Protein–DNA Interactions: Structure and Energetics

Secondary article

Tom K Kerppola, *Howard Hughes Medical Institute and University of Michigan Medical School, Ann Arbor, Michigan, USA*

Article Contents

DNA-binding proteins recognize specific DNA sequences by a combination of molecular interactions. Protein–DNA complex formation is frequently accompanied by conformational changes in one or both components of the complex. Protein–DNA interactions differ in several respects from most other ligand–receptor interactions in cells, and these characteristics place special requirements on the energetics and dynamics of protein–DNA interactions and explain many of the special properties of these complexes.

DNA-binding Specificity

The biological functions of various DNA-binding proteins require different degrees of sequence selectivity in DNA recognition. Proteins that function at all sites in the genome, such as those that participate in DNA replication, general recombination or DNA packaging, usually exhibit small differences in binding affinity at different nucleotide sequences. In contrast, proteins that function at only a small number of sites in the genome, such as those that participate in transcription regulation, site-specific recombination or DNA restriction, generally exhibit preferential binding to selected nucleotide sequences or structures. The ratio between the affinity of binding to specific DNA recognition sequences and to nonspecific sites (selectivity ratio) varies over a range of more than 100-fold for different DNA-binding proteins. There is no obligate relationship between DNA-binding affinity and sequence selectivity. Specific interactions can be weak and nonspecific interactions can be strong. To achieve the broad range of sequence selectivities and binding affinities, different proteins make use of different combinations of sequence-selective and nonselective interactions. (***See*** Protein–DNA Interactions)

The binding affinities of each protein at different DNA sequence elements form a continuum between the specific and nonspecific DNA binding affinities. Some proteins exhibit broad sequence selectivity, in which case single or multiple base pair substitutions in the optimal recognition sequence have only minor effects on binding affinity. Other proteins exhibit narrow sequence selectivity, in which case single base pair substitutions cause large reductions in binding affinity. The biological functions of proteins at different sites can be affected by differences in binding affinity. Nevertheless, natural regulatory elements often contain suboptimal recognition sequences, indicating that sites of intermediate affinity can also have important biological functions. The occupancy of sites of different affinities in the genome can vary as a function of the protein concentration. Thus, differences in binding affinity can tune the biological response to the strength of the signal through variations in protein concentration. Other factors that influence the distribution of proteins among different sites in the genome include interactions with other DNA-binding proteins and accessibility of the sites. (***See*** Protein–DNA Complexes: Specific; Protein–DNA Complexes: Nonspecific)

In contrast to most other ligand–receptor interactions, it is neither possible nor advantageous for DNA-binding proteins to exclude nonspecific binding sites from their recognition surfaces. Nonspecific DNA binding can facilitate the search for specific DNA-recognition sequences. This is due to the covalent linkage between specific and nonspecific binding sites on DNA. Many proteins can shift rapidly between adjacent nonspecific binding sites, and thereby slide along the DNA helix. This enables a protein to sample more binding sites in a given time than would be possible were it necessary for the protein to dissociate from DNA prior to transfer to a new site. Some proteins are also able to undergo direct transfer between two DNA molecules without the need to dissociate from DNA. These facilitated diffusion mechanisms make it kinetically possible for proteins to locate their specific binding sites embedded in a large excess of nonspecific sites.

Structural studies of protein–DNA complexes together with analysis of the energetics of protein–DNA interactions have provided a detailed understanding of the molecular basis of specific DNA recognition. The sequence selectivity of protein binding to DNA results from a combination of interactions with functional groups on the nucleotide bases ('direct read-out') and interactions with the deoxyribose and phosphate moieties as well as water molecules whose positions or conformations depend on the base sequence ('indirect read-out'). Proteins also make DNA contacts that are not affected by the nucleotide

sequence of the binding site. Such nonselective interactions increase the affinity of the interaction but reduce the selectivity ratio. Both direct and indirect read-out as well as nonselective contacts involve several types of molecular interactions, including hydrogen bonding, electrostatic interactions, the hydrophobic effect, and van der Waal's forces. These interactions are also affected by the steric fit between the protein and DNA surfaces. In some cases, DNA-binding proteins recognize base pairs that are outside the structurally defined contact interface. Recognition of such flanking base pairs may be mediated by long-range electrostatic interactions or by sequence-dependent variations in DNA structure. (***See*** Protein–Ligand Interactions: General Description)

The overall affinity of protein–DNA interaction is determined by the difference in free energy between the free components and the complex together with the solvent molecules associated with each state. The sum total differences in intermolecular and intramolecular interactions define the changes in enthalpy (ΔH) and entropy (ΔS) resulting from the protein–DNA interaction (see **Table 1**). These determine the change in Gibbs free energy (ΔG) caused by complex formation, and the affinity of the interaction (K_d) at each temperature. The enthalpy and entropy changes for DNA binding by many proteins have opposing effects on the free energy of complex formation. Formation of some protein–DNA complexes is enthalpy driven, whereas formation of the majority is entropy driven. Thus, protein–DNA complexes vary both in the molecular interactions that stabilize complex formation and in the thermodynamic consequences of those interactions. (***See*** Thermodynamics in Biochemistry; Binding Constants: Measurement and Biological Range; Protein–Ligand Interactions: Energetic Contributions and Shape Complementarity)

