Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · SEPTEMBER 2013

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2013.08.025 · Source: PubMed

2 AUTHORS, INCLUDING:

Kaspars Tars
Latvian Biomedical Research and Study Ce…

55 PUBLICATIONS **716** CITATIONS

SEE PROFILE

Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene

Janis Rumnieks and Kaspars Tars

Biomedical Research and Study Center, *Ratsupites 1, Riga LV1067, Latvia*

Correspondence to Kaspars Tars: *Biomedical Research and Study Center, Ratsupites 1, Riga LV1067, Latvia. kaspars@biomed.lu.lv*

http://dx.doi.org/10.1016/j.jmb.2013.08.025

Edited by F. Allain

Abstract

The coat proteins of single-stranded RNA bacteriophages specifically recognize and bind to a hairpin structure in their genome at the beginning of the replicase gene. The interaction serves to repress the synthesis of the replicase enzyme late in infection and contributes to the specific encapsidation of phage RNA. While this mechanism is conserved throughout the *Leviviridae* family, the coat protein and operator sequences from different phages show remarkable variation, serving as prime examples for the co-evolution of protein and RNA structure. To better understand the protein–RNA interactions in this virus family, we have determined the three-dimensional structure of the coat protein from bacteriophage Qβ bound to its cognate translational operator. The RNA binding mode of Qβ coat protein shares several features with that of the widely studied phage MS2, but only one nucleotide base in the hairpin loop makes sequence-specific contacts with the protein. Unlike in other RNA phages, the Qβ coat protein does not utilize an adenine-recognition pocket for binding a bulged adenine base in the hairpin stem but instead uses a stacking interaction with a tyrosine side chain to accommodate the base. The extended loop between β strands E and F of Qβ coat protein makes contacts with the lower part of the RNA stem, explaining the greater length dependence of the RNA helix for optimal binding to the protein. Consequently, the complex structure allows the proposal of a mechanism by which the Qβ coat protein recognizes and discriminates in favor of its cognate RNA.

Introduction

For bacteriophages of the *Leviviridae* family, the single-stranded RNA genome does not merely encode phage proteins but also forms extensive secondary and tertiary structures that are critical for RNA replication, regulation of phage protein synthesis and assembly of virus particles [1]. The function of three out of the four phage proteins—replicase, maturation and coat—is intricately linked with specific RNA structures that they recognize and bind to at some point in the viral life cycle [2–4]. The *Leviviridae* coat protein adopts a fold observed only in this virus family with an N-terminal β hairpin, a five-stranded antiparallel β sheet and two C-terminal α helices [5]. The helices from two coat protein molecules interlock to form a very stable dimer with a continuous ten-stranded β sheet that lines the interior of the capsid and forms the RNA-binding surface of the protein. Although the primary role of the coat protein is to encapsulate the genome, it also acts as a translational repressor that regulates the synthesis of the replicase. The operator is an RNA sequence of approximately 20 nucleotides at the beginning of the replicase gene that folds into a stem–loop structure and comprises the initiation codon of the gene [6,7]. Specific binding of the coat protein to the RNA hairpin effectively shuts down the translation of the replicase when the coat protein accumulates in the infected cells [8] and marks the genome for packaging into capsids [9]. This regulatory mechanism is highly conserved within the *Leviviridae* family, but similarities in operator hairpins are limited to a stem structure of seven to eight base pairs with an unpaired base in it, whereas the number and identity of nucleotides in the loop as well as the position of the bulged nucleotide vary from phage to phage. The

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

bulged nucleotide is usually an adenosine and is located on the 5′ side of the stem, and the loops of all of the studied operators contain another adenosine that is critical for specific interactions with the coat proteins.

The interaction between the coat protein and the operator of phage MS2 has been extensively studied both biochemically [10–12] and structurally [13–17], making it one of the best characterized protein–RNA interactions to date. Three-dimensional structures of coat protein–operator complexes have also been solved for the phages PRR1 [18] and PP7 [19]. The MS2 and PRR1 operators differ primarily in the size of the loop, and the respective coat proteins utilize a very similar RNA binding mode. The recognition mechanism in both cases involves the binding of two adenine bases, namely, the unpaired one in the stem and another in the loop, to symmetrical adenine-recognizing pockets in the protein dimer. The complex is further stabilized by aromatic stacking that extends from the helical RNA stem via two bases in the hairpin loop to a conserved tyrosine side chain in the coat protein. The operator of phage PP7 is remarkably different from MS2 and uses a distinct RNA binding mode. Nonetheless, the PP7 coat protein also uses symmetrical pockets to bind two adenine bases in the bulge and the loop despite the fact that the pockets are very different from those found in MS2.

The bacteriophage Qβ is distantly related to MS2 with their coat proteins only about 20% identical. Both coat proteins preferentially bind their cognate translational operators, which are also rather different (Fig. 1). For strong binding to the MS2 coat protein, the operator helix needs to be at least five base pairs long and contain an unpaired purine nucleotide two base pairs prior to a four-nucleotide loop with adenosines as the first and last nucleotides and a pyrimidine nucleotide at the penultimate position [12]. For high-affinity binding to the Qβ coat protein, the operator requires a three-nucleotide loop and an eight-base-pair stem with a bulged nucleotide four base pairs from the loop [20]. The only critical nucleotide in the loop is an adenosine at the last position, whereas the unpaired adenosine in the stem can be mutated or removed altogether with a rather minor decrease in affinity [21]. Despite the differences, several facts suggest that the RNA binding modes of MS2 and Qβ coat proteins are nevertheless related. Although the overall sequence identity is low, the three-dimensional structure of the two proteins is very similar, and many of the residues that are involved in RNA binding in MS2 are conserved in Qβ [22]. Furthermore, MS2 and Qβ coat protein mutants that are able to tightly bind the operator of the other phage have been isolated [21,23], but analogous experiments were unsuccessful with PP7 [24].

