Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein

2512–2518 Nucleic Acids Research, 1995, Vol. 23, No. 13

Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein

Peter G. Stockley*, Nicola J. Stonehouse, James B. Murray, Shona T. S. Goodman+, Simon J. Talbot§, Chris J. Adams, Lars Liljas[1] and Karin Valegård[1]

Department of Genetics, University of Leeds, Leeds LS2 9JT, UK and [1]Department of Molecular Biology, Biomedical Centre, Uppsala University, Box 590, S-751 24 Uppsala, Sweden

Received March 2, 1995; Revised and Accepted May 22, 1995

ABSTRACT

We present the results of *in vitro* binding studies aimed at defining the key recognition elements on the MS2 RNA translational operator (TR) essential for complex formation with coat protein. We have used chemically synthesized operators carrying modified functional groups at defined nucleotide positions, which are essential for recognition by the phage coat protein. These experiments have been complemented with modification-binding interference assays. The results confirm that the complexes which form between TR and RNA-free phage capsids, the X-ray structure of which has recently been reported at 3.0 Å, are identical to those which form in solution between TR and a single coat protein dimer. There are also effects on operator affinity which cannot be explained simply by the alteration of direct RNA–protein contacts and may reflect changes in the conformational equilibrium of the unliganded operator. The results also provide support for the approach of using modified oligoribonucleotides to investigate the details of RNA–ligand interactions.

INTRODUCTION

The translational repression complexes which form between RNA bacteriophage coat proteins and their genomic RNAs to regulate expression of phage replicase cistrons have long been studied as examples of sequence-specific RNA recognition (1). Uhlenbeck and his colleagues have shown that for the Group I phage R17(MS2) the information required for sequence-specific recognition by a dimer of phage coat protein is contained entirely in a 19 nucleotide (nt) operator fragment capable of forming a stem–loop structure with a bulged adenine on the 5′ leg of the stem (Fig. 1). As well as acting as a translational repression complex, the 19mer–coat protein dimer interaction appears to act as an assembly initiation complex (2), triggering self-assembly of the quasi-equivalent shell (3). Sequence variation experiments have been used to define a consensus sequence for coat protein binding. Remarkably, the identity of the nucleotides at only a few positions is vital, provided that the Watson–Crick base pairing of the stem is maintained. These important recognition sites are the loop residues A –4, U –5 and A –7 and the bulged A –10 (numbering relative to the start of the replicase cistron +1; see Fig. 1 for details). Indeed, even some of these key residues can be varied with retention of binding. A –10 can be substituted by guanosine with some loss of affinity, but substitution of U –5 by C yields an operator with a significantly higher affinity for coat protein than the wild-type sequence (4).

Previous attempts to explain the importance of these residues have involved arguments about: (i) intercalation of the purine at –10 into the base paired stem, thus kinking the structure; (ii) the ability of adenines to stack on the underlying base paired stem and the possibility of non-Watson–Crick hydrogen bonding between the adenines at –4 and –7 (5,6); (iii) the possibility of covalent Michael adduct formation with a protein cysteine thiol group for the pyrimidine at –5 (7). Suggestions (i) and (ii) were based largely upon model building, whilst proposal (iii) was based on the inactivation of phage coat protein for RNA binding by modification of residue Cys46 by thiol reagents or by treatment with 5-halogenated pyrimidines (4,8). It has been proposed that such a covalent adduct mechanism explains the increased affinity of the C –5 variant, since cytidine is more reactive to nucleophilic attack at C6 (from a protein thiol group) than uridine. Biologically, the presence of uridine rather than cytidine at position –5 in the wild-type operator makes sense in terms of the regulatory role of the translational repression complex, the C –5 variant having a significantly decreased off-rate (a half-life of 408 min versus 0.70 min for U –5) (9).

Below we describe the effects on coat protein affinity of systematic chemical substitutions or modifications at defined sugar or base positions within the 19mer and correlate them with the RNA–protein interactions observed in the coat protein–operator crystal structure (10).

