Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 9244–9249, August 1998
Biochemistry

A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein

(coliphage/phosphorodithioate/methylphosphonate/phosphorothioate/RNA–protein interaction)

HANS E. JOHANSSON*†, DAGMAR DERTINGER*, KAREN A. LECUYER‡, LINDA S. BEHLEN, CHARLES H. GREEF§, AND OLKE C. UHLENBECK¶

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215

Contributed by Olke C. Uhlenbeck, May 29, 1998

ABSTRACT **Most mutations in the sequence of the RNA hairpin that specifically binds MS2 coat protein either reduce the binding affinity or have no effect. However, one RNA mutation, a uracil to cytosine change in the loop, has the unusual property of increasing the binding affinity to the protein by nearly 100-fold. Guided by the structure of the protein–RNA complex, we used a series of protein mutations and RNA modifications to evaluate the thermodynamic basis for the improved affinity: The tight binding of the cytosine mutation is due to (*i*) the amino group of the cytosine residue making an intra-RNA hydrogen bond that increases the propensity of the free RNA to adopt the structure seen in the complex and (*ii*) the increased affinity of hydrogen bonds between the protein and a phosphate two bases away from the cytosine residue. The data are in good agreement with a recent comparison of the cocrystal structures of the two complexes, where small differences in the two structures are seen at the thermodynamically important sites.**

The bacteriophage MS2 coat protein specifically binds to a small RNA hairpin in its genomic RNA (1–4). The complex represses translation of the early replicase gene and identifies the RNA for future packaging into virus particles by forming an initiation complex. This protein–RNA interaction has been extensively studied as a model for the rapidly expanding class of proteins that bind RNA hairpins (5–9). During the biochemical dissection of the MS2 RNA binding site, one of the uracil residues in the loop was identified as a major determinant of sequence-specific binding. When this uracil was replaced with a purine, binding affinity was 10- to 100-fold weaker (10). Surprisingly, when the uracil was changed to a cytosine, binding was 50- to 100-fold tighter, primarily because of a slower dissociation rate (11). Experiments with RNAs containing modified pyrimidines at this site confirmed that the exocyclic amino group of cytosine is critical for the tighter binding (2). An early explanation was that a cysteine in the protein formed a transient covalent Michael adduct with the RNA at this site (12). However, this model was not supported by the fact that the 5-fluorouracil-containing hairpin bound normally and that no ^{3}H exchange from a labeled uracil-containing hairpin was observed (13). In addition, mutation of either of the two cysteines in the protein to serine had no effect on binding (14).

Potential sources of binding energy for this interaction have been identified from the crystal structure of the capsid bound to the RNA hairpin (15). The critical cytosine residue is stacked between an adenine and a tyrosine and its exocyclic amino group forms a hydrogen bond with a phosphate (Fig. 1).

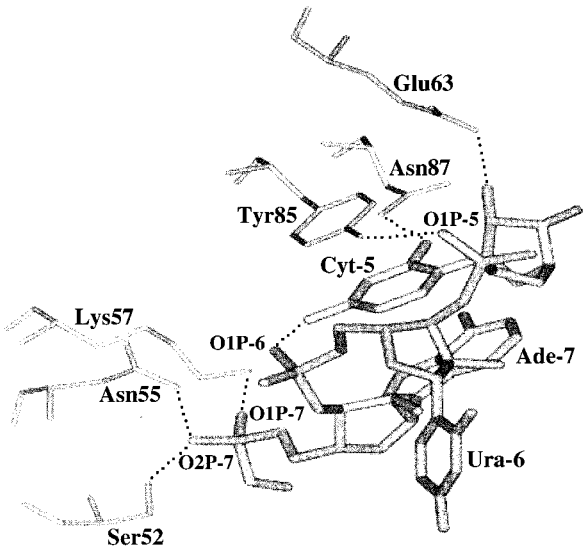

FIG. 1. Structure in the neighborhood of residue −5 of the MS2 replicase operator in the crystal structure of the cytosine-containing RNA (C-RNA)–coat protein complex (15, 16).

This finding led to the suggestion that the weaker binding of the uracil-containing RNA (U-RNA) was because it could bind only as a hydroxyl in the rare enol tautomer (2). However, a recent comparison of the cocrystal structures of hairpins containing cytosine or uracil bound to the protein revealed that although the structures were very similar, small differences in the position of this particular RNA phosphate relative to the protein placed the uracil residue too far from the phosphate to make the tautomerization hypothesis likely (16). As the relative contributions of the individual RNA–protein contacts to the overall free energy of binding are unknown, the thermodynamic basis for the sequence specificity at this site remains unclear. In this work, a series of protein mutants and modified RNAs are used with the goal of understanding how the sequence specificity at this site is achieved.

MATERIALS AND METHODS

All experiments presented have been performed with MS2 coat proteins carrying a double mutation [Val-75 → Glu

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*H.E.J. and D.D. contributed equally to this work.