The mechanism by which the MS2 coat protein discriminates between the MS2 and Qβ operators is well understood. Genetic studies have shown that amino acid changes at residues 87 and 89 of the MS2 coat protein confer an ability to bind the Qβ operator with high affinity [23]. The molecular mechanism for this discrimination has been elucidated by solving crystal structures of the mutant coat proteins bound to the Qβ operator [25]. In the wild-type MS2 coat protein, Asn87 forms a hydrogen bond with the –5 uracil base in the cognate operator, while the Qβ operator has the bulkier adenine base in the equivalent +7 position, which results in a steric clash with the asparagine side chain. Mutation of the asparagine to a serine or alanine decreases the affinity for the MS2 operator because the hydrogen

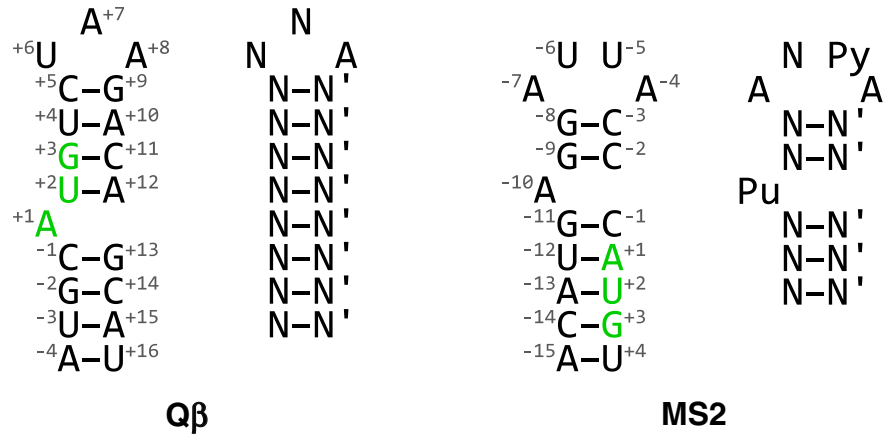

Fig. 1. Secondary structure of the Qβ and MS2 operators. For both phages, the wild-type operator sequence is shown on the left and the minimal sequence requirements for binding to the coat protein are shown on the right (Py, pyrimidine; Pu, purine; N, any nucleotide; N′, a nucleotide complementary to N). For the wild-type operators, the initiation codons of the replicase are marked in green, and nucleotide positions relative to the start of the replicase ORF are indicated as superscript numbers next to the bases.

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

bond interaction is lost but improves binding of the Qβ operator by allowing sufficient space to accommodate the adenine base. A second mutation of Glu89 to a lysine eliminates an unfavorable electrostatic repulsion with the phosphate backbone of the Qβ operator and instead provides an additional contact that further improves the binding.

Although a genetic study of the Qβ coat protein [21] demonstrated that the RNA binding modes of the Qβ and MS2 coat proteins are similar, the molecular mechanism that allows the Qβ coat protein to recognize and discriminate its cognate operator has remained unknown. To address this issue, we solved the crystal structure of the Qβ coat protein in complex with its operator hairpin, which we present here and compare to the coat protein–RNA complexes found in other RNA phages.

Results and Discussion

Design and structure determination of assembly-deficient Qβ coat protein in complex with RNA

Previous work with MS2 that led to numerous protein–RNA complex structures relied on the ability to soak small RNA hairpins into pre-crystallized capsids via pores that are present at their 3-fold and 5-fold symmetry axes. However, the same approach failed when applied to Qβ, which was attributed to the fact that the FG loops from neighboring Qβ coat protein dimers are covalently linked to each other with disulfide bonds that could in turn restrict RNA diffusion into capsids. To address this issue, we crystallized Qβ capsids assembled from modified coat proteins that had cysteines in the FG loop mutated to glycines and used these crystals for the RNA soaking experiments. Unfortunately, still no bound RNA was detected in the electron density maps, suggesting that the crystallization conditions (0.4 M NaCl at pH 7.5) could be suboptimal for RNA binding and that the approach of soaking capsid crystals with RNA would not be successful with Qβ.

The structure of the PP7 coat protein in complex with its operator was determined via a different approach, namely, by crystallizing the RNA together with coat protein dimers that were lacking the FG loops and therefore incapable of assembling into capsids. However, our initial attempts to truncate the FG loop in Qβ resulted in a largely insoluble protein; therefore, an approach was devised to introduce other amino acid changes into the coat protein that would prevent it from assembling into particles. Examination of the Qβ capsid structure suggested Asn129 as a good candidate for mutagenesis as its side chain forms two hydrogen bonds with the main chain of the adjacent dimer; thus, introduction of a bulkier side chain at this position would both destroy the bonding and cause a steric clash with the nearby chain. A similar situation was observed for Pro42 in the CD loop where substitution with a longer side chain would likely result in a collision with the neighboring dimer. Mutation of the two residues to arginines (Pro42Arg, Asn129Arg) in the cysteine-less mutant (Cys74Gly, Cys80Gly) indeed resulted in a protein that produced a highly soluble and homogenous dimeric species suitable for structural studies. The coat protein–RNA complex was obtained by mixing purified dimers and RNA in a molar ratio of 1:1.2, and the mixture was immediately subjected to crystallization. Crystals that diffracted to 2.4 Å resolution were obtained, and the structure was solved by molecular replacement.

Quality of the model

The final model (Fig. 2a) contains one Qβ coat protein dimer (chains A and B) and one RNA molecule (chain R). There are no crystal contacts close to the protein–RNA interface, suggesting that the model represents a biologically relevant structure. The unassembled dimer adopts a conformation highly similar to that found in the crystallized phage capsids [22], with a root-mean-square deviation (rmsd) of C^{α} atoms of 0.8 Å. Notably, the EF loops of the assembly-deficient dimer make contacts with RNA and can be reliably modeled, whereas they were only partly visible in the virus structure. In contrast, the FG loops (residues 74–84 of chain A and residues 75–83 of chain B) are disordered in the unassembled dimer and were not included in the final model. Electron density for the whole RNA molecule (20 nucleotides) was clearly visible (Fig. 2b), and the complete hairpin was modeled without breaks. Interestingly, the stems of two neighboring RNA hairpins stack end-to-end in the crystal in a somewhat similar manner as in the PP7 coat protein–RNA structure. This arrangement likely restricts their movement and contributes to the well-defined electron density observed for the RNA. The final model also includes six zinc ions from the crystallization solution. One of them is tetrahedrally coordinated between Asp102 and Glu103 of two adjacent dimers where it provides an important crystal contact, while the others are located in the proximity of the RNA.