Hydrogen Bonding

The nucleotide base pairs present numerous functional groups that can serve as hydrogen bond donors or acceptors. These include groups on the bases and deoxyribose rings as well as the phosphates. Likewise, the peptide backbone as well as most amino acid side-chains can form hydrogen bonds. Hydrogen bonding can contribute to direct read-out through contacts with base pairs as well as to indirect read-out and nonselective interactions through contacts with deoxyriboses and phosphates. Hydrogen-bonding interactions can also be mediated by water molecules whose positions and orientations are fixed by simultaneous hydrogen bonding to both the protein and the DNA. Since many groups on proteins and DNA contain multiple hydrogen-bond donors and acceptors, hydrogen bonds frequently form networks that connect multiple amino acid residues and nucleotide base pairs. Since the length and geometry of a hydrogen bond are restricted, interactions within such hydrogen-bonding networks can be highly cooperative. (***See*** Nucleic Acids: General Properties; Water: Structure and Properties)

Contacts with the nucleotide bases generally require that the protein enter into the major or minor groove to access the edges of the stacked base pairs. The four base pairs present distinct arrangements of hydrogen-bond donors and acceptors in the major groove (**Figure 1**). Thus, the base pairs can be specifically recognized through direct hydrogen-bonding interactions within the major groove. Many secondary structural elements are well suited to position amino acid residues within the grooves of the DNA helix. The diameter of the α helix is similar to the contour of the major groove. Thus, the side-chains of amino acid residues within an α helix that passes through the major groove can

Table 1 Protein–DNA interactions

Free energy of protein–DNA complex formation

$\Delta G = \Delta H - T\Delta S$

where ΔG = change in Gibbs free energy; ΔH = change in enthalpy; ΔS = change in entropy; T = absolute temperature

Heat capacity

$C_p = \Delta H/\Delta T$

Electrostatic potential Φ at a distance r from point charge of ze with Debye-Hückel screening

$$\Phi = \frac{ze}{\varepsilon r}\exp(-\kappa r)$$

$$\kappa^2 = \frac{8\pi N_0 e^2}{1000\varepsilon kT} I$$

where e = charge on the electron; ε = extinction coefficient; N_0 = Avogadro's number; k = Boltzmann's constant; I = ionic strength

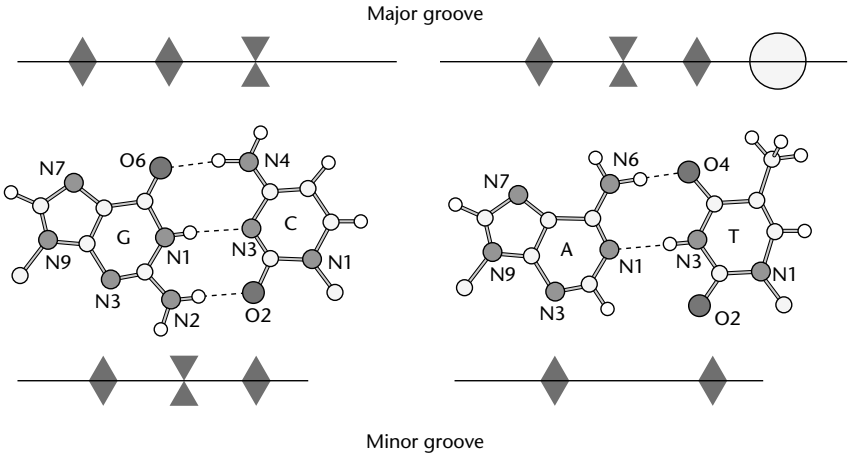

Figure 1 Functional groups available for direct recognition in the major and minor grooves on guanine (G)–cytosine (C) and adenine (A)–thymine (T) base pairs. Hydrogen bond acceptors are indicated by diamonds and hydrogen bond donors are indicated by hourglass shapes. The methyl group on thymine is indicated by a circle. Reproduced from Steitz (1990) © Cambridge University Press.

contact the edges of base pairs at the bottom of the major groove. Bases in the major groove can also be contacted by amino acid residues in β ribbons, at the edges of β sheets, within turns or in loops connecting elements of secondary structure. The arrangement of hydrogen-bond donors and acceptors in the minor groove is similar for G:C and C:G as well as for A:T and T:A base pairs (**Figure 1**). In addition, the minor groove is deeper and narrower than the major groove, and therefore the base pairs are less accessible. However, bases within the unperturbed minor groove can be contacted by amino acid residues in extended peptide conformations, β ribbons, β turns or the ends of helices. In addition, bending and unwinding of the DNA helix can make the minor groove more accessible to direct contacts by amino acid residues in other secondary structural elements. (***See*** Protein–Nucleic Acid Interaction: Major Groove Recognition Determinants)

