

23.2. Protein–ligand interactions

BY A. E. HODEL AND F. A. QUIOCHO

23.2.1. Introduction

There are currently over a thousand unique protein-bound ligands described in the Protein Data Bank, illustrating the enormous variety of small molecules with which proteins interact. These ligands serve as cofactors in protein-mediated reactions, substrates in these reactions, and elements that maintain or alter protein structure or macromolecular assembly. The specific binding of small molecules to proteins is a primary means by which living systems interact and exchange information with their environment. The atomic details of protein–ligand interactions are often quite similar to the intramolecular interactions observed within a protein molecule. Examples of all the various non-covalent interactions described in Parts 20 and 22, such as hydrogen bonds, van der Waals forces and other electrostatic phenomena, are observed between proteins and their small-molecule ligands.

Through the diverse interactions observed between proteins and their ligands, a few fundamental patterns of recognition emerge. In general, ligand binding requires that the protein partially or fully sequesters the ligand from the solvent. This demands that the energy of interaction between the protein and the ligand must be strong enough to overcome the interactions between both species and the solvent as well as the translational and rotational entropy which is lost upon fixing the orientation of the ligand relative to the protein. The protein achieves this level of interaction by presenting a binding site that is complementary to the ligand both in shape and electrostatic functionality. Beyond this generalization, each ligand has its own unique and complicated story. Rather than attempt to summarize the enormous subject of protein–ligand interactions in a comprehensive manner, we will instead illustrate several of the unique interactions observed between proteins and other molecules.

23.2.2. Protein–carbohydrate interactions

The interaction between proteins and carbohydrates provides a prototypic example of how proteins specifically recognize small organic ligands. Protein-mediated recognition of carbohydrates is crucial in a diverse array of processes, including the transport, biosynthesis and storage of carbohydrates as an energy source, signal transduction through carbohydrate messengers, and cell–cell recognition and adhesion (Rademacher *et al.*, 1988). Many aspects of protein–carbohydrate recognition are observed in the interactions of proteins with other small-molecule ligands. For example, carbohydrate recognition is largely conferred through a combination of hydrogen bonding and van der Waals interactions with the protein (Quiucho, 1986). These interactions are generally presented in a binding site that is highly complementary to the target ligand in shape and functionality. Carbohydrate-binding proteins also occasionally employ ordered water molecules and bound metal ions to facilitate ligand recognition (Quiucho *et al.*, 1989; Vyas, 1991; Weis & Drickamer, 1996). There are also many examples of ‘induced fit’ recognition of carbohydrates, where the binding of the ligand induces a conformational change in the protein. This phenomenon of ligand-induced conformational change was observed in one of the first ligand–protein interactions visualized by X-ray crystallography, namely the binding of a carbohydrate substrate to lysozyme (Blake *et al.*, 1967). Since these early studies, a wide variety of carbohydrate–protein structures have been determined, and a number of general themes have emerged from this continually active field (Vyas, 1991). These themes are further applicable in the general study of protein–ligand recognition.

Two general classes of carbohydrate-binding modes have been observed in their complexes with proteins (Quiucho, 1986; Vyas,

1991). The first group of proteins completely sequester the carbohydrate from the surrounding solvent. These proteins, including the periplasmic proteins and the catalytic site of glycogen phosphorylase, tend to have a high affinity for their ligand ($K_d \approx 10^{-6}$ – 10^{-7}). The high affinity of these proteins can be attributed to the entropic gain of desolvating the protein and carbohydrate surfaces as well as the exceptionally high degree of functional complementarity in the ligand-binding site. The periplasmic proteins nearly saturate the hydrogen-bonding potential of their carbohydrate ligands (Quiucho, 1986). The second group of proteins bind their ligands in shallow clefts in the solvent-exposed protein surface. This mode of binding, observed in the lectins, lysozyme and the storage site of glycogen phosphorylase, generally has a lower binding constant than the first class of binding proteins ($K_d \approx 10^{-3}$ – 10^{-6}). Some members of this second class of proteins, such as the lectins, are able to increase the affinity for their ligands dramatically by clustering a number of low-affinity sites through the oligomerization of the polypeptides (Weis & Drickamer, 1996). The specific arrangement of lectin oligomers allows them to discriminate between large cell-surface arrays of polysaccharides with high affinity and selectivity.

23.2.2.1. Carbohydrate recognition at the atomic level

An early review of protein–carbohydrate interactions revealed several atomic level interactions that continue to appear ubiquitously in the structures of protein–carbohydrate complexes (Quiucho, 1986). A generic cyclic sugar in mono- or oligosaccharides appears to be recognized as a disc that displays two flat non-polar surfaces surrounded by a ring of polar hydroxyls. Proteins recognize these features by hydrogen bonding to the ring of polar hydroxyls while stacking flat aromatic side chains against the non-polar disc faces.

Cooperative hydrogen bonding, where the hydroxyl group of a carbohydrate participates as both a donor and acceptor of hydrogen bonds, is often observed in the direct interactions between proteins and carbohydrate ligands (Fig. 23.2.2.1). The sp^3 -hybridized oxygen atom of a carbohydrate hydroxyl may act as both an acceptor of two hydrogen bonds through the two lone pairs of electrons as well as a donor of a single hydrogen bond. Cooperative

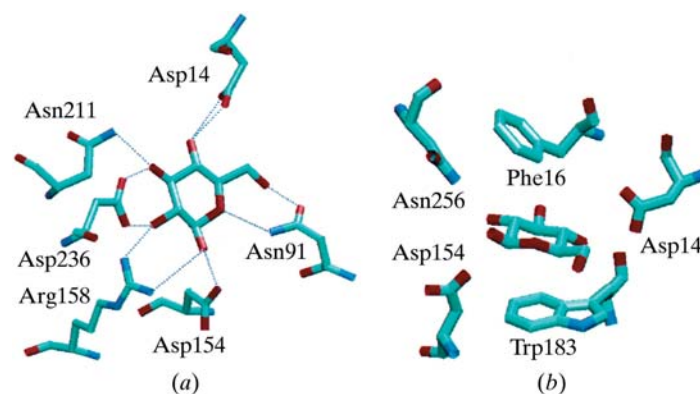


Fig. 23.2.2.1. The atomic recognition of carbohydrates by a protein. The environment of glucose bound to the galactose/glucose binding protein is shown (Vyas *et al.*, 1988). (a) The ring of hydrogen bonds around the polar edge of the sugar molecule. Note the ‘bidentate’ hydrogen bonds with Asp236, Arg158 and Asn91. As the hydroxyl groups 1, 2 and 3 simultaneously donate and accept hydrogen bonds from protein side chains, they are involved in ‘cooperative’ hydrogen bonds. (b) Two aromatic protein side chains, Phe16 and Trp183, ‘stack’ on the non-polar faces of the carbohydrate.

23. STRUCTURAL ANALYSIS AND CLASSIFICATION

Table 23.2.3.1. *Metal ions associated with proteins*

Metal ion	Concentration in blood plasma (mM)	Common cofactors	Hard/soft classification	Common coordination number and geometry	Preferred ligand atom
Na ⁺	138		Hard	6	O
K ⁺	4		Hard	8	O
Ca ²⁺	3		Hard	8	O
Mg ²⁺	1		Hard	6	O
Fe	0.02	Haem	Intermediate	6 (octahedral)	N
Zn ²⁺	0.02		Intermediate	4 (tetrahedral), 6 (octahedral)	N, S
Cu ²⁺	0.015		Soft	4 (tetrahedral)	S
Co ²⁺	0.002		Hard	6 (octahedral)	O
Mn ²⁺	0		Hard	6 (octahedral)	O
Ni ²⁺	0		Intermediate	6 (octahedral)	N
Mo	0	Pterin	Intermediate	6 (octahedral)	S
W	0	Pterin	Intermediate	6 (octahedral)	S
V	0			5 (trigonal bipyramidal)	

hydrogen bonding generally follows a simple pattern in which the carbohydrate hydroxyl accepts a hydrogen bond from a protein amide group while simultaneously donating a hydrogen bond to a protein carbonyl oxygen. Hydrogen bonding to protein hydroxyl groups is observed only infrequently. This pattern is thought to be a result, in part, of the entropic cost of fixing a freely rotating protein hydroxyl group while simultaneously fixing the ligand hydroxyl group. Amides and carbonyls are usually fixed in a planar geometry and thus do not require as much energy to compensate for their loss of entropy in ligand binding.

The vicinal hydroxyl groups of carbohydrates provide an ideal geometry for the formation of 'bidentate' hydrogen bonds, where the pair of hydroxyls interacts with two functional groups of a single amino-acid side chain or the main-chain amide groups of two consecutive residues (Fig. 23.2.2.1). These interactions occur when the adjacent carbohydrate hydroxyls are either both equatorial, or one is equatorial and the other axial. The interatomic distance for the carbohydrate hydroxyl oxygens is ~ 2.8 Å in these cases, allowing for a bidentate interaction with the planar side chains of aspartate, asparagine, glutamate, glutamine and arginine. Bidentate hydrogen bonds have not been observed for consecutive axial hydroxyls where the oxygen–oxygen distance increases to ~ 3.7 Å.