†Present address: Department of Medical Genetics, Biomedical Center, Uppsala University, Box 589, Uppsala, Sweden S-75123.
‡Present address: Department of Biochemistry, University of Connecticut Health Science Center, Farmington, CT 06032.
§Present address: Nanogen, 10398 Pacific Center Court, San Diego, CA 92121.
¶To whom reprint requests should be addressed. e-mail: Olke.Uhlenbeck@Colorado.edu.

(Val75Glu) and Ala-81 → Gly (Ala81Gly)] that prevents aggregation but does not affect RNA binding (17). Cloning, overexpression, and purification of this protein and the coat proteins additionally containing the Tyr-85 → Phe (Tyr85 Phe), Tyr-85 → His (Tyr85 His), and Tyr-85 → Ser (Tyr85 Ser) mutations have been described (18). The Asn-87 → Ala (Asn87Ala) and Glu-63 → Gln (Glu63Gln) mutations were prepared in the same way as the Tyr-85 mutations.

Transcription of RNAs *1–4* was performed as described earlier (19), using [α-^{32}P]CTP. RNAs *5–8* were synthesized by using phosphoramidites from Glen Research (Sterling, VA) and Chem-Genes (Waltham, MA). The deprotection was done as in Usman *et al*. (20), but TE (10 mM Tris·Cl, pH 8.0/1 mM EDTA) was used instead of 0.05 M NH_4OAc for the quenching of tetra-*n*-butylammonium fluoride. Specific phosphorothioate linkages in RNAs *7* and *8* were introduced during the standard synthesis by using Beaucage reagent according to Glen Research (ref. 21 and Glen Research Technical Bulletin). The two stereoisomers were purified by HPLC on a C_{18} column. The earlier-eluting isomer corresponds to the R_P phosphorothioate and the late-eluting isomer to the S_P phosphorothioate (22–24). The position of the phosphorothioate linkage was confirmed by specific cleavage at that phosphate in 1 mM I_2 during a 5-min incubation at room temperature (25). Synthesis and HPLC purification of RNAs *7* and *8* containing a phosphorodithioate at −6 were carried out as described in Greef *et al*. (26). Synthesis and purification of RNAs *7* and *8* containing a methylphosphonate linkage were performed as described in Pritchard *et al.* (27) by using a deoxymethylphosphonate synthon. Again the first peak in the HPLC run corresponds to the R_P isomer, and the second, to the S_P isomer, as described earlier (27–30). Confirmation of methylphosphonate incorporation at the intended location was achieved by partial alkaline hydrolysis (29), but with a buffer consisting of 50 mM $NaHCO_3/Na_2CO_3$ (pH 9), 1 mM EDTA, and 0.25 mg/ml *Escherichia coli* tRNA. The methylphosphonate linkage can, by this method, easily be identified by two cleavage products arising at the site of incorporation, one with a 3′-OH group and one with a 3′ singly charged methylphosphonate group.

Synthetic RNAs were 5′-^{32}P-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP or 3′-^{32}P-labeled with T4 RNA ligase and 5′-[^{32}P]pCp. The labeled RNAs were purified by electrophoresis on denaturing 20% polyacrylamide gels, followed by passive elution from the gel matrix into distilled water. Integrity of the RNAs after purification was confirmed by electrophoresis.

Protein excess filter binding assays were used to obtain equilibrium dissociation constants (K_d). Reaction mixtures were incubated in microtiter dishes and then filtered on a modified 96-well filter apparatus (31) with filters from Schleicher & Schuell (Nitro-ME) or Millipore (Multiscreen MAHA N45). Typical binding curves contained <10 pmol of labeled RNA and sequential 1.8-fold dilutions of coat protein. The binding buffer was 100 mM K–Hepes, pH 7.5/10 mM $MgCl_2$/80 mM KCl unless noted otherwise. For many of the complexes involving C-RNAs, the K_d was <1.0 nM. In these cases, protein excess binding experiments yield unreliable K_d values because at the very low protein concentrations required to obtain a complete curve, the protein loses activity (11). In these cases, protein binding affinity was obtained by first determining k_{off}. An equilibrium complex was formed somewhat above K_d, a large excess of nonradioactive C-RNA was added, and aliquots were filtered at time intervals (11). The K_d was then calculated from k_{off} using $k_{on} = 1.0 \times 10^8\ M^{-1}\cdot min^{-1}$, the average of several very similar values of k_{on} determined with a variety of RNAs that bind to MS2 coat protein (11, 32, 33). For several of the modified RNAs, k_{on} was redetermined and found to be within 2-fold of the previously published values. In those cases where the K_d was >1.0 nM, the K_d value calculated from an experimental k_{off} was in good agreement with the value determined directly.

RESULTS

Structures. The local environment of the residues under scrutiny in this study as deduced from the crystal structure of the protein–RNA complex (15) is shown in Fig. 1. The critical uracil or cytosine residue at position −5 is stacked between the adenine at position −7 and Tyr-85. The hydroxyl of the tyrosine is held in place by a hydrogen bond to the O1 of phosphate −5. The uracil or cytosine ring is held in place by the interaction of the exocyclic 2-oxygen with the amide of Asn-87. The cytosine exocyclic 4-amino group also forms a hydrogen bond with O1 of phosphate −6. The adenine ring of nucleotide −7 in both complexes does not make any hydrogen bonds, but is stacked on the first base pair of the hairpin helix. However, the O1 of phosphate −7 makes a charged hydrogen bond with Lys-57, and O2 makes hydrogen bonds with Asn-55 and Ser-52. An additional nearby interaction, which may be important for the specificity of binding, is the hydrogen bond between the 2′ hydroxyl of ribose −5 and Glu-63.

