy) that create a complementary interface.

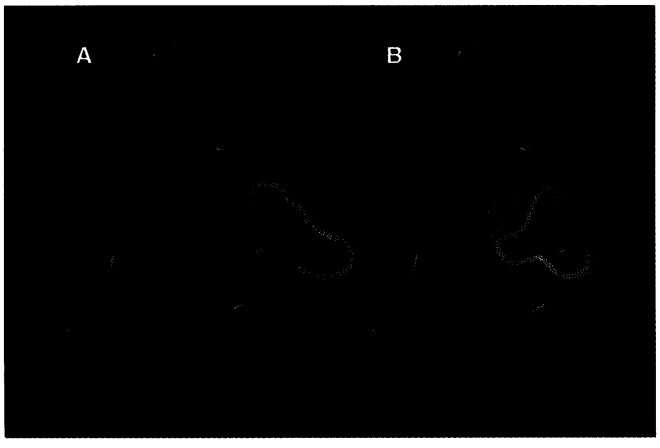

Fig. 2. Ribbon model of avidin-biotin "induced fit" interaction. (**A**) Model of the uncomplexed avidin monomer in solution (green). Residues (36–44) (dashed loop in yellow) are disordered in the free crystal structure (*49*) and are inferred to be in a flexible coil state of high conformational entropy in solution. (**B**) Avidin-biotin complex. Ordering of the looped region (yellow) upon binding encloses biotin (in purple) in a "hydrophobic box" (*49*).

In 12 out of 14 cases (Table 2) the observed heat capacity changes agree within experimental uncertainty with those predicted by Eq. 1. The only exceptions are the dimerization of α-chymotrypsin (α CT) and the tetramerization of melittin. The thermodynamics of α CT dimerization may be affected by autolysis. A proposal to explain the discrepancy for melittin is discussed below. From Table 2 we conclude that ΔC°_{assoc} can generally be quantitatively described in terms of contributions from changes in water-accessible nonpolar and polar surface area. Changes in vibrational modes or other effects of association apparently contribute little to ΔC°_{assoc}.

The large negative heat capacity changes of protein-ligand association and of protein folding indicate that the observed entropy changes of both processes must include a significant contribution from the hydrophobic effect in the physiological temperature range (Eqs. 2 and 3). However, other contributions to the entropy changes of these processes should differ significantly. Folding of single-subunit proteins involves the intramolecular conversion of a random polypeptide chain into an ordered structure, which reduces its conformational entropy. This unfavorable entropic contribution to folding increases in direct proportion to the number of amino acid residues. Unless folding is coupled to association, protein-ligand binding does not involve a reduction in conformational entropy. Instead, another unfavorable entropic term resulting from the reduction in the available rotational and translational degrees of freedom of the protein and ligand on complexation is predicted to contribute (ΔS°_{rt}: *42–44*). Since the dependences of ΔS°_{rt} on temperature and on molecular mass are predicted to be logarithmic, ΔS°_{rt} should be relatively insensitive to T_S and to the molecular mass of the protein and ligand involved and thus similar for all associations (*43*). Hence, for protein-ligand association at the characteristic temperature T_S where the net entropy change ΔS°_{assoc} is zero, Eq. 4 is replaced by

$$\Delta S^\circ_{assoc} = 0 = \Delta S^\circ_{HE}(T_S) + \Delta S^\circ_{rt} + \Delta S^\circ_{other} \quad (5)$$

where ΔS°_{other} includes any other entropic contribution in addition to those enumerated in Eq. 5, and is assumed to be temperature-independent (*40*).

In applying Eq. 5 to analyze entropic effects in the protein associations listed in Table 2, we focus initially on the set of rigid body associations where ΔS°_{other} may be anticipated to be small and which therefore provide a means of evaluating ΔS°_{rt} from experimental thermodynamic data. The proposed contributions to the entropy changes for protein-ligand associations which do not involve large conformational changes (Fig. 1) are summarized in Table 3. At T_S, we find

that ΔS°_{HE} for these processes is compensated by an entropy change $\Delta S^\circ_{rt} + \Delta S^\circ_{other}$ in the range −40 to −60 e.u. (Eq. 5). The magnitude and sign of this term and its relative insensitivity to the size of the associating species agree well with the statistical mechanical estimate of $\Delta S^\circ_{rt} \cong -50$ e.u. for protein-protein association (*43, 45*). We conclude that for these rigid body associations $\Delta S^\circ_{other} = 0$ and that ΔS°_{assoc} is entirely determined by ΔS°_{HE} and ΔS°_{rt}.

The set of protein-ligand and protein-protein associations where evidence indicates that folding of one or more associating species is coupled to binding (Fig. 2) is collected in Table 4. To dissect the entropy changes for these associations we use the average value $\Delta S^\circ_{rt} \cong -50$ e.u. from Table 3. Comparison of Table 3 and Table 4 shows that ΔC°_{assoc} of processes with coupled folding is generally larger in magnitude than for rigid body associations and that the observed T_S is smaller. Consequently, values of $\Delta S^\circ_{HE}(T_S)$ estimated from Eq. 2 are much larger than ΔS°_{HE} of rigid body associations and much larger than the magnitude of ΔS°_{rt}. For the observed entropy change to be zero at T_S, Eq. 5 predicts that there must be a significant entropy change ΔS°_{other}. For these associations, we propose that ΔS°_{other} corresponds to the change in conformational entropy (ΔS°_c) on binding.