Carbohydrates often present a disc-like face of non-polar aliphatic hydrogen atoms which proteins recognize through the use of aromatic side chains. The protein aromatic groups are 'stacked' on the flat face of the carbohydrate, thus generating both specificity and binding energy through van der Waals interactions. Tryptophan, the aromatic amino acid with the largest surface area and highest electronegativity, is the most common side chain employed in van der Waals 'stacking' with carbohydrates. The infrequent use of aliphatic groups in the binding of the non-polar carbohydrate faces suggests that the aromatic moieties are employed in a specific manner. The electron-rich electron clouds of the aromatic side chains may provide a strong electrostatic interaction with the aliphatic carbohydrate protons that could not be satisfied by protein aliphatic groups. The anionic character of aromatic side chains is observed in a number of protein–intramolecular (Chapter 22.2) and protein–ligand interactions (see below).

23.2.3. Metals

Metal ions provide a number of important functions in their diverse and ubiquitous interactions with proteins. The most common

function for a protein-bound metal ion is the stabilization and orientation of the protein tertiary structure through coordination to specific protein functional groups. In addition to this structural role, metal ions are also often directly involved in enzyme catalysis and protein function. Examples of these functions include redox reactions, the activation of chemical bonds and the binding of specific ligands. Myoglobin, the first protein structure determined by X-ray crystallography, specifically binds molecular oxygen through an iron ion of a haem cofactor. Myoglobin provides a prototypic example of a protein and a metal ion providing a unique and specific functionality through their combination.

23.2.3.1. *Metals important in protein function and structure*

A number of metals are relatively abundant and available in living systems (Table 23.2.3.1) (Glusker, 1991). The most common ions include sodium, potassium, magnesium and calcium. Along with these ions, a large variety of trace metals are also found coordinated to proteins. The structures of protein complexes with some of these trace ions, including iron, zinc and copper, have been studied extensively for some time (Glusker, 1991). More recently, the structures of protein complexes with more unusual ions, such as nickel, vanadium and tungsten, have been determined (Volbeda *et al.*, 1996).

Specificity in the interactions between proteins and metal ions is conferred through each ion's preference for the coordinating atoms and the geometry of the binding site. All four of the more common metals, *i.e.* sodium, potassium, magnesium and calcium, are classified as 'hard' metals, referring to the polarizability of the electron cloud of the ion. The nucleus of a hard metal has a relatively tight hold on the surrounding electrons. These ions lack easily excitable unshared electrons and have a low polarizability. The interactions between these metals and their ligands tend to have the character of ionic interactions rather than the more covalent nature preferred by the 'soft' metals. In general, the hard metals prefer to coordinate with hard acids, such as the oxygen atoms of hydroxyls, carbonyls and carboxyls.

The soft metals have a high polarizability, large ionic radius and several unshared valence electrons. They generally prefer to coordinate with soft acids, such as the thiol and thiol ether groups of cysteine and methionine. The loosely held valence electrons of soft metals tend to favour partially covalent π -bonding with their coordinated ligands. These outer-shell electrons can be donated to the empty outer orbitals of the ligand atom. The partially covalent nature of these bonds yields more stable complexes than the ionic

23.2. PROTEIN-LIGAND INTERACTIONS

complexes of the hard metals. This partial covalent bond also polarizes the ligand coordinated to the metal and can thus activate adjacent atoms to nucleophilic attack.

A large number of the transition metals, including zinc and iron, form ions that have intermediate polarizability with regard to hard and soft metals. These ions mainly prefer nitrogen ligands like the imidazole side chain of histidine or the central nitrogens of the haem cofactor.

The geometry of the metal-binding site in a protein depends on a combination of the radial size of the ion as well as the polarizability of the metal. The number of coordinating ligands around the metal is primarily correlated with the relative size of the ion, where as many anions as possible are packed around the cationic metal without leaving any cavities (Orgel, 1966). This leads to a relatively simple correlation between the ratio of the radii of the cation and the anion ($r_{\text{cation}}/r_{\text{anion}}$) with the coordination number. Beyond this simple geometric constraint, the coordination number is also influenced by the repulsion between the closely packed anion ligands. This repulsion can be tempered by the distortions in the cation's electron cloud, leading to a dependency between the coordination number and the polarizability of the metal ion. Table 23.2.3.1 gives the most common coordination numbers and geometries for the listed metal ions. For a more comprehensive description of possible coordination geometries, see Glusker (1991).

A short example of the diversity of metal functions in protein complexes is found in a comparison between the calcium-binding proteins calmodulin and staphylococcal nuclease. Calmodulin functions in signal transduction by binding to a wide variety of proteins in a calcium-dependent manner. In the absence of calcium, calmodulin adopts a conformation where two loosely folded domains are connected by a flexible α -helix analogous to two balls tied together by a string. In the presence of Ca^{2+} , each of the two domains of calmodulin binds to a single metal ion. The binding of Ca^{2+} to the two calmodulin domains induces a large conformational change in the protein, which confers a high affinity for peptide ligands. Crystallographic studies show that the two calcium-bound domains form a clamp that closes on the target peptide ligand (Meador *et al.*, 1995). Thus, in this case, the metal ion plays an indirect role as a structural element in the protein function.

In the case of staphylococcal nuclease, calcium binding appears to play a more direct role in the catalytic function of the protein. A Ca^{2+} ion binds at the active site and coordinates with protein side chains, water molecules and the substrate phosphate group. The addition of calcium affects the nuclease reaction both in the binding of the substrate and directly in the catalytic step. Although calcium increases the K_m of the nucleic acid substrate, this effect can be reproduced with a large number of other metal ions (Tucker *et al.*, 1979). The effect on catalysis, however, is specific to Ca^{2+} ions. In a proposed mechanism, Ca^{2+} directly contributes to catalysis by activating a water-derived hydroxide ion for nucleophilic attack on the phosphorus atom of the nucleic acid backbone (Cotton *et al.*, 1979).

23.2.4. Protein-nucleic acid interactions

23.2.4.1. The DNA double helix

DNA provides one of the more compelling protein 'ligands' for biophysical study, as the sequence-specific binding of proteins to the DNA double helix mediates the interaction between the environment surrounding the living cell and the information 'programmed' into the cell within its genome. A classic example of such a process is the response of the bacteria *Escherichia coli* to

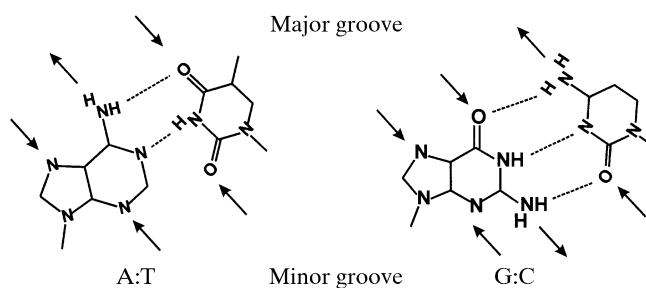


Fig. 23.2.4.1. A schematic diagram of the base pairs of DNA showing the hydrogen-bonding groups which may be used in the sequence-specific recognition of DNA. The major groove is at the top of the figure and the minor groove at the bottom. Arrows point towards hydrogen-bond acceptors and away from donors.

the nutrients in the surrounding media through the regulation of gene expression. A simple case of this interaction is found in the biosynthesis of the amino acid tryptophan. The transcription of the genes necessary for the synthesis of tryptophan is suppressed when tryptophan is present in the environment. This process is mediated by the tryptophan-dependent sequence-specific binding of the trp repressor protein to the *trp* operon within the genes encoding the metabolic enzymes (Joachimiak *et al.*, 1983). In the absence of tryptophan, the affinity of the aporepressor for the *trp* operon is dramatically reduced. Thus, when tryptophan is not available in the environment, transcription of the biosynthetic genes proceeds. In mammalian cells, the analogous process is observed in the activation of gene expression through hormones, cytokines and other stimuli.

Although DNA has often been considered to be a long, nearly featureless cylindrical double helix, proteins have evolved with exquisite specificity for their cognate DNA sequences. This apparent contradiction can be reconciled with the acknowledgement of two recently appreciated properties of DNA (Harrington & Winicov, 1994). First, the local structure of DNA is actually highly variable and dependent on the specific sequence of the base pairs in the helical ladder. Second, the DNA double helix is a relatively soft structure that is easily deformed into concerted bends, kinks and other distortions. DNA-binding proteins thus recognize their cognate sequences both by utilizing the unique local structure of the double helix and by inducing distortions into the helix which facilitate recognition.

