Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: Implications for binding requirements

ELIN GRAHN,¹ NICOLA J. STONEHOUSE,² JAMES B. MURRAY,^{2,3}
SJOERD VAN DEN WORM,^{1,4} KARIN VALEGÅRD,^{1,5} KERSTIN FRIDBORG,¹
PETER G. STOCKLEY,² and LARS LILJAS¹

¹Department of Molecular Biology, Uppsala University, Box 590, SE-751 24 Uppsala, Sweden

²School of Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

ABSTRACT

The coat protein of bacteriophage MS2 is known to bind specifically to an RNA hairpin formed within the MS2 genome. Structurally this hairpin is built up by an RNA double helix interrupted by one unpaired nucleotide and closed by a four-nucleotide loop. We have performed crystallographic studies of complexes between MS2 coat protein capsids and four RNA hairpin variants in order to evaluate the minimal requirements for tight binding to the coat protein and to obtain more information about the three-dimensional structure of these hairpins. An RNA fragment including the four loop nucleotides and a two-base-pair stem but without the unpaired nucleotide is sufficient for binding to the coat protein shell under the conditions used in this study. In contrast, an RNA fragment containing a stem with the unpaired nucleotide but missing the loop nucleotides does not bind to the protein shell.

Keywords: bacteriophage MS2; crystal structure; RNA binding; RNA–protein interactions

INTRODUCTION

The *Escherichia coli* phage MS2 belongs to the leviviridae family and contains a positive, single-stranded, 3,569-nt-long RNA molecule. The genome encodes four proteins: the A (or maturation) protein, the coat protein, the lysis protein, and a replicase subunit (Fiers, 1979). The $T = 3$ icosahedral shell protecting the RNA molecule is built up from 180 copies of the coat protein subunit. The subunits exist as three different conformers denoted A, B, and C, which form two structurally different dimers, A/B and C/C in the capsid (Valegård et al., 1990; Golmohammadi et al., 1993). The major

conformational difference between the conformers lies in the FG-loop connecting the β -strands F and G (Fig. 1A).

At the 5' end of the replicase gene, the RNA genome can form a hairpin, which has a dual function in vivo. Binding of a dimer of viral coat protein to this site represses the translation of the replicase, because the hairpin includes the Shine-Dalgarno (SD) sequence and the start codon of the replicase gene (Fig. 1B). The RNA–coat protein complex also functions as a nucleation point for assembly and hence encapsidation of the correct RNA molecule. The RNA hairpin comprises a 7-bp, double-helical stem closed by a 4-nt loop and interrupted by one unpaired residue on the 5' leg of the stem, between bp 5 and 6. To achieve tight binding of the hairpin to the coat protein, the base at position –4 (see Fig. 1B for numbering) must be an adenine, –5 a pyrimidine, and –7 and –10 purines (Witherell et al., 1991).

We have used X-ray crystallography to study RNA–protein complexes between four differently modified RNA operator hairpins and MS2 coat protein. Crystals were grown of capsids formed by recombinant MS2

Reprint requests to: Lars Liljas, Department of Molecular Biology, Uppsala University, Box 590, SE-751 24 Uppsala, Sweden; e-mail: lars@xray.bmc.uu.se.

³Present address: Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA.

⁴Present address: Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, PO Box 9502, NL-2300 RA Leiden, The Netherlands.

⁵Present address: Department of Biochemistry, Uppsala University, Box 576, SE-751 23 Uppsala, Sweden.

coat protein overexpressed in *E. coli*. RNA molecules were synthesized chemically and soaked into the MS2 crystals. Previously, we have published the X-ray structures of MS2 coat protein–RNA complexes for the wild-type RNA operator hairpin and for the C-5 variant operator where the wild-type uracil at position –5 is replaced by a cytosine (Valegård et al., 1994; Valegård et al., 1997). This single-nucleotide substitution results in an RNA hairpin binding to the MS2 coat protein several times stronger than the wild-type (Lowary & Uhlenbeck, 1987; Stockley et al., 1995; Lago et al., 1998). Recently, the X-ray structures of two complexes between coat proteins with single amino acid substitutions and wild-type hairpin have been published (van den Worm et al., 1998). Although the hairpins in all these cases have been 19 nt in length, the 5' and 3' ends of the RNA have been disordered, allowing at most 16 nt, from –13 to +3, to be modeled in the electron density. The crystal structure of an aptamer–coat protein complex, however, shows clear electron density for all 14 nt (Convery et al., 1998).