Hydrogen-bonding interactions with deoxyribose and phosphate groups can contribute to both sequence-dependent as well as nonselective DNA binding. Since the strength of hydrogen bonding depends on the distance between and geometry of the donor and of acceptor, such interactions can sense differences in DNA conformation or flexibility. The base sequence affects both the average conformation and the dynamic flexibility of the DNA helix. Thus, hydrogen-bonding interactions that depend on the conformation of the DNA helix can contribute to the sequence-dependence of DNA binding. In addition, hydrogen-bonding networks that involve both nucleotide bases and deoxyriboses and phosphates can increase the energetic contribution of hydrogen bonding to sequence-selective DNA binding.

The amino acid residues that form hydrogen bonds to specific base pairs vary between different protein–DNA complexes. Thus, there is no universal code for protein–DNA recognition. However, some hydrogen-bonding interactions between individual amino acid residues and base pairs have been observed at high frequency. These include a bidentate interaction between the guanidium group of arginine with N7 and O6 of guanine, and interactions between the hydrogen bond donor and acceptor of glutamine with N6 and N7 of adenine (**Figure 1**). In addition, among structurally related DNA-binding proteins, differences in DNA binding specificity often correlate with specific amino acid substitutions. Thus, in the context of a particular protein fold, specific amino acid residues have dominant effects on the recognition of individual base pairs. Consequently, it is possible to predict the changes in DNA binding specificity caused by individual amino acid substitutions, and in some cases to design modified proteins with defined DNA-binding specificities. (***See*** DNA-binding Enzymes: Structural Themes)

Hydrogen-bonding interactions make a significant contribution to the enthalpy change (ΔH) of protein–nucleic acid complex formation. This enthalpy change reflects the net difference between the enthalpy of the hydrogen bonds formed and the enthalpies of the hydrogen bonds with water molecules and intramolecular bonds that are broken during complex formation. Hydrogen bonding also influences the entropy change (ΔS) of DNA binding by constraining both local and global vibrational, translational and rotational degrees of freedom of both the protein and DNA. In addition, changes in water structure caused by hydrogen bonding can both increase the entropy as a result of the release of water molecules from the hydrogen-bonded surfaces and decrease the entropy as a result of trapping of ordered water within the protein–DNA interface. The relative magnitudes of these factors differ among various protein–DNA complexes.

Electrostatic Interactions

Owing to the acidic character of the phosphate groups, DNA has high negative charge at physiological pH. Consequently, charge interactions are a major determinant of the affinity of protein–nucleic acid interactions. The major charged groups in DNA are the phosphates. Other potential sites of electrostatic interactions in DNA are the π orbitals of unstacked base pairs. Proteins contain both negatively (glutamate and aspartate) and positively (lysine and arginine) charged residues, as well as residues whose charges depend on their local environments (histidine, cysteine and tyrosine). The number and positions of these residues in the complex determine the contribution of electrostatic interactions to protein–DNA complex formation. The amino and carboxyl terminal ends of the peptide chain as well as a helix dipoles also influence the charge distribution of the complex. (***See*** Protein–Ligand Interactions: Molecular Basis)

Electrostatic interactions are generally considered to contribute primarily to the nonspecific DNA-binding affinity of proteins. However, since the strength of electrostatic interactions depends on the distance between the charged groups, electrostatic interactions with phosphates can also be affected by DNA structure and therefore by base sequence (Ramirez-Carrozzi and Kerppola, 2001a). Conversely, DNA structure can be affected by electrostatic interactions with charged residues in the protein. Positively charged groups can bend DNA towards the protein, whereas negatively charged groups can bend DNA away from the protein (Ramirez-Carrozzi and Kerppola, 2001a). The structure of DNA can also be affected by neutralization of phosphates on one face of the DNA helix. The resulting charge imbalance on opposite sides of DNA can cause the DNA to bend towards the uncharged surface. Thus, electrostatic interactions contribute to the interdependence between protein binding and DNA structure. (***See*** DNA Structure Changes Coupled to Protein Binding)

The strength of electrostatic interactions is affected by the concentration and valence of ions in solution. The ions have two distinct effects on electrostatic interactions between proteins and DNA. First, because of the high charge density of DNA, it is associated with counterions that partially neutralize the charge of the phosphates (Manning, 1978). The valence of the counterions is the major determinant of the extent of charge neutralization. Owing to the high charge density of DNA, the extent of counterion condensation is relatively insensitive to the concentration of ions in bulk solution. Interactions between charged groups on proteins and the phosphates releases these counterions from DNA. Thus, counterion condensation reduces the absolute enthalpy change (ΔH) of protein–nucleic acid interactions, but increases the entropy change (ΔS). The effect of salt concentration on the affinity of protein binding to DNA can be used to estimate the number of counterions displaced from DNA during complex formation. Second, the presence of ions in the medium separating the charged groups reduces the strength of their interaction through Debye–Hückel screening (see **Table 1**). The strength of charge interactions in the presence of dissolved ions decreases exponentially with the distance between the charged groups. Nevertheless, because of the high charge density of DNA, electrostatic interactions can have long-range effects on protein–DNA interactions and do not require direct contact between the charged groups. Thus, electrostatic interactions can mediate effects of cofactor binding to sites on DNA-binding proteins distal from the DNA contact surface on their binding affinities. (***See*** Protein–DNA Interactions: Polyelectrolyte Effects)