The most intuitive features of the double helix that are important in sequence-specific recognition are the unique surfaces presented by the bases in the helix grooves. DNA is primarily found in a B-form helix that presents a wide, accessible major groove and a deep, narrow minor groove. An analysis of the arrangement of hydrogen-bonding functional groups presented by DNA bases (Fig. 23.2.4.1) suggests that the sequence-specific recognition of the DNA helix is best facilitated through the major groove, where each of the four possible base-pair combinations present unique hydrogen-bonding patterns (Steitz, 1990). The majority of sequence-specific DNA-binding proteins of known structure appear to utilize this direct readout of the major groove by inserting a portion of an α -helix, a two-stranded β -hairpin, or even a peptide coil which presents complementary hydrogen-bonding arrangements with the DNA bases (Pabo & Sauer, 1992; Steitz, 1990). The narrow surface of the minor groove presents some characteristic hydrogen-bonding patterns; however, the absolute identity of each base pair is ambiguously represented in these patterns (Fig. 23.2.4.1). The similar position of hydrogen-bonding groups in the minor groove would make it hard to distinguish AT base pairs from TA base pairs and GC base pairs from CG base pairs. Although there are proteins that recognize DNA through the minor groove,

23. STRUCTURAL ANALYSIS AND CLASSIFICATION

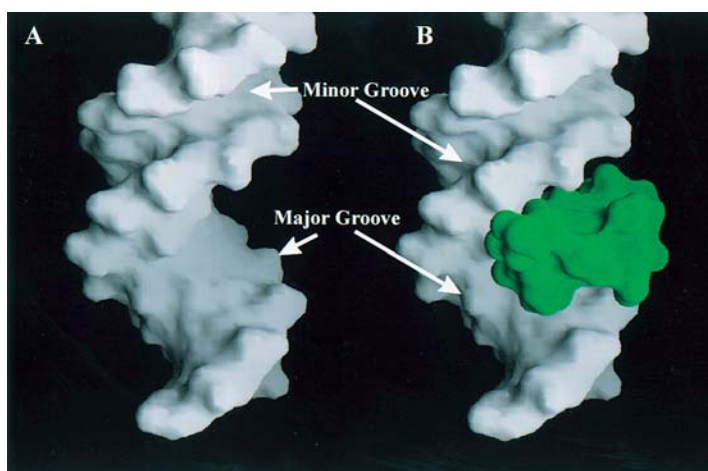


Fig. 23.2.4.2. (a) A space-filling model of B-DNA showing the relative accessibility of the major and minor grooves. (b) A helix of the 434 repressor bound in the major groove of the helix, illustrating how the dimensions of a protein α -helix are compatible for reading the major groove of B-DNA (Shimon & Harrison, 1993).

such as the TATA-box binding protein, the recognition of their target is completed through dramatic distortion of the DNA helix through intercalation (see below).

α -Helices are the most frequently observed structural motif for recognition in the major groove of DNA (Pabo & Sauer, 1992). The overall shape and dimensions of the α -helix are geometrically suited for binding in the major groove of a B-DNA helix (Fig. 23.2.4.2). The exact orientations of helices in various protein–DNA complexes are quite variable. Most helices bind in the major groove at an angle of approximately 30 (15°) from the plane normal to the DNA helical axis (Fig. 23.2.4.3). However, the numerous variants to this rule would include the trp repressor/operator complex, where only the N-terminal end of the ‘recognition’ helix is inserted into the major groove (Otwinowski *et al.*, 1988). Interactions observed between these inserted elements and the DNA bases include the common direct hydrogen bond between the protein side chain and base, the less common hydrogen bond between the protein

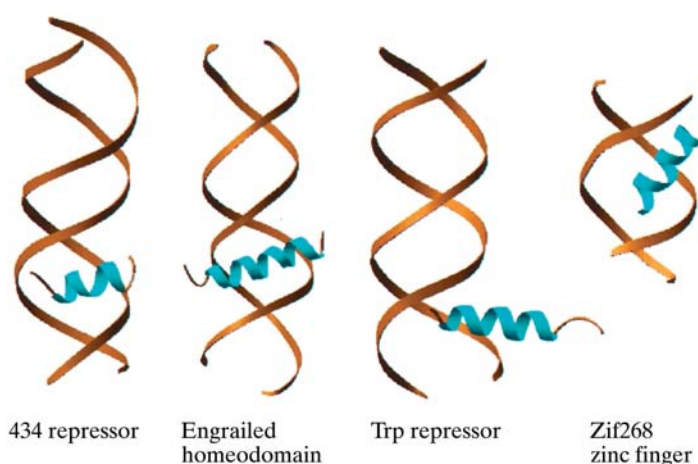


Fig. 23.2.4.3. A comparison of the orientations of α -helices bound in the major groove, taking examples from four DNA-binding proteins: the 434 repressor (Shimon & Harrison, 1993), the engrailed homeodomain (Kissinger *et al.*, 1990), the trp repressor (Otwinowski *et al.*, 1988) and the Zif268 zinc finger (Pavletich & Pabo, 1991). The DNA backbone is shown as a brown ribbon, whereas the protein helix is shown as a blue ribbon.

backbone and base, indirect but specific hydrogen bonding through water molecules, and hydrophobic interactions.

There appears to be no simple correlation between the primary sequence of the peptide segments which make specific base contacts and the DNA sequence that those segments recognize (Pabo & Sauer, 1992; Steitz, 1990). Examples of every polar protein side chain participating in specific hydrogen bonds with DNA bases have been observed, but each amino acid does not show any preference for any one particular base. What is observed is that conserved residues within families of DNA-binding proteins tend to make conserved base-specific interactions in DNA–protein complexes. Strikingly, this subset of interactions which are conserved within protein families include cooperative hydrogen bonding reminiscent of the pairs of hydrogen bonds often observed in carbohydrate–protein complexes. These interactions, which include the pairing of arginine with guanine and glutamine or asparagine with adenine, were predicted early on by Seeman *et al.* (1976).

Although the elements of protein structure in direct contact with the DNA bases play a prominent role in sequence specificity, these elements are not sufficient to impart the specificity of the DNA-binding protein. This statement is supported by the variety of orientations in which the ‘recognition’ helices bind to the major groove. The structural context of the recognition elements and the overall docking of the protein to the DNA helix play as important a role in specificity as the direct base interactions.

The contacts between the protein and the ribose–phosphate backbone of the DNA appear to be one of the more important aspects of the ‘indirect readout’ of the DNA sequence (Pabo & Sauer, 1992). On average, more than half of the interactions between protein and DNA in complex structures involve the backbone of the DNA helix. Thus, the sheer number of interactions suggests that these contacts serve an important function in recognition. Although several of the protein–DNA–backbone contacts observed involve salt bridges between the phosphates and basic protein side chains, these interactions are not as highly represented as one might expect. This could be a result of the high degree of flexibility inherent in the long side chains of arginine and lysine. Instead, examples of every basic and neutral residue and occasionally even acidic residue with some hydrogen-bonding potential interacting with the phosphate backbone have been observed. These contacts may contribute to specificity through two mechanisms. First, they can establish the exact orientation of the base-specific contacts relative to the ‘rungs’ in the phosphate backbone. Second, they may read the base sequence indirectly through sequence-specific backbone distortions or flexibility. There are numerous examples of DNA–protein complexes with highly distorted DNA helices. There is also evidence that certain DNA sequences inherently confer bends within the B-form helix. Thus, it is conceivable that protein interactions with the DNA backbone may confer specificity by selecting for a specific distorted conformation of the helix.

The most dramatic distortion of the DNA helix has been observed in DNA–protein complexes where the protein induces a kink or bend through the intercalation of the DNA helix at the minor groove (Werner *et al.*, 1996). Intercalation involves the insertion of a hydrophobic protein side chain into the helix, disrupting the stacking of two adjacent base pairs, and, in some cases, the side chain itself then stacks with one of the base pairs. Examples of this mode of binding include the complexes of the TATA-box binding protein (TBP), the PurR repressor and the human oncogene *ETS1* with their cognate DNA partners (Werner *et al.*, 1996). The *ETS1*–DNA complex provides the only current example of complete intercalation of the DNA extending from the minor groove to the major groove. A tryptophan side chain extends into the helix from the minor groove and stacks with one of the displaced base pairs. The remaining base pair contacts the ring system of the tryptophan

23.2. PROTEIN-LIGAND INTERACTIONS

edge in forming a pseudo-hydrogen bond between the indole hydrogens and the π -rings of the DNA bases. In *ETSI*, the deformation of the DNA helix resulting from protein intercalation results in the kinking of the helical axis from 45° to about 60° .

Examples of protein intercalation of the DNA helix from the major groove are found in proteins, such as the methyltransferases, that perform chemistry on the bases of the DNA. To perform their enzymatic function, these proteins must extract the target base from the DNA helix and 'flip' the base out into the enzyme active site (Cheng, 1995). The resulting void in the DNA is then filled by protein side chains that partially satisfy the hydrogen-bonding and van der Waals interactions that were broken when the target base was flipped. Although there are only a few known structures of DNA-protein complexes with extra-helical bases, base flipping is thought to be a relatively common feature of DNA-modifying enzymes.

23.2.4.2. Single-stranded sequence-nonspecific DNA-protein interactions

There have been a few reports of single-stranded DNA-protein complex structures, all of which involve the sequence-nonspecific recognition of DNA. In the binding of a tetranucleotide to the exonuclease active site of the DNA polymerase I Klenow fragment (Freemont *et al.*, 1988), extensive hydrogen-bonding interactions between the sugar-phosphate backbone and the protein are observed. This provides the most intuitive mechanism for sequence-nonspecific nucleic acid binding, where the protein simply recognizes the phosphate backbone of a single-stranded coil. The protein also appears to form a few hydrophobic interactions with the DNA bases; however, these interactions, which include the partial intercalation between two bases, are thought to be nonspecific.

The structure of replication protein A complexed with single-stranded DNA does not exhibit the intuitive nonspecific mechanism of recognition found in the Klenow fragment (Bochkarev *et al.*, 1997). In this structure, the DNA is extended with its bases splayed out over the surface of the protein. The bases form several pairwise stacking interactions that are interrupted by intercalating protein side chains. Contrary to the sequence-nonspecific nature of recognition, numerous hydrogen bonds are found between the protein and the bases of the DNA strand. These base-dependent contacts require that the protein-DNA interactions must be flexible and plastic in order to accommodate different base sequences.