There are two structurally interesting parts of the operator hairpin, the loop region and the unpaired nucleotide (Fig. 1B). These are also the key recognition targets for the coat protein. The residue A-4 in the loop region binds to a pocket in the A subunit in the A/B dimer, whereas the unpaired residue A-10 binds to the equivalent pocket in the B subunit. At the C/C dimer, the RNA hairpin can bind in both orientations. We have dissected the wild-type RNA hairpin into two asymmetric “half-sites” (Lago et al., 1998). One short RNA hairpin (called “Loop”) encompassing the four loop nucleotides and a two-base-pair stem, but missing the unpaired nucleotide was synthesized according to the sequence in Figure 1B. Two RNA fragments (“A-bulge” and “F-bulge”) encompassing the unpaired nucleotide but lacking the four loop nucleotides were also synthesized. The difference between the A-bulge and the F-bulge lies in the linkers that keep the two strands together covalently (Fig. 1B). These three RNA molecules were made to determine if either of the two half-sites, the unpaired nucleotide or the loop, alone is sufficient for the presentation of an essential recognition surface needed for strong binding to the coat protein.

A fourth RNA hairpin (“Clamp”) was synthesized with two additional base pairs compared to the longest one used in earlier studies (Fig. 1B). Earlier studies have shown that the last 1 or 2 bp give a weaker electron density than the rest of the hairpin, suggesting increased flexibility at the bottom of the stem (Valegård et al., 1994, 1997; van den Worm et al., 1998). The structure of a longer RNA fragment, where two extra base pairs function as a clamp, allows us to distinguish between two types of flexibility: rigid body motion of the stem, which would lead to similar density irrespective of stem length, or breathing of the base pairs at the end

of the stem, which would allow us to see more base pairs in the longer fragment.

RESULTS

Quality of the models

Electron density maps were calculated to 2.6, 2.9, 2.75, and 2.6 Å resolution for the Clamp, Loop, A-bulge and F-bulge complexes, respectively. All maps are of high quality because of the phase refinement possible because of the tenfold noncrystallographic symmetry. Only the maps of the Clamp and the Loop complexes showed interpretable density for the RNA, and it was possible to build RNA models for these two complexes as evidenced by Figure 2. At the asymmetric A/B dimer, the RNA hairpin binds in a unique orientation relative to the protein. This gives clear electron density for the RNA. At the symmetric C/C dimer, however, the RNA hairpin can bind in two ways to the protein, giving two overlapping electron densities for the RNA, making the interpretation more difficult. Both the Clamp and the Loop complex models include the three coat protein conformers (A, B, and C), RNA bound to the A/B as well as the C/C dimers, and approximately 160 water molecules. The model of the Loop RNA complex includes all 8 nt, whereas the model of the Clamp RNA hairpin complex includes all but the last 2 bp, that is, 19 of the 23 nt. The same numbers of nucleotides are included for the A/B and the C/C binding RNA molecules in both models. The quality of our protein models has been tested using the program Procheck (Laskowski et al., 1993). The Ramachandran plots show no residues in disallowed regions. The RNA models were checked by calculating the torsion angles for all residues and comparing the results with the corresponding values for tRNA^{Phe} (Holbrook et al., 1978). All residues have torsion angles close to the ones observed in the tRNA. The crystallographic R factors of the final models of Clamp and Loop were 0.207 and 0.182, respectively.

Loop

The Loop RNA molecule contains the 4 nt in the C-5 variant loop and a 2-bp stem, with the sequence according to Figure 1B.