If conformational changes that involve folding (changes in 2° or 3° structure) dominate ΔS°_{other}, then division of ΔS°_{other} by −5.6 e.u. yields a thermodynamic estimate of the number of residues involved in the folding transition (*46*) (designated $\mathfrak{R}^{th}$):

$$\mathfrak{R}^{th} \equiv \Delta S^\circ_{other}/-5.6 \text{ e.u.} \qquad (6)$$

The last two columns of Table 4 compare $\mathfrak{R}^{th}$ with our structural estimates of extents of coupled folding ($\mathfrak{R}^{str}$) based on biophysical [circular dichroism (CD), NMR, x-ray] characterizations of free and bound states. For the first six entries in Table 4, local folding of the protein or ligand accompanies association. The uncomplexed peptide ligands S-peptide (*47*) and angiotensin II (*48*) are completely unfolded in solution. Similarly, in the absence of ligand, local regions of avidin (*49*) (Fig. 2), trp repressor (*11*), and glucocorticoid receptor (*13*) are disordered or unfolded in solution. All of these regions are folded in the crystal structure or NMR structure of the complexes. In these six systems, the folding transition creates all or part of the binding site and is driven by binding free energy. Values of ΔS°_{other} for these cases range from −18 to −185 e.u. and predict folding of 3 to 33 residues on binding. In all cases these thermodynamic predictions agree with structural estimates, regardless of whether the ligand or regions of the protein fold.

The last four processes in Table 4 involve assembly of protein oligomers from monomeric subunits that are known to be completely unfolded in solution. In these cases, binding is accompanied by conformational changes both at the subunit interface and distant from it. Calculated values of ΔS°_{other} (−280 to −580 e.u.) are larger in magnitude than for the other processes in Table 4, indicating more extensive folding in agreement with structural data. For these proteins, the number of residues predicted to fold from ΔS°_{other} is greater than the number of residues that fold to create the subunit interfaces. The striking characteristic of these four systems is that binding free energy from subunit-subunit interactions drives nucleation of folding which propagates beyond the interface. For these four cases, $\mathfrak{R}^{th}$ is somewhat smaller than $\mathfrak{R}^{str}$ as estimated from the crystal structure of the complex. Hence these protein complexes appear to be less folded in solution under the conditions of the association experiments than in the crystal. Evidence of this is available for melittin (*50*).

Coupling of folding to binding in site-specific protein-DNA interactions. Most site-specific protein-DNA associations are accompanied by large negative ΔC°_{assoc} (*18*) (Table 5). By analogy with the above analyses of folding and of protein-ligand associations, we assume that ΔC°_{assoc} for protein-DNA interactions is entirely due to changes in water-accessible surface. To examine our previous proposal that the large negative ΔC°_{assoc} of site-specific binding of proteins to DNA is in general too large in magnitude to result from a rigid body association (*18*), we dissect ΔS°_{assoc} of protein-DNA interactions to examine whether folding transitions are coupled to binding.

In Table 5 we list experimental values or

Table 3. Entropic contributions to "rigid body" associations.

Process	$-\Delta C^\circ_{assoc}$ (cal mol^{-1}K^{-1})	T_S* (K)	$\Delta S^\circ_{HE}(T_S)$† (e.u.)
Soybean inhibitor + trypsin → complex	440 (*83*)	349	(*60*)
Subtilisin inhibitor + subtilisin monomer → complex	240 (*71*)	339	41
Subtilisin inhibitor + α chymotrypsin monomer → complex	270 (*84*)	343	(*43*)
FK506 + FKBP-12 → complex	260 (*73*)	289	60
			Average: 50 ± 10

*Values of T_S calculated from values of ΔC°_{assoc}, ΔS°_{assoc} in references cited in column 2. †$\Delta S^\circ_{HE}(T_S)$ calculated from Eq. 2 with values for ΔA_{np} from Table 2. Values of $\Delta S^\circ_{HE}(T_S)$ in parentheses are calculated from Eq. 3 for systems lacking structural data to evaluate ΔA_{np}.

Table 4. Entropic contributions where folding is coupled to association: predictions of the number of residues participating in the folding transition.

Process (structural references)	T_S* (K)	$\Delta S^\circ_{HE}(T_S)$† (e.u.)	ΔS°_{rt}‡ (e.u.)	ΔS°_{other}§ (e.u.)	$\mathfrak{R}^{th}$‖	$\mathfrak{R}^{str}$¶
Angiotensin II (*48*) + antibody Fab 131 (*85*) → complex (*85*)	312	68	−50	−18	3	8
Avidin (*49*) + biotin → complex (*49*)	291	85	−50	−35	6	9
S-peptide (*47*) + S-protein (*47*) → ribonuclease S (*86*)	253#	145	−50	−95	17	15
L-tryptophan + apo Trp R monomer (*11*) → complex (*12*)	263	127	−50	−77	14	17**
Holo Trp R dimer (*11*) + *trp* operator DNA → complex (*12*)	319	147	−50	−97	17	16
2 GR DBD (*13*) + DNA → complex (*14*)	308	285	−100	−185	33	40
3 glucagon (*81*) → trimer (*87*)	271	364	−100	−264	47	48–72
4 melittin (*82*) → tetramer (*88*)	313	477	−150	−327	58	104
2 arc repressor (*78*) → dimer (*77*)	289	525	−50	−475	85	80–92
2 λ cro repressor (*80*) → dimer (*79*)	287	620	−50	−570	102	120

*References for data used to calculate T_S are the same as those for ΔC°_{assoc} in Table 2. †$\Delta S^\circ_{HE}(T_S)$ evaluated from Eq. 2 with values for ΔA_{np} from Table 2. ‡Table 3. §Eq. 5. ‖Eq. 6. Propagated uncertainties in $\mathfrak{R}^{th}$ increase from ±15 percent for λ cro repressor to ±50 percent for angiotensin II, and are typically ±25 percent. ¶$\mathfrak{R}^{str}$ represents the difference between the number of residues folded in the crystal structure of the complex and the number of residues observed to be folded in the free species by NMR, x-ray, or CD as referenced in column 1. #T_S estimated from values of ΔC°_{assoc} (273) and ΔS°_{assoc} (273) obtained from the temperature dependence of ΔC°_{assoc} given in (*47*), based on the assumption that S protein is completely native at 273 K. **Number of residues folded in the complex based on the NMR structure.

estimates of ΔC^{o}_{assoc} and T_S for all protein-DNA interactions that have been studied by van't Hoff or calorimetric analysis and the corresponding values of ΔS^{o}_{HE} calculated from Eq. 3 (*51*). The estimate of ΔS^{o}_{rt} in Table 3 is assumed to be appropriate for these interactions. In addition, entries in Table 5 for which the effect of salt concentration on binding constants has been determined include an estimate of the temperature-independent entropy change due to the polyelectrolyte effect [ΔS^{o}_{PE} (*52*); relative to a 1 M salt concentration reference state] at the salt concentration where ΔC^{o}_{assoc} and T_S were determined.