The comparison of the refined x-ray crystal structures of the cytosine −5 and uracil −5 complexes revealed virtually identical structures (16). While the position of residue −5 with respect to the protein is identical in the two complexes, the positions of phosphates −6 and −7 in the U-RNA are shifted ≈0.5 Å away from the pyrimidine ring compared with the C-RNA. As a result, the distance between the 4-amino nitrogen of cytosine −5 to O1 of phosphate −6 is 2.8 Å, whereas the corresponding distance from the 4-carbonyl oxygen of uracil to O1 is 3.3 Å. Thus, cytosine −5 can make a direct hydrogen bond, whereas uracil −5 cannot. This is the only difference in the hydrogen bonding pattern between the two structures. The shift in position of phosphate −7 between the uracil −5 and cytosine −5 complexes results in a small difference in the position of the side chain of Lys-57 that hydrogen bonds to it. However, the positions of the side chains of Ser-52 and Asn-55, which also hydrogen bond to phosphate −7, do not change. Several additional subtle differences exist between the two complexes, including a change in the position of Arg-49 and a change in the length of the hydrogen bond between Lys-43 and phosphate −4.

The environment of cytosine −5 in the structure of the free RNA determined by NMR is very different from that in the complex (34, 35). The cytosine ring no longer stacks on the adenine −7 but protrudes into the solvent. No data are available for the structure of the uncomplexed hairpin containing a uracil at position −5, although such a structure would be expected to be similar to the C-RNA.

RNAs. Four different RNA hairpin background sequences are used in this study (Fig. 2). Although their lengths and base pair sequences differ, they all contain the consensus sequences needed for tight protein binding (36). RNAs *1–4* have been used to study the complex (37, 38). Hairpin *4*, used to solve the structure of the free RNA by NMR (34), has a slightly different sequence and additional residues compared with hairpins *1–3*. However, these changes are not in the region that contacts the protein (15) and do not affect K_d (36). Thus, the essential difference between RNAs *1–4* is the identity of the nucleotide at position −5. RNAs *5* and *6* have also been used in earlier studies (39). RNAs *7* and *8* were designed to be the minimal substrate needed for protein binding. The small size facilitates separation of phosphorothioate and methylphosphonate stereoisomers. When these RNAs possess a 5′-phosphate, they contain all the contacts required for protein binding, but bind 3-fold tighter than the longer hairpins because of the increased affinity of Arg-49 to the 5′-terminal phosphomonoester (H.E.J., D.D., C.H.G., and O.C.U., unpublished data). As the dissociation rate constant of the 5′-[^{32}P]RNA *8* is inconve-

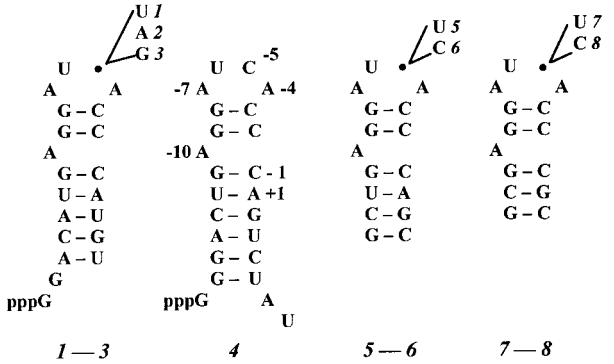

FIG. 2. RNA oligonucleotides used in protein binding experiments. Residue numbering (shown in *4*) is based on replicase gene (1).

niently slow ($t_{1/2} > 24$ hr), measurements with *7* and *8* variants were carried out with RNAs labeled on their 3′ end with 5′-[^{32}P]pCp and containing a 5′-terminal hydroxyl that slightly reduces affinities to the protein because of the missing contact between Arg-49 and phosphate −13.

Protein Mutations. To examine the contribution of Asn-87 to the sequence specificity, the Asn87Ala mutation was prepared. With a truncated side chain, the mutant protein is no longer able to make the hydrogen bond to the 2-oxygen of cytosine or uracil at position −5. Surprisingly, the mutant protein binds both pyrimidine-RNAs (*1* and *4*) with a K_d similar to that of the native protein (Table 1), suggesting that this contact does not contribute substantially to the binding energy of these RNAs. However, the Asn87Ala protein does bind to RNAs containing adenine and guanine at −5 (*2* and *3*) somewhat differently. As had been reported (10), the A- and G-RNAs bind the native protein less well (14-fold and 65-fold) than the U-RNA, presumably because of the difficulty in accommodating the bulky purine rings. However, the Asn87Ala protein binds the A- and G-RNAs somewhat better (3-fold and 5-fold) than the native protein, suggesting that at least part of the detrimental effect of the steric clash is mitigated by the smaller amino acid side chain. A similar conclusion was made by Lim *et al.* (40) on the basis of experiments with the Asn87Ser protein. However, this mutant protein binds the wild-type RNA with a much lower affinity, suggesting that the serine side chain makes a deleterious contact with the RNA.