Well-defined electron density was observed in the wild-type binding site on the coat protein, that is the A subunit. No RNA density was seen on the B subunit. In the 4-nt loop region, the electron density is strong and continuous, but it is weaker for the 2-bp stem. The RNA model (Fig. 2A) is very similar to the corresponding part of the C-5 variant structure (Valegård et al., 1994, 1997). The uridine at –6, however, is modeled differently here. In all complexes studied so far, this base has a very weak density most likely because of dis-

order. Previously, this uracil has been modeled in the electron density with the glycosyl bond between sugar and base in the anti conformation. In the Loop electron density, however, this base was observed in the syn conformation. The χ -angle ($O_4-C_1-N_1-C_2$ for pyrimidines) describing the glycosyl bond is 15° for nucleotide -6 in the Loop RNA model and 259° for the same nucleotide in the Clamp RNA model.

The twofold symmetry at the C/C dimer gives two possibilities for interaction with the RNA hairpin. When longer RNA molecules are used, the resulting electron density shows the two orientations overlapped. Since the Loop RNA is only eight residues long, two molecules can bind to each C/C dimer without causing any overlap in the electron density. The shortest distance between two RNA molecules bound to one C/C dimer is 5.4 Å and the electron density for each of these two RNA regions is of the same strength as for the RNA bound to the A/B dimer. This indicates that two Loop RNA molecules can bind simultaneously to one C/C dimer. At the A/B dimer, the Loop RNA binds in only one orientation relative to the protein in the same way as for all other RNA molecules soaked into the MS2 capsids so far. We have previously suggested that the unique orientation of the RNA hairpin binding to the A/B dimer is because of a contact between O2P of U-6 and the side chain of Ala B80 that is too close in the alternative twofold related orientation (Valegård et al., 1997). If the RNA molecule is rotated using an operator optimally relating the A and B subunits (the quasi twofold rotation), this distance would be 1.9 Å, whereas the observed distance between this oxygen and C β of Ala A80 is 11.2 Å (Fig. 3). The observed unique binding of the short Loop RNA to the A/B dimer indicates that it is perhaps one of the eight residues in this hairpin that is responsible for the exclusion of the other possible binding orientation, which is consistent with our hypothesis.

Bulge

The Bulge RNA molecules were constructed with the aim of forming a stem containing the unpaired adenosine, but without the 4-nt loop. Initially, crystals were soaked with two RNA fragments corresponding to the two strands of the stem, but with no covalent link. No electron density was observed at the expected RNA-binding position (unpubl. results). Since there was a possibility that the two segments would not form a duplex to a sufficient extent for binding, two different covalent connections between the two strands (Fig. 1B) were tested, thereby increasing the probability of forming double-stranded stems inside the coat protein particle. The possibility that only single-stranded RNA can enter the particles made it necessary to use connections that are flexible enough to permit both opening and reclosing of the double-stranded parts of these

molecules. In the A-bulge case, the connection was a triethylene glycol group. The F-bulge connection was made of two tetrahydrofuran derivatives.

No continuous electron density was observed for either the A-bulge or the F-bulge complexes at the wild-type binding site at the coat protein. This was not due to poor quality of the electron density, because the density of the coat protein was the same as observed in other cases. Unconnected electron density was observed that could be due to RNA, but the distribution of this electron density was not significantly different from what has been observed for "empty" MS2 capsids (unpubl. results), where it appears that cellular RNA fragments have bound unspecifically and differently to the dimers of the capsid, resulting in weak noncontinuous electron density. The lack of continuous electron density indicates that the F-bulge and A-bulge RNA molecules do not form strong interactions with the protein capsid and maybe they do not bind at all.

Clamp

The Clamp RNA hairpin consists of 23 nt with a sequence shown in Figure 1B. The molecule is identical to the C-5 variant RNA used in earlier studies (Valegård et al., 1994, 1997), but with two additional GC base pairs at the bottom of the stem.

Interpretable electron density was observed for 19 of the 23 nt in the Clamp RNA-MS2 coat protein complex. No electron density at all was observed for the two extra base pairs added at the bottom of the stem. An overview of the complex between the Clamp RNA and the A/B dimer of MS2 coat protein is shown in Figure 4.