Structure of Qβ coat protein–operator complex

The experimentally observed structure of the RNA hairpin is consistent with the predicted secondary structure and consists of an eight-base-pair stem, a three-nucleotide loop and an unpaired adenosine in the stem. The stem adopts a canonical A-form helical conformation with ribose puckers in the C3′-*endo* conformation except for loop nucleotides

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

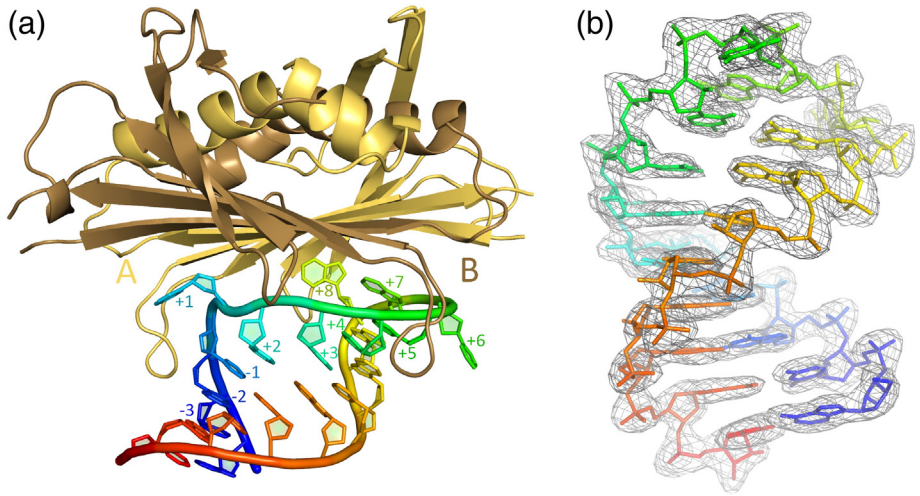

Fig. 2. Three-dimensional structure of the Qβ coat protein–operator complex. (a) Overall structure of the complex. The coat protein dimer is represented in light orange (monomer A) and light brown (monomer B), and the RNA is rainbow-colored blue (5′ end) to red (3′ end). Nucleotide positions relative to the first nucleotide of the replicase initiation codon are indicated next to the bases. (b) A close-up view of the RNA hairpin. The RNA is shown in a stick representation colored as in (a) and shown as modeled into a $2F_o - F_c$ electron density contoured at 1.1 σ. This figure and Figs. 3–5 were prepared using PyMOL [26].

A+7 and A+8, which adopted more of a C2′-*endo* conformation. The overall conformation of the operator hairpin (nucleotides +2 to +12) in the cognate complex is very similar to that observed for the Qβ operator in complex with the MS2 coat protein mutant [25] with an rmsd of 0.8 Å (Fig. 3). The majority of the contacts between the protein and RNA are sequence-independent interactions between the sugar-phosphate backbone of the RNA and the EF loop and β strand F of both coat protein monomers (Table 1). The adenine base of the A+8 nucleotide fits into an adenine-binding pocket formed by Val32, Thr49, Ser51, Gln65 and Lys67 of chain A in the coat protein dimer. The base of the A+7 nucleotide is stacked between C+5 in the stem and the aromatic side chain of Tyr89 of the A chain. In addition, the hydroxyl group of the tyrosine forms a stabilizing hydrogen bond with an oxygen atom in the phosphate backbone. In the crystallized complex, there is also a zinc ion from the crystallization solution coordinated between the OD2 oxygen of Asp91 of the A monomer, the N1 nitrogen of A+7 and two water molecules. This interaction is not physiologically relevant because both atoms would act as hydrogen acceptors at physiological pH, and the AspA91 side chain would not be able to form a hydrogen bond with the adenine base under these conditions. The base of the last loop nucleotide, U+6, points away from the protein and does not make any contacts with it. The unpaired A+1 nucleotide bulges out from the stem and stacks with Tyr89 in chain B of the coat protein. There seem to be no additional stabilizing interactions involving the base, but the phosphate oxygen of A+1 forms an electrostatic interaction with the side chain of Lys63 in

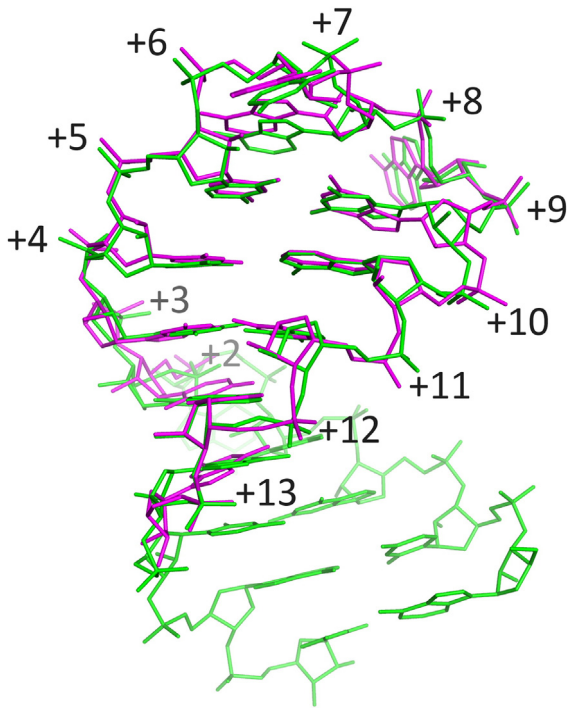

Fig. 3. Structure of the Qβ operator bound to the Qβ coat protein and the MS2 coat protein mutant. Although the upper part of the hairpin adopts a remarkably similar conformation in both cases, the lower part, including the bulged adenosine, is disordered in the complex with the MS2 mutant. Nucleotide numbers as of Fig. 1 are indicated next to the phosphates. The operator hairpin from the cognate Qβ complex is represented in green while that bound to the Asn87Ser MS2 coat protein mutant (PDB entry 1ZSE) is represented in magenta.