Summation of contributions from $\Delta S^{o}_{HE}(T_S)$, ΔS^{o}_{rt}, and ΔS^{o}_{PE} (Table 5) indicates the existence in general of an additional large entropic term ΔS^{o}_{other} that opposes site-specific protein-DNA association:

$$\Delta S^{o}_{assoc}(T_S) = 0 = \Delta S^{o}_{HE}(T_S) + \Delta S^{o}_{rt} + \Delta S^{o}_{PE} + \Delta S^{o}_{other} \quad (7)$$

By analogy with the analysis of protein-ligand interactions discussed above, we propose that ΔS^{o}_{other} (Eq. 7) results primarily from folding or other conformational changes in the protein or DNA (or both) on binding. When ΔS^{o}_{other} is large in magnitude, as is generally the case, we propose that it reflects local (and in some cases global) folding transitions in the protein that are coupled to binding. Thermodynamic estimates of the number of residues that fold upon binding ($\mathfrak{N}^{th}$) calculated from ΔS^{o}_{other} by Eq. 6 are given in Table 5. By analogy with the protein-ligand interactions in Table 4, we propose that the specific DNA sequence serves as a template for local folding transitions in the protein (driven by binding free energy). Where $\mathfrak{N}^{th}$ is small, binding is likely to occur via a rigid body association.

High resolution structures have been determined for uncomplexed trp repressor protein (TrpR) in solution (*11*) and in the crystalline state (*53*), and for the specific DNA complex (*12*) in the crystalline state. As such TrpR provides a key test of the prediction of $\mathfrak{N}^{th}$ from ΔS^{o}_{other}. If the value of ΔS^{o}_{other} (−94 e.u.) for binding of TrpR to its operator is analyzed in terms of local folding, we find that ~16 residues fold when the two subunits of TrpR contact DNA. This thermodynamic estimate of $\mathfrak{N}$ is consistent with the current interpretation of NMR studies of the uncomplexed TrpR: The D helices (eight residues each) of both monomers appear disordered in solution (*11*). Both D helices are folded in the crystal structure of the TrpR dimer-operator complex, thereby creating part of the site for recognition of the operator (*12*). Although these helices are ordered in the crystal structure of TrpR, Sigler and coworkers refer to helix D–turn–helix E as a "flexible reading head" due to the high crystallographic B factors of residues in this region and the observation that conformational differences between different crystal forms of TrpR occur here (*54*). The agreement between $\mathfrak{N}^{th}$ and $\mathfrak{N}^{str}$ from NMR experiments is consistent with the observation that the predicted and experimental values of the heat capacity change for DNA binding agree only when a folding transition in helix D is included in the surface area calculation (*55*) (Table 2).

The solution structure of the DNA binding domain of glucocorticoid receptor (GR DBD) has been determined by NMR (*13*), and the solid-state structure of the complex of GR DBD with two specific half sites spaced four base pairs apart has been determined by x-ray crystallography (*14*). Uncomplexed GR DBD exists as a monomer up to mM concentrations (*13*). However, in its DNA complex, GR DBD exists as a dimer, with one monomer bound specifically and the other nonspecifically (*14*). Observed conformational differences between the free and bound monomers provide a rationale for why dimerization is linked to binding: The second Zn finger, which provides the dimer interface in the complex, contains an alpha helix and beta sheet which are not detected in the free state by NMR (*13, 14*). These structural studies suggest that binding of the GR DBD to a DNA half site is coupled both to local folding and to dimerization, consistent with the large negative ΔC^{o}_{assoc} in Tables 2 and 5. If we model the DNA binding equilibria as involving a 2:1 stoichiometry, on the basis of both thermodynamic (*56, 57*) and structural studies (*13, 14*), we calculate that approximately 18 residues per monomer fold on binding, a thermodynamic estimate which agrees with the size of the second Zn finger "loop" in the monomeric free state. As in the case of TrpR, this agreement is consistent with the observation in Table 2 that predicted and experimental values of ΔC^{o}_{assoc} agree when dimerization and folding equilibria are included in the overall DNA-binding equilibria.

Structural information in the solid state is available for both free and bound states of λ cro repressor protein (λ cro) and the catabolite activator protein (CAP). In both cases, changes in 4° structure of the protein are coupled to DNA binding. One monomer of λ cro appears to rotate approximate-

Table 5. Entropic contributions to protein-DNA interactions. ND, not determined.