The electrostatic potential in the vicinity of proteins is also affected by the large difference in dielectric constants between the interior of proteins ($\varepsilon \approx 4$) and in aqueous solution ($\varepsilon = 80$). Therefore, the overall shape of the protein affects the electrostatic potential distribution (Honig and Nicholls, 1995). This effect can mediate electrostatic focusing where high electrostatic potential is concentrated in narrow clefts in a protein. The DNA contact surface of DNA-binding proteins typically has high positive electrostatic potential. This charge complementarity may help steer the protein into the appropriate alignment with DNA during the initial encounter. Electrostatic interactions are also a major determinant of nonspecific DNA binding and can facilitate localization of specific binding sites.

Solvation and the Hydrophobic Effect

Macromolecules in aqueous solution are associated with a large number of solvent molecules whose properties are affected by their interactions with the macromolecule. Protein–DNA interactions involve large changes in the organization of water near the macromolecular surfaces. The water structure contributes to both the entropy and enthalpy changes of protein–DNA complex formation. These changes are caused by differences between water–macromolecule and water–water interactions. Water associated with polar surfaces is hydrogen-bonded to donor and acceptor groups on the surface. Removal of such water requires breaking of the hydrogen bonds and their replacement with hydrogen bonds to other water molecules. This desolvation of polar surfaces generally increases the enthalpy. In contrast, water associated with nonpolar surfaces has fewer hydrogen-bonding interactions. Transfer of such water into the bulk phase increases hydrogen–bonding interactions and reduces the enthalpy. (***See*** Cell Macromolecules; Cell Biophysics)

An even larger contribution to the energetics of many protein–DNA interactions is made by the entropy changes caused by the displacement of bound water. Water

molecules near nonpolar surfaces are more constrained in motion and orientation than water molecules in bulk solution. Thus, removal of nonpolar surfaces from solution through complex formation releases water molecules into bulk solution, resulting in higher entropy. This effect promotes the association of hydrophobic groups, and is therefore called the hydrophobic effect. The release of water is the major factor contributing to increased entropy (ΔS) during protein–DNA complex formation. In the majority of cases, this entropy increase is larger than the reduction in entropy caused by the decrease in the local and global vibrational, translational and rotational degrees of freedom of the macromolecule caused by complex formation. Thus, water is an important participant in protein–DNA complex formation. (***See*** Hydrophobic Effect)

The majority of specific protein–DNA interactions are characterized by large negative heat capacity changes (ΔC_P) of complex formation. The heat capacity represents the temperature-dependence of the enthalpy of a system (see **Table 1**). The change in heat capacity upon protein–DNA interaction is caused at least in part by the removal of nonpolar surfaces from contact with water. However, the changes in nonpolar surface area in structurally characterized protein–DNA complexes appear to be smaller than those predicted by the change in heat capacity. This apparent discrepancy has been explained either by postulating that DNA binding is coupled to other structural changes that reduce the exposed nonpolar surface area (Spolar and Record, 1994) or by invoking other factors that affect heat capacity, such as the vibrational and rotational degrees of freedom of water molecules in the protein–DNA interface (Ladbury *et al.*, 1994). The large increase in entropy and decrease in heat capacity accompanying DNA binding by many proteins indicate that protein–DNA complex formation is thermodynamically analogous to protein folding. (***See*** Protein Folding: Overview of Pathways)

The hydrophobic effect can contribute to sequence-selective DNA recognition. In the small number of complexes bound to nonspecific sites for which high-resolution structural information is available and in molecular dynamics simulations, the protein–DNA interface in nonspecific complexes is more hydrated. In addition, nonspecific DNA binding by several proteins causes smaller changes in heat capacity. This suggests that the nonpolar surfaces of such complexes are still associated with water molecules and that molecular motions at the interface are not constrained. In contrast, binding to specific sites causes a large reduction in heat capacity, consistent with a more rigid interface. Formation of a specific contact interface is coupled to the release of water molecules that remain bound in nonspecific complexes. The release of bound water is therefore a hallmark of specific complex formation and can contribute to sequence-specific DNA recognition.