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

Table 1. Hydrogen bonds and electrostatic interactions between protein and RNA in the Qβ coat protein–RNA operator complex

RNA		Protein		Distance (Å)
Residue	Atom	Residue	Atom	
U–3	O2′	AsnA58	ND2	2.7
G–2	OP1	ArgA59	NH1	3.8
C–1	OP1	LysA63	NZ	3.3
	OP2	LysA63	NZ	3.4
A+1	OP2	LysA63	NZ	3.0
G+3	OP1	LysB67	NZ	2.9
U+4	OP1	ArgB59	NE	2.7
	OP2	LysB63	NZ	2.4
C+5	OP1	LysB60	N	2.9
		AsnB61	N	2.8
	OP2	LysB63	NZ	2.8
U+6	OP1	LysB60	NZ	2.9
	OP2	AsnB61	ND2	2.8
A+7	OP2	TyrA89	OH	2.4
A+8	O2′	AsnA30	OD1	3.0
	N1	SerA51	OG	2.8
	N6	ThrA49	OG1	3.2
		GlnA65	O	2.9
	N7	ThrA49	OG1	2.6

the A chain, and additional contacts with sugars and phosphates of C–1, G–2 and U–3 nucleotides in the lower part of the stem stabilize the hairpin in the observed orientation.

Comparison of RNA binding between Qβ and MS2

The 970-Å^2 interface between the Qβ coat protein and its operator hairpin is close to the value reported for PP7 (950 Å^2) and slightly larger than that of MS2 (830 Å^2), but the overall structure of the complex is undoubtedly more similar to that of MS2. The top part of the Qβ hairpin that faces the protein (nucleotides +3 to +8) adopts a conformation that is remarkably similar to that of the MS2 operator (nucleotides –9 to –4, respectively) with an rmsd of 1.1 Å, which supports the hypothesis that the two proteins share a similar RNA binding mode. The number of hydrogen bonds and electrostatic interactions between the protein and RNA is similar in Qβ and MS2; however, in MS2, a higher proportion of the interactions involve contacts with the nucleotide bases rather than the sugar-phosphate backbone (Fig. 4a). The adenine-binding pocket of the Qβ coat protein is almost identical with that of MS2, and all of the base–protein interactions within the pocket are the same in the two phages. However, the nearby interaction between LysA43 and the phosphate backbone in MS2 is not preserved as the equivalent ArgA47 in Qβ is too far away from the RNA (4.4 Å) to make any significant contribution to the interaction. The similarities in RNA binding of the two proteins extend to the A+7 nucleotide, which in Qβ is stacked with TyrA89 while in MS2 an analogous interaction is found between U–5 and TyrA85, and a contact between the hydroxyl of the tyrosine and a phosphate of the RNA backbone is also conserved. Like U–6 of MS2, the U+6 in Qβ points away from the protein and does not make contacts with it. Finally, residues AsnB61 and LysB63, which make interactions with the sugar-phosphate backbone in Qβ, are conserved and provide the same function in MS2.

Away from the hairpin loop, the differences in protein–RNA interactions in the two phages become more pronounced. In the lower part of the hairpin, only a single electrostatic interaction exists between Arg49 of the A monomer and the –13 phosphate in MS2, but in Qβ the arginine residue is not conserved and interactions involving AsnA58, ArgA59 and LysA63 take place instead. The additional contacts are possible due to an extended EF loop that, in Qβ, is two residues longer than in MS2. However, the most profound difference between Qβ and other RNA phages involves the interaction with the bulged adenosine in the stem of the hairpin. In MS2, the bulged A–10 base fits into the same pocket as A–4 in the other monomer, albeit in a different orientation; however, in Qβ, the other adenine-binding pocket is empty, and the A+1 base is stacked with Tyr89 of the other monomer (Fig. 4b). This configuration has not been observed in any other coat protein–RNA complex and thus represents a novel mechanism for accommodating an unpaired base in the stem.

RNA binding discrimination of Qβ coat protein

The conformation of the β sheet that makes up the RNA-binding surface of the coat protein is very similar in MS2 and Qβ with an rmsd of 0.9 Å when C^{α} atoms from strands D, E, F and G of the two proteins are superimposed. In the superimposed protein–RNA complexes, the A+8/A–4 bases, the adenine-binding pockets and other conserved RNA-binding residues align remarkably well. A possible RNA discrimination mechanism for the Qβ coat protein can therefore be modeled with some confidence by combining protein coordinates from the Qβ complex with RNA coordinates from the fitted MS2 complex.

In the modeled Qβ coat protein–MS2 operator complex, the A–10 and A–4 bases fit very well into the adenine-binding pockets of the Qβ coat protein, and many of the interactions with the RNA backbone in the upper stem seem to be preserved. AsnB61 and LysB63 in Qβ occupy positions equivalent to AsnB55 and LysB57 in MS2, and although LysB60 of the Qβ coat protein is not conserved in MS2, there is no reason to exclude an interaction with the MS2 operator. There appear to be some differences regarding the interactions involving Arg49, which is found in the wild-type MS2 complex but is not conserved in Qβ. In the wild-type MS2 complex, Arg49 in the A monomer forms a salt bridge with the –13 phosphate, but this interaction is lost with the Qβ