23.2.4.3. RNA

Although RNA and DNA are chemically similar, RNA presents a much greater variety of shapes and surfaces compared to the relatively simple B-form helix of DNA. Generally single-stranded, RNA often forms secondary structures driven by the base pairing of complementary stretches of sequence within the same strand. The formation of base-paired regions can result in stem loops, bulges and helices which can further assemble into more complicated tertiary structures, such as that observed for transfer RNAs. Protein-mediated recognition of RNA often depends as much on the three-dimensional structure presented by these secondary structures as on the specific identity of the base sequence.

Very little information is currently available on the structural details of protein-RNA interactions (Nagai, 1996). Only a handful of protein-RNA complex structures have been determined. These fall into three basic categories, depending on the secondary structure of the RNA: four tRNA-protein complexes, two stem-loop-protein complexes and a capped single-stranded RNA-protein complex.

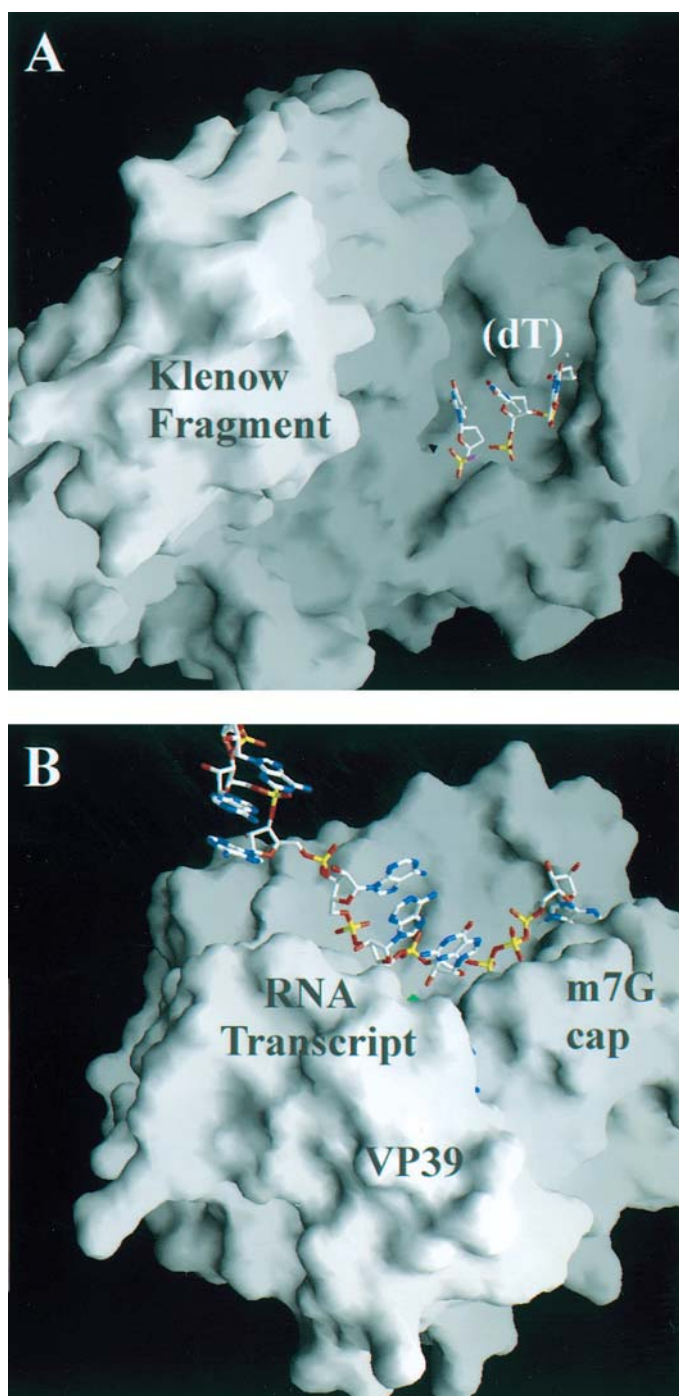


Fig. 23.2.4.4. The sequence-nonspecific recognition of single-stranded nucleic acid. (a) Oligo(dT) bound in the exonuclease active site of DNA polymerase I Klenow fragment (Freemont *et al.*, 1988). (b) A short capped RNA transcript bound to the VP39 RNA methyltransferase (Hodel *et al.*, 1998). Both proteins primarily interact with the backbone of the nucleic acid.

23.2.4.4. Transfer RNA

In the four known structures of tRNA bound to their aminoacyl tRNA synthetases (Cusack *et al.*, 1996a,b; Goldgur *et al.*, 1997; Rould *et al.*, 1991), the effects of RNA's preference for A-form helices on recognition are immediately apparent. The proteins make numerous contacts in the shallow and exposed minor grooves of the RNA helices. This contrasts with the extensive use of the major groove in the recognition of B-form DNA helices. Beyond this generalization, the details of tRNA recognition differ in each

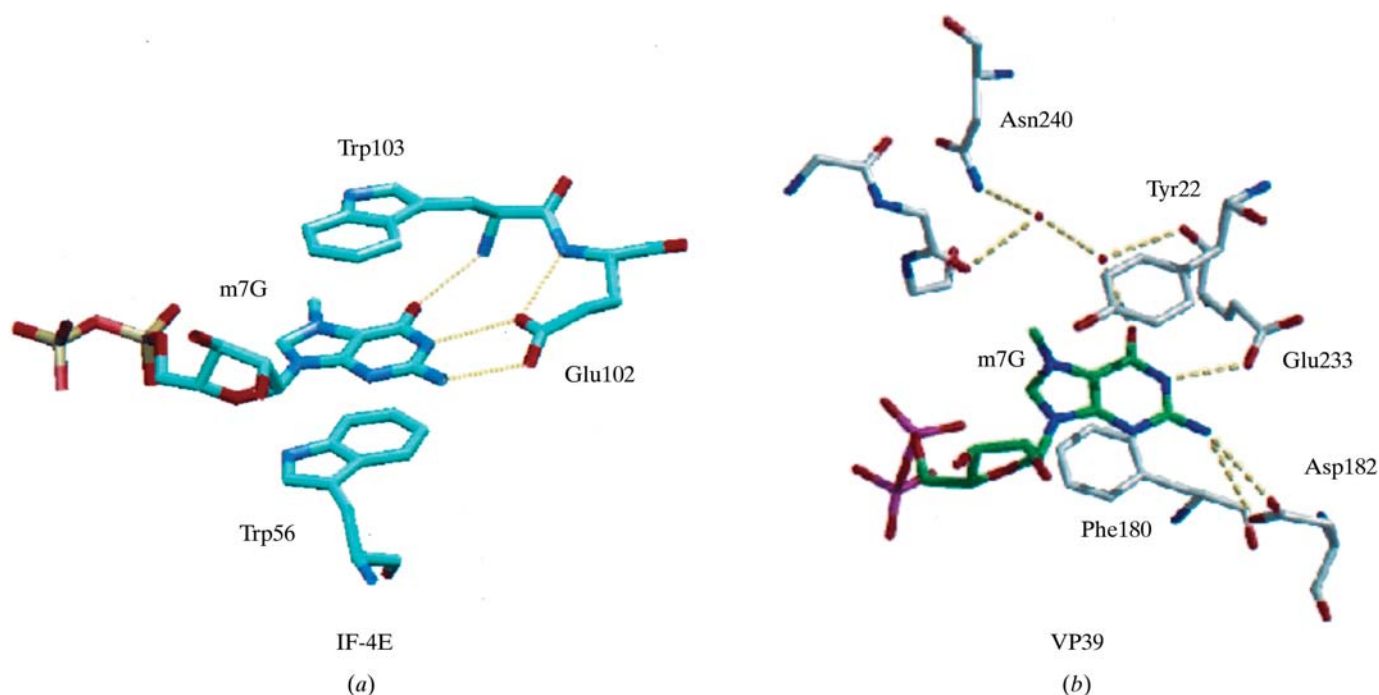


Fig. 23.2.4.5. The specific recognition of the messenger RNA 7-methylguanosine cap. (a) The residues contacting the m^7G base in the cap-binding protein, IF-4E (Marcotrigiano *et al.*, 1997). (b) The residues interacting with the cap in the vaccinia RNA methyltransferase VP39 (Hodel *et al.*, 1997). Both proteins bind to the charged, methylated base by stacking aromatic amino acids on both sides of the base.

specific case. Comparison of the protein-bound tRNA to the structure of free tRNA reveals that the proteins tend to distort the RNA conformation and partially unwind the helices near the anticodon loop. In one case, namely the structure of glutamyl-tRNA synthetase (Rould *et al.*, 1991), the final base pair near the acceptor stem of the tRNA is broken, and the CCA acceptor makes a dramatic hairpin turn into the enzyme active site.

23.2.4.5. Stem loops

One fascinating observation in viewing the structures of RNA-binding proteins, even in the absence of RNA, is that aside from the tRNA-binding synthetases, they all appear to have evolved from or towards a very similar general fold (Burd & Dreyfuss, 1994). This fold, exemplified by the RNP domain found in numerous RNA-binding proteins, consists of a β -sheet surrounded on one side by α -helices and solvent-exposed on the opposing face. This general folding architecture is found in RNP domains, ribosome proteins, K-homologous domains (KH), double-stranded RNA-binding domains and cold shock proteins. Although each of these subsets of RNA-binding domains has a different topology and most probably bind to RNA with different surfaces, they all appear to have this alpha-beta-solvent architecture.