Van der Waal's Forces and Steric Fit

Specific protein–DNA complex formation requires a close fit between the surfaces of the protein and DNA molecules. The close apposition of neutral atoms causes a correlation between their electron distributions, resulting in an attractive van der Waal's force. The magnitude of the force depends on the distance between the interacting surfaces and the area of their interaction. Thus, large neutral surfaces, such as thymine methyl groups on DNA are sites of potential van der Waal's interactions with uncharged side-chains on proteins. Van der Waal's interactions with deoxyribose rings can also contribute to indirect read-out or nonselective DNA binding. In some cases, disruption of the stacking of bases in the DNA helix allows contacts between the faces of the bases and hydrophobic residues, causing a large van der Waal's force. In general, van der Waal's forces between proteins and DNA are relatively small and influence mainly DNA binding specificity rather than binding affinity.

If atoms are brought into too close proximity, their electron distributions will overlap, causing a strong repulsive force. Thus, in order for proteins to interact with DNA it is necessary not only for them to make favourable contacts but also to avoid unfavourable ones. It is theoretically possible to create a specific protein–DNA interaction by combining nonselective favourable interactions with sequence-selective unfavourable interactions. Although this may not happen to the exclusion of favourable sequence-selective interactions, some of the close contacts in protein–DNA complexes confer sequence specificity by preventing interactions at other sites rather than by stabilizing the complex relative to the free components. Thus, unfavourable interactions at nonspecific binding sites are an important mechanism contributing to sequence-selective DNA recognition.

The overall free energy change of sequence-selective DNA binding reflects the balance between large favourable and unfavourable terms (**Figure 2**) (Jen-Jacobson, 1997). The magnitudes of both favourable and unfavourable terms are larger at specific binding sites owing to the more intimate interaction between the protein and DNA. The magnitudes of all terms at nonspecific binding sites are smaller, and the net change in free energy is smaller. Consequently, it is not possible to identify one type of interaction that is responsible for the sequence selectivity of DNA binding. The magnitudes of many of the unfavourable terms and their structural basis are poorly defined. Nevertheless, many of the favourable terms must be directly coupled to unfavourable terms, preventing the formation of protein–DNA complexes with only favourable interactions. (***See*** Protein–Ligand Interactions: Induced Fit)

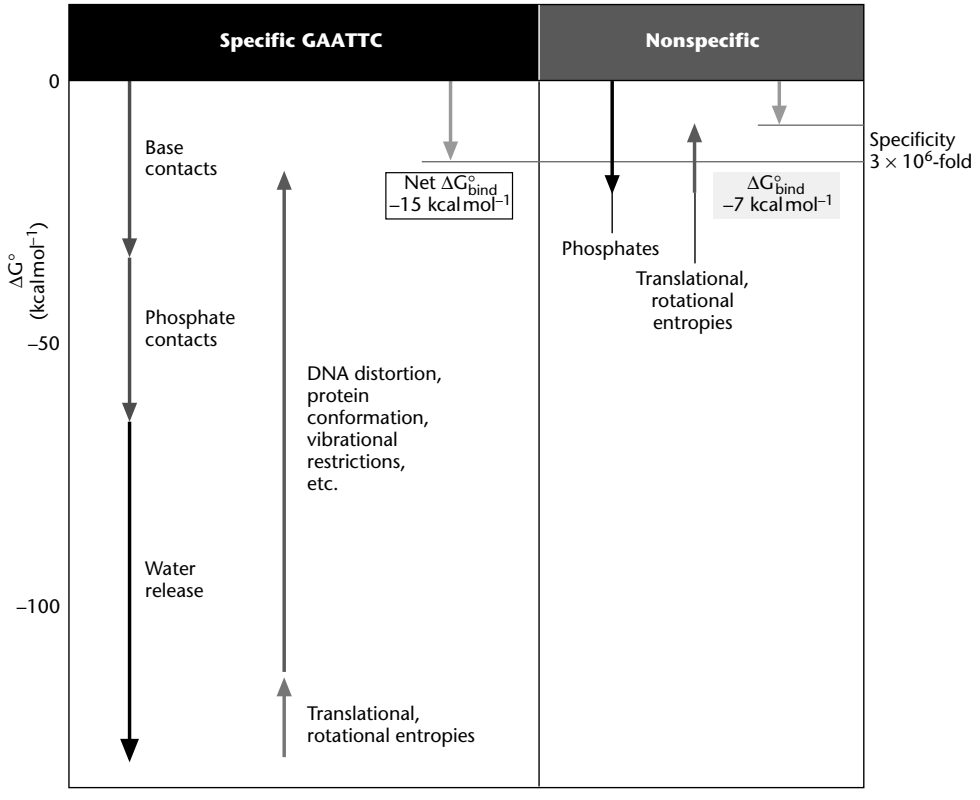

Figure 2 Estimated favourable (black) and unfavourable (grey) energetic terms that contribute to the overall stabilities of complexes formed by EcoRI at specific (left) and nonspecific (right) binding sites. The contribution of DNA distortion, protein conformational changes, vibrational restrictions etc. was calculated based on the difference between the other terms. The various favourable and unfavourable contributions to binding energy are interdependent. The difference between the specific and nonspecific binding affinities represents the specificity ratio. Reproduced from Jen-Jacobson (1997) © John Wiley & Sons, Inc.

