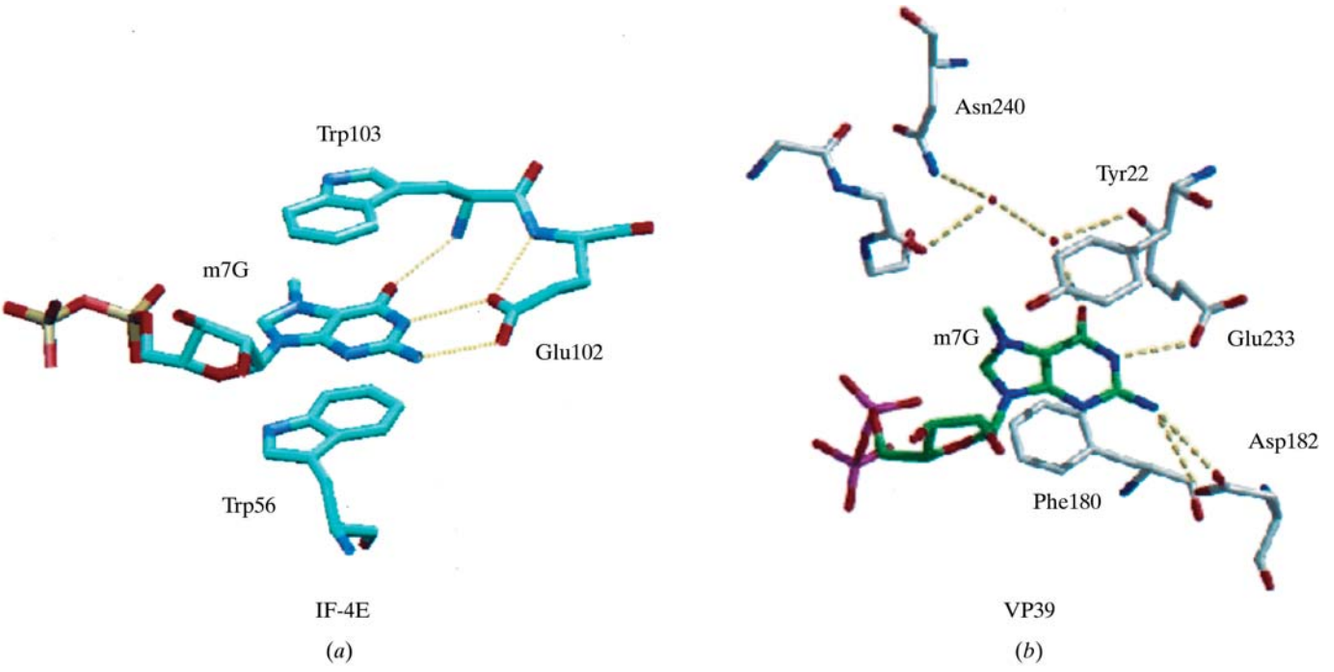

Fig. 23.2.4.5. The specific recognition of the messenger RNA 7-methylguanosine cap. (*a*) The residues contacting the m^7G base in the cap-binding protein, IF-4E (Marcotrigiano *et al.*, 1997). (*b*) The residues interacting with the cap in the vaccinia RNA methyltransferase VP39 (Hodel *et al.*, 1997).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

as to how the proteins can discriminate between the charged methylated m^7G base and the unmodified guanosine base. The m^7G base is stacked between aromatic protein side chains and hydrogen bonded to acidic protein residues (Fig. 23.2.4.5). One long-held hypothesis is that IF-4E, with dual tryptophan residues, binds specifically to the positively charged form of the base through a charge-transfer complex (Ueda, Iyo, Doi, Inoue & Ishida, 1991). The formation of a charge-transfer complex is evident in small-molecule studies and spectroscopic studies with IF-4E (Ueda, Iyo, Doi, Inoue, Ishida *et al.*, 1991). However, VP39 performs the same discrimination with the much less electronegative phenylalanine and tyrosine side chains (Hodel *et al.*, 1997). So far, no charge-transfer complex has been observed in VP39.

The recognition of charged methylated bases is important not only in mRNA processing, but also in the repair and recognition of DNA damaged by alkylating carcinogens. The mechanism by which the charged m^7G base is recognized is probably similar to how other positively charged bases, such as 3-methyladenosine, O2-methylcytosine and O2-methylthymidine, are recognized. In fact, the *E. coli* DNA repair enzyme, AlkA, will catalyse the glycolysis of all of these bases (Lindahl, 1982). The structure of AlkA is known, but only in the absence of a substrate (Labahn *et al.*, 1996). In this structure, a number of solvent-exposed tryptophan residues are found at the putative active site. This observation suggests that AlkA may recognize positively charged bases through an aromatic 'sandwich', much like that found in IF-4E and VP39.

23.2.5. Phosphate and sulfate

Novel features of molecular recognition and electrostatic interactions of these two tetrahedral oxyanions have emerged from our crystallographic and functional studies of the phosphate-binding protein (PBP) and sulfate-binding protein (SBP), which serve as extremely specific initial receptors for ATP-binding cassette (ABC)-type active transport or permease in bacterial cells. The complexes of these proteins have K_d values in the low μM range. Although phosphate and sulfate are structurally similar, at physiological pH PBP and SBP exhibit no overlap in specificity (Medveczky & Rosenberg, 1971; Pardee, 1966; Jacobson & Quiocho, 1988). This stringent specificity prevents one tetrahedral oxyanion nutrient from becoming an inhibitor of transport for the other. The specificity of the PBP-dependent phosphate transport system is also shared by other phosphate transport systems in eukaryotic cells and across brush borders and into mitochondria.