MATERIALS AND METHODS

Preparation of MS2 coat protein

MS2 coat protein was prepared from wild-type bacteriophage or from empty recombinant capsids (11,12) and disassembled coat

* To whom correspondence should be addressed

Present addresses: +9 Malthouse Court, Derham, Norfolk NR20 4UA, UK and §Institute for Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

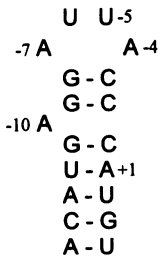


Figure 1. Sequence of the MS2 translational operator (TR) stem–loop. Numbering is relative to the replicase start codon, AUG; A is +1. Dashes indicate the position of Watson–Crick base pairing.

protein subunits purified using the cold acetic acid procedure of Sugiyama *et al.* (13). Concentrations of coat protein were determined by measuring the optical densities at 280 and 260 nm and using the following equation (14):

protein concentration (mg/ml) = $1.55\ A_{280} - 0.76_{260}$.

Coat protein solutions were stored at 4°C until use.

Filter binding assays

Filter binding assays were performed as described by Carey and Uhlenbeck (15). Coat protein was serially diluted in 1 mM acetic acid before addition to a constant, low concentration of labelled RNA (~10 pM) in TMK buffer (50 mM Tris–HCl, pH 8.0, 10 mM $MgCl_2$, 100 mM KCl). Coat protein concentrations ranged from 10^{-9} to 10^{-6} M. Incubations were on ice for 20 min prior to filtration through 0.45 μm nitrocellulose filters (Millipore). Filters were added to 5 ml aqueous scintillant (NEN) and counted in a liquid scintillation counter. Assays were reproducible within a factor of two for individual operator fragments and retention efficiencies were between 60 and 90% of input RNA. Under these conditions of protein excess over input RNA, the protein concentration required to produce 50% maximal retention on the filters approximates the apparent equilibrium dissociation constant, $K_{d'}$. Values of $K_{d'}$ were between 10 and 20 nM and the relative affinity of each RNA variant was determined from multiple filter binding assays carried out at the same time and under identical conditions to the wild-type measurement.

Chemical synthesis, purification and characterization

Chemical synthesis of stem–loop variants was carried out using commercially available 2′-silyl phosphoramidite reagents (ChemGenes Inc., Waltham, MA). Variant nucleotides were either purchased from commercial suppliers or synthesized in house (16–18). Deoxypurine phosphoramidite (19) was a gift from Professor B. A. Connolly (University of Newcastle). Following deprotection, RNAs were purified to apparent homogeneity, either across denaturing polyacrylamide gels or via reverse-phase HPLC (20). Translational operator (TR) fragments prepared similarly have been used to produce crystals of the operator–capsid and have been studied extensively by two-dimensional NMR spectroscopy without any sign of significant chemical heterogeneity. In order to ensure that all the molecules used in binding assays contained the expected variant chemical groups, variants eluted from the reverse-phase HPLC were characterized by sequencing with specific RNAses (21), as described by Talbot *et al.* (9), which identifies the site of deoxy substitution, and in a few cases by complete base composition analysis as described (20). Thio-containing operators were identified by the characteristic absorbance peak between 320 and 350 nm. Samples analysed by circular dichroism spectroscopy and/or UV thermal melting studies suggested that the fragments were essentially monomeric stem–loops under the conditions of the binding assays and that incorporation of variant bases did not lead to gross conformational changes, e.g. the wild-type and C –5 TR variants both had a T_m of 48.2°C. Variants were 5′ radio-labelled and then used in nitrocellulose filter binding assays with wild-type recombinant MS2 coat protein. Since some of the variants used had limited chemical stability over periods of several weeks, affinities were always determined as soon as possible after final purification and compared with the affinity of the wild-type operator as a control.

Binding interference experiments

SP6 RNA polymerase transcription. RNA run-off transcripts were prepared from a linearized plasmid (pSP64-BL) encompassing the wild-type operator sequence (22), essentially as described by Melton *et al.* (23). *In vitro* transcription was stopped by addition of EDTA to a final concentration of 10 mM and the RNA transcripts dephosphorylated by treatment with 10 U calf alkaline phosphatase at 37°C for 30 min in the same buffer. The reaction was stopped by extraction with a 50:50 mixture of phenol and chloroform, followed by precipitation of the RNA by addition of NaCl to 100 mM and 3 vol ethanol. RNAs for diethylpyrocarbonate (DEPC) experiments were radiolabelled with ^{32}P at the 5′-end before modification.