Two proteins with this architecture have been co-crystallized with their specific RNA stem-loop ligands (Nagai *et al.*, 1995; van den Worm *et al.*, 1998). In both cases, the loop of the RNA binds to the open face of the β -sheet where solvent-exposed aromatic amino-acid side chains stack with the extrahelical bases of the RNA. Unpaired bases from the RNA also form numerous specific hydrogen bonds with protein side chains and polar backbone groups, imparting sequence specificity in the interaction. These structures suggest that the flat, open face of a β -sheet provides a good surface for RNA binding, where the extrahelical bases can make extensive and specific contacts with the protein.

23.2.4.6. Single-stranded sequence-nonspecific RNA-protein interactions

There is a single example of a single-stranded RNA-protein complex which is sequence-nonspecific. The structure of the vaccinia RNA methyltransferase VP39 bound to a 5' m^7G -capped RNA hexamer reveals a mechanism of nonspecific recognition reminiscent of the Klenow fragment-DNA tetramer complex (Hodel *et al.*, 1998). The RNA forms two short single-stranded helices of three bases each. The first of these helices binds in the active site of VP39 solely through hydrogen bonds between the protein and the ribose-phosphate backbone. The bases of the RNA strand stack together as trimers, but do not form any interactions with the protein (Fig. 23.2.4.4). Like the Klenow-DNA complex, this observation suggests an intuitive mechanism for sequence-nonspecific nucleic acid binding, where the single-stranded RNA forms short transient helices driven by intramolecular stacking interactions. The protein then recognizes and stabilizes the helical backbone conformation formed by this transient stacking without interacting with the bases themselves.

23.2.4.7. The recognition of alkylated bases

The complex of VP39 with capped RNA also illustrates a final example of the diversity of protein-ligand interactions in the specific recognition of the 7-methylguanosine cap. When guanosine is methylated at the N7 position, a positive charge is introduced to the π -ring system of the base. Eukaryotic cells utilize the methylation of a guanosine base at the N7 position as a tag or cap for the 5' end of messenger RNA. The $m^7G(5')$ ppp mRNA cap is specifically recognized in the splicing of the first intron in nascent transcripts, in the transport of mRNA through the nuclear envelope and in the translation of the message by the ribosome (Varani, 1997). Two structures of specific m^7G binding proteins are now known: VP39 and the ribosomal cap-binding protein IF-4E, (Hodel *et al.*, 1997; Marcotrigiano *et al.*, 1997). Each structure offers clues

as to how the proteins can discriminate between the charged methylated m^7G base and the unmodified guanosine base. The m^7G base is stacked between aromatic protein side chains and hydrogen bonded to acidic protein residues (Fig. 23.2.4.5). One long-held hypothesis is that IF-4E, with dual tryptophan residues, binds specifically to the positively charged form of the base through a charge-transfer complex (Ueda, Iyo, Doi, Inoue & Ishida, 1991). The formation of a charge-transfer complex is evident in small-molecule studies and spectroscopic studies with IF-4E (Ueda, Iyo, Doi, Inoue, Ishida *et al.*, 1991). However, VP39 performs the same discrimination with the much less electronegative phenylalanine and tyrosine side chains (Hodel *et al.*, 1997). So far, no charge-transfer complex has been observed in VP39.

The recognition of charged methylated bases is important not only in mRNA processing, but also in the repair and recognition of DNA damaged by alkylating carcinogens. The mechanism by which the charged m^7G base is recognized is probably similar to how other positively charged bases, such as 3-methyladenosine, O2-methylcytosine and O2-methylthymidine, are recognized. In fact, the *E. coli* DNA repair enzyme, AlkA, will catalyse the glycolysis of all of these bases (Lindahl, 1982). The structure of AlkA is known, but only in the absence of a substrate (Labahn *et al.*, 1996). In this structure, a number of solvent-exposed tryptophan residues are found at the putative active site. This observation suggests that AlkA may recognize positively charged bases through an aromatic 'sandwich', much like that found in IF-4E and VP39.

23.2.5. Phosphate and sulfate

Novel features of molecular recognition and electrostatic interactions of these two tetrahedral oxyanions have emerged from our crystallographic and functional studies of the phosphate-binding protein (PBP) and sulfate-binding protein (SBP), which serve as extremely specific initial receptors for ATP-binding cassette (ABC)-type active transport or permease in bacterial cells. The complexes of these proteins have K_d values in the low μM range. Although phosphate and sulfate are structurally similar, at physiological pH PBP and SBP exhibit no overlap in specificity (Medveczky & Rosenberg, 1971; Pardee, 1966; Jacobson & Quioco, 1988). This stringent specificity prevents one tetrahedral oxyanion nutrient from becoming an inhibitor of transport for the other. The specificity of the PBP-dependent phosphate transport system is also shared by other phosphate transport systems in eukaryotic cells and across brush borders and into mitochondria.

As described below, discrimination between anions is based solely on the protonation state of the ligand. Sulfate, a conjugate base of a strong acid, is completely ionized at pH values above 3, whereas phosphate, a conjugate base of a weak acid, remains protonated up to pH 13.

The structure of the PBP-phosphate complex was initially determined at 1.7 Å resolution (Luecke & Quioco, 1990). The resolution has been pushed to an ultra high resolution of 0.98 Å, the first reported for a protein with a molecular weight as high as 34 kDa with a bound ligand (Wang *et al.*, 1997). The bound phosphate is completely desolvated and sequestered in the protein cleft between two domains. It makes 12 hydrogen bonds with the proteins (11 with donor groups and one with an acceptor group), as well as one salt link to an Arg that is in turn salt-linked to an Asp residue (Fig. 23.2.5.1). The distances of the 12 hydrogen bonds between phosphate and PBP obtained from the ultra high resolution structure range from 2.432 to 2.906 Å (Wang *et al.*, 1997). The Asp56 carboxylate, the lone acceptor group, plays two key roles in conferring the exquisite specificity of PBP. It recognizes, by way of the hydrogen bond, a proton on the phosphate and presumably

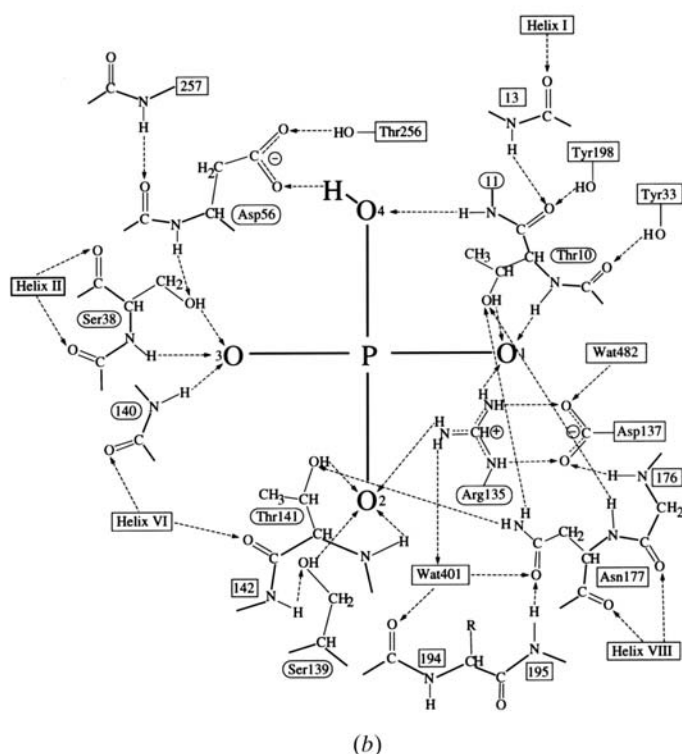
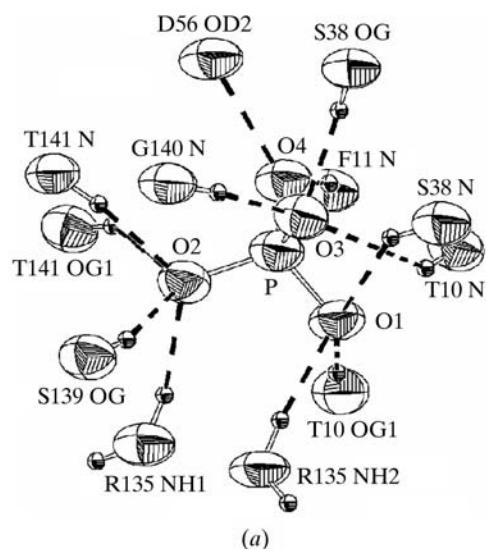


Fig. 23.2.5.1. 12 hydrogen-bonding interactions between the phosphate-binding protein (PBP) and phosphate. (a) Displacement ellipsoids of the atoms involved in the interactions from the 0.98 Å atomic structure (Wang *et al.*, 1997). (b) Schematic diagram of the interactions, including additional hydrogen bonds.

disallows, by charge repulsion, the binding of a fully ionized sulfate dianion (Luecke & Quioco, 1990).

