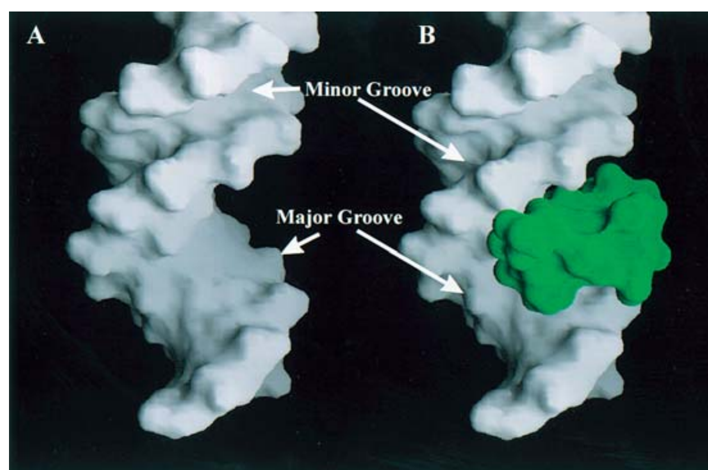

Fig. 23.2.4.2. (*a*) A space-filling model of B-DNA showing the relative accessibility of the major and minor grooves. (*b*) A helix of the 434 repressor bound in the major groove of the helix, illustrating how the dimensions of a protein α-helix are compatible for reading the major groove of B-DNA (Shimon & 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 backbone and base, indirect but specific hydrogen bonding through water molecules, and hydrophobic interactions.

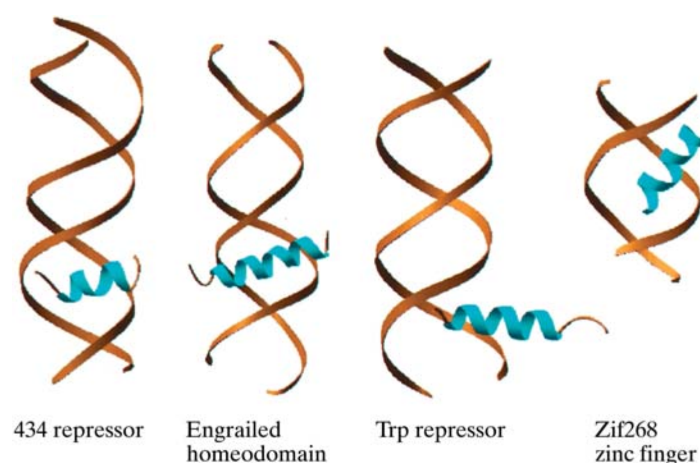

Fig. 23.2.4.3. A comparison of the orientations of α-helices bound in the major groove, taking examples from four DNA-binding proteins: the 434 repressor (Shimon & Harrison, 1993), the engrailed homeodomain (Kissinger *et al.*, 1990), the trp repressor (Otwinowski *et al.*, 1988) and the Zif268 zinc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

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

Copyright © 2006 International Union of Crystallography

references