The conformation of the MS2 coat protein in the capsid is very similar to the conformations in the wild-type and C-5 complexes with one exception; the sidechain of Lys 61 in the A conformer has a different orientation compared to the earlier models. This new orientation gives possibilities for new contacts with the -5 nt in the RNA molecule (Fig. 5). In Lys 61, the distance from atom N ζ to O2 is 3.1 Å, to O2' is 2.6 Å, and to O3' is 3.2 Å in cytidine -5. In the C-5 variant model, this lysine is stacked against the base of A-4, but the amino group does not form any direct contacts with the RNA. In all complexes, the lysine in the B conformer has a single conformation stacked on the base of A-10, forming salt links to the phosphate oxygens of nucleotides -10 and -11. In the C-5 variant and most other complexes studied, the conformation of the lysine is similar in the A, B, and C conformers, in contrast to the Clamp complex. In the wild-type RNA complex, however, the electron density for Lys A61 is weak and there is evidence for both conformations for the lysine, although only one conformation was modeled. Lysine residues are often flexible, but we have not been able to find any explanation for the specific conformational variability of this particular residue.

FIGURE 1. A: Schematic representation of the three coat protein dimers around the quasi threefold axis (q3) in the MS2 coat protein capsid viewed from the inside of the particle. The A subunits are shown in yellow, the B subunits in green, and the C subunits in blue. The structurally different FG-loops are marked for the three conformers A, B, and C. This figure and Figures 3, 4, and 6 were made using Molscript (Kraulis, 1991). **B:** The possible secondary structures of the wild-type MS2 hairpin that binds specifically to the MS2 coat protein, the C-5 variant where the U at -5 is replaced by a C, and the four RNA molecules used in this study. The wild-type hairpin includes the Shine-Dalgarno sequence (circled) and the start codon for the replicase gene (boxed). The numbering of the nucleotides is relative to the first nucleotide in the start codon, assigned +1. In the A-bulge and F-bulge RNA molecules, the two polynucleotide strands are connected by a triethylene glycol group (indicated by a*) and by two tetrahydrofuran derivatives (indicated by F), respectively. The Loop RNA consists of the four C-variant loop nucleotides and a 2-bp stem. The Clamp RNA is identical to the C-5 variant but has two additional GC base pairs at the bottom of the stem, making the molecule 23 nt in length.

FIGURE 2. Stereo drawings of the RNA models built in this study with the electron density maps shown in grey. **A:** The 8-nt Loop RNA model where positions of the nucleotides -2 and -9 are shown. **B:** The 23-nt Clamp RNA model where positions of the nucleotides -15, -10, and +4 are shown. These drawings and Figure 5 were made using Bobscript (Esnouf, 1997).

FIGURE 3. The distances between the sidechain of Ala 80 in the protein and O2P in the -6 nucleotide. **A:** In the A subunit the observed distance is 11.2 Å. **B:** After applying the quasi twofold rotation on the RNA molecule, the distance to C β atom of Ala 80 in the B subunit would be only 1.9 Å.

In the wild-type RNA model (Valegård et al., 1997), a salt link is observed between Arg 49 in the A conformer and O2P of residue -13 in the RNA. In the Clamp model the phosphate -13 is moved 1.0 Å away from the protein, changing the distance between N η 1 in Arg 49 and O2P in A-13 from 3.2 Å to 4.2 Å, which is too long for a salt link to be formed. The orientation of Arg49 is similar in all electron density maps calculated so far. Probably, this bond is formed only when a short RNA fragment with flexible ends is added to the protein capsid. In such a case, the phosphate can adopt a position to form this salt link, but in the presence of additional base pairs, as in the Clamp complex, this is

FIGURE 4. Schematic representation showing how the Clamp RNA binds to an A/B coat protein dimer (A in yellow, B in green, and RNA in red). The positions of A-4 in the loop and the A-10 bulge are indicated as well as the 3' and 5' ends of the RNA. **A:** Side view of the complex. **B:** View along the quasi twofold axis from the inside of the coat protein capsid.