The contribution of Tyr-85 to the sequence specificity was evaluated by using three available mutations at this site, Tyr85Phe, Tyr85Ser, and Tyr85His (18). While all three mutations substantially decrease the affinity to both the U- and C-RNAs (Table 1), the specificity for cytosine over uracil is reduced from 75-fold for the native protein to 12-, 27-, and 9-fold for the Tyr85Phe, Tyr85Ser, and Tyr85His mutations, respectively. In other words, mutation of the tyrosine ring has a larger detrimental effect on the affinity of C-RNA than U-RNA. The binding of the mutant proteins to the A- and G-RNAs was too weak to obtain accurate data.

Unlike the affinity of the native protein, that of the Tyr85His protein to U-RNA increases with decreasing pH until pH 6.5, at which point the affinity is similar to that of the native interaction (18). This observation was interpreted by proposing that the protonated form of histidine could make a hydrogen bond with phosphate −5 in a manner similar to the hydroxyl of tyrosine, thereby positioning the amino acid on top of the uracil ring. Interestingly, when the affinities of the C-, A-, and G-RNAs to the Tyr85His protein were determined at pH 6.5, they were very similar to the corresponding affinities of the native protein at pH 8.5. Thus, it appears that the contribution of residue 85 to the specificity is due not to its identity but to its ability to properly position a hydrogen bond donor.

RNA Backbone Modifications. It is possible that the hydrogen bond between Glu-63 and the 2′-hydroxyl of ribose −5 is contributing to the tighter binding of the C-RNA. This contact is near the part of the structure that differs between cytosine and uracil, and the hydrogen bond length is slightly (0.2 Å) shorter in the U-RNA structure. To evaluate the contribution of this contact to the specificity of U-RNA vs. C-RNA, the conservative Glu63Gln mutant protein was prepared. Surprisingly, this protein had virtually no binding activity with either C- or U-RNA. While the protein appeared normal during expression and purification in comparison with the wild-type protein, it may have an alternate conformation because of the disruption of the ion pair between Glu-63 and the nearby Arg-83 in the free protein that is highly conserved in *E. coli* RNA bacteriophage (41). As this protein mutation could not be used to assess the importance of this interaction on sequence specificity, the effect of removing the 2′-hydroxyl group of ribose −5 on the RNA was examined. As has been reported (39), the affinity of the −5 deoxyuridine-containing RNA is ≈50-fold weaker than the same sequence containing only ribose sugars. As shown in Table 2, the affinity of the −5 deoxycytidine-containing RNA is similarly reduced. Thus, it appears that the contact between Glu-63 and the 2′-hydroxyl of position −5 contributes little to the binding specificity of U-RNA vs. C-RNA.

The two phosphates −6 and −7, whose positions differed between the x-ray structure of the cytosine and uracil complexes, were chemically modified to assess their contribution to the binding specificity (Fig. 3). A control phosphate, −3, that does not contact the protein was also modified. As shown in Table 2, the modification of phosphate −7 of the U-RNA to a phosphorothioate has a small, but consistent, effect on the binding affinity. The R_P isomer shows ≈4-fold tighter binding than the uracil-modified RNA, whereas the S_P isomer binds 4-fold weaker. These effects are presumably a consequence of the poor hydrogen bonding capacity of sulfur and the asymmetric charge distribution of the phosphorothioate (42) affecting the interaction of Lys-57 with O1 and Asn-55 and Ser-52 with O2. In contrast, modification of phosphate −7 of the C-RNA to a phosphorothioate results in an entirely different effect on the binding affinity. The R_P isomer binds the coat protein with the same affinity as the unmodified RNA,

Table 1. Binding affinities of RNAs to coat protein mutants

	K_d, nM					
RNA	Native	Asn87Ala	Tyr85Phe	Tyr85Ser	Tyr85His	Tyr85His (pH 6.5)
1 (U)	4.6*, 2.8†	3.5, 1.7†	170*	300*	33*	2.6, 4.0†
2 (A)	66	23	$>10^3$	10^3	57	26
3 (G)	300	56	$>10^3$	$>10^3$	$>10^3$	200
4 (C)	0.038†	0.028†	7.5	11	3.5	0.045†

K_d values were measured in 80 mM KCl/10 mM $MgCl_2$/100 mM K–Hepes, pH 7.5, at 0°C except for Tyr85His (pH 6.5), for which 100 mM K–Mes, pH 6.5, substitutes for Hepes.
*From LeCuyer *et al.* (18).
†Calculated by dividing the experimental off-rate with an on-rate of 10^8 $M^{-1}\cdot min^{-1}$.

