

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

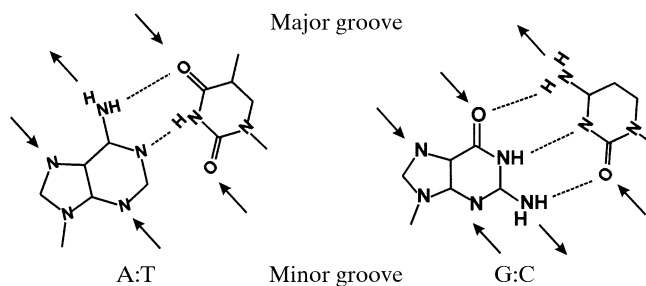


Fig. 23.2.4.1. A schematic diagram of the base pairs of DNA showing the hydrogen-bonding groups which may be used in the sequence-specific recognition of DNA. The major groove is at the top of the figure and the minor groove at the bottom. Arrows point towards hydrogen-bond acceptors and away from donors.

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,

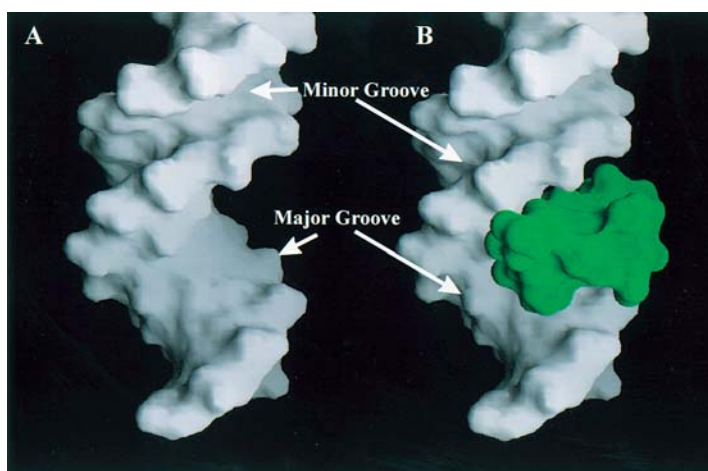


Fig. 23.2.4.2. (a) A space-filling model of B-DNA showing the relative accessibility of the major and minor grooves. (b) A helix of the 434 repressor bound in the major groove of the helix, illustrating how the dimensions of a protein α -helix are compatible for reading the major groove of B-DNA (Shimon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

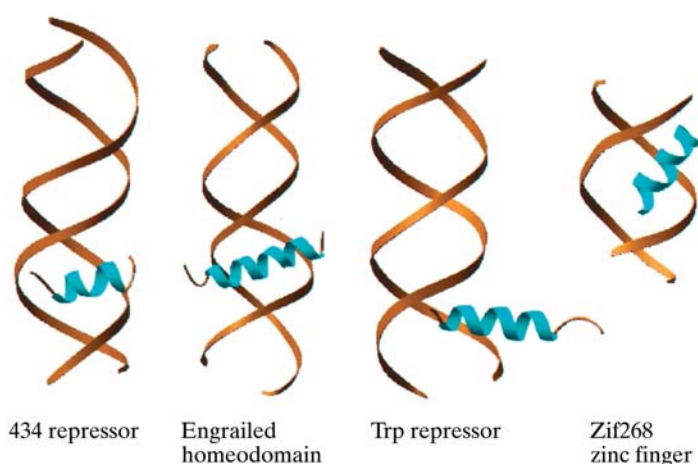


Fig. 23.2.4.3. A comparison of the orientations of α -helices bound in the major groove, taking examples from four DNA-binding proteins: the 434 repressor (Shimon & Harrison, 1993), the engrailed homeodomain (Kissinger *et al.*, 1990), the trp repressor (Otwinowski *et al.*, 1988) and the Zif268 zinc 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

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.

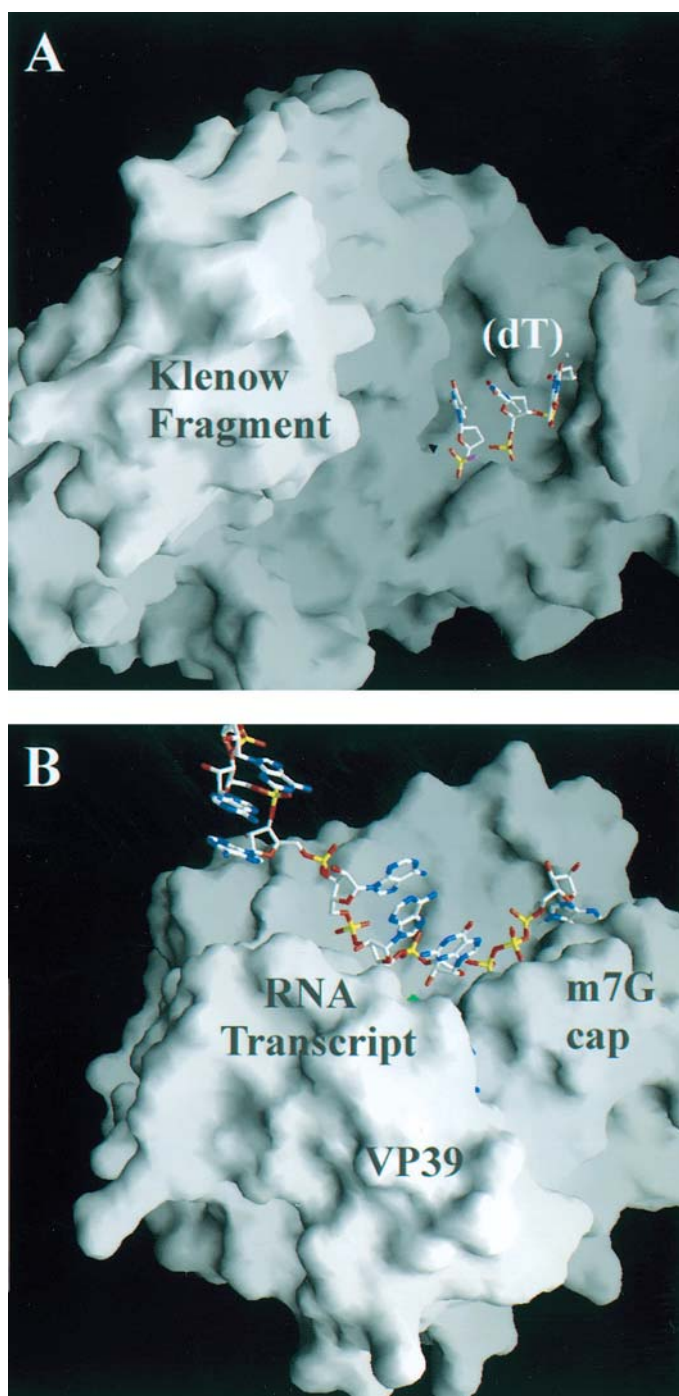


Fig. 23.2.4.4. The sequence-nonspecific recognition of single-stranded nucleic acid. (a) Oligo(dT) bound in the exonuclease active site of DNA polymerase I Klenow fragment (Freemont *et al.*, 1988). (b) A short capped RNA transcript bound to the VP39 RNA methyltransferase (Hodel *et al.*, 1998). Both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