FIGURE 5. A detail of the A conformer Lys 61 sidechain and the bases -5 and -4 . **A:** The situation in the Clamp model and **B:** the situation in the previous published C-5 variant model (Valegård et al., 1994, 1997). Dashed lines indicate three possible new bonds between A61 in the protein and C-5 in the Clamp RNA. The electron density is shown in grey.

not possible. In the Clamp model, no direct contacts exist between the coat protein and nucleotides -12 to -15 or -3 to $+4$.

Except for the differences discussed above, the conformation of the Clamp RNA is very similar to that of the C-5 variant (Valegård et al., 1994, 1997). The rms deviation of the 13 common phosphorus atoms is 0.24 Å, calculated using the program LSQMAN (Kleywegt, 1996). The direct contacts between the protein and the RNA involve the base-specific stacking of bases A-4 and A-10 in pockets formed by Val 29, Thr 45, Ser 47, Thr 59, and Lys 61 in the A and B subunits, respectively, and stacking of base C-5 on Tyr A85, with Asn A87 hydrogen bonded to O2 of the base. Polar contacts between protein side chains and RNA backbone are made by Glu A63 to O2' of the C-5 ribose, Tyr A85 to O1P of -5 , Ser B51 to O1P and O2P of -7 , Ser B52 and Asn B55 to O2P of -7 , Arg B49 and Lys B57 to O2P of -8 , Arg B49 to O1P of -8 , and Lys B61 to O2P of -10 and -11 . All these contacts are present in both the C-5 and the Clamp complexes. The higher affinity for a cytosine at position -5 compared to the wild-type uracil was shown to be due to an intra-RNA hydrogen bond between N4 of cytidine -5 and one of the phosphate oxygens of U-6 (Valegård et al., 1997). This interaction is also present in the Clamp model, where the distance for this hydrogen bond is 2.9 Å.

The RNA hairpin binds with its helical axis forming an angle of about 30° with the β -sheet plane of the dimer. Thus the RNA molecules form a second shell inside the protein rather than pointing into the center of the capsid. Figure 6 shows the situation at the fivefold axis

FIGURE 6. Five Clamp RNA models including 19 of the 23 nt around the fivefold axis of the viral capsid. The bottom two base pairs are not included in this model, but would overlap the neighboring molecule if they continued the helix. This view is along the five-fold axis (5) from the inside of the coat protein capsid.

where five Clamp RNA models meet. The electron density corresponds to nucleotides -15 to $+4$ since no electron density was observed for the last 2 bp. The shortest distance from an atom in the modeled RNA to its closest neighbor is $7.7 \text{ \AA}$. If the last 2 bp continued the RNA helix, a consequence would be a collision with the neighboring molecule. Because no overlapping electron density is observed at this position, we conclude that these 2 bp are flexible, and thus do not give any electron density at all.

DISCUSSION

The MS2 coat protein capsid system used in this study is excellent for probing both RNA–protein interactions and RNA structure by means of X-ray crystallography. The structure can be determined relatively straightforwardly by soaking RNA molecules into pre-made crystals of MS2 coat protein capsids and using molecular replacement to obtain the initial phases.

The RNA hairpin binding to the MS2 coat protein dimer has two nonhelical elements: the 4-nt loop and the unpaired nucleotide at position -10 (Fig. 1B). The dissection of these two elements into two different RNA molecules allowed us to study whether one of these features (in the context of a stem) is sufficient for binding. From the observed RNA electron density in the Loop complex and the missing RNA electron density in the Bulge complexes, we conclude that the Loop RNA binds to the MS2 coat protein capsid under the conditions used while the Bulge RNA molecules do not.

The Loop RNA consists of only eight residues, but can still form a stable complex with the MS2 coat protein in the form of an RNA-free capsid. This shows that the interactions formed between the protein and the nucleotides in the Loop molecule (i.e., the loop region and the upper part of the stem) are sufficient for binding and that the unpaired nucleotide at position -10 is not needed for complex formation. This is in contrast to the results of both Wu & Uhlenbeck (1987), who showed that an RNA hairpin where A-10 is base paired does not bind to the coat protein dimer in solution, and Talbot et al. (1990), who achieved the same results for an RNA hairpin with a deletion at position -10 . These studies were based on nitrocellulose filter binding assays (Carey & Uhlenbeck, 1983), probably reflecting an increased off-rate for these hairpin variants, leading to apparent nonbinding in an assay involving separation of equilibrium mixtures. Equilibrium filtration experiments (Lago et al., 1998) indicate that the Loop RNA binds with an affinity of about one eighteenth of that of the C-5 variant hairpin. The lower affinity can probably be explained by the missing contacts, which would be contributed by A-10 in the C-5 variant hairpin. The loop nucleotides alone seem to be sufficient for the hairpin to adopt the three-dimensional structure needed for protein recognition and binding. Upon binding to the protein, this region becomes the best defined of the hairpin (i.e., the lowest temperature factors and the clearest electron density).