Chemical modification. Operator RNA was modified with DEPC, dimethylsulphate (DMS) or 1-cyclohexyl-3-(2-morpholino-ethyl)-carbodiimide metho-*p*-toluenesulphonate (CMCT) as described (22). Modified RNAs (<10 pM) were resuspended in 475 μl TMK buffer and incubated on ice for 10 min before addition of MS2 coat protein to a final concentration of 1 μM. After incubation for a further 2 h on ice the mixtures were filtered through nitrocellulose filters (Millipore HAWP 0.45 μM) for 5 s with the filtrate being collected in sterile, siliconized test tubes. RNA bound to the filter in RNA–coat protein complexes was extracted with 200 μl 10 mM Tris–HCl, pH 8.0, containing 0.1% (w/v) SDS and 200 μl phenol/chloroform mixture. RNA from both the filter and the filtrate was then precipitated by addition of sodium acetate to 0.3 M and 3 vol ethanol. The RNA modified with DEPC was cleaved with aniline; CMCT- and DMS-treated RNAs were analysed by primer extension.

Primer extension.

(i) Hybridization of the primer TR transcript RNA was resuspended in 3 μl containing (0.5 pmol) 5′-^{32}P of a DNA primer complementary to the 3′-end of the transcript and 1 μl 5× hybridization buffer (250 mM HEPES–KOH, pH 7.0, 250 mM potassium borate, 0.5 M KCl). The mixture was incubated at 90°C for 60 s and then allowed to cool slowly to 42°C over 60 min.

(ii) Extension. The extension mixture contained 1 μl DNA–RNA hybrid, 1 μl 100 mM DTT, 2 μl 2.5× extension mix (125 mM Tris–HCl, pH 8.5, 125 mM KCl, 25 mM $MgCl_2$, 250 μM each dGTP, dATP, dTTP and dCTP) and 2 U AMV reverse transcriptase. Dideoxy sequencing reactions (24) contained 1 μl

dideoxynucleotide stock (25 μM) in addition. After incubation for 30 min at 42°C, 1 μl chase mix (1 mM each dGTP, dATP, dTTP and dCTP) was added and incubation continued for 15 min at 42°C. Reactions were stopped by addition of 5 μl denaturing loading buffer.

Analysis of RNA fragments by polyacrylamide gel electrophoresis. RNA fragments resulting from binding interference experiments and DNAs from primer extensions were analysed by polyacrylamide gel electrophoresis. Samples were heated to 90°C for 60 s and quenched on ice immediately prior to loading onto a 40 cm long 20% (w/v) polyacrylamide–7 M urea gel (0.5 mm thick) and electrophoresed at 30 mA constant current for 3–4 h, as described by Sanger and Coulson (25). Enzymatic RNA sequencing ladders (21) were electrophoresed in parallel in order to assign the bands of chemically/enzymatically cleaved RNA. The gels were autoradiographed at –70°C for 1–2 days.

Analysis of autoradiograms. Relative intensities of bands on the autoradiograms were either determined directly by excision of radioactive material from the gel followed by counting in a liquid scintillation counter or by laser densitometry. In all cases the intensity of a band at a particular position was determined relative to the amount of full-length material at the top of the gel.

RESULTS AND DISCUSSION

Synthesis, modification and characterization of operator variants

Previously we have shown that 2′-deoxyadenosine substitutions at –4, –7 or –10 are silent in terms of coat protein affinity (9). At each site, therefore, we produced variants containing 2′-deoxyinosine (dI), purine-2′-deoxyriboside (dPu) and 2′-deoxy-7-deazaadenosine (d7deazaA) to examine recognition of both exocyclic and ring functional groups. Substituents at position –5 were aimed initially at trapping the proposed Michael adduct and later at dissecting the role(s) of the exocyclic groups at pyrimidine ring positions 4 and 5. The variants used were: 2′-deoxy-uridine (dU), 2′-deoxy-thymidine (T, d5MeU), 2′-deoxy-5-fluorouridine (d5FU), 2′-deoxy-5-bromouridine (d5BrU), 2′-deoxy-5-iodouridine (d5IU), 2′-deoxy-4-thiothymidine (4ST, d4SMeU), ribo-5-bromouridine (5BrU), ribo-5-cyanouridine (5CNU), ribo-4-thiouridine (4SU) and 2-pyrimidinone riboside (4HC). All the variants were bound by the coat protein under the assay conditions with the exception of the A –10 deletion, which did not bind at all, confirming that the complexes formed by the other variants were sequence-specific. The complex formed with the wild-type operator was also resistant to competition with excess tRNA (9). The relative affinities of the variant operators used are listed in Table 1.

