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 $(r_{\rm cation}/r_{\rm anion})$ 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 staphyloccocal 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

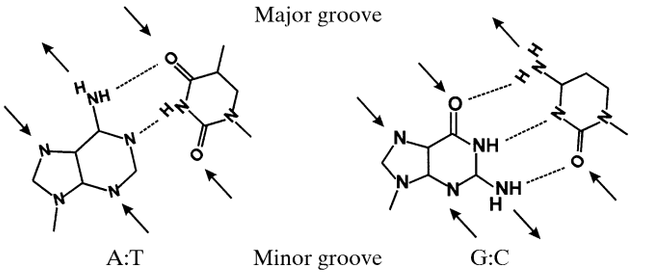

Fig. 23.2.4.1. A schematic diagram of the base pairs of DNA showing the hydrogen-bonding groups which may be used in the sequence-specific recognition of DNA. The major groove is at the top of the figure and the minor groove at the bottom. Arrows point towards hydrogen-bond acceptors and away from donors.

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,

references