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

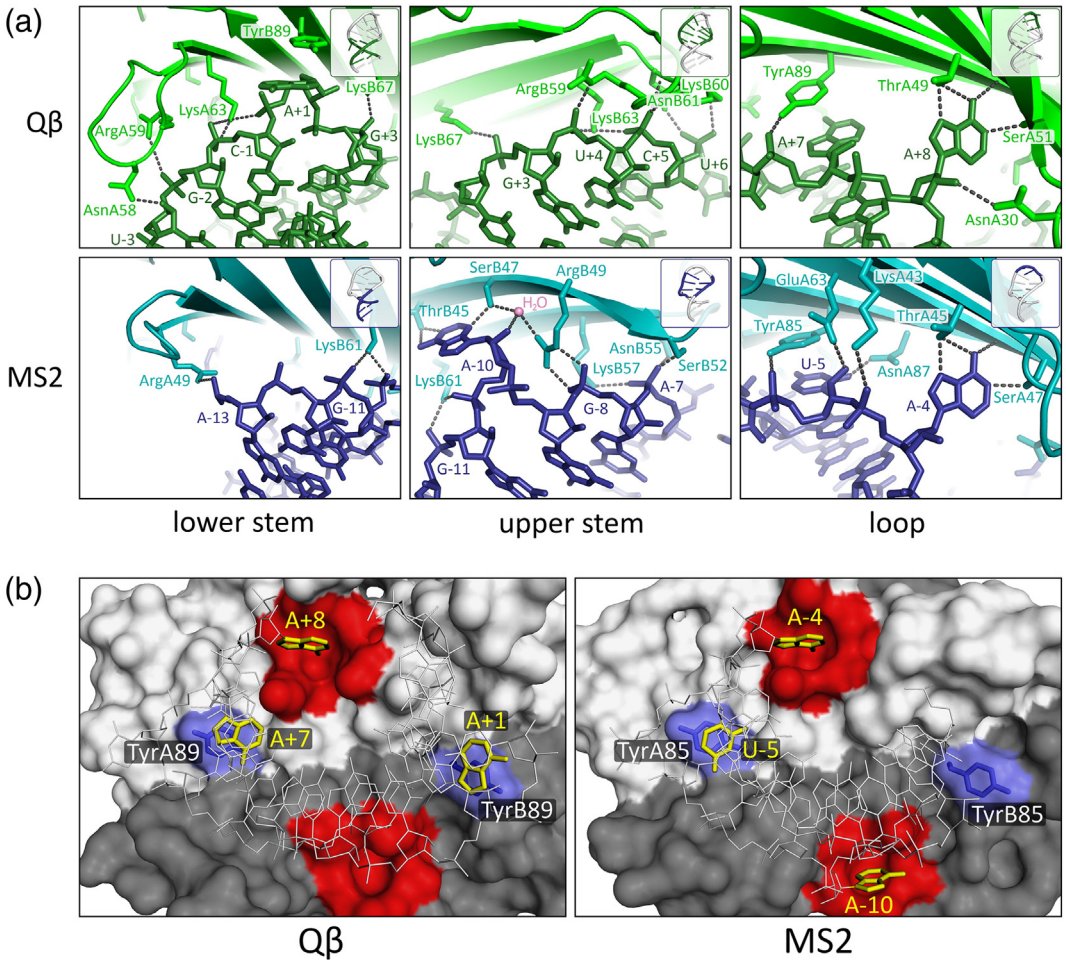

Fig. 4. Differences in binding of the Qβ and MS2 coat proteins to their cognate operators. (a) Close-up views of the protein–RNA interactions in Qβ and MS2. Hydrogen bonds and electrostatic interactions in the lower and upper parts of the stem and the hairpin loops are indicated as gray broken lines. Side chains of interacting amino acid residues and nucleotides are labeled as in Table 1. The insets on top right highlight the approximate region of the operator hairpin that is visible in the particular close-up. (b) Comparison of protein–RNA interactions in Qβ and MS2 involving the loop and the bulged adenosine. The solvent-accessible surfaces of Qβ and MS2 coat protein dimers are shown in different shades of gray as for A and B monomers. The adenine-binding pockets are shown in red, while the tyrosine residues that stack with RNA bases are colored blue. The RNA is shown in light gray as a stick model except for the bases that occupy the adenine-binding pockets or stack with the tyrosine side chains, which are shown in yellow. In Qβ, only one of the symmetrical adenine-binding pockets is occupied and tyrosines from both monomers participate in base stacking. In contrast, both pockets are occupied by adenine bases in MS2, while only a single tyrosine is involved in base stacking.

coat protein, which has a serine residue at the equivalent position. In the B monomer, Arg49 forms a salt bridge with the –8 phosphate and additionally coordinates a water molecule that forms a hydrogen bond with the O2′ atom of the A–10 ribose. In Qβ, the side chain of ArgB59 lies in approximately the same place as ArgB49 in MS2 and partly serves the same function by providing an electrostatic interaction with the phosphate of U+4. This interaction would likely be preserved in the complex with the MS2 operator, but the arginine side chain would be too far away from the A–10 nucleotide to allow interactions similar to those observed in the cognate MS2 complex. Consequently, this might contribute to the weaker binding of the MS2 operator to the Qβ coat protein.

Another reason for the poor binding of the MS2 operator likely involves the –5 uracil base in the loop. The side chain of TyrA89 that stacks with A+7 in Qβ is tilted by approximately 20° compared to TyrA85 in MS2. This orientation is observed both in complex with the RNA and in assembled capsids and is unlikely to switch to an MS2-like conformation due to the proximity of the GlnA69 and GlnA87 side chains. As a result, planes going through the U–5 base of the MS2 operator and the side chain of TyrA89 in Qβ coat protein would not be parallel, which could lead

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

to impaired binding of the RNA. In addition, the interaction between U–5 and AsnA87 that is present in the cognate MS2 complex is lost. The corresponding amino acid in Qβ is AspA89, and repulsion between the acidic side chain and the O2 carbonyl of the uracil base would prevent an analogous interaction with the Qβ coat protein. This is consistent with the observation that the interaction between an Asp91Asn Qβ coat protein mutant and the MS2 operator is 20 times stronger than with wild-type Qβ coat protein [21]. In contrast to aspartic acid, the asparagine side chain would permit formation of a hydrogen bond between the protein and RNA and result in the observed improvement in binding.

It is interesting to note that the Qβ coat protein is able to bind the operator of the closely related phage SP with the same affinity as the cognate one [27]. A notable difference between the two hairpins is that the SP operator contains a C-A base pair in the upper part of the stem. It was further demonstrated [27] that Qβ coat protein can tolerate several other base-pair mismatches in the stretch between the bulged adenosine and the loop, suggesting that the integrity of the upper part of the helix is not critical for high-affinity binding. Accommodation of a non-Watson–Crick base pair in an RNA hairpin has been visualized in the crystal structure of an RNA aptamer bound to the MS2 coat protein [28], which showed that a non-canonical G-A base pair does not result in the disruption of the helical stem. In the absence of similar experimentally determined structures for Qβ, it seems reasonable to assume that single-base-pair mismatches in the Qβ operator would be tolerated in a similar manner as in the MS2 aptamer. Apparently, the interactions between the protein and the RNA backbone on the 5′ side of the stem are sufficiently strong to hold the RNA in the protein-bound conformation and render perfect base pairing in the stem redundant.