Table 2. Specificity of RNAs with backbone modifications

RNA molecule	Contacts affected by modification	Modified phosphate oxygen	K_d, nM U-RNA	K_d, nM C-RNA	Specificity
RNAs *5* and *6*					
Unmodified			10, 1.7*	0.04*	146
Deoxy −5	Glu-63		270, 23*	1.8*	81
RNAs *7* and *8*					
Unmodified			13, 4.1*	0.04*	214
R_P-thioP-3	None	O1P	19, 4.8*	0.04*	298
S_P-thioP-3	None	O2P	9.4, 5.7*	0.03*	252
R_P-thioP-7	Lys-57	O1P	2.7, 1.4*	0.05*	41
S_P-thioP-7	Asn-55, Ser-52	O2P	51, 15*	1.3*	25
S_P-methylP-7	Lys-57	O1P	88, 16*	26*	2.0
R_P-methylP-7	Asn-55, Ser-52	O2P	295, 34*	18*	9.1
R_P-thioP-6	Cytosine-5	O1P	0.6, 0.2*	0.02*	20
S_P-thioP-6	None	O2P	7.9, 2.1*	0.03*	167
DithioP-6	Cytosine-5	O1P, O2P	0.4, 0.06*	0.04*	5.8
S_P-methylP-6	Cytosine-5	O1P	17, 6.2*	23*	0.5
R_P-methylP-6	None	O2P	34, 9.2*	3.3*	6.5

K_d values were measured in 80 mM KCl/10 mM $MgCl_2$/100 mM K–Hepes, pH 7.5, at 0°C. Specificity is defined as K_d (U-RNA)/K_d (C-RNA) and was calculated by using an average of the kinetic and thermodynamic determinations of the U-RNA.
*Calculated by dividing the experimental off-rate by 10^8 $M^{-1}\cdot min^{-1}$.

whereas the S_P isomer binds 33-fold weaker. Thus, although the same three amino acids contact phosphate −7 in the cytosine-loop structure as in the uracil-loop structure, the effect of the phosphorothioate substitution in those two sequence contexts is different. This finding implies that these contacts do not contribute to the overall binding affinity of the two RNAs in the same way and therefore are partially responsible for the binding specificity.

The effect is much more dramatic when phosphate −7 is replaced by either isomer of methylphosphonate. As the ribomethylphosphonate linkage is not stable, a deoxyribomethylphosphonate linkage was inserted. Fortunately, the coat protein does not contact the 2′-hydroxyl of ribose −8 and substitution by a deoxyribose at this site does not alter the binding affinity (39). In the U-RNA, replacement of O1 by a methyl group weakens binding by ≈6-fold, whereas substitution of a methyl group for O2 weakens binding ≈20-fold. The complexity of the interactions at phosphate −7 prevents a detailed rationalization of these effects, but it is clear that the inability of the methyl group to be a hydrogen bond acceptor and the loss of the negative charge on the methylphosphonate are much more disruptive than the phosphorothioate substitutions. The methylphosphonate substitutions of phosphate −7 have even a greater effect on C-RNA binding. In this case, substitution of a methyl group for either O1 or O2 weakens the binding affinity by ≈500-fold, resulting in a nearly complete loss of specificity. In other words, substitution of a methylphosphonate for phosphate −7 eliminates sequence-specific binding at the distal −5 site. It could be that the methylphosphonate substitution causes an unexpectedly large rearrangement in the structure of the complex. However, because these data are consistent with the data from the more conservative phosphorothioate substitutions at the same site, this possibility is less likely. Thus, it appears that the RNA–protein contacts at phosphate −7 also contribute thermodynamically to the sequence specificity.

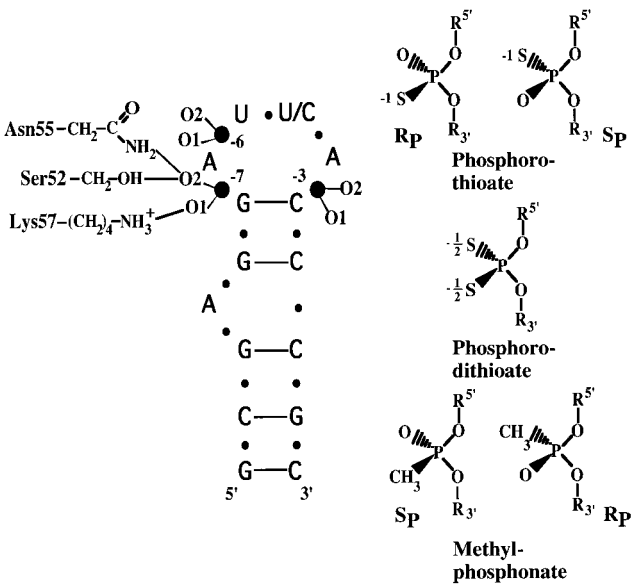

FIG. 3. Phosphate modifications of *7* and *8*. The sites of modification (●) and the contacts made with MS2 coat protein are indicated on the secondary structure. The chemical structures and stereochemistry of the phosphate modifications are shown on the right. Note that rules of nomenclature (46) require that when an oxygen is replaced by a sulfur to give the R_P phosphorothioate, the same oxygen substituted with a methyl group yields the S_P methylphosphonate.