The SBP binding-site cleft is also tailor-made for sulfate (Pflugrath & Quioco, 1985). In keeping with the stringent specificity of SBP for fully ionized tetrahedral oxyanions (Pardee, 1966; Jacobson & Quioco, 1988), the bound sulfate, which is also completely dehydrated and buried, is held in place by seven hydrogen bonds made entirely with donor groups from uncharged polar residues of the protein (Fig. 23.2.5.2) (Pflugrath & Quioco, 1985). The absence of a hydrogen-bond acceptor group accounts for the inability of SBP to bind phosphate. Interestingly, the absence of a salt link and the formation of five fewer hydrogen bonds with the

23. STRUCTURAL ANALYSIS AND CLASSIFICATION

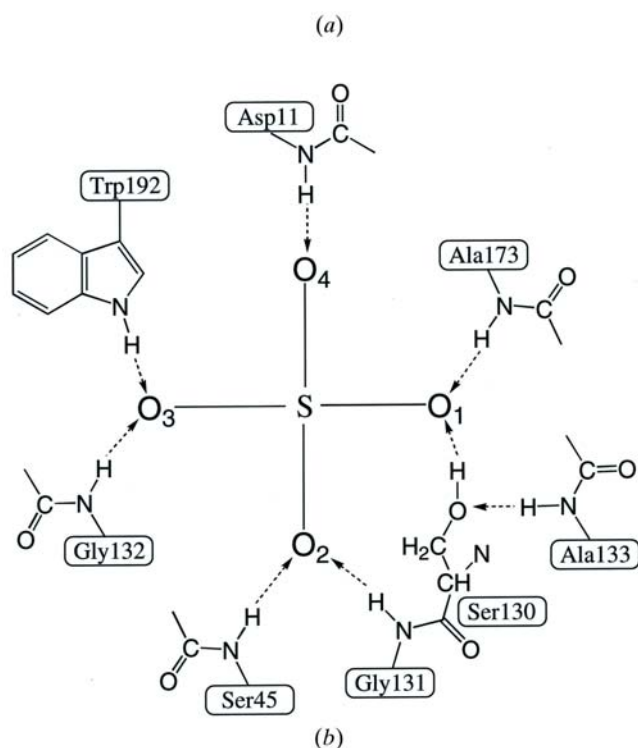
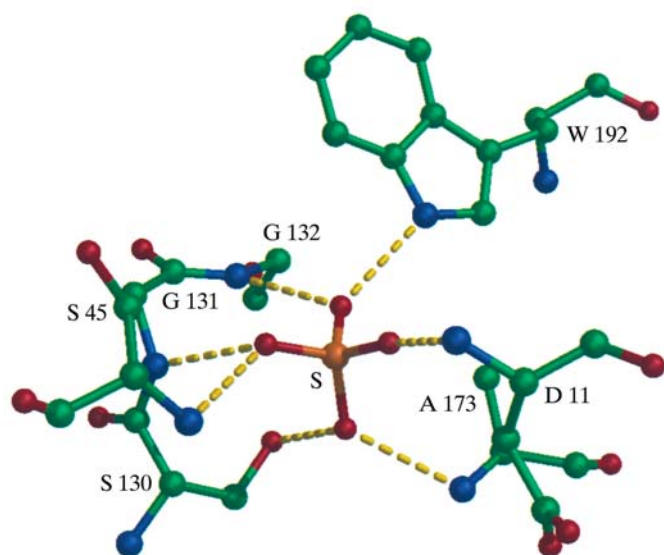


Fig. 23.2.5.2. Seven hydrogen-bonding interactions between the sulfate-binding protein (SBP) and sulfate. (a) Interactions based on the 1.7 Å structure (J. Sack & F. A. Quiocho, unpublished data). (b) Schematic diagram of the interaction.

bound sulfate (Fig. 23.2.5.2b) than with the bound phosphate (Fig. 23.2.5.1b) do not make the affinity of the SBP–sulfate complex any weaker than that of the PBP–phosphate complex. In fact, the sulfate binds 10–20 times more tightly to SBP (Pardee, 1966; Jacobson & Quiocho, 1988). Also, the hydration energies of both anions are likely to be similar.

The ability of PBP and SBP to differentiate each oxyanion ligand through the presence or absence of proton(s) is an extremely high level of sophistication in molecular recognition. The importance of complete hydrogen bonding in recognition of buried ligands is powerfully demonstrated in PBP and SBP. As the sulfate is fully ionized (*i.e.* possesses no hydrogen at physiological pH), repulsion

occurs at Asp56 of PBP specifically for this dianion. On the other hand, SBP is unable to bind phosphate because it contains no hydrogen-bond acceptor in the binding site. Significantly, despite the potential for a large number of matched hydrogen-bonding pairs, a single mismatched hydrogen bond (*e.g.* a fully ionized sulfate providing no proton for interaction with Asp56 of PBP and no acceptor group in SBP for a phosphate proton) represents a binding energy barrier of 6–7 kcal mol⁻¹ (1 kcal mol⁻¹ = 4.184 kJ mol⁻¹).

23.2.5.1. Dominant role of local dipoles in stabilization of isolated charges

A novel finding of further paramount importance and wide implication is how the isolated charges of the protein-bound phosphate and sulfate are stabilized. No counter-charged residues or cations are associated with the sulfate completely buried in SBP. Although a salt link involving Arg135 is formed with the phosphate bound to PBP, it is shared with an Asp residue (Fig. 23.2.5.1b). Moreover, site-directed mutagenesis studies indicate that phosphate binding is quite insensitive to modulation of the salt link (Yao *et al.*, 1996). These findings are a powerful demonstration of how a protein is able to stabilize the charges by means other than salt links. Experimental and computational studies indicate that local dipoles, including the hydrogen-bonding groups and the backbone NH groups from the first turn of helices, immediately surrounding the sulfate and phosphate are responsible for charge stabilization (Pflugrath & Quiocho, 1985; Quiocho *et al.*, 1987; Aqvist *et al.*, 1991; He & Quiocho, 1993; Yao *et al.*, 1996; Ledvina *et al.*, 1996). Helix macrodipoles play little or no role in charge stabilization of the anions. The same principle of charge stabilization by local dipoles also applies for the following buried uncompensated ionic groups: Arg151 of the arabinose-binding protein (Quiocho *et al.*, 1987), the zwitterionic leucine ligand bound to the leucine/isoleucine/valine-binding protein (Quiocho *et al.*, 1987), the potassium in the pore of the potassium channel (Doyle *et al.*, 1998) and Arg56 of synaptobrevin-II in a SNARE complex (Sutton *et al.*, 1998).

23.2.5.2. Short hydrogen bonds

The ultra high resolution refined structure of the PBP–phosphate complex is the first to show structurally the formation of an extremely short hydrogen bond (2.432 Å) between the Asp56 carboxylate of PBP and phosphate. Although this short hydrogen bond is within the proposed range of low-barrier hydrogen bonds with estimated energies of 12–24 kcal mol⁻¹ (Hibbert & Emsley, 1990), its contribution to phosphate binding affinity has been assessed to be no better than that of a normal hydrogen bond (Wang *et al.*, 1997). Thus, a unique role for short hydrogen bonds in biological systems, such as in enzyme catalysis (Gerlt & Gassman, 1993; Cleland & Kreevoy, 1994), remains controversial.

23.2.5.3. Non-complementary negative electrostatic surface potential of protein sites specific for anions

The presence of an uncompensated negatively charged Asp56 is unusual for an anion-binding site, as observed in PBP. In fact, a related discovery of profound ramification is that the binding-cleft region of PBP has an intense negative electrostatic surface potential (Fig. 23.2.5.3a) (Ledvina *et al.*, 1996). Non-complementarity between the surface potential of a binding region and an anion ligand is not unique to PBP. We have reported similar findings for SBP, a DNA-binding protein, and, even more dramatically, for the redox protein flavodoxin (Fig. 23.2.5.3b) (Ledvina *et al.*, 1996). Evidently, for proteins such as these, which rely on hydrogen-