In addition to the synthetic variants, we used binding interference assays to probe recognition at positions –4, –5, –7 and –10. Wild-type operator fragments produced by *in vitro* transcription were modified briefly using functional group-specific reagents, the modified RNAs recovered and used in filter binding assays with concentrations of coat protein expected to produce ~80% retention of input RNA, equivalent to 10 times the apparent equilibrium constant. Equivalent aliquots of unfiltered, filter-bound and filtrate (unbound) RNAs were recovered and the extent of modification at each site determined by direct cleavage or reverse transcriptase primer extention followed by denaturing polyacrylamide electrophoresis. Comparison of the relative band intensities in all three lanes allowed rough quantitation of the degree of interference (Fig. 2 and Table 2). At the protein concentrations used it is likely that minor interference effects would not be detected. The reagents used were DEPC, DMS and CMT. These react at N7 of adenine, N1 of adenine and N3 of uridine respectively (26).

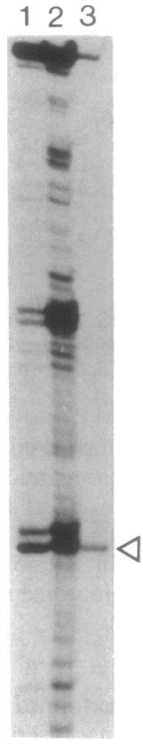


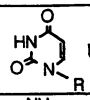
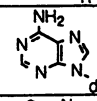
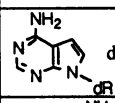
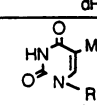
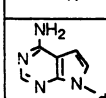
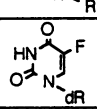
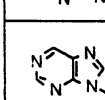
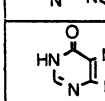
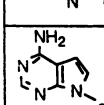
Figure 2. Binding interference of CMCT-modified RNA operators. Lane 1, unfiltered control RNA; lane 2, filter-bound RNA; lane 3, filtrate RNA. The positions of modification were identified by reverse transcriptase dideoxy sequencing of unmodified RNA followed by electrophoresis on the same gel. Interference effects at specific sites were estimated by normalizing the amount of radioactivity in each lane. The arrowhead indicates the position of U –5.

Recognition of the A –4 and A –10 residues

As the data in Table 1 show, removal of the 2′-OH group at positions –4 or –10 is silent in terms of coat protein affinity. However, substitution of the exocyclic NH_2 group or elimination of N7 at –4 have large effects on affinity. In contrast, the same set of modifications at –10 show only modest effects on affinity. These results are consistent with the binding interference data (Table 2), which show that modification of N1 at either site or N7 at –4 results in strong interference, whilst modification of N7 at –10 does not.

These results correlate reasonably well with expectations based on the coat protein–operator crystal structure (Fig. 3a–c). Although residues A –4 and A –10 are quasi-2-fold related and each is surrounded by the same amino acid residues on each subunit, the details of the interaction at each site differ. At A –4 the OH groups of ThrA45 and SerA47 make hydrogen bond contacts to the N7 and N1 ring positions respectively. The side chains of LysA61 and ValA29 help to define the edges of the

Table 1. Summary of coat protein affinity measurements for the variant RNA operators

Operator Tested	Modification	Relative Affinity	Operator Tested	Modification	Relative Affinity
WT	 U (-5)	1.0		 5BrU	11.1 and 1.3*,[1]
-4 variants	 dA	1.0[1]		 5CNU	0.2[2]
	 dPu	< 0.1		 4SU	0.8
	 dI	0.1		 4HC	0.5
	 d7deazaA	<< 0.1	-7 variants	 dA	1.0[1]
-5 variants	 C	6.0[1]		 dPu	0.5
	 dU	0.3[2]		 dI	0.1
	 T	0.1[2]		 d7deazaA	0.6
	 d5FU	< 0.2[2]	-10 variants	 dA	1.0[1]
	 d5BrU	0.6[2]		 dPu	0.6
	 d5IU	0.5[2]		 dI	0.5
	 4ST	<< 0.1[2]		 d7deazaA	0.7
				Δ	nd[1]

Δ indicates the position of nucleotide delection; R, ribose; dR, deoxyribose; nd indicates that binding was not detectable; *binding was multi-phasic
[1,2]As described previously in Talbot *et al.* (9) or Stockley *et al.* (28) respectively.