When phosphate −3 is replaced with a phosphorothioate, little effect on the binding affinity is observed for either isomer for both U-RNA and C-RNA. As this residue does not contact the protein, this control experiment shows that phosphorothioate substitution does not intrinsically or indirectly alter binding affinity and specificity.

The phosphate modifications at position −6 also showed significant effects on the specificity. In the case of the U-RNA, the S_P phosphorothioate substitution has little, if any, effect on the affinity to the protein, whereas the R_P isomer increases binding affinity ≈20-fold. This tighter binding appears to be the result of the sulfur substitution rather than the asymmetric charge distribution of the phosphorothioate, because the introduction of a phosphorodithioate at this position shows a similar tighter binding affinity despite its symmetric charge distribution. The source of this tighter binding is unclear because the O1 of phosphate −6 does not appear to make any direct contacts in the U-RNA crystal structure. Perhaps a nearby amino acid such as Asn-55 contributes to the binding energy in the modified RNA, but not in the unmodified RNA. The corresponding substitution of the two phosphorothioate isomers at −6 in the C-RNA shows only slightly tighter binding

than the control RNA. The phosphorodithioate-substituted analogue binds as well as the control RNA. Thus, the added stability resulting from the O1 sulfur substitution observed in the U-RNA is not seen for the C-RNA. This may be because of the different positioning of the two RNAs on the protein or because the amino group of the cytosine in some way counteracts the effect of the sulfur substitution. In any case, the decrease in specificity caused by the O1 sulfur substitution of phosphate −6 reflects selective stabilization of the U-RNA by the sulfur by some unknown mechanism and not by any effect due to the amino group of cytosine.

In contrast to the sulfur substitutions, both isomers of the methylphosphonate analogue at position −6 show relatively little effect on the affinity of the U-RNA, which is consistent with the absence of any direct contacts with this phosphate or the 2′ hydroxyl group of ribose −7 in the crystal structure of the complex. However, both methylphosphonate isomers greatly reduce the affinity of the C-RNA such that the specificity is virtually lost. Presumably, the amino group of cytosine cannot form a hydrogen bond to the methyl group in the R_P isomer and the strength of a potential hydrogen bond to the uncharged oxygen in the S_P isomer is expected to be greatly reduced. Thus, it appears that part of the specificity for protein binding is due to the intramolecular hydrogen bond in the RNA.

DISCUSSION

In an attempt to understand the sequence specificity at a single site of an RNA–protein interaction, we have prepared protein mutations and RNA modifications and determined the K_d of the interactions. As is always the case in such a structure–function study, there is a concern that complexes involving mutant proteins or modified RNAs may have significantly different structures involving different hydrogen bonding patterns that would greatly complicate the interpretation of the binding data. This kind of rearrangement has been seen in the structure of mutants of the T4 lysozyme (43) and mutants of an RNA hairpin (44). In the absence of cocrystal structures of the modified complexes, the possibility of altered hydrogen bonding patterns cannot be entirely discounted. However, the possibility has been minimized by choosing mutations and modifications that are chemically modest, often only involving either single atom changes or functional group deletions.

A related concern in performing structure–function studies on an intermolecular complex is that the mutant proteins or modified RNAs may adopt a different structure in their unbound state and thereby indirectly affect the binding affinity. Fortunately, all of the relevant amino acid side chains protrude from the β-sheet on the inside of the virus and do not appear to be constrained. Indeed, in one structure of MS2 coat protein, the orientation of the side chain of Tyr-85 is somewhat different (45). Similarly, the relevant nucleotide at position −5 and phosphates at −6 and −7 protrude into the solvent in the free RNA. Thus, it is reasonable to expect that the effects of mutations and modifications could primarily be understood in terms of the structure of the protein–RNA complex.

From experiments described here and elsewhere, it is now possible to understand how this part of the native U-RNA–coat protein complex contributes to the overall binding affinity. Although the crystal structure shows that uracil −5 is held in place by the 2-carbonyl, forming a direct hydrogen bond to Asn-87, RNA modification and protein mutagenesis experiments suggest that this interaction is thermodynamically insignificant. Disrupting the 2-carbonyl interaction by replacing the asparagine with an alanine has little effect on RNA binding. As expected, disrupting the 4-carbonyl interaction by either changing the oxygen to a proton or a sulfur (2) or changing phosphate −6 to a methylphosphonate does not change protein binding affinity. Thus, uracil −5 appears to contribute to the binding energy primarily by stacking with Tyr-85 above and adenine −7 below. The binding affinity is not greatly affected by the stacking partners. Modification of adenine −7 to a purine or hypoxanthine only slightly affects protein binding (2), and even the Tyr85His protein binds with similar affinity, provided that the histidine ring is protonated to make the hydrogen bond with phosphate −5 (18). Additional nearby contacts between the protein and the RNA that were identified to contribute to the U-RNA binding affinity in this region are the 2′-hydroxyl of ribose −5 with Glu-63 (39) and the O1 and O2 oxygens of phosphate −7 with Lys-57, Asn-55, and Ser-52.

