edge in forming a pseudo-hydrogen bond between the indole hydrogens and the π-rings of the DNA bases. In *ETS1*, 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.

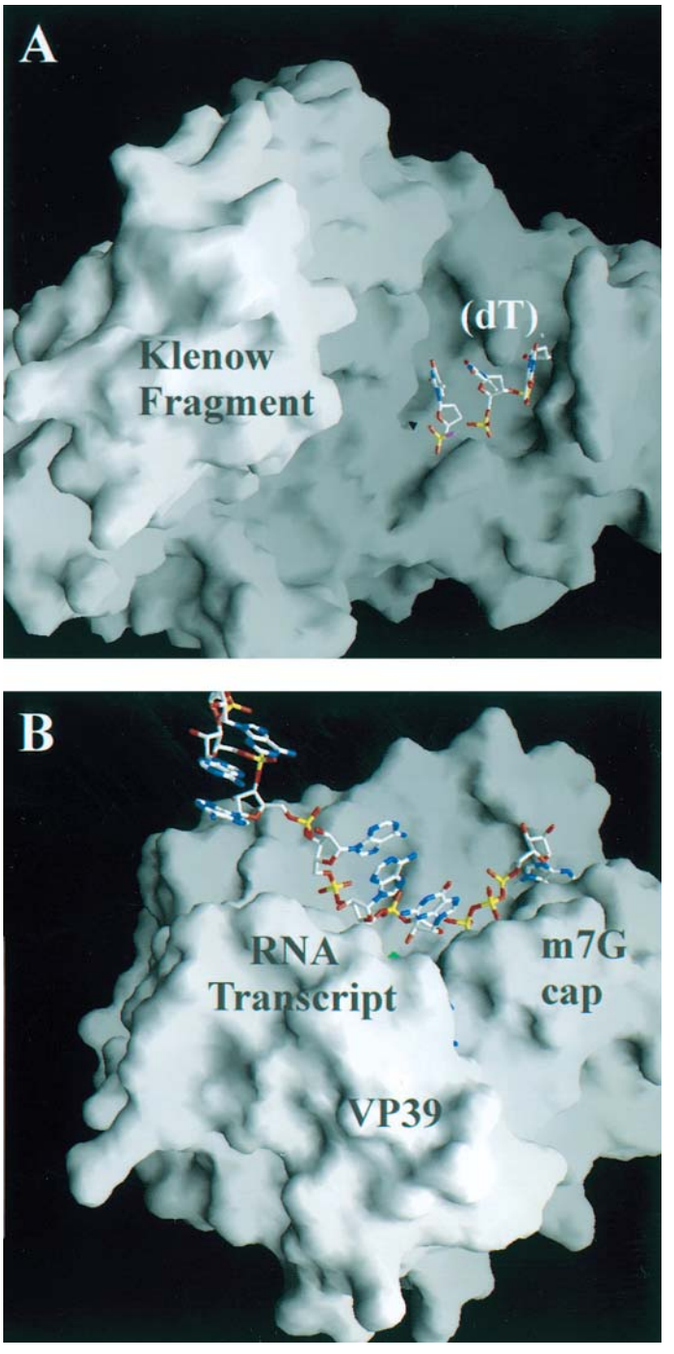

Fig. 23.2.4.4. The sequence-nonspecific recognition of single-stranded nucleic acid. (*a*) Oligo(dT) bound in the exonuclease active site of DNA polymerase I Klenow fragment (Freemont *et al.*, 1988). (*b*) A short capped RNA transcript bound to the VP39 RNA methyltransferase (Hodel *et al.*, 1998). 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.*, 1996*a*,*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

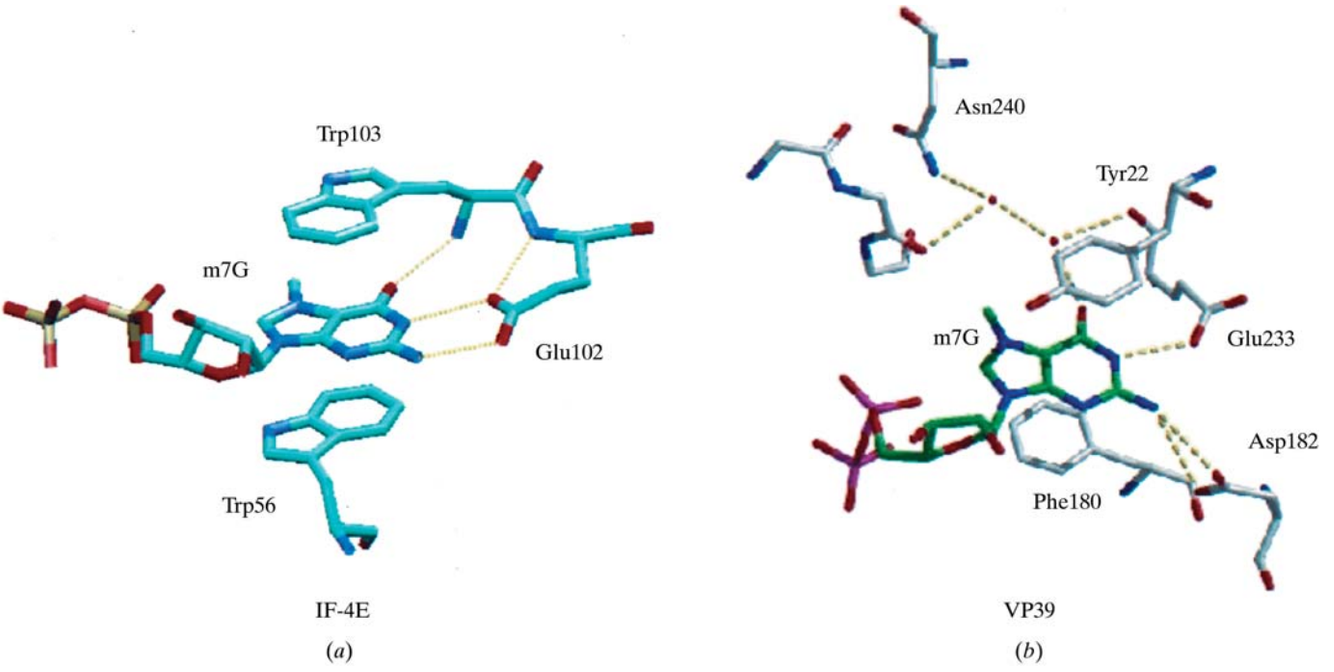

Fig. 23.2.4.5. The specific recognition of the messenger RNA 7-methylguanosine cap. (*a*) The residues contacting the m^7G base in the cap-binding protein, IF-4E (Marcotrigiano *et al.*, 1997). (*b*) The residues interacting with the cap in the vaccinia RNA methyltransferase VP39 (Hodel *et al.*, 1997). 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 anti-codon 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 extra-helical 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

references