binding pocket. The exocyclic amino group points into this pocket, interacting with oxygen atoms in the protein. This correlates well with the low affinities of dPu and dI, where the amino group is replaced or eliminated. Elimination of N7 removes the possibility of forming the hydrogen bond to ThrA45, which results in still lower affinities. At A –10 the orientation of the base is different and the exocyclic amino group, as well as N7, points away from the pocket. This explains the small effects of substitutions and modifications at these positions. The N1 of A –10 also forms a hydrogen bond, but with ThrB45. The binding interference data are consistent with this. The side chain of SerB47 appears to be involved in a water-mediated hydrogen bond to the 2′-OH position of the sugar, whilst the ε-amino group of LysB61 also makes a contact to the 5′-phosphate oxygen of the same nucleotide. It is less easy to rationalize the affinity of dA –10. It is possible that the water-mediated hydrogen bond to the 2′-OH group at –10 either does not form in solution or contributes very little to the free energy of complex formation. The reduced affinities may therefore reflect altered interactions with solvent molecules or changes in the conformational equilibrium between liganded and free forms of the operator (22).

Recognition at A –7

At –7 substitution by dA is silent, whilst the dPu and d7deazaA variants are mildly deleterious. dI substitution leads to a 10-fold drop in affinity and modification at both N1 or N7 interferes with binding. The explanation for these effects is not readily apparent

from inspection of the crystal structure (Fig. 3c). The A –7 residue is tightly stacked between C –5 and G –8, which presumably helps to orientate the base at C –5. It is not contacted directly by the protein, suggesting that deleterious effects of substitutions may be due to disruption of solvent structure or displacement of the conformational equilibrium of the unliganded RNA. The modification data are consistent with the structural data, since the bulkier reaction products could not easily be accommodated in the stacked structure. DEPC modification also leads to opening of the adenine five-membered ring, which would reduce the ability of the base to stack with its neighbours.

Table 2. Summary of binding-interference data

Residue	Atom modified	Interference with CP binding
A –17	N7	–
	N1	–
A –16	N7	–
	N1	–
A –15	N7	–
	N1	–
A –10	N7	–
	N1	++
A –7	N7	++
	N1	+
U –6	N3	–
U –5	N3	++
A –4	N7	+
	N1	++

The table lists the coat protein (CP) binding interference effects observed. Data were analysed by densitometry of the appropriate lanes from the autoradiograph and compared with unfiltered controls. –, No interference; +, moderate interference; ++, strong interference of binding.

Recognition at C(U) –5

Table 1 lists the affinities of a series of variants at the –5 position. Substitution of dU had only a modest effect on the affinity for coat protein, consistent with the observation in the crystal that all of the interactions at –5 involve contacts to the base or phosphate backbone, rather than the sugar residue. Previously we reported that the 5BrU –5 variant, which would have been expected to trap a Michael adduct, produced an RNA variant having a temperature-dependent, biphasic binding curve with ~20% of the sample having an affinity significantly higher than the C –5 derivative (9), a result incompatible with formation of a covalent bond. It has been argued that the bulky bromine atom might have introduced additional, favourable van der Waals contacts with the protein. Similar increases in affinity were not observed with 5CNU –5. Although 5CNU –5 retains an electron-withdrawing function at C –5, the cyano group is relatively long and narrow, providing less radial bulk than the bromine atom. In the case of T –5 (d5MeU –5), the methyl group provides a large radial bulk, but has electron-donating properties. The affinity values of the 5-halogenated derivatives suggest that the bromine atom provides the optimum size for van der Waals contacts. One consequence of