Effects of hairpin loop size and bulged nucleotides on RNA binding

In Qβ, the size of the hairpin loop plays an important role for optimal binding of the cognate operator, as the addition of an extra nucleotide in the loop severely reduces the affinity [20]. When the stems of the MS2 and Qβ operators are superimposed (residues +10 to +15 for Qβ and residues –3 to +3 in MS2), substantial differences in loop conformations are evident due to the extra base pair at the top of the Qβ hairpin (Fig. 5a). However, in the superimposed protein–RNA complexes of the two phages, the smallest conformational differences are observed in the region

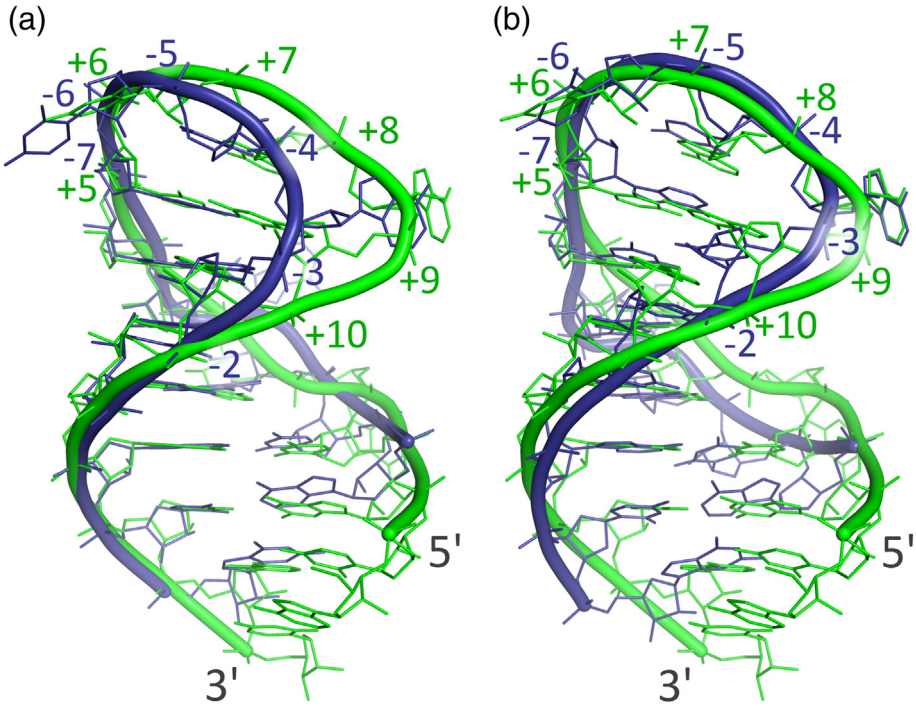

Fig. 5. Conformational differences of the Qβ and MS2 operators. Superimposition of the helical stems (a) demonstrates the differences in hairpin loop conformations of the two operators. Superimposition of the RNA-binding residues of the two cognate protein–RNA complexes (b) results in different relative orientations of the stems that, in turn, cause the phosphate backbones of the two RNAs to follow different paths. The Qβ (green) and MS2 (blue) operators are shown as stick models with the phosphate backbones represented by ribbon traces. Nucleotides in the loop and the upper part of stem are numbered as of Fig. 1 and indicated next to the phosphates.

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

comprising the loop and two preceding nucleotides and not in the stems (Fig. 5b). Consequently, the different-sized loops impose different relative orientations of the RNA stems that appear to play some role in optimal binding of the RNA. Biochemical studies have shown that the Qβ coat protein requires a longer RNA stem than MS2 for high-affinity binding [20], which is likely necessary to compensate for the lack of some of the interactions in the upper stem. The length dependence is explained by the EF loops, which are longer in the Qβ coat protein than their MS2 counterparts and make contacts with the lower stem; however, binding of a hairpin with a three-nucleotide loop would position the phosphate backbone in a more favorable orientation regarding the interactions than the binding of a four-nucleotide loop. The conformation with a three-nucleotide loop also restricts the ability to accommodate bulged nucleotides in the stem except those at a position four nucleotides prior to the loop; in this case, an additional stacking interaction with the protein that further stabilizes the complex is formed. However, the unpaired adenosine is not critical for binding and results in only 1.5- to 5-fold reduction in affinity when absent [20,21]. Removal of the bulged adenosine would eliminate only a single stacking interaction since there are no additional contacts between the protein and the base and would indeed result in a rather minor decrease in affinity. The lack of the unpaired base apparently does not impose significant conformational changes to the stem and still permits the EF loop to bind the lower part of the RNA hairpin, although the interactions are probably somewhat different from those in the wild-type complex. For the MS2 operator bound to the Qβ coat protein, the combined effects of binding a four-nucleotide loop and the requirement to accommodate the −10 adenine in its binding pocket would cause the lower stem to adopt an orientation that is not optimal for interacting with the EF loop.

Interactions with the lower stem are also impaired for the Qβ operator bound to the MS2 coat protein mutant because the lower part of the hairpin, including the bulged adenosine, was disordered and not visible in the three-dimensional structure of the complex [25] (Fig. 3). In this case, the size of the hairpin loop does not seem to play a very important role because the MS2 coat protein can bind a three-nucleotide loop almost as well as a four-nucleotide one given that the distance between the −10 and −4 adenosines is preserved [29]. In the Qβ operator, however, the distance is greater by one nucleotide, which would prevent the bulged adenosine from being accommodated in the MS2 pocket and would not allow favorable stacking interactions with the tyrosine side chain. We believe that this, together with the shorter EF loop found in the MS2 coat protein that cannot interact with the lower part of the stem, explains the observed lack of interactions with the lower part of the Qβ operator.