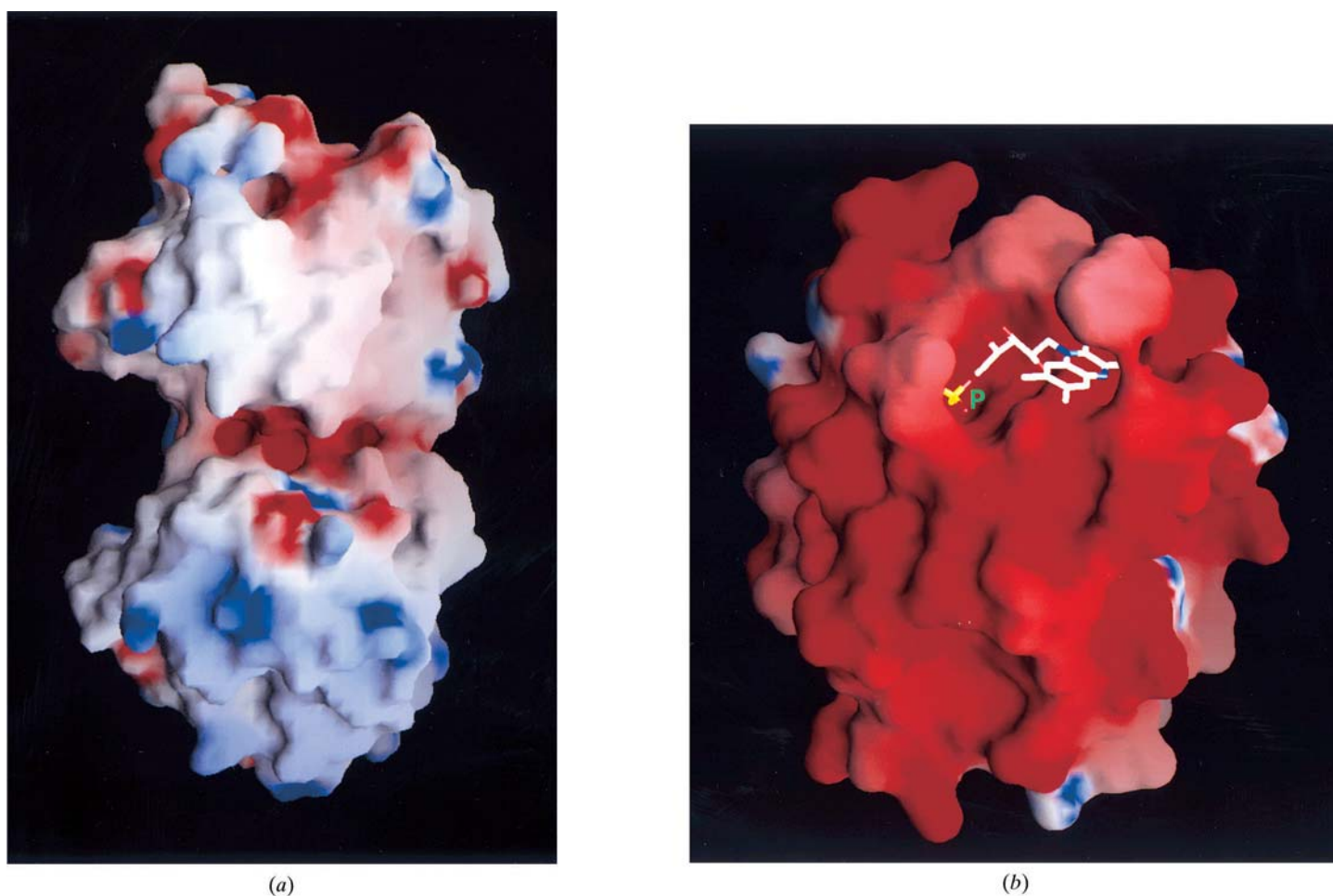


Fig. 23.2.5.3. Electrostatic surface potential of (a) the phosphate-binding protein and (b) flavodoxin. The molecular surface electrostatic potentials, calculated and displayed using *GRASP* (Nicholls *et al.*, 1991), are -10 kT (red), neutral (white) and $+10 \text{ kT}$ (blue) [see Ledvina *et al.* (1996) for more details]. (a) Wild-type phosphate-binding protein based on the X-ray structure of the open cleft, unliganded form (Ledvina *et al.*, 1996). The phosphate-binding site is located in the cleft (with negative surface potential) in the middle of the molecule and between the two domains. (b) Flavodoxin with bound flavin mononucleotide (FMN). The phosphoryl group (P) of the FMN is bound in a pocket with intense negatively charge surface potential. The surface potential was calculated without the bound flavin mononucleotide using the structure from the Protein Data Bank (PDB code: 2fox).

bonding interactions with only uncharged polar residues for anion binding and electrostatic balance, a non-complementary surface potential is not a barrier to binding. This conclusion is supported by very recent fast kinetic studies of binding of phosphate to PBP and the effect of ionic strength on binding (Ledvina *et al.*, 1998).

Acknowledgements

FAQ is an HHMI Investigator. The work carried out in his laboratory is supported in part by grants from NIH and the Welch Foundation.

23. STRUCTURAL ANALYSIS AND CLASSIFICATION

23.1 (cont.)

- tive family members into similar folds. *Structure Fold. Des.* **1**, 209–220.
- Swindells, M. B. (1995). A procedure for detecting structural domains in proteins. *Protein Sci.* **4**, 103–112.
- Taylor, W. R. & Orengo, C. A. (1989). Protein structure alignment. *J. Mol. Biol.* **208**, 1–22.
- Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* **15**, 5739–5751.
- Vriend, G. & Sander, C. (1991). Detection of common three-dimensional substructures in proteins. *Proteins*, **11**, 552–558.
- Wernisch, L., Hunting, M. & Wodak, J. (1999). Identification of structural domains in proteins by a graph heuristic. *Proteins*, **35**, 338–352.
- Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl Acad. Sci. USA*, **70**, 697–701.
- Wodak, J. & Janin, J. (1981). Location of structural domains in proteins. *Biochemistry*, **20**, 6544–6552.
- Zehfus, M. H. (1994). Binary discontinuous compact protein domains. *Protein Eng.* **7**, 335–340.
- Zehfus, M. H. (1997). Identification of compact, hydrophobically stabilized domains and modules containing multiple peptide chains. *Protein Sci.* **6**, 1210–1219.
- Zehfus, M. H. & Rose, G. D. (1986). Compact units in proteins. *Biochemistry*, **25**, 5759–5765.

23.2

- Åqvist, J., Luecke, H., Quioco, F. A. & Warshel, A. (1991). Dipoles localized at helix termini of proteins stabilize charges. *Proc. Natl Acad. Sci. USA*, **88**, 2026–2030.
- Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). On the conformation of the hen egg-white lysozyme molecule. *Proc. R. Soc. London Ser. B Biol. Sci.* **167**, 365–377.
- Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature (London)*, **385**, 176–181.
- Burd, C. G. & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.
- Cheng, X. (1995). DNA modification by methyltransferases. *Curr. Opin. Struct. Biol.* **5**, 4–10.
- Cleland, W. W. & Kreevoy, M. M. (1994). Low-barrier hydrogen bonds and enzymic catalysis. *Science*, **264**, 1887–1890.
- Cotton, F. A., Hazen, E. E. Jr & Legg, M. J. (1979). *Staphylococcal nuclease: proposed mechanism of action based on structure of enzyme-thymidine 3',5'-bisphosphate-calcium ion complex at 1.5-Å resolution.* *Proc. Natl Acad. Sci. USA*, **76**, 2551–2555.
- Cusack, S., Yaremchuk, A. & Tukalo, M. (1996a). The crystal structure of the ternary complex of *T. thermophilus* seryl-tRNA synthetase with tRNA(Ser) and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J.* **15**, 2834–2842.
- Cusack, S., Yaremchuk, A. & Tukalo, M. (1996b). The crystal structures of *T. thermophilus* lysyl-tRNA synthetase complexed with *E. coli* tRNA(Lys) and a *T. thermophilus* tRNA(Lys) transcript: anticodon recognition and conformational changes upon binding of a lysyl-adenylate analogue. *EMBO J.* **15**, 6321–6334.
- Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A. L., Gulbis, J. M., Cohen, S. L., Chait, B. T. & MacKinnon, R. (1998). *Science*, **280**, 68–77.
- Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. & Steitz, T. A. (1988). Cocystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl Acad. Sci. USA*, **85**, 8924–8928.
- Gerlt, J. A. & Gassman, P. G. (1993). *Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl-transfer reaction, and displacement reactions of phosphodiesteres.* *Biochemistry*, **32**, 1943–1952.
- Glusker, J. P. (1991). Structural aspects of metal liganding to functional groups in proteins. *Adv. Protein Chem.* **42**, 1–76.
- Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S. & Safro, M. (1997). The crystal structure of phenylalanyl-tRNA synthetase from *Thermus Thermophilus* complexed with cognate tRNAPhe. *Structure*, **5**, 59–68.
- Harrington, R. E. & Winicov, I. (1994). New concepts in protein-DNA recognition: sequence-directed DNA bending and flexibility. *Prog. Nucleic Acid Res. Mol. Biol.* **47**, 195–270.
- He, J. J. & Quioco, F. A. (1993). Dominant role of local dipoles in stabilizing uncompensated charges on a sulphate sequestered in a periplasmic active transport. *Protein Sci.* **2**, 1643–1647.
- Hibbert, F. & Emsley, J. (1990). Hydrogen bonding and chemical reactivity. *Adv. Phys. Org. Chem.* **226**, 255–379.
- Hodel, A. E., Gershon, P. D. & Quioco, F. A. (1998). Structural basis for sequence non-specific recognition of 5'-capped mRNA by a cap-modifying enzyme. *Mol. Cell*, **1**, 443–447.
- Hodel, A. E., Gershon, P. D., Shi, X., Wang, S. M. & Quioco, F. A. (1997). Specific protein recognition of an mRNA cap through its alkylated base. (Letter.) *Nature Struct. Biol.* **4**, 350–354.
- Jacobson, B. L. & Quioco, F. A. (1988). Sulphate-binding protein dislikes protonated oxyacids: a molecular explanation. *J. Mol. Biol.* **204**, 783–787.
- Joachimiak, A., Schevitz, R. W., Kelley, R. L., Yanofsky, C. & Sigler, P. B. (1983). Functional inferences from crystals of *Escherichia coli* trp repressor. *J. Biol. Chem.* **258**, 12641–12643.
- Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, **63**, 579–590.
- Labahn, J., Schärer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L. & Ellenberger, T. E. (1996). Structural basis for the excision repair of alkylation-damaged DNA. *Cell*, **86**, 321–329.
- Ledvina, P. S., Tsai, A.-H., Wang, Z., Koehl, E. & Quioco, F. A. (1998). Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface potential: equilibrium, kinetic and crystallographic studies. *Protein Sci.* **7**, 2550–2559.
- Ledvina, P. S., Yao, N., Choudhary, A. & Quioco, F. A. (1996). Negative electrostatic surface potential of protein sites specific for anionic ligands. *Proc. Natl Acad. Sci. USA*, **93**, 6786–6791.
- Lindahl, T. (1982). DNA repair enzymes. *Annu. Rev. Biochem.* **51**, 61–87.
- Luecke, H. & Quioco, F. A. (1990). High specificity of a phosphate transport protein determined by hydrogen bonds. *Nature (London)*, **347**, 402–406.
- Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. (1997). X-ray studies of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Nucleic Acids Symp. Ser.* pp. 8–11.
- Meador, W. E., George, S. E., Means, A. R. & Quioco, F. A. (1995). X-ray analysis reveals conformational adaptation of the linker in functional calmodulin mutants. (Letter.) *Nature Struct. Biol.* **2**, 943–945.
- Medveczky, N. & Rosenberg, H. (1971). Phosphate transport in *Escherichia coli*. *Biochim. Biophys. Acta*, **241**, 494–506.
- Nagai, K. (1996). RNA-protein complexes. *Curr. Opin. Struct. Biol.* **6**, 53–61.
- Nagai, K., Oubridge, C., Ito, N., Jessen, T. H., Avis, J. & Evans, P. (1995). Crystal structure of the U1A spliceosomal protein complexed with its cognate RNA hairpin. *Nucleic Acids Symp. Ser.* 1–2.
- Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association; insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Orgel, L. E. (1966). *An introduction to transition-metal chemistry. Ligand-field theory*, 2nd ed. London: Methuen and New York: Wiley.