Complexation process	$-\Delta C^{o}_{assoc}$ (cal mol^{-1} K^{-1})	T_S* (K)	$\Delta S^{o}_{HE}(T_S)$† (e.u.)	ΔS^{o}_{rt}‡ (e.u.)	ΔS^{o}_{PE}§ (e.u.)	$-\Delta S^{o}_{other}$‖ (e.u.)	$\mathfrak{N}^{th}$¶	$\mathfrak{N}^{str}$
λ cro repressor + nonspecific DNA	0 (*66*)							
λ cI repressor + O_R1	0 (*91*)							
λ cro repressor + O_R3	360 (*66*)	339	63	−50	ND	~13	2	
Holo TrpR + *trp* operator	540 ± 100 (*55*)	319	139	−50	ND	~89	16	16 (*11, 12*)
CAP(cAMP)$_2$ + DNA consensus site	~680#	~344	~106	−50	24**	~80	~14	
LacR + O_1	900 ± 100††	312	258	−50	27‡‡	235	42	
LacR + O^{sym}	1300 ± 300 (*18*)	305	413	−50	25§§	388	69	
Eco RI + -GAATTC- site	1500 ± 800 (*18*)	309	450	−50	34‖‖	434	78	
GalR + O_E	2300 (*58*)	296	824	−50	14¶¶	788	140	
RNAP + λP_R promoter	2400 ± 300 (*19*)	322	587	−50	54##	591	105	
MntR + O_{mnt}	3100 ± 620***	292	1170	−50	11†††	1131	200	
2 GR DBD + DNA	1000 ± 200 (*56*)	308	305	−100	ND	205	37	40 (*13, 14*)

*T_S calculated from ΔC^{o}_{assoc} and ΔS^{o}_{assoc} from the reference cited in column 2. †Eq. 3. ‡Table 3. §The contribution to the observed entropy change from the polyelectrolyte effect (PE) was estimated from the expression $\Delta S^{o}_{PE} = -0.88\, z\, R \ln [MX]$, where the value of z represents the apparent valence of the binding site on the ligand (0.88 z is the number of cations thermodynamically released on DNA binding) (*20, 89*). [MX] is the salt concentration at which the temperature dependence of the observed binding constant was measured. For systems where the temperature dependence was investigated in buffers containing both monovalent (M^+) and divalent cations (M^{2+}), ΔS^{o}_{PE} was estimated using the expression $\Delta S^{o}_{PE} = -z\, R[0.88 \ln[M^+] + \ln\{0.5[1 + (1 + 12[M^+]^{-1.7}[M^{2+}])^{0.5}]\}]$ (*90*). ‖Eq. 7. ¶Eq. 6. Uncertainty in $\mathfrak{N}^{th}$ is primarily determined by the error in ΔC^{o}_{assoc}. Propagated uncertainties range from 20 to 50 percent. . #Value estimated numerically from equilibrium constants at only three temperatures (*92*). Since the range of this data does not include the temperature where the equilibrium constant is predicted to be a maximum, the uncertainty in this ΔC^{o}_{assoc} probably exceeds that of the others tabulated. **z = 8 (*92*). ††Data of (*93*), analyzed according to (*18*). ‡‡z = 8 (*93*). §§z = 9 (*59*). ‖‖z = 8 (*94*). ¶¶z = 4, obtained fitting the data of (*58*) to Eq. 3 in (*90*) to correct for the presence of 6 mM M^{2+} (Mg^{2+} and Ca^{2+}) in buffers used to determine the salt-dependence of K_{obs}. ##z = 20 (*95*). ***Data of (*96*) analyzed according to (*18*). †††z = 5 (*96*).

ly 40° relative to the other (6); the relative orientations of the NH_2- and COOH-terminal domains in each CAP subunit change from being "open" to "closed" (2, 5). ΔC°_{assoc} for λ cro binding was determined calorimetrically in a concentration regime where it exists as a stably folded dimer. Hence the small magnitude of ΔS°_{other} (~18 e.u.) indicates the absence of extensive coupled local folding, and may reflect the entropic cost of 4° subunit rearrangements. For CAP binding, our estimate of ΔS°_{other} is significantly larger in magnitude than ΔS°_{other} for λ cro, possibly because of extensive coupled changes in DNA conformation upon binding CAP (2). Alternatively, the magnitude of ΔS°_{other} may indicate that key regions of the CAP binding interface are disordered in solution, even though (as in TrpR) these regions are folded in the crystal structure of the free protein.

In contrast to the above cases, values of ΔS°_{other} for association of both gal and mnt repressor proteins with their DNA operators are so large in magnitude that they cannot correspond simply to local folding or restricted 4° rearrangements. The Gal repressor (GalR) undergoes a monomer-dimer equilibrium in the concentration range where DNA binding was investigated (58). This coupled process may contribute to the large magnitude of ΔC°_{assoc} and hence to ΔS°_{other}. Little is known about equilibria coupled to the DNA binding of mnt repressor (MntR). On the basis of the thermodynamic data and the homology to arc repressor protein, we propose that ΔC°_{assoc} for MntR reflects coupled monomer-dimer-tetramer association and folding of the monomer (Tables 2 and 4) as part of DNA binding in the concentration range where ΔC°_{assoc} was investigated.

For binding of Eco RI endonuclease (Eco RI), LacR, and RNA polymerase (RNAP), the magnitude of ΔS°_{other} falls in a midrange of the values (Table 5). Although the binding thermodynamics of these proteins have been characterized extensively in solution (18–20, 59), complete high-resolution structural information does not exist for any of these systems. Values of ΔS°_{other} either reflect exceedingly large changes in 4° structure or substantial (but local, that is, nonglobal) folding upon binding. For LacR, $\mathfrak{R}^{th}$ is less than the number of residues in the two DNA-binding domains, and may indicate folding of part of these regions on binding to operator. Possibly this folding is more extensive in binding to the symmetric operator than to the nonsymmetric wild-type (O_1) operator, thereby accounting for the apparent differences in ΔC°_{assoc} and ΔS°_{other} for binding to these sites. For dimeric Eco RI, $\mathfrak{R}^{th}$ is comparable to the size of the two inner and outer arms that wrap around the DNA (3). In the case of RNAP, the conformational transitions predicted by ΔS°_{other} possibly represent both local folding of key recognition regions of the sigma subunit (60) as well as 4° conformational changes in RNAP which are thought to occur in the process of opening one to one-and-a-half helical turns of DNA at the transcription start site (21).