RNA recognition mechanisms among *Leviviridae* phages

Including the Qβ structure presented here, the three-dimensional structures of coat protein–operator complexes are now known for four different RNA phages. Despite some profound differences, a number of common themes can also be recognized. One such feature that has been observed in all phage operator structures is that some of the nucleotide bases in the loop stack with bases in the helical stem. In MS2, PRR1 and Qβ, the nucleotide stack further extends to the aromatic side chain of a conserved tyrosine residue, whereas in PP7, a van der Waals interaction with a valine residue takes place. The aromatic stacking is likely important for constraining the loop nucleotides in an appropriate position to bind the protein and is therefore conserved during evolution. Another RNA recognition strategy shared between all phages involves sequence-specific interactions between nucleotide bases and a complementary RNA-binding surface of the protein. In all of the studied phages, binding of an adenine base in the loop into an adenine-recognition pocket in the coat protein is critical for the operator–coat protein interaction, but the importance of other base-specific interactions varies. For the PP7 coat protein, base-specific interactions play a fundamental role in operator recognition and involve four nucleotides in the loop and the bulge, while the sugar-phosphate backbone does not make any contacts with the protein outside of these regions. In MS2, the situation is somewhat similar in that three bases make direct contact with the protein; however, the RNA backbone also makes significant interactions with the protein in the stretch between the bulged adenosine and the loop. In Qβ, the loop adenine is the only nucleotide that makes base-specific contacts with the coat protein while the majority of interactions between the protein and RNA involve the sugar-phosphate backbone. Despite the smaller amount of sequence-specific information, the Qβ coat protein is still able to discriminate its cognate operator, which demonstrates how co-evolution of the protein and RNA can result in a highly specific interaction based on the conformation of the phosphate backbone rather than numerous sequence-specific contacts with bases. The three very different modes of accommodating an unpaired base in PP7, MS2 and Qβ further demonstrate the notable flexibility of protein–RNA interactions in evolutionarily related viruses.

Nevertheless, the overall binding mode of the Qβ coat protein to its operator is clearly similar to those of MS2 and PRR1, which suggests that this particular mechanism is conserved among the conjugative plasmid-dependent *Leviviridae* phages. Outside this group, the PP7 coat protein is the only one that still has some traces of sequence identity with MS2 and Qβ,

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

but its RNA recognition mechanism is very different. Two other *Leviviridae* phages that are remarkably different from the rest have been identified and sequenced: *Acinetobacter* phage AP205 and *Caulobacter* phage ϕCb5. Their coat protein sequences share no recognizable similarities with those of MS2, Qβ, PP7 or each other. For phage AP205, a putative operator hairpin at the beginning of the replicase gene has been identified, which, unlike other phages, has a bulged uridine located on the 3′ side of the stem [30]. An operator hairpin could not be reliably identified in the genome of phage ϕCb5, raising the question of whether it exists at all [31]. The three-dimensional structure of the ϕCb5 virion revealed strong electron density for RNA bases between the dimers, which indicates a very different RNA packaging and recognition mechanism [32]. Therefore, further studies on protein–RNA interactions of the small RNA phages have the potential to provide even more discoveries about the biology, evolution and structure of these fascinating viruses.

Materials and Methods

Preparation of coat protein and RNA

Plasmid p205 encoding the Qβ coat protein with cysteines in the FG loop mutated to glycines was kindly provided by Dr. Indulis Cielēns. Using p205 as a template, we PCR-amplified the coat protein coding sequence with forward primer 5′-CAGGATCCATGGCAAAATTAGAGACTGTTAC-3′ and reverse primer 5′-TATGAAGCTTAATACGCTGGGCGCAGCTGATCAA-3′ to introduce the Asn129Arg amino acid substitution and cloned it into the pET28a expression vector (Novagen). The resulting plasmid was used as a template for site-directed mutagenesis by PCR using primers 5′-CAAGCGGGTGCAGTTCGTGCGCTGGAGAAGCGT-3′ and 5′-ACGCTTCTCCAGCGCACGAACTGCACCCGCTTG-3′ to introduce the additional Pro42Arg mutation. The resulting plasmid was named pET28-Qβ150 and used to produce the assembly-deficient coat protein dimer for crystallization.

For protein production, *Escherichia coli* BL21(DE3) cells containing pET28-Qβ150 were grown in LB medium supplemented with 30 μg/ml kanamycin with aeration at 37 °C. When the OD_{590} of the culture reached 0.5, IPTG was added to a final concentration of 1 mM, and the bacteria were grown for another 4 h and harvested by centrifugation. To purify the protein, we resuspended cells in buffer containing 40 mM Tris–HCl (pH 8.0), 200 mM NaCl, 20 mM $MgSO_4$, 0.1% Triton X-100, 0.1 mg/ml DNase and 1 mg/ml lysozyme and lysed them by three freeze–thaw cycles. The lysate was clarified by centrifugation at 18,500*g*, and the supernatant was loaded on a 1-ml HiTrap SP FF column (GE Healthcare) equilibrated with buffer A [20 mM Tris–HCl (pH 8.0) and 200 mM NaCl]. After extensive washing with buffer A, we eluted bound proteins with a 10-ml gradient of 0–100% buffer B [20 mM Tris–HCl (pH 8.0) and 1 M NaCl] and collected them in 1-ml fractions. Individual fractions containing coat protein were diluted to 5 ml with buffer A and loaded on a Mono S 5/50 GL column (GE Healthcare) equilibrated with buffer A. Bound proteins were eluted with a 15-ml gradient of 0–50% buffer B, corresponding to 200–600 mM NaCl. Fractions containing coat protein and no major contaminants were pooled, concentrated to 500 μl with an Amicon Ultra 10K spin unit (Millipore) and loaded on a Superdex 200 10/300 GL gel-filtration column (GE Healthcare) equilibrated with buffer C [50 mM 4-morpholineethanesulfonic acid (pH 6.0) and 50 mM NaCl]. Fractions containing coat protein were pooled, concentrated and stored at 4 °C until crystallization.

An HPLC-purified RNA oligonucleotide with the sequence 5′-AUGCAUGUCUAAGACAGCAU-3′ corresponding to the wild-type Qβ translation operator was purchased from Metabion AG.

Table 2. Crystallographic data collection, scaling and refinement statistics

Data collection and scaling	
Space group	$P6_522$
Cell parameters (Å)	
a	75.84
c	303.49
Wavelength (Å)	1.0000
Resolution (Å)	38–2.40
Highest-resolution bin (Å)	2.53–2.40
R_{merge}	0.09 (0.61)
Total number of observations	73,407
Number of unique reflections	21,316
$I/\sigma(I)$	9.7 (2.2)
Completeness (%)	99.9 (100.0)
Multiplicity	3.4 (3.5)
Refinement	
Number of reflections in work set	20,180
Number of reflections in test set	1089
R_{work}	0.25
R_{free}	0.29
B-factor (Å^2)	
Protein atoms	33.1
RNA atoms	39.8
From Wilson plot	34.9
Number of atoms	
Protein	1854
RNA	422
Solvent	93
rmsd from ideal	
Bond lengths (Å)	0.016
Bond angles (°)	1.670
Ramachandran plot (%)	
Residues in favored regions	96.2
Residues in allowed regions	100.0

Values in parentheses are given for the highest-resolution shell.