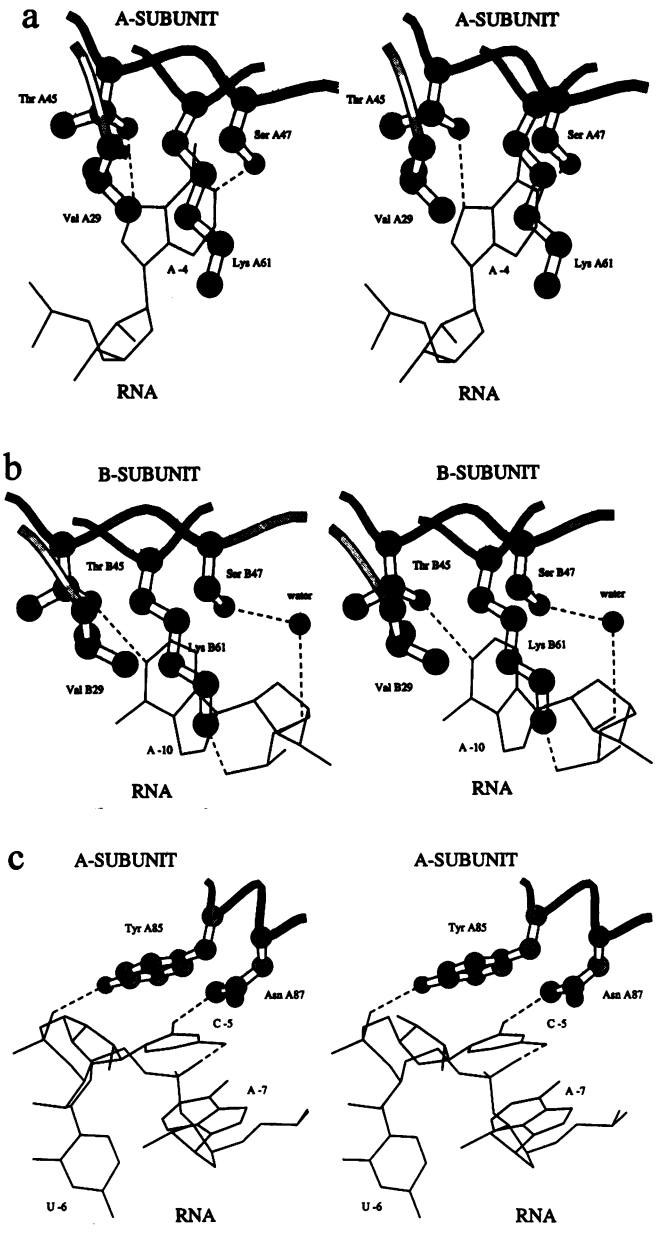


Figure 3. Molscript (33) diagrams showing sequence-specific RNA–protein interactions. A, adenine; C, cytidine. (**a**) At A –4 the hydroxyl groups of ThrA45 and SerA47 make hydrogen bond contacts (dashed lines) to the N7 and N1 positions of the base respectively. (**b**) At A –10 the orientation of the base is different. The hydroxyl group of ThrB45 makes a hydrogen bond to N1. The hydroxyl group of SerB47 makes a water-mediated hydrogen bond to the 2′-OH position of the sugar. LysB61 also makes a contact to the 5′-phosphate of the same nucleotide. (**c**) There is a hydrogen bond contact to the O2 of C –5 from the amide of AsnA87, whilst the N4 of C –5 and the side chain of TyrA85 are hydrogen bonded to the O1P group of U –6 and O2P of C –5 respectively.

halogenation at C –5 might be to alter the tautomeric equilibrium between keto and enol forms of the base, an effect which has been reported in model compounds (27). The 4ST, 4SU and 4HC derivatives were produced to examine this possibility. Their relatively low affinities suggest that the identity of the exocyclic group is indeed important and led us to suggest that recognition might be due to hydrogen bond formation by this group via

donation (28). Hence the higher affinity of C versus U, the latter presumably binding in its rare tautomeric form. This is consistent with the contacts seen in the C –5 operator complex, where the exocyclic NH_2 group hydrogen bonds to a neighbouring phosphate oxygen and presumably helps orientate the base so that it can readily stack between A –7 and TyrA85. The derivatives at C –5 would thus have affinities which are a combination of the ease of formation of the enol tautomer and the ability of the base to stack on the neighbouring aromatic rings, making a direct test of this hypothesis difficult. However, if the wild-type operator fragment binds to empty protein shells in a similar way to the C –5 variant, it should be possible to test this idea directly by determining the three-dimensional structure of complexes and measuring the distance between the –6 phosphate oxygen and –5 uridine C-O groups.

The coat protein–operator structure contains the high affinity C –5 variant, which is stacked between TyrA85 and A –7. There is a direct hydrogen bond between the O2 of the pyrimidine ring and the amide of AsnA87, whilst the exocyclic NH_2 group is hydrogen bonded to the O1P group of U –6 (Fig. 3c). The side chain of TyrA85 is hydrogen bonded to the O2P group of C –5. There is no covalent adduct to this residue, as previously discussed (1). The side chain of Cys46 is actually on the opposite side of the β-sheet from the face which binds to the RNA. Since the residues on either side of Cys46 (Thr45 and Ser47) form key elements of the recognition pockets for A –4 and A –10, it is plausible to attribute the chemical sensitivity of the protein to thiol reagents to local disruption of the conformation of the polypeptide chain in this region. This is consistent with the results of experiments with mutant coat proteins in which both protein Cys residues (46 and 101) were substituted by residues incapable of Michael adduct formation. Both *in vivo* (29–31) and *in vitro* (32,28) such mutants are still able to bind the operator RNA and to discriminate in favour of cytidine. Clearly this discrimination cannot be the result of the contact from AsnA87, suggesting that the interaction of the exocyclic amino group at C4 is the critical one.