Structural evidence for coupled conformational changes. High-resolution structural data in solution and in the crystal provide evidence for folding and 4° conformational changes coupled to site-specific binding in other systems where the thermodynamics of binding have not yet been extensively investigated. For example, changes in 4° structure of Eco RV endonuclease similar to those observed for λ cro and CAP are observed by comparison of free and bound crystal structures (4). In addition to this 4° conformational change, folding of two short loops in Eco RV appears to be coupled to DNA binding. One loop (residues 68 to 71), which contacts the minor groove and the sugar-phosphate backbone, forms a β-turn in the presence of both cognate and noncognate DNA, but is poorly ordered in the free protein (4). Another loop (residues 182 to 187), which is ordered and interacts with the base pairs in the major groove in the cognate complex, is disordered in both the uncomplexed protein and in the complex with noncognate DNA (4).

In addition, folding transitions upon site-specific DNA binding are observed for the NH_2-terminal domains of λ cI repressor (λ cI) and of *Antennapedia* (Antp) homeodomain, and for loops of TrpR and GCN4 transcriptional activator protein. Both λ cI (8) and Antp homeodomain (7) have flexible NH_2-terminal regions in solution which order upon DNA binding. In the low-temperature crystal structure of its DNA complex, the NH_2-terminus of λ cI wraps around the DNA, forming hydrogen bonds with base pairs in the major groove and the sugar-phosphate backbone of the consensus half site (8). Interestingly, the NH_2-terminus of the symmetry related monomer does not appear ordered in the nonconsensus DNA half site, reminiscent of the lack of ordering of the loop in Eco RV in the major groove of the noncognate DNA. The NH_2-terminal domain of Antp homeodomain folds to make contacts with the minor groove: NMR studies indicate that residues 1–6 in the free protein do not have a defined structure, but nuclear Overhauser effects exist between arginine 5 and the sugar in the minor groove of the complex (7).

As discussed above, formation of well-defined structure appears to accompany the DNA binding of TrpR. Site-specific DNA binding of the yeast transcription factor GCN4 is characterized by a similar transition. In the absence of DNA, CD and NMR studies indicate that the basic region which is necessary for binding DNA is largely disordered (9, 61). In the crystal structure of its complex with the specific AP-1 DNA site, this region forms an α helix and makes extensive contacts to base pairs and to the sugar-phosphate backbone of the major groove (10). For GCN4 this extensive folding transition should be readily detected in ΔC°_{assoc} and ΔS°_{other}.

Role of adaptability in site-specific recognition. All of the protein folding (2°, 3°) and the 4° conformational transitions observed in proteins as part of site-specific binding to DNA create key parts of the contact interface. In these cases, recognition does not result from a simple alignment of rigid, complementary surfaces on the protein and DNA. Instead, these protein–nucleic acid interactions are generally described by an "induced fit" model at the level of either 2°, 3°, or 4° structure. This induced fit is analogous to Koshland's hypothesis for enzyme-substrate recognition (17), as pointed out for GCN4 by Frankel and Kim (15) and by Alber (16). The implications of ligand-induced conformational changes for enzyme-substrate specificity have been and continue to be actively debated (21, 62). However, the thermodynamic consequences of induced fit for the specificity of protein-DNA recognition are only beginning to be considered (1, 10, 15, 16, 63). For protein-DNA interactions in Table 5, conformational changes resulting from folding appear to be the only way to interpret the excess in $|\Delta C^\circ_{assoc}|$ and the deficit in ΔS°_{assoc}, as compared to the values expected for rigid body associations (18).

If regions of a DNA-binding protein are unfolded in the absence of a specific DNA site, the formation of well-defined 2° or 3° structure upon binding requires expenditure of free energy, and therefore must be driven by binding free energy (that is, the formation of a more extensive complementary interface, burying more macromolecular surface and releasing more water and ions). Formation of a highly stable specific complex involves maximizing complementarity and favorable interactions while minimizing juxtaposition of noncomplementary surfaces (18, 20, 64). If key regions of proteins are unstructured in solution and change conformation upon binding to a specific DNA sequence, then it is likely that the final conformation will be a function of the DNA sequence and that both the driving force (binding free energy) and the driven process (folding) will be functions of DNA sequence (16). Observed differences in protein conformation in specific and nonspecific complexes with DNA provide examples of this proposal for sequence-depen-

dent induced fit. When bound to noncognate DNA, regions of λ cI (8) and Eco RV (4) apparently do not order, and the α-helical content of the basic region of GCN4 is less than when bound to a specific site (65). Nonspecific binding of λ cro is not accompanied by a detectable ΔC°_{assoc} (66), indicating that little nonpolar surface is buried either in the interface or in coupled local folding transitions (Eq. 1). In general, however, nonspecific binding may be coupled to folding transitions to create a favorable coulombic or other nonspecific interface.

Only limited thermodynamic or structural analyses of effects of single- (or multi-) site variants in protein-DNA complexes have been performed. (In vivo characterizations of DNA sequence variants by expression assays and in vitro characterizations by measurements of relative binding affinity at one set of solution conditions are more extensive; these have generally been interpreted in terms of additivity.) Adaptability of LacR in specific DNA binding is suggested by thermodynamic experiments in which various thermodynamic functions [for example, the (salt) derivative of the binding constant; the van't Hoff enthalpy] exhibit global changes in response to an (O^c) base substitution at a single site (22). Adaptability of Eco RI is suggested by altered patterns of phosphate ethylation interference and large energetic penalties associated with binding single base pair substitution variants (23). In these cases, the observed thermodynamic or structural behavior has been interpreted in terms of induced fit or adaptability of the protein, so that changes in DNA sequence result in global changes in structure and contacts in the protein-DNA interface.