The observed nonbindings of the Bulge RNA molecules indicate that these molecules either lack the ability to obtain the correct conformation for binding or form interactions with the protein that are too weak. Equilibrium filtration experiments, on the other hand, show that the affinity for the A-bulge hairpin variant is about the same as for the Loop hairpin (Lago et al., 1998). The solution structure of a hairpin similar to the C-5 variant has been studied by NMR (Borer et al., 1995). According to this study, the loop region is not well defined, suggesting a relatively high degree of flexibility. The unpaired A-10 is mainly stacked between the adjacent bases of the helical stem, and the bulged out conformation is not detectable. As indicated by this NMR study, it is possible that the probability for the unpaired nucleotide at -10 to bulge out of the RNA double helix is low for the molecule in solution. When the RNA molecule is bound and oriented by the interaction in the loop region, this probability increases because of the correct positioning of the bulged residue and its pocket.

The expected protein–RNA interactions of the Bulge and the Loop complexes can be deduced from the interactions in the wild-type or C-5 variant complexes (Valegård et al., 1997). Two of the bases in the loop region, -4 and -5 , form specific contacts with the protein. The protein residues in the A-subunit involved in the formation of the A-4 binding pocket are the same as

the ones in the B-subunit that form the A-10 binding pocket, although the two bases bind in different orientations and thus form different contacts with the protein. Nucleotide -5 in the loop has several interactions with the protein including the stacking of the base on Tyr A85, and these interactions have no counterparts in the bulge region. The nonspecific interactions between the RNA backbone and the coat protein include six hydrogen bonds and three salt links in the Loop RNA region and five hydrogen bonds and two salt links in the Bulge RNA region. The predicted protein surface area buried upon RNA binding is about 420 Å² for the Loop RNA and 320 Å² for the Bulge RNA (data from the wild-type RNA complex). The larger buried surface area together with the slightly larger number of polar RNA-protein interactions in the Loop case compared with the Bulge case may account for the observed differences in binding.

Previous RNA complexes have shown little electron density for the last 1 or 2 bp in the RNA stem (Valegård et al., 1997; van den Worm et al., 1998), probably because of higher degrees of disorder at the ends compared with the rest of the RNA molecule. Since more base pairs are observable in the Clamp RNA model compared with shorter hairpins, this disorder is probably caused by the breathing of base pairs at the bottom of the stem rather than bending of the stem. RNA models built into electron density maps will therefore be shorter than the actual molecules used. The structure of the Clamp RNA complex suggests that hairpins of at most eight ordered base pairs can be bound, because longer hairpins would collide at the fivefold axes. No electron density at all is observed for the eighth and ninth base pairs in the Clamp complex. The absence of electron density for these 2 bp is probably caused by increased flexibility for avoiding contacts with the neighboring RNA molecule at the fivefold axis. RNA molecules with longer base paired regions than the Clamp RNA would either not bind at all, bind with low occupancy because of steric exclusion, or be forced to change their conformations.

MATERIALS AND METHODS

The four RNA fragments with sequences shown in Figure 1B were chemically synthesized and purified as described by

TABLE 2. Statistics for model refinement.