Crystallization and data collection

The concentration of coat protein was quantified spectrophotometrically assuming that one absorption unit at λ = 280 nm corresponds to a protein concentration of 2.37 mg/ml, as calculated with the ProtParam utility on the ExPASy server [33]. To quantify the RNA, we used data provided by the supplier. The coat protein (10.5 mg/ml in buffer C) and RNA (20 mg/ml in diethylpyrocarbonate-

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

treated water) were mixed immediately before crystallization at a molar ratio of coat protein dimer to RNA operator of 1:1.2, corresponding to a final concentration of 9.2 mg/ml of protein and 2.48 mg/ml of RNA. The complex was crystallized using the sitting-drop method by mixing 1 μl of the coat protein–RNA complex with 1 μl of reservoir solution [0.1 M sodium acetate (pH 4.5), 0.2 M zinc acetate and 9% polyethylene glycol 3000] and incubating at room temperature (293 K). Hexagonal bipyramid-shaped crystals appeared overnight and grew for a few days, reaching maximum dimensions of 0.15 mm. Prior to data collection, the crystals were cryoprotected by briefly soaking them in a reservoir solution containing 30% ethylene glycol and flash-frozen in liquid nitrogen. Data were collected at MAX-Lab beamline I911-3 (Lund, Sweden). The crystal parameters and data collection statistics are presented in Table 2.

Structure determination

Diffraction data were indexed using MOSFLM [34] and scaled using Scala [35] from the CCP4 suite [36]. Molecular replacement was performed with MOLREP [37] using the coordinates of a coat protein dimer in the AB conformation from the crystal structure of Qβ bacteriophage (PDB entry 1QBE) as the search model. The solution was further refined using REFMAC [38]. Examination of the resulting electron density map revealed a region of unassigned density below the RNA-binding surface of the coat protein dimer that could be readily interpreted as RNA. To facilitate the modeling of the RNA stem–loop, we performed another round of molecular replacement in MOLREP using the coat protein dimer as the fixed input model and the partial Qβ operator (chain R from PDB entry 1ZSE) as the search model. The resulting model was subjected to several rounds of model building in Coot [39], refinement in REFMAC and validation using the MolProbity server [40]. Refinement and validation statistics are shown in Table 2.

Analysis and superimposition of atomic coordinates

The interface areas between the protein and RNA were calculated with PISA [41]. The rmsd values were calculated using the program LSQMAN [42]. The Qβ coat protein dimer in the AB conformation from the capsid structure (PDB entry 1QBE) was compared to that in the unassembled state by superimposing the C^{α} atoms of the two dimers with a distance cutoff of 3.5 Å. This resulted in the fit of 238 out of the 239 corresponding atoms, namely, residues 1–73 and 85–132 of chain A and residues 1–55, 61–74 and 85–132 of chain B in the unassembled dimer to the equivalent residues of monomers B and A in the capsid structure. To superimpose the RNAs, we used the backbone atoms C4′, P, C1′, C2′, C3′, O2′, O3′ and O4′ in all cases with residue ranges as indicated in the text. The cognate Qβ and MS2 coat protein–operator complexes were superimposed by explicitly fitting the Qβ coat protein C^{α} atoms of residues 33–37, 46–54, 62–71 and 87–97 to those of residues 30–34, 42–50, 56–65 and 83–93, respectively, in the MS2 coat protein–RNA complex (PDB entry 1ZDI).

Accession numbers

The atomic coordinates and structure factors of the Qβ coat protein–operator complex have been deposited in the Protein Data Bank with the accession code 4L8H.

Acknowledgments

We thank Dr. Andris Kazāks for help with protein purification and Ināra Akopjana for excellent technical assistance. We also thank the personnel at MAX-Lab for their help during our stay at the synchrotron. We are also grateful to Prof. Lars Liljas for reading and commenting on the manuscript. Our studies were supported by grant 09.1294 from the Latvian Research Council and grant 2DP/2.1.1.1.0/10/APIA/VIAA/052 from the European Regional Development Fund.

Received 2 July 2013;
Received in revised form 29 August 2013;
Accepted 30 August 2013
Available online xxxx

Keywords:
Leviviridae;
allolevivirus;
protein–RNA interaction;
RNA recognition;
translational repression

References

[1] van Duin J, Tsareva N. Single-stranded RNA phages. In: Calendar R, editor. The Bacteriophages. New York, NY: Oxford University Press; 2006. p. 175–96.
[2] Shiba T, Suzuki Y. Localization of A protein in the RNA-A protein complex of RNA phage MS2. Biochim Biophys Acta 1981;654:249–55.
[3] Bernardi A, Spahr PF. Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. Proc Natl Acad Sci USA 1972;69:3033–7.
[4] Meyer F, Weber H, Weissmann C. Interactions of Qβ replicase with Qβ RNA. J Mol Biol 1981;153:631–60.
[5] Valegård K, Liljas L, Fridborg K, Unge T. The three-dimensional structure of the bacterial virus MS2. Nature 1990;345:36–41.
[6] Gralla J, Steitz JA, Crothers DM. Direct physical evidence for secondary structure in an isolated fragment of R17 bacteriophage mRNA. Nature 1974;248:204–8.
[7] Weber H. The binding site for coat protein on bacteriophage Qβ RNA. Biochim Biophys Acta 1976;418:175–83.
[8] Nathans D, Oeschger MP, Polmar SK, Eggen K. Regulation of protein synthesis directed by coliphage MS2 RNA. I. Phage protein and RNA synthesis in cells infected with suppressible mutants. J Mol Biol 1969;39:279–92.

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025

[9] Beckett D, Wu HN, Uhlenbeck OC. Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. J Mol Biol 1988;204:939–47.
[10] Carey J, Cameron V, De Haseth PL, Uhlenbeck OC. Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. Biochemistry 1983;22:2601–10.
[11] Uhlenbeck OC, Carey J, Romaniuk PJ, Lowary PT, Beckett D. Interaction of R17 coat protein with its RNA binding site for translational repression. J Biomol Struct Dyn 1983;1:539–