Applications of the model. Our use of model compound transfer data and application of principles derived from protein folding thermodynamics lead us to propose that site-specific DNA recognition by proteins is in general not a simple rigid body association whereby a specific site is recognized by a static "lock and key" interaction which simply matches up protein and DNA surfaces containing a pre-existing pattern of hydrogen bond donors and acceptors. Instead, as indicated by the large negative ΔC°_{assoc} which accompany these reactions and by our dissection of the observed entropy change, site-specific binding in general involves coupled changes in the 2°, 3°, and/or 4° structure of the protein. Although in some cases the tightly packed networks of functional groups on amino acid side chains and DNA bases seen in protein-DNA interfaces (1) may result from simple "docking" of preexisting complementary surfaces, a mounting body of structural and thermodynamic data argue that the solution structure of the complex differs strikingly from those of the uncomplexed species. All these conformational changes are driven by binding free energy, and therefore are examples of "induced fit." Of the conformational changes in the protein and DNA observed to date, folding transitions in the protein appear to be the most prevalent, to have the most characteristic thermodynamic signature (ΔC°, ΔS°), and to possess the most possibilities for adaptability or induced fit.

REFERENCES AND NOTES

1. C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **61**, 1053 (1992); S. C. Harrison and A. K. Aggarwal, *ibid.* **59**, 933 (1990).
2. S. C. Schultz, G. C. Shields, T. A. Steitz, *Science* **253**, 1001 (1991).
3. J. M. Rosenberg, *Curr. Opin. Struct. Biol.* **1**, 104 (1991); Y. Kim, J. C. Grable, R. Love, P. J. Greene, J. M. Rosenberg, *Science* **249**, 1307 (1990); J. McClarin *et al.*, *ibid.* **234**, 1526 (1986).
4. F. K. Winkler *et al.*, *EMBO J.* **12**, 1781 (1993).
5. I. T. Weber and T. A. Steitz, *J. Mol. Biol.* **198**, 311 (1987).
6. R. G. Brennan, S. L. Roderick, Y. Takeda, B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8165 (1990); R. G. Brennan, *Curr. Opin. Struct. Biol.* **1**, 80 (1991).
7. G. Otting *et al.*, *EMBO J.* **9**, 3085 (1990).
8. N. D. Clarke, L. J. Beamer, H. R. Goldberg, C. Berkower, C. O. Pabo, *Science* **254**, 267 (1991).
9. V. Saudek *et al.*, *Biochemistry* **30**, 1310 (1991).
10. T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell* **71**, 1223 (1992).
11. D. Zhao, C. H. Arrowsmith, X. Jia, O. Jardetzky, *J. Mol. Biol.* **229**, 735 (1993); C. H. Arrowsmith, J. Czaplicki, S. B. Iyer, O. Jardetzky, *J. Am. Chem. Soc.* **113**, 4020 (1991); C. H. Arrowsmith, R. Pachter, R. Altman, O. Jardetzky, *Eur. J. Biochem.* **202**, 53 (1991). A more quantitative evaluation of the degree of disorder is being performed (O. Jardetzky, personal communication).
12. Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
13. T. Härd *et al.*, *Science* **249**, 157 (1990); T. Härd *et al.*, *Biochemistry* **29**, 9015 (1990).
14. B. F. Luisi *et al.*, *Nature* **352**, 497 (1991).
15. A. D. Frankel and P. S. Kim, *Cell* **65**, 717 (1991).
16. T. Alber, *Curr. Biol.* **3**, 182 (1993).
17. D. E. Koshland Jr., *Proc. Natl. Acad. Sci. U.S.A.* **44**, 98 (1958).
18. J. H. Ha, R. S. Spolar, M. T. Record Jr., *J. Mol. Biol.* **209**, 801 (1989).
19. J.-H. Roe, R. R. Burgess, M. T. Record Jr., *ibid.* **184**, 441 (1985).
20. M. T. Record Jr., J.-H. Ha, M. Fisher, *Methods Enzymol.* **208**, 291 (1991).
21. T. Alber, thesis, Massachusetts Institute of Technology (1981); T. Alber, W. A. Gilbert, D. R. Ponzi, G. A. Petsko, *Mobility and Function in Proteins and Nucleic Acids* (Pitman, London, 1982), pp. 4–24.
22. M. C. Mossing and M. T. Record Jr., *J. Mol. Biol.* **186**, 295 (1985).
23. D. R. Lesser, M. R. Kurpiewski, L. Jen-Jacobson, *Science* **250**, 776 (1990).
24. P. L. Privalov, *Adv. Prot. Chem.* **33**, 167 (1979).
25. ——— and S. J. Gill, *ibid.* **39**, 191 (1988).
26. R. S. Spolar, J. R. Livingstone, M. T. Record Jr., *Biochemistry* **31**, 3947 (1992).
27. J. R. Livingstone, R. S. Spolar, M. T. Record Jr., *ibid.* **30**, 4237 (1991).
28. R. S. Spolar, J. H. Ha, M. T. Record Jr., *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8382 (1989).
29. K. P. Murphy, P. L. Privalov, S. J. Gill, *Science* **247**, 559 (1990).