	Clamp	Loop
R-factor (%)	20.7	18.2
Number of H ₂ O molecules	156	170
Rms—deviation of:		
bond lengths (Å)	0.018	0.019
bond angles (°)	1.8	1.9
Average B—factor of:		
protein (Å ²)	14.8	15.3
RNA (Å ²)	109.7	67.7
H ₂ O (Å ²)	37.7	37.6
Occupancy of:		
A/B RNA	0.5	0.6
C/C RNA	0.5	0.6

Murray et al. (1994). The MS2 coat protein was overexpressed in *E. coli* and the self-assembled, RNA-free protein capsids formed were purified according to Valegård et al. (1986) and Mastico et al. (1993). Crystals in the spacegroup R32, with cell dimensions $a = b = 288.0$ Å and $c = 653.0$ Å, were grown from the MS2 coat protein capsids using the hanging drop technique under conditions described earlier (Valegård et al., 1986). RNA fragments were added to the empty capsid crystals to a final concentration of 2 mg/mL (Valegård et al., 1994; van den Worm et al., 1998). Diffraction data for the Loop complex were collected at station 9.6 at the CLRC Daresbury Laboratory, United Kingdom. The wavelength was 0.87 Å and the temperature 5 °C. Data sets for the A-bulge, the F-bulge, and the Clamp complexes were collected at station ID2 at ESRF, Grenoble, France. The wavelength was 1.00 Å and the temperature 10 °C. In all cases, the oscillation angle was 0.5° and a MAR Research image plate detector was used. The data sets for A-bulge, F-bulge, and Clamp complexes were processed and scaled using the HKL package (Otwinowski & Minor, 1996). Loop data was processed with Denzo (Otwinowski & Minor, 1996) and scaled using the CCP4 package (Collaborative Computational Project, 1994).

Table 1 presents statistics from the scaling of the data for the four complexes. Initial phases were obtained by molecular replacement using the model of the empty capsid of MS2 coat protein. The phases were refined by cyclic averaging using the noncrystallographic symmetry and the program RAVE (Kleywegt & Jones, 1994). Models of Loop RNA and Clamp RNA complexes were built using the program O (Jones et al., 1991). The program X-PLOR (Brünger, 1990) was used for positional refinement using the conjugate gradient-type minimization and for refinement of restrained temperature factors of the two models. The values of occupancies for the RNA molecules were chosen to give temperature factors of the RNA atoms participating in hydrogen bonds to the protein in the same range as in earlier studies. For the Loop model, the occupancies for both A/B and C/C binding RNA molecules were therefore set to 0.6 and for the Clamp model these values were set to 0.5.

Statistics from the refinement are shown in Table 2. The final crystallographic R factors for the two models were 0.182 for the Loop and 0.207 for the Clamp. Free R factors would be similar to the crystallographic R factors because of the tenfold noncrystallographic symmetry and were not used to

TABLE 1. Statistics from data collection and scaling.

	Resolution (Å)	Scaling R-factor (%)	Number of reflections	Completeness (%)	Number of crystals
A-bulge	20-2.75	14.3	90,565	33.7	2
F-bulge	20-2.6	14.2	94,999	29.9	3
Clamp	16-2.6	13.0	149,789	47.4	7
Loop	25-2.9	11.4	168,328	74.1	4

monitor refinement (Kleywegt & Brünger, 1996). Coordinates for the Clamp and Loop models have been submitted to the Brookhaven Data Bank (ID codes: 1zdk and 1zdj, respectively).

ACKNOWLEDGMENTS

We thank Dr. Roshan Golmohammadi for help with data collection and Stephanie Fonseca for help with protein purification. This work was supported by the Swedish Natural Science Research Council, the UK BBSRC, the Leverhulme Trust and the Wellcome Trust. N.J.S. is an MRC fellow.