Protein Conformation

The structures of many DNA-binding proteins are altered upon binding to DNA. These structural changes range from reorientation of side-chains to folding or refolding of the DNA binding domain. Many of the structural changes are sequence-dependent, and occur only upon binding to specific DNA recognition sites (Dlakić *et al*., 2001). Coupling of protein folding to specific DNA binding may be mediated by base contacts that are required to stabilize the folded conformation. Alternatively, protein folding may be induced by interaction with a specific DNA structure. The most common large-scale structural change induced by DNA binding is the stabilization of a helical conformation (Patel *et al*., 1990). The DNA-binding domains of many eukaryotic transcription factors are only partially folded in the absence of DNA, but they are induced to form an α helical DNA recognition interface upon contact with the specific DNA binding site. This α helical structural transition can be coupled to structural changes in other parts of the proteins. DNA binding can also result in unfolding of protein domains and disruption of intramolecular interactions within a protein. Such intramolecular interactions can regulate DNA-binding activity and other aspects of protein function. The thermodynamic consequences of the interdependence between DNA binding and protein folding remain to be elucidated. It is not known how the decrease in configurational entropy at specific versus nonspecific binding sites affects binding selectivity. Nevertheless, it is clear that DNA binding by many proteins is a dynamic process rather than a simple docking of rigid components. (***See*** Proteins: Fundamental Chemical Properties)

DNA binding also regulates the oligomerization state of many proteins. Most DNA-binding proteins bind palindromic recognition elements as dimers or higher-order oligomers. These dimers can be preformed in solution or the proteins may dimerize on DNA. Coupling of DNA binding to dimerization can effect cooperative DNA binding and a nonlinear response to variations in protein concentration. In the case of proteins that can form heterodimers, the sequence of the binding site can influence the choice of dimerization partner. Furthermore, the recognition sequence can influence the orientation of

binding by asymmetric heterodimers (Leonard and Kerppola, 1998). Thus, DNA binding frequently influences the quaternary structure of protein complexes even in cases where their tertiary structures are unaffected by DNA binding. DNA binding can also stabilize interactions between proteins by increasing their local concentrations. (***See*** Haemoglobin: Cooperativity in Protein–Ligand Interactions; Protein Quaternary Structure: Subunit–Subunit Interactions)

Changes in protein conformation upon DNA binding can influence the biological function of the protein. The DNA-binding site can function as an allosteric effector of the protein and alter its function. This may prevent nonproductive functions of the protein when not bound to DNA. In addition, the conformation of the protein can vary at different recognition sequences, allowing the possibility of differential regulation of protein function at different binding sites. One example of the regulation of protein function by the sequence of the binding site is the influence of the orientation of protein binding on interactions with other proteins bound to adjacent sites (Diebold *et al*., 1998; Ramirez-Carrozzi and Kerppola, 2001b). (***See*** Enzyme Activity: Allosteric Regulation)

DNA Conformation

DNA is not a passive substrate in the formation of protein–DNA complexes. The structure of DNA varies because of both sequence-dependent differences and protein-induced changes in conformation. The sequence-dependent variability includes variations in the alignment of adjacent base pairs, which cause bending and twisting of the DNA helix. Protein-induced changes can also include unstacking of base pairs and disruption of hydrogen bonding. These changes in DNA structure serve many purposes, including specific DNA recognition, assembly of multiprotein complexes, access to functional groups on DNA, and duplication and interpretation of the genetic information. (***See*** DNA Topology: Fundamentals)

The sequence-dependent structure and flexibility of DNA influence its interactions with proteins. The average structure of DNA is well described by the regular double helix proposed by Watson and Crick. However, there are marked departures from this average structure caused by both different base sequences and thermal fluctuations. Each oligonucleotide segment in DNA has a unique structure. The local differences in structure and flexibility contribute to sequence-selective DNA binding by proteins. These differences in structure and flexibility can be affected both by base sequence as well as by interactions with proteins and other ligands.

Many proteins induce dramatic changes in DNA structure. The most common are bending and twisting of the DNA helix. Three major mechanisms of protein-induced DNA bending have been described. First, proteins such as CAP and MATa1/α2 heterodimers contain a curved interaction interface that wraps the DNA partially or completely around the protein. These interaction interfaces are formed by different combinations of hydrogen bonding and electrostatic, hydrophobic, and van der Waal's interactions. Second, proteins such as the TATA binding protein (TBP) and LEF-1 insert large hydrophobic amino acid side-chains between the base pairs in the minor groove. This causes partial unstacking of the base pairs and kinking of the DNA helix towards the major groove at the position of the inserted side-chain. Third, proteins such as Fos/Jun heterodimers and SRF bend DNA through electrostatic interactions. Positively charged amino acid residues bend DNA towards the residues, whereas negatively charged residues bend DNA away from the residues (Leonard *et al*., 1997). Members of the same protein family can induce distinct DNA bends and even bend DNA in opposite directions (Kerppola and Curran, 1991). (***See*** DNA Structure Changes Coupled to Protein Binding; SRY and DNA-bending Proteins)

In order for a protein to induce DNA bending, the gain in free energy from protein interactions with the bent DNA conformation must exceed the unfavourable free energy of bending the DNA. The sequence of the DNA-binding site can influence the extent of DNA bending by some proteins, whereas other proteins appear to bend all sites equally. DNA bending may therefore be a necessary condition for DNA binding by some proteins and it can affect the sequence selectivity of DNA binding by others.