30. P. L. Privalov and G. I. Makhatadze, *J. Mol. Biol.* **224**, 715 (1992); K. P. Murphy and S. J. Gill, *ibid.* **222**, 699 (1991).
31. J. M. Sturtevant, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2236 (1977).
32. R. L. Baldwin, *ibid.* **83**, 8069 (1986).
33. C. Tanford, *The Hydrophobic Effect* (Wiley, New York, 1980).
34. K. A. Dill, *Biochemistry* **29**, 7133 (1990).
35. The applicability of the "liquid hydrocarbon" model does not require that the protein interior be like hexane, for example. Instead, the thermodynamics of transfer from water to a pure liquid phase reference the interaction of a nonpolar surface with water to one involving only van der Waals contacts. Since the thermodynamics of transfer from water to gas phase or water to solid phase contain additional contributions from the thermodynamics of vaporization or fusion, we conclude that the "liquid hydrocarbon" model provides the most direct means of evaluating the thermodynamic consequences of removal of nonpolar macromolecular surface from water in the process of interest.
36. S. Miller, J. Janin, A. M. Lesk, C. Chothia, *J. Mol. Biol.* **196**, 641 (1987).
37. Using a different analysis, previous investigators have observed a similar entropic contribution to protein stability that scales with size (*24, 25, 29*).
38. Using a statistical thermodynamic analysis, Dill and colleagues [K. A. Dill, *Biochemistry* **24**, 1501 (1985); ———, D. O. V. Alonso, K. Hutchinson, *ibid.* **28**, 5439 (1989)] estimated that the backbone conformational entropy change is −2 to −3 e.u. per residue, relatively independent of temperature in the physiological temperature range.
39. Monte Carlo calculations on a subset of side chains [T. P. Creamer and G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5937 (1992)] yield an average value of −3.6 ± 1.7 e.u. per residue; observed rotamer distributions of all side chains in high-resolution protein crystal structures [S. D. Pickett and M. J. E. Sternberg, *J. Mol. Biol.* **231**, 825 (1993)] weighted by their average occurrence in proteins [T. E. Creighton, *Proteins* (Freeman, New York, 1983)] yields a value of approximately −3.5 e.u. per residue. Both values are based on the assumption that side chains adopt a single conformation in the native state.
40. Because the burial of polar surface contributes to ΔC° (Eq. 1), it must contribute to ΔS°_{other} over some temperature range. We do not detect a significant polar entropy change in our analysis and conclude that, in the temperature range of T_S in Tables 3, 4, and 5, contributions to ΔS°_{other} from changes in polar surface are not significant.
41. An analysis of protein vibrational modes indicates that the vibrational entropies of unfolded and folded states are large but similar, and hence that the change in vibrational entropy on folding is negligible [M. Karplus, T. Ichiye, B. M. Pettitt, *Biophys. J.* **52**, 1083 (1987)].
42. M. I. Page and W. P. Jencks, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1678 (1971).
43. A. V. Finkelstein and J. Janin, *Protein Engineering* **3**, 1 (1989); J. Janin and C. Chothia, *Biochemistry* **17**, 2943 (1978).
44. H. P. Erickson, *J. Mol. Biol.* **206**, 465 (1989).
45. Correspondence between this value and the prediction may be fortuitous because the ideal gas derivation neglects differences in hydration between the free reactants and the complex. Hence effects from translational and rotational entropy of released water are not accounted for in this derivation. (Release of water from nonpolar surface is presumably incorporated in ΔS°_{HE}.)
46. Although we expect that the backbone contribution to ΔS°_{c} for folding an internal loop is smaller in magnitude than for folding a terminal region of the same length, the analysis treats both types of regions equally.
47. R. Varadarajan, P. R. Connelly, J. M. Sturtevant, F. M. Richards, *Biochemistry* **31**, 1421 (1992).
48. K. P. Murphy, D. Xie, K. C. Garcia, L. M. Amzel, E. Freire, *Proteins: Struct. Funct. Genet.* **15**, 113 (1993).
49. O. Livnah, E. A. Bayer, M. Wilchek, J. L. Sussman, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5076 (1993).
50. CD measurements of tetrameric melittin in solution yield a mean residue ellipticity at 222 nM of −14,800 deg cm^2 dmol^{-1} at 25°C (*82*). Using the method of J. M. Scholtz, H. Qian, E. J. York, J. M. Stewart, and R. L. Baldwin [*Biopolymers* **31**, 1463 (1991)], we calculate that melittin is ~45 percent