Protein mutagenesis provides some insight into the ability of MS2 coat protein to discriminate against RNAs with a purine at −5. Converting Asn-87 to Ala improves the ability of both purine-RNAs to bind MS2 coat protein, but they still bind much less tightly than the pyrimidine-RNAs. Presumably, the small side chain can more easily accommodate the bulky purine ring. As pointed out in Lim *et al.* (40), this explanation may also account for the presence of a serine at the corresponding position in the closely related coat protein of bacteriophage GA, whose operator contains adenine at −5. However, in the absence of structural data, it is unclear exactly how the purine ring is accommodated in the −5 site.

As proposed by Valegård *et al.* (16), the much tighter binding of the C-RNA is partly due to the intramolecular hydrogen bond between the amino group of cytosine and phosphate −6. When this bond is disrupted by introducing either methylphosphonate isomer at −6, the binding affinity of the C-RNA is reduced. As the same modification does not greatly affect the U-RNA, it is clear that the hydrogen bond contributes significantly to specificity. However, several additional modifications differentially affect the binding of C-RNA, including phosphorothioate and methylphosphonate modifications of phosphate −7 and, to a lesser extent, modifications of Tyr-85. This observation suggests that although the C-RNA and U-RNA bind the protein in very similar ways, the manner in which the individual contacts contribute to the total binding free energy is different. In other words, the binding specificity is achieved by differences in the strength of a number of contacts in the interacting surfaces and not simply at the site where the mutation occurs.

In general, the biochemical data identifying the thermodynamic source of the sequence specificity agreed quite well with interactions identified in the complex structure. The parts of the structure where the largest differences between the C-RNA and U-RNA complexes are observed (phosphates −6 and −7) are also the sites that significantly contribute to the specificity. In contrast, disruption of the nearby hydrogen bond between the 2′-hydroxyl of ribose −5 and Glu-63, which was similar in the two structures, showed no effect on the binding specificity of U-RNA vs. C-RNA. However, it is important to point out that not all the contacts between the protein and the RNA have yet been tested biochemically for their contribution to the specificity. It is possible that other sites of contact may contribute to the tighter binding of the C-RNA. One candidate is the nearby ion pair between Lys-43 and phosphate −4 because the distance is 0.5 Å closer in the C-RNA than in the U-RNA. Even more distal contacts such as those between phosphates −10 and −11 and Lys-61 may contribute to specificity despite the fact that no large structural differences are observed. A more complete thermodynamic understanding of the specificity, therefore, awaits the accumulation of a larger set of protein mutations and RNA modifications.

It is interesting that the tighter binding of the C-RNA to MS2 coat protein is not entirely a result of different contacts between the RNA and the protein, but is also partly due to stabilization by a hydrogen bond between two residues within the RNA molecule. This fact suggests that the change of the binding equilibrium may partially be due to a thermodynamic

effect on the free RNA. Although several possible explanations can be proposed, perhaps the simplest is that the uracil at −5 is quite dynamic in the free RNA and exerts an unfavorable entropy change upon the protein binding equilibrium. In the case of the cytosine, the intramolecular hydrogen bond may help organize the free RNA into the structure more amenable to protein binding, thereby reducing the unfavorable entropy change. The available NMR structure of the free C-RNA (34, 35) does confirm that this residue is dynamic, but shows no indication of a close contact between the amino group and the phosphate. Presumably, the hydrogen bond is not sufficient to fully organize the structure in the absence of protein. It would, therefore, be interesting to compare the structure and dynamics of the uracil −5 and cytosine −7 residues in the free RNA to see whether a difference can be detected.

We thank Drs. J. Williamson and K. Hall for their comments on this manuscript. We are grateful to Dr. N. Baidya for help with the synthesis of oligonucleotides and to Drs. K. Valegård and L. Liljas for providing coordinates of the MS2 cocrystal structures before publication. This work was supported by National Institutes of Health Grant GM36944 to O.C.U. and postdoctoral fellowships from the Swedish Natural Science Research Council to H.E.J. and the American Cancer Society to K.A.L., and by a graduate student fellowship from the Gottlieb Daimler- and Karl Benz-Foundation to D.D.