The bendability of DNA depends on its nucleotide sequence. Thus, protein-induced DNA bending can contribute to sequence-specific DNA recognition. Furthermore, since DNA bendability can be altered through protein binding, DNA binding by proteins that induce DNA bending can be regulated by other DNA-binding proteins.

Changes in DNA twist are required to align sites on DNA that are separated by a nonintegral number of helical turns. Untwisting of DNA also facilitates unhindered access to the minor groove and is necessary for separation of the strands prior to DNA replication and transcription. Untwisting frequently accompanies DNA bending as in the complex formed by TBP at the TATA box. Large changes in twist caused by the binding of multiple proteins or other ligands can alter the writhe of topologically closed DNA. Topoisomerases regulate the superhelical density of DNA, which can influence the DNA-binding affinity of proteins that alter twist. Changes in the superhelical density of DNA also influence the energy requirement for strand separation. Thus, changes in DNA twist can both influence DNA-binding selectivity as well as prime DNA for functions that require further alteration of its structure. (***See*** DNA Topology: Supercoiling and Linking; Topoisomerases)

Access to the functional groups that mediate hydrogen bonding between the two strands as well as enzymatic reactions that use DNA as a substrate require dramatic

changes in DNA structure. Many enzymes that modify or repair DNA can flip out nucleotide bases from the DNA helix. A large hydrophobic amino acid side-chain is often inserted into the DNA helix to stabilize the flipped conformation. The transcription and replication complexes must unwind the DNA helix and separate the two strands. The prokaryotic RNA polymerases accomplish this through a multistep process that involves protein binding to the displaced single strands. Consequently, despite its apparent uniformity, DNA is a dynamic molecule that must assume different conformations in order to fulfil its biological functions. Proteins that interact with DNA have evolved the ability to modify its structure as well as to recognize structural variations in the DNA helix. (***See*** Base Flipping; Prokaryotic DNA-binding Proteins)

Multiprotein Complexes

Individual DNA-binding proteins do not operate in isolation but act in concert with many other proteins that influence their functions. This cooperativity is essential for specific recognition of unique sites within the genomes of eukaryotic organisms. The sizes of these genomes exceed the selectivity ratio of even the most highly selective DNA-binding proteins by several orders of magnitude. The selectivity ratios for eukaryotic DNA-binding proteins are generally no higher than those for prokaryotic proteins. Instead, multiple DNA-binding proteins function together to increase selectivity through cooperative interactions. Selective DNA binding is also favoured by sequestration of part of the genome into forms of chromatin that are accessible to only a subset of DNA-binding proteins. (***See*** Heterochromatin and Euchromatin; Nucleosomes: Structure and Function)

Cooperative DNA binding by unrelated DNA-binding proteins is generally mediated by stabilization of complexes on DNA rather than through interactions prior to DNA binding. Thus, although closely related DNA-binding proteins often form stable oligomers in the absence of DNA, unrelated proteins usually do not bind as a preformed complex. Exceptions include the proteins that are components of transcription and replication complexes that function as stable assemblies. Most regulatory proteins cooperate with different proteins at different regulatory elements. Interactions that take place after DNA binding allow proteins to cooperate with many other proteins without interference from competing interactions. The interactions among different proteins may require conformational changes induced by DNA binding, or the affinities of the protein–protein interactions may be too low to occur in the absence of DNA. In some cases, interactions between DNA-binding proteins can occur in the absence of DNA. Such interactions can link signalling pathways that influence one protein with those that influence the other. Interactions between DNA-binding proteins and proteins that do not bind to DNA independently can recruit non-DNA-binding proteins to specific sites on DNA. Cooperative functions between DNA-binding proteins can also involve joint recruitment of such proteins to DNA. Thus, cooperative DNA binding is an important mechanism contributing to the specificity of biological regulatory mechanisms.