As described below, discrimination between anions is based solely on the protonation state of the ligand. Sulfate, a conjugate base of a strong acid, is completely ionized at pH values above 3, whereas phosphate, a conjugate base of a weak acid, remains protonated up to pH 13.

The structure of the PBP–phosphate complex was initially determined at 1.7 Å resolution (Luecke & Quiocho, 1990). The resolution has been pushed to an ultra high resolution of 0.98 Å, the first reported for a protein with a molecular weight as high as 34 kDa with a bound ligand (Wang *et al.*, 1997). The bound phosphate is completely desolvated and sequestered in the protein cleft between two domains. It makes 12 hydrogen bonds with the proteins (11 with donor groups and one with an acceptor group), as well as one salt link to an Arg that is in turn salt-linked to an Asp residue (Fig. 23.2.5.1). The distances of the 12 hydrogen bonds between phosphate and PBP obtained from the ultra high resolution structure range from 2.432 to 2.906 Å (Wang *et al.*, 1997). The Asp56 carboxylate, the lone acceptor group, plays two key roles in conferring the exquisite specificity of PBP. It recognizes, by way of the hydrogen bond, a proton on the phosphate and presumably

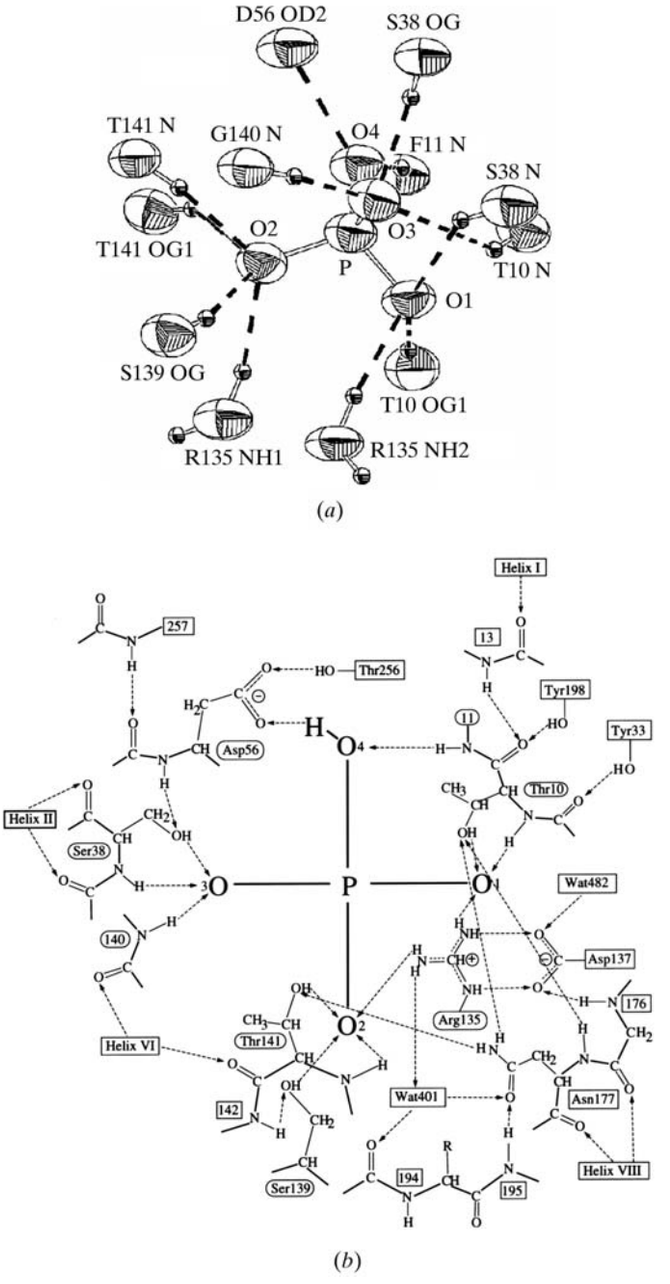

Fig. 23.2.5.1. 12 hydrogen-bonding interactions between the phosphate-binding protein (PBP) and phosphate. (*a*) Displacement ellipsoids of the atoms involved in the interactions from the 0.98 Å atomic structure (Wang *et al.*, 1997). (*b*) Schematic diagram of the interactions, including additional hydrogen bonds.

disallows, by charge repulsion, the binding of a fully ionized sulfate dianion (Luecke & Quiocho, 1990).

The SBP binding-site cleft is also tailor-made for sulfate (Pflugrath & Quiocho, 1985). In keeping with the stringent specificity of SBP for fully ionized tetrahedral oxyanions (Pardee, 1966; Jacobson & Quiocho, 1988), the bound sulfate, which is also completely dehydrated and buried, is held in place by seven hydrogen bonds made entirely with donor groups from uncharged polar residues of the protein (Fig. 23.2.5.2) (Pflugrath & Quiocho, 1985). The absence of a hydrogen-bond acceptor group accounts for the inability of SBP to bind phosphate. Interestingly, the absence of a salt link and the formation of five fewer hydrogen bonds with the

references