REFERENCES

23.2 (cont.)

- Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). *Crystal structure of trp repressor/operator complex at atomic resolution. Nature (London)*, **335**, 321–329; erratum (1988), **335**, 837.
- Pabo, C. O. & Sauer, R. T. (1992). *Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem.* **61**, 1053–1095.
- Pardee, A. B. (1966). *Purification and properties of a sulfate-binding protein from Salmonella typhimurium. J. Biol. Chem.* **241**, 5886–5892.
- Pavletich, N. P. & Pabo, C. O. (1991). *Zinc finger–DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. Science*, **252**, 809–817.
- Pflugrath, J. W. & Quioco, F. A. (1985). *Sulphate sequestered in the sulphate-binding protein of Salmonella typhimurium is bound solely by hydrogen bonds. Nature (London)*, **314**, 257–260.
- Quioco, F. A. (1986). *Carbohydrate-binding proteins: tertiary structures and protein–sugar interactions. Annu. Rev. Biochem.* **55**, 287–315.
- Quioco, F. A., Sack, J. S. & Vyas, N. K. (1989). *Substrate specificity and affinity of a protein modulated by bound water molecules. Nature (London)*, **340**, 404–407.
- Quioco, F. A., Wilson, D. K. & Vyas, N. K. (1987). *Stabilization of charges on isolated charged groups sequestered in proteins by polarized peptide units. Nature (London)*, **329**, 561–564.
- Rademacher, T. W., Parekh, R. B. & Dwek, R. A. (1988). *Glycobiology. Annu. Rev. Biochem.* **57**, 785–838.
- Rould, M. A., Perona, J. J. & Steitz, T. A. (1991). *Structural basis of anticodon loop recognition by glutamyl-tRNA synthetase. Nature (London)*, **352**, 213–218.
- Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). *Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl Acad. Sci. USA*, **73**, 804–808.
- Shimon, L. J. & Harrison, S. C. (1993). *The phage 434 OR2/R1-69 complex at 2.5 Å resolution. J. Mol. Biol.* **232**, 826–838.
- Steitz, T. A. (1990). *Structural studies of protein–nucleic acid interaction: the sources of sequence-specific binding. Q. Rev. Biophys.* **23**, 205–280.
- Sutton, R. B., Fasshauer, D., Jan, R. & Brunger, A. T. (1998). *Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. Nature (London)*, **395**, 347–353.
- Tucker, P. W., Hazen, E. E. Jr & Cotton, F. A. (1979). *Staphylococcal nuclease reviewed: a prototypic study in contemporary enzymology. III. Correlation of the three-dimensional structure with the mechanisms of enzymatic action. Mol. Cell. Biochem.* **23**, 67–86.
- Ueda, H., Iyo, H., Doi, M., Inoue, M. & Ishida, T. (1991). *Cooperative stacking and hydrogen bond pairing interactions of fragment peptide in cap binding protein with mRNA cap structure. Biochim. Biophys. Acta*, **1075**, 181–186.
- Ueda, H., Iyo, H., Doi, M., Inoue, M., Ishida, T., Morioka, H., Tanaka, T., Nishikawa, S. & Uesugi, S. (1991). *Combination of Trp and Glu residues for recognition of mRNA cap structure. Analysis of m7G base recognition site of human cap binding protein (IF-4E) by site-directed mutagenesis. FEBS Lett.* **280**, 207–210.
- Varani, G. (1997). *A cap for all occasions. Structure*, **5**, 855–858.
- Volbeda, A., Fontecilla-Camps, J. C. & Frey, M. (1996). *Novel metal sites in protein structures. Curr. Opin. Struct. Biol.* **6**, 804–812.
- Vyas, N. K. (1991). *Atomic features of protein–carbohydrate interactions. Curr. Opin. Struct. Biol.* **1**, 732–740.
- Vyas, N. K., Vyas, M. N. & Quioco, F. A. (1988). *Sugar and signal-transducer binding sites of the Escherichia coli galactose chemoreceptor protein. Science*, **242**, 1290–1295.
- Wang, Z., Luecke, H., Yao, N. & Quioco, F. A. (1997). *Nature Struct. Biol.* **4**, 519–522.
- Weis, W. I. & Drickamer, K. (1996). *Structural basis of lectin–carbohydrate recognition. Annu. Rev. Biochem.* **65**, 441–473.
- Werner, M. H., Gronenborn, A. M. & Clore, G. M. (1996). *Intercalation, DNA kinking, and the control of transcription. Science*, **271**, 778–784; erratum (1996), **272**, 19.
- Worm, S. H. van den, Stonehouse, N. J., Valegard, K., Murray, J. B., Walton, C., Fridborg, K., Stockley, P. G. & Liljas, L. (1998). *Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. Nucleic Acids Res.* **26**, 1345–1351.
- Yao, N., Ledvina, P. S., Choudhary, A. & Quioco, F. A. (1996). *Modulation of a salt link does not affect binding of phosphate to its specific active transport receptor. Biochemistry*, **35**, 2079–2085.

23.3

- Altona, C., Geise, H. J. & Romers, C. (1968). *Conformation of non-aromatic ring compounds, XXIV. On the geometry of the perhydrophenanthrene skeleton in some steroids. Tetrahedron*, **24**, 13–32.
- Altona, C. & Sundaralingam, M. (1972). *Conformational analysis of the sugar ring in nucleosides and nucleotides. J. Am. Chem. Soc.* **94**, 8205–8212.
- Ansevin, A. T. & Wang, A. H. (1990). *Evidence for a new Z-type left-handed DNA helix. Nucleic Acids Res.* **18**, 6119–6126.
- Arnott, S. (1970). *The geometry of nucleic acids. Prog. Biophys. Mol. Biol.* **21**, 265–319.
- Babcock, M. S. & Olson, W. K. (1994). *The effect of mathematics and coordinate system on comparability and ‘dependencies’ of nucleic acid structure parameters. J. Mol. Biol.* **237**, 98–124.
- Babcock, M. S., Pednault, E. & Olson, W. (1993). *Nucleic acid structure analysis: a users guide to a collection of new analysis programs. J. Biomol. Struct. Dyn.* **11**, 597–628.
- Babcock, M. S., Pednault, E. & Olson, W. (1994). *Nucleic acid structure analysis. Mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. J. Mol. Biol.* **237**, 125–156.
- Basham, B., Eichman, B. F. & Ho, P. S. (1998). *The single-crystal structures of Z-DNA. In Oxford handbook of nucleic acid structure*, edited by S. Neidle, ch. 7, pp. 200–252. Oxford University Press.
- Berman, H. M. (1996). *Crystal studies of B-DNA: the answers and the questions. Biopolymers Nucleic Acid Sci.* **44**, 23–44.
- Bugg, C. E., Thomas, J. M., Sundaralingam, M. & Rao, S. T. (1971). *Stereochemistry of nucleic acids and their constituents. X. Solid-state base-stacking patterns in nucleic acid constituents and polynucleotides. Biopolymers*, **10**, 175–219.
- Crick, F. H. C. & Watson, J. D. (1954). *The complementary structure of deoxyribonucleic acid. Proc. R. Soc. London Ser. A*, **223**, 80–96.
- Crothers, D. M. & Drak, J. (1992). *Global features of DNA structure by comparative gel electrophoresis. Methods Enzymol.* **212**, 46–71.
- Crothers, D. M., Haran, T. E. & Nadeau, J. G. (1990). *Intrinsically bent DNA. J. Biol. Chem.* **265**, 7093–7096.
- Davies, D. B. (1978). *Conformations of nucleosides and nucleotides. Prog. Nucl. Magn. Reson. Spectros.* **12**, 135–186.
- Dickerson, R. E. (1972). *The structure and history of an ancient protein. Sci. Am.* **226** (April), 58–72.
- Dickerson, R. E. (1983). *The DNA helix and how it is read. Sci. Am.* **249** (December), 94–111.
- Dickerson, R. E. (1985). *Helix polymorphism and information flow in DNA. In Proceedings of the Robert A. Welch Foundation conferences on chemical research XXIX: Genetic chemistry; the molecular basis of heredity*, pp. 38–79.
- Dickerson, R. E. (1992). *DNA structure from A to Z. Methods Enzymol.* **211**, 67–111.
- Dickerson, R. E. (1997a). *Obituary: Irving Geis, 1908–1997. Structure*, **5**, 1247–1249.
- Dickerson, R. E. (1997b). *Irving Geis, molecular artist, 1908–1997. Protein Sci.* **6**, 2843–2844.
- Dickerson, R. E. (1997c). *Biology in pictures: molecular artistry. Curr. Biol.* **7**, R720–R741.