α-helical under the solution conditions where thermodynamics of association were measured. In contrast, x-ray structural analysis indicates that melittin is 92 percent α-helical in the crystalline state (*88*). Tetramerization of melittin in solution buries less surface than that calculated based on the crystal structure, consistent with the observation (Table 2) that ΔC°_{assoc} predicted from Eq. 1 using the crystal structure to calculate ΔA_{np} and ΔA_p is twofold larger in magnitude than the experimental measurement of ΔC°_{assoc} in solution (*82*).

51. Use of Eq. 3 assumes that the ratio of nonpolar to polar surface buried in protein-DNA complexes is similar to that observed in the process of folding the globular proteins in Table 1 (63:37). Interfaces of the four protein-DNA complexes investigated by Sigler and coworkers bury ratios of nonpolar: polar surface ranging from 65:35 to 40:60 [Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature* **365**, 512 (1993)]. Since use of the lower ratio (40:60) would lead to an increase in the magnitude of ΔA_{np} predicted from ΔC°_{assoc}, use of Eq. 3 appears reasonable and conservative.
52. When a protein binds a DNA site, coulombic interactions with positively charged amino acid side chains reduce the local phosphate charge density of DNA, collapsing the pre-existing ion gradients in the vicinity of DNA. The concomitant release of thermodynamically bound counterions from a region of high concentration to bulk solution results in a free energy contribution which becomes more favorable as the bulk salt concentration is reduced. This free energy contribution is proposed to be primarily entropic (*20*).
53. R. W. Schevitz, Z. Otwinowski, A. Joachimiak, C. L. Lawson, P. B. Sigler, *Nature* **317**, 782 (1985).
54. R.-G. Zhang *et al.*, *ibid.* **327**, 591 (1987); C. L. Lawson *et al.*, *Proteins* **3**, 18 (1988).
55. L. Jin, J. Yang, J. Carey, *Biochemistry* **32**, 7302 (1993).
56. T. Lundbäck, C. Cairns, J.-Å. Gustafsson, J. Carlstedt-Duke, T. Härd, *ibid.*, p. 5074.
57. T. Härd *et al.*, *ibid.* **29**, 5358 (1990).
58. M. Brenowitz, E. Jamison, A. Majumdar, S. Adhya, *ibid.*, p. 3374.
59. J. H. Ha, M. W. Capp, M. D. Hohenwalter, M. Baskerville, M. T. Record Jr., *J. Mol. Biol.* **228**, 252 (1992).
60. C. A. Gross, M. Lonetto, R. Losick, in *Transcriptional Regulation*, S. L. McKnight and K. R. Yamamoto, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992), pp. 129–176.
61. M. A. Weiss *et al.*, *Nature* **347**, 575 (1990); M. A. Weiss, *Biochemistry* **29**, 8020 (1990); K. T. O'Neil, J. D. Shuman, C. Ampe, W. F. DeGrado, *ibid.* **30**, 9030 (1991).
62. D. Herschlag, *Bioorg. Chem.* **16**, 62 (1988); C. B. Post and W. J. Ray Jr., in preparation.
63. A. D. Frankel, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11653 (1992); J. Kim, D. Tzamarias, T. Ellenberger, S. C. Harrison, K. Struhl, *ibid.* **90**, 4513 (1993).
64. M. T. Record Jr., in *Unusual DNA Structures*, R. D. Wells and S. C. Harvey, Eds. (Springer-Verlag, New York, 1988), pp. 237–249; ——— and R. S. Spolar, in *The Biology of Nonspecific DNA-Protein Interactions*, A. Revzin, Ed. (CRC Press, Ann Arbor, MI, 1990), pp. 33–69.
65. K. T. O'Neil, R. H. Hoess, W. F. DeGrado, *Science* **249**, 774 (1990).
66. Y. Takeda, P. D. Ross, C. P. Mudd, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8180 (1992).
67. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977); E. E. Abola, F. C. Bernstein, S. H. Bryant, T. F. Koetzle, J. Wang, *Crystallographic Databases-Information Content, Software Systems, Scientific Applications* (Data Commission of the International Union of Crystallography, Cambridge, 1987).
68. P. Alexander, S. Fahnestock, T. Lee, J. Orban, P. Bryan, *Biochemistry* **31**, 3597 (1992).
69. K.-S. Kim *et al.*, *Protein Science* **2**, 588 (1993).
70. V. V. Filimonov, W. Pfeil, T. N. Tsalkova, P. L. Privalov, *Biophys. Chem.* **8**, 117 (1978).
71. K. Takahasi and H. Fukada, *Biochemistry* **24**, 297 (1985).
72. K. Takahasi, J. L. Casey, J. M. Sturtevant, *ibid.* **20**, 4693 (1981).
73. P. R. Connelly and J. A. Thomson, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4781 (1992).
74. K. C. Aune, L. C. Goldsmith, S. N. Timasheff, *Biochemistry* **10**, 1617 (1971).
75. J. Sussman, in preparation.
76. J. Suurkuusk and I. Wadsö, *Eur. J. Biochem.* **28**, 438 (1972).
77. U. Obeysekare *et al.*, in preparation.
78. J. U. Bowie and R. T. Sauer, *Biochemistry* **28**, 7139 (1989).
79. D. H. Ohlendorf, D. E. Tronrud, B. W. Matthews, unpublished data.
80. Yu. V. Griko, V. V. Rogov, P. L. Privalov, *Biochemistry* **31**, 12701 (1992).
81. S. Formisano, M. L. Johnson, H. Edelhoch, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3340 (1977).
82. W. Wilcox and D. Eisenberg, *Protein Sci.* **1**, 641 (1992).
83. R. J. Baugh and C. G. Trowbridge, *J. Biol. Chem.* **247**, 7498 (1972).
84. H. Fukada, K. Takahashi, J. M. Sturtevant, *Biochemistry* **24**, 5109 (1985).
85. K. C. Garcia, P. M. Ronco, P. J. Verroust, A. T. Brünger, L. M. Amzel, *Science* **257**, 502 (1992).
86. E. E. Kim, R. Varadarajan, H. W. Wyckoff, F. M. Richards, *Biochemistry* **31**, 12304 (1992).
87. K. Sasaki, S. Dockerill, D. A. Adamiak, I. J. Tickle, T. Blundell, *Nature* **257**, 751 (1975).
88. T. C. Terwilliger and D. Eisenberg, *J. Biol. Chem.* **257**, 6016 (1982).
89. M. T. Record Jr., T. M. Lohman, P. deHaseth, *J. Mol. Biol.* **107**, 145 (1976).
90. M. T. Record Jr., P. L. deHaseth, T. M. Lohman, *Biochemistry* **16**, 4791 (1977).
91. K. S. Koblan and G. K. Ackers, *ibid.* **31**, 57 (1992).
92. R. H. Ebright, Y. W. Ebright, A. Gunasekra, *Nucleic Acids. Res.* **17**, 10295 (1989).
93. P. A. Whitson, J. S. Olson, K. S. Matthews, *Biochemistry* **25**, 3852 (1986).
94. B. J. Terry, W. E. Jack, R. A. Rubin, P. Modrich, *J. Biol. Chem.* **258**, 9820 (1983); L. Jen-Jacobson *et al.*, *ibid.*, p. 14638.
95. J.-H. Roe and M. T. Record Jr., *Biochemistry* **24**, 4721 (1985).
96. A. K. Vershon, S.-M. Liao, W. R. McClure, R. T. Sauer, *J. Mol. Biol.* **195**, 311 (1987).
97. We thank B. E. Raumann, R. T. Sauer, D. E. Tronrud, and B. W. Matthews for coordinates in advance of publication; J. L. Sussman for coordinates in advance of deposition in the Protein Data Bank; J. Carey for a copy of her manuscript in advance of publication; T. Alber, R. L. Baldwin, T. R. Cech, T. M. Lohman, O. Jardetsky, P. S. Kim, R. Raines, P. von Hippel, and the referees for their comments and critique of the manuscript; C. Bingman, J. Bond, A. Edison, H. Guttman, and K. Prehoda for assistance with computational aspects of this work; A. Steinberg of the Biochemistry Media Center for assistance with Figs. 1 and 2; J. L. Markley for the use of his IRIS 4D workstation; and S. Aiello for preparation of the manuscript. Supported by NIH grant GM23467.

23 August 1993; accepted 30 December 1993