1. Witherell, G. W., Gott, J. M. & Uhlenbeck, O. C. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* **40,** 185–220.
2. Stockley, P. G., Stonehouse, N. J., Murray, J. B., Goodman, S. T. S., Talbot, S. J., Adams, C. J., Liljas, L. & Valegård, K. (1995) *Nucleic Acids Res.* **23,** 2512–2518.
3. Peabody, D. S. (1997) *Mol. Gen. Genet.* **254,** 358–364.
4. Johansson, H. E., Liljas, L. & Uhlenbeck, O. C. (1997) *Semin. Virol.* **8,** 176–185.
5. Scherley, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J. & Mattaj, I. W. (1989) *EMBO J.* **8,** 4163–4170.
6. Jaffrey, S. R., Haile, D. J., Klausner, R. D. & Harford, J. B. (1993) *Nucleic Acids Res.* **21,** 4627–4631.
7. Reusken, C. B. E. M. & Bol, J. F. (1996) *Nucleic Acids Res.* **24,** 2660–2665.
8. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N. & Summers, M. F. (1998) *Science* **279,** 384–388.
9. Wang, Z. F., Whitfield, M. L., Ingledue, T. C., III, Dominski, Z. & Marzluff, W. F. (1996) *Genes Dev.* **10,** 3028–3040.
10. Carey, J. C., Lowary, P. T. & Uhlenbeck, O. C. (1983) *Biochemistry* **22,** 4723–4730.
11. Lowary, P. T. & Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15,** 10483–10493.
12. Romaniuk, P. J. & Uhlenbeck, O. C. (1985) *Biochemistry* **24,** 4239–4244.
13. Milligan, J. F. (1988) Ph.D. Thesis (University of Illinois at Urbana–Champaign).
14. Peabody, D. S. (1989) *Nucleic Acids Res.* **17,** 6017–6027.
15. Valegård, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994) *Nature (London)* **371,** 623–626.
16. Valegård, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G. & Liljas, L. (1997) *J. Mol. Biol.* **270,** 724–738.
17. LeCuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. (1995) *Biochemistry* **34,** 10600–10606.
18. LeCuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. (1996) *EMBO J.* **15,** 6847–6853.
19. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15,** 8783–8798.
20. Usman, N., Ogilvie, K. K., Jiang, M.-Y. & Cedergren, R. J. (1987) *J. Am. Chem. Soc.* **109,** 7845–7854.
21. Iyer, R. P., Phillips, L. R., Egan, W., Regan, J. B. & Beaucage, S. L. (1990) *J. Org. Chem.* **55,** 4693–4699.
22. Burgers, P. M. J. & Eckstein, F. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4798–4800.
23. Slim, G. & Gait, M. J. (1991) *Nucleic Acids Res.* **19,** 1183–1188.
24. Thorogood, H., Grasby, J. A. & Connolly, B. A. (1996) *J. Biol. Chem.* **271,** 8855–8862.
25. Gish, G. & Eckstein, F. (1988) *Science* **240,** 1520–1522.
26. Greef, C. H., Seeberger, P. H. & Caruthers, M. H. (1996) *Tetrahedron Lett.* **37,** 4451–4454.
27. Pritchard, C. E., Grasby, J. A., Hamy, F., Zacharek, A. M., Singh, M., Karn, J. & Gait, M. J. (1994) *Nucleic Acids Res.* **22,** 2592–2600.
28. Mayer, A. N. & Barany, F. (1994) *J. Biol. Chem.* **269,** 29067–29076.
29. Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Bulter, P. J. G., Karn, J. & Gait, M. J. (1993) *J. Mol. Biol.* **230,** 111–123.
30. Lebedev, A. V., Frauendorf, A., Vyazovkina, E. V. & Engels, J. W. (1993) *Tetrahedron* **49,** 1043–1052.
31. Wong, I. & Lohman, T. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5428–5432.
32. Carey, J. C. & Uhlenbeck, O. C. (1983) *Biochemistry* **22,** 2610–2615.
33. Talbot, S. J., Goodman, S., Bates, S. R., Fishwick, C. W. & Stockley, P. G. (1990) *Nucleic Acids Res.* **18,** 3521–3528.
34. Borer, P. N., Lin, Y., Wang, S., Roggenbuck, M. W., Gott, J. M., Uhlenbeck, O. C. & Pelczer, I. (1995) *Biochemistry* **34,** 6488–6503.
35. Kerwood, D. J. & Borer, P. N. (1996) *Magn. Reson. Chem.* **34,** S136–S146.
36. Romaniuk, P. J., Lowary, P., Wu, H.-N., Stormo, G. & Uhlenbeck, O. C. (1987) *Biochemistry* **26,** 1563–1568.
37. Milligan, J. F. & Uhlenbeck, O. C. (1989) *Biochemistry* **28,** 2849–2855.
38. Wu, H.-N. & Uhlenbeck, O. C. (1987) *Biochemistry* **26,** 8221–8227.
39. Baidya, N. & Uhlenbeck, O. C. (1995) *Biochemistry* **34,** 12363–12368.
40. Lim, F., Spingola, M. & Peabody, D. S. (1994) *J. Biol. Chem.* **269,** 9006–9010.
41. Tars, K., Bundule, M., Fridborg, K. & Liljas, L. (1997) *J. Mol. Biol.* **271,** 759–773.
42. Frey, P. A. & Sammons, R. D. (1985) *Science* **228,** 541–545.
43. Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P. & Matthews, B. W. (1987) *Nature (London)* **330,** 41–46.
44. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M. & Pardi, A. (1996) *J. Mol. Biol.* **264,** 968–980.
45. Ni, C.-Z., Syed, R., Kodandapani, R., Wickersham, J., Peabody, D. S. & Ely, K. R. (1995) *Structure (London)* **3,** 255–263.
46. Cahn, R. S., Ingold, S. C. & Prelog, V. (1966) *Angew. Chem. Int. Ed. Engl.* **5,** 385–415.