Conclusions

Although crystallographic data allow the intermolecular contacts in complexes to be identified, they cannot be used to determine the importance of particular contacts for assembly or overall stability. Such data come from solution binding studies, such as those reported here, guided by the crystallographic results. In this case, the crystals of the operator–capsid complex were obtained by adding TR fragments to preformed crystals of RNA-free, recombinant capsids (10). The capsids are composed of 180 copies of the coat protein subunit organized into a $T = 3$ quasi-equivalent surface lattice formed from two types of non-covalent dimer, A/B and C/C. Remarkably, the TR fragments were able to penetrate to the centre of the capsid, where they bound to all the dimers in the shell. Binding to C/C dimers was in either orientation, as expected, but binding to A/B dimers was in a unique orientation allowing the details of the sequence-specific RNA–protein interaction to be seen in the resultant electron density maps. Although the protein subunits in A/B dimers are conformationally distinct, the principal differences are relatively remote from the RNA binding site, which appears at the present resolution (3.0 Å) to be identical in both subunits. Thus it is not possible to understand the specificity of binding from the structural data alone.

The functional studies in solution presented here complement the structural data and show unambiguously that the complex seen in the crystal is the same as the biologically relevant one, i.e. the one that forms in solution between TR and a single coat protein dimer. In summary, substitution or modification of RNA functional groups involved in direct hydrogen bond contacts with the protein were deleterious, with the exception of the water-mediated contact to the 2′-OH at –10. Large effects on relative binding affinity (>10-fold reduction in affinity compared with wild-type) were observed when direct contacts were eliminated. Milder effects (<10-fold reduction in affinity compared with wild-type) could not be interpreted in terms of the interactions seen in the crystal and may be the result of local disruptions to the operator structure, to effects on the conformational equilibria at the loop residues and at –10 or simply to experimental error. The coat protein–operator crystal structure is consistent with our earlier proposal that recognition at position –5 involves hydrogen bond donation from the exocyclic atom/group at C4. The biphasic binding curve for 5BrU –5 was reported previously (9). We proposed then that its unusual affinities could be the result of the effects of halogen substitution on tautomer equilibrium or to the steric effects of the bulky bromine atom on a conformational equilibrium. Definitive explanations of the effects of substitutions at position –5 may well emerge from the crystal structure of the coat protein wild-type operator complex (Valegård *et al.*, in progress).

These data illustrate very clearly the subtlety of RNA–protein recognition events. For the MS2 translational repression complex to form, both the protein and the RNA stem–loop must adopt defined conformations from a number of possible states. Sequence specificity primarily involves single-stranded nucleotide residues which are located in separate binding pockets on the surface of the β-sheet of the coat protein subunits. Stacking interactions between bases and protein tyrosine side chains and contacts to the phosphodiester backbone from both main chain and side chain groups are also important. Since the origin of the unique TR interaction with A/B dimers is still unknown, it is not yet clear how the TR–coat protein complex can catalyse the self-assembly of the phage shell. NMR studies on the unliganded TR fragment in solution suggest that the RNA conformer seen bound to the protein in the capsid structure is only a minor proportion of the species present, suggesting that assembly must involve conformational change in both the RNA and the coat protein subunits.

Note added in proof

The refined crystal strucutre of the C –5 operator complex has recently been completed and reveals an additional RNA–protein interaction in the loop region. GluA63 is hydrogen bonded to the 2′-OH group at –5, consistent with the reduced affinity of the dU variant at this site (Table 1).

ACKNOWLEDGEMENTS

We thank Kay Collier and Dr Hemant K. Tewary for assistance with the purification and characterization of some of the RNA variants and Dr Colin W. G. Fishwick for helpful discussions about the altered tautomer hypothesis. We also thank Prof.

Bernard Connolly (University of Newcastle) for the gift of purine 2′-deoxyriboside phosphoramidite. This work was supported by the UK SERC, The Wellcome Trust and the Swedish Natural Science Research Council.