Cooperative DNA binding is generally mediated by protein–protein interactions. The juxtaposition of interaction surfaces often requires a specific positioning of binding sites within regulatory regions. Interactions between proteins that bind to separate sites on DNA frequently require DNA bending. In addition, asymmetric protein complexes must generally bind in a specific orientation to allow juxtaposition of the molecular surfaces that mediate the interaction. Multiprotein complex formation therefore frequently requires conformational changes in proteins and DNA. Thus, protein–protein interactions can be modulated by changes in DNA bending or protein conformation. (***See*** Protein–Protein Interactions)

The structural basis of cooperative DNA binding has been defined for a small number of multiprotein complexes. In these complexes, a relatively small interaction interface with little water-excluded surface area is formed. In the Fos–Jun–NFAT1 complex at the ARRE2 site, the DNA is bent and the conformations of the proteins are altered (Chen *et al*., 1998). Furthermore, the orientation of Fos-Jun heterodimer binding is constrained by interaction with NFAT1 (Diebold *et al*., 1998). Formation of the MCM1–Matα2 complex at the STE6 site is also facilitated by DNA bending (Tan and Richmond, 1998). The interaction interface is formed by refolding of a region of Matα2 that is helical in the absence of MCM1 to form a strand within a β sheet formed by MCM1. In the GABPα/β complex, there are no direct contacts between GABPβ and DNA, although the interaction stabilizes DNA binding by the complex relative to GABPα alone. Cooperative DNA binding may therefore be mediated by an allosteric mechanism that stabilizes DNA binding via long-range interactions. Each multiprotein complex that has been structurally characterized forms a unique interaction interface. Changes in both protein and DNA conformation have been observed in some of the complexes. Further studies of the structural basis and energetics of multiprotein complex formation are necessary to define the various mechanisms whereby DNA binding proteins cooperate to carry out their biological functions. (***See*** Protein–DNA Interactions)

References

Chen L, Glover JNM, Hogan PG, Rao A and Harrison SC (1998) Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **392**: 42.

Dlakić M, Grinberg AV, Leonard DA and Kerppola TK (2001) DNA sequence-dependent folding determines the divergence in binding specificities between Maf and other bZip proteins. *EMBO Journal* **20**: 828–840.

Diebold RJ, Rajaram N, Leonard DA and Kerppola TK (1998) Molecular basis of cooperative DNA bending and oriented heterodimer binding in the NFAT1-Fos-Jun-ARRE2 complex. *Proceedings of the National Academy of Sciences of the USA* **95**: 7915.

Honig B and Nicholls A (1995) Classical electrostatics in biology and chemistry. *Science* **268**: 1144–1149.

Jen-Jacobson L (1997) Protein–DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers* **44**: 153–180.

Kerppola TK and Curran T (1991) DNA bending by Fos and Jun: the flexible hinge model. *Science* **254**: 1210–1214.

Ladbury JE, Wright JG, Sturtevant JM and Sigler PB (1994) A thermodynamic study of the trp repressor-operator interaction. *Journal of Molecular Biology* **238**: 669–681.

Leonard DA and Kerppola TK (1998) DNA bending determines Fos-Jun heterodimer orientation. *Nature Structural Biology* **5**: 877.

Leonard DA, Rajaram N and Kerppola TK (1997) Structural basis of DNA bending and oriented heterodimer binding by Fos and Jun. *Proceedings of the National Academy of Sciences of the USA* **94**: 4913–4918.

Manning GS (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quarterly Review of Biophysics* **11**: 179–246.

Patel L, Abate C and Curran T (1990) Altered protein conformation on DNA binding by Fos and Jun. *Nature* **347**: 572–575.

Ramirez-Carrozzi VR and Kerppola TK (2001a) Long-range electrostatic interactions influence the orientation of Fos–Jun binding at AP-1 sites. *Journal of Molecular Biology* **305**: 411–427.

Ramirez-Carrozzi VR and Kerppola TK (2001b) Dynamics of Fos–Jun–NFAT1 complexes. *Proceedings of the National Academy of Sciences of the USA* **98**: 4893–4898.

Spolar RS and Record MT Jr (1994) Coupling of local folding to site-specific binding of DNA proteins. *Science* **263**: 777–784.

Tan S and Richmond TJ (1998) Crystal structure of the yeast MAT-alpha2/MCM1/DNA ternary complex. *Nature* **391**: 660.

Further Reading

Batchelor AH, Piper DE, de la Brousse FC, McKnight SL and Wolberger C (1998) The structure of GABPα/β: an ETS domain–ankyrin heterodimer bound to DNA. *Science* **279**(5353): 1037–1041.

Honig B and Nicholls A (1995) Classical electrostatics in biology and chemistry. *Science* **268**: 1144–1149.

Jen-Jacobson L (1997) Protein–DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers* **44**: 153–180.

Kerppola TK (1998) Transcriptional cooperativity: bending over backwards and doing the flip. *Structure* **6**: 549–554.

Manning GS (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quarterly Review of Biophysics* **11**: 179–246.

Martin AM, Sam MD, Reich NO and Perona JJ (1999) Structural and genetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease. *Nature Structural Biology* **6**: 269–277.

Patel L, Abate C and Curran T (1990) Altered protein conformation on DNA binding by Fos and Jun. *Nature* **347**: 572–575.

Steitz TA (1990) Structural studies of protein–nucleic acid interaction: the sources of sequence-specific binding. *Quarterly Reviews of Biophysics* **23**(3): 205–280.

von Hippel PH and Berg OG (1989) Facilitated target location in biological systems. *Journal of Biological Chemistry* **264**(2): 675–678.