Manuscript accepted without revision September 28, 1998

REFERENCES

- Borer PN, Lin Y, Wang S, Roggenbuck MW, Gott JM, Uhlenbeck OC, Pelczar I. 1995. Proton NMR and structural features of a 24-nucleotide RNA hairpin. *Biochemistry* 34:6488–6503.
- Brünger AT. 1990. *XPLOR Manual*. New Haven, Connecticut: Yale University Press.
- Carey J, Uhlenbeck OC. 1983. Kinetic and thermodynamic characterization of the R17 coat protein–RNA interaction. *Biochemistry* 22:2610–2615.
- Collaborative Computational Project. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Cryst D50*:760–763.
- Convery MA, Rowsell S, Stonehouse NJ, Ellington AD, Hirao I, Murray JB, Peabody DS, Phillips SEV, Stockley PG. 1998. Crystal structure of an RNA aptamer–protein complex at 2.8 Å resolution. *Nature Struct Biol* 5:133–139.
- Esnouf RM. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph* 15:133–138.
- Fiers W. 1979. Structure and function of RNA bacteriophages. In: Fraenkel-Conrat H, Wagner RR, eds. *Comprehensive Virology*. Vol. 13. New York: Plenum Publishing Corporation. pp 69–204.
- Golmohammadi R, Valegård K, Fridborg K, Liljas L. 1993. The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J Mol Biol* 234:620–639.
- Holbrook SR, Sussman JL, Warrant RW, Kim S-H. 1978. Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J Mol Biol* 123:631–660.
- Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst A47*:110–119.
- Kleywegt GJ. 1996. Use of non-crystallographic symmetry in protein structure refinement. *Acta Cryst* 52:842–857.
- Kleywegt GJ, Brünger AT. 1996. Cross-validation in crystallography. *Structure* 4:899–903.
- Kleywegt GJ, Jones TA. 1994. Halloween ... masks and bones. In: Bailey S, Hubbard R, Waller D, eds. *From first map to final model. Proceedings of the CCP4 Study Weekend*. Daresbury, United Kingdom: SERC Daresbury Laboratory. pp 59–66.
- Kraulis PJ. 1991. Molscript: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24:946–950.
- Lago H, Fonseca SA, Murray JB, Stonehouse NJ, Stockley PG. 1998. Dissecting the key recognition features of the MS2 bacteriophage translational repression complex. *Nucleic Acids Res* 26:1337–1344.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291.
- Lowary PT, Uhlenbeck OC. 1987. An RNA mutation that increases the affinity of an RNA–protein interaction. *Nucleic Acids Res* 15:10483–10493.
- Mastico RA, Talbot SJ, Stockley PG. 1993. Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid. *J Gen Virol* 74:541–548.
- Murray JB, Collier AK, Arnold JRP. 1994. A general purification procedure for chemically synthesized oligoribonucleotides. *Anal Biochem* 218:177–184.
- Otwinowski Z, Minor W. 1996. Processing of X-ray diffraction data collected in oscillation mode. In: Carter CW, Jr, Sweet RM, eds. *Methods in enzymology*. Vol. 276. New York: Academic Press. pp 307–326.
- Stockley PG, Stonehouse NJ, Murray JB, Goodman STS, Talbot SJ, Adams CJ, Liljas L, Valegård K. 1995. Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic Acids Res* 23:2512–2518.
- Talbot SJ, Goodman S, Bates SRE, Fishwick CWG, Stockley PG. 1990. Use of synthetic oligoribonucleotides to probe RNA–protein interactions in the MS2 translational operator complex. *Nucleic Acids Res* 18:3521–3528.
- Valegård K, Liljas L, Fridborg K, Unge T. 1990. The three-dimensional structure of the bacterial virus MS2. *Nature* 345:36–41.
- Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. 1994. Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature* 371:623–626.
- Valegård K, Murray JB, Stonehouse NJ, van den Worm S, Stockley PG, Liljas L. 1997. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein–RNA interactions. *J Mol Biol* 270:724–738.
- Valegård K, Unge T, Montelius I, Strandberg B, Fiers W. 1986. Purification, crystallization and preliminary X-ray data of the bacteriophage MS2. *J Mol Biol* 190:587–591.
- van den Worm S, Stonehouse NJ, Valegård K, Murray JB, Walton C, Stockley PG, Liljas L. 1998. Crystal structures of MS2 coat protein mutant in complex with wild-type RNA operator fragments. *Nucleic Acids Res* 26:1345–1351.
- Witherell GW, Gott JM, Uhlenbeck OC. 1991. Specific interaction between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol* 40:185–220.
- Wu HN, Uhlenbeck OC. 1987. Role of a bulged A residue in a specific RNA–protein interaction. *Biochemistry* 26:8221–8227.