REFERENCES

1 Witherell,G.W., Gott,J.M. and Uhlenbeck,O.C. (1991) *Prog. Nucleic Acid Res. Mol. Biol.*, **40**, 185–220.
2 Sorger,P.K., Stockley,P.G. and Harrison,S.C. (1986) *J. Mol. Biol.*, **191**, 639–658.
3 Beckett,D. and Uhlenbeck,O.C. (1988) *J. Mol. Biol.*, **204**, 927–938.
4 Romaniuk,P.J. and Uhlenbeck,O.C. (1985) *Biochemistry*, **24**, 4239–4244.
5 Romaniuk,P.J., Lowary,P., Wu,H.N., Stormo,G. and Uhlenbeck,O.C. (1987) *Biochemistry*, **26**, 1563–1568.
6 Haneef,I., Talbot,S.J. and Stockley,P.G. (1989) *J. Mol. Graphics*, **7**, 186–195.
7 Uhlenbeck,O.C., Carey,J., Romaniuk,P.J., Lowary,P.T. and Beckett,D. (1983) *J. Biomol. Struct. Dynam.*, **1**, 539–552.
8 Berzin,V.M., Tsimanis,A.Y., Gren,E.Y., Zagorski,W. and Szafranski,P. (1981) *Bioorg. Khim*, **7**, 482–486.
9 Talbot,S.J., Goodman,S., Bates,S.R.E., Fishwick,C.W.G. and Stockley,P.G. (1990) *Nucleic Acids Res.*, **18**, 3521–3528.
10 Valegård,K., Murray,J.B., Stockley,P.G., Stonehouse,N.J. and Liljas,L. (1994) *Nature*, **371**, 623–626.
11 Walton,C. (1993) PhD Thesis, University of Leeds, Leeds, UK.
12 Mastico,R.A., Talbot,S.J. and Stockley,P.G. (1993) *J. Gen. Virol.*, **74**, 541–548.
13 Sugiyama,T., Hebert,R.R. and Hartman,K.A. (1967) *J. Mol. Biol.*, **25**, 455–463.
14 Layne,E. (1957) *Methods Enzymol.*, **3**, 447–454.
15 Carey,J. and Uhlenbeck,O.C. (1983) *Biochemistry*, **22**, 2610–2615.
16 Goodman,S.T.S. (1992) PhD Thesis, University of Leeds, Leeds, UK.
17 Adams,C.J., Murray,J.B., Arnold,J.R.P. and Stockley,P.G. (1994) *Tetrahedron Lett.*, **35**, 765–768.
18 Adams,C.J., Murray,J.B., Arnold,J.R.P. and Stockley,P.G. (1994) *Tetrahedron Lett.*, **35**, 1597–1600.
19 Connolly,B.A. (1991) In Eckstein,F. (ed.), *Oligonucleotides and Analogues: A Practical Approach*. IRL Press, Oxford, UK, pp. 155–183.
20 Murray,J.B., Collier,A.K. and Arnold,J.R.P. (1994) *Anal. Biochem.*, **218**, 177–184.
21 Donis-Keller,H., Maxam,A. and Gilbert,W. (1977) *Nucleic Acids Res.*, **4**, 2527–2538.
22 Talbot,S.J., Medina,G., Fishwick,C.W.G., Haneef,I. and Stockley,P.G. (1991) *FEBS Lett.*, **283**, 159–164.
23 Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
24 Sanger,F.G., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
25 Sanger,F.G. and Coulson,A.R. (1979) *FEBS Lett.*, **87**, 107–110.
26 Ehreshmann,C., Baudin,F., Mougel,M., Romby,P., Ebel,J.P. and Ehreshmann,B. (1987) *Nucleic Acids Res.*, **15**, 9108–9128.
27 Saenger,W. (1984) In *Principles of Nucleic Acid Structure*. Springer-Verlag, Berlin, Germany.
28 Stockley,P.G., Stonehouse,N.J., Walton,C., Walters,D.A., Medina,G., Macedo,J.M.B., Hill,H.R., Goodman,S.T.S., Talbot,S.J., Tewary,H.K., Golmohammadi,R., Liljas,L. and Valegård,K. (1993). *Biochem. Soc. Trans.*, **21**, 627–633.
29 Peabody,D.S. (1989) *Nucleic Acids Res.*, **17**, 6017–6027.
30 Peabody,D.S. (1990) *J. Biol. Chem.*, **265**, 5684–5689.
31 Peabody,D.S. and Ely,K.R. (1992) *Nucleic Acids Res.*, **20**, 1649–1655.
32 Macedo,J.M.B. (1992) PhD Thesis, University of Leeds, Leeds, UK.
33 Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.