The major groove at an angle of approximately 30°. Protein–DNA complexes are quite variable. Most helices bind in the major groove (Fig. 23.3.4.2). The exact orientations of helices in various triples are suited for binding in the major groove of a B-DNA helix, although there are proteins that recognize DNA through the minor groove, 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. 23.3.4.2). 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° from the plane normal to the DNA helical axis (Fig. 23.3.4.3). 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., 1988a,b). 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 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

---

**Figure 23.3.4.1**
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.

**Figure 23.3.4.2**
(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).

**Figure 23.3.4.3**
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., 1988a,b) and the Zif268 zinc finger (Pabo & Pabo, 1991). The DNA backbone is shown as a brown ribbon, whereas the protein helix is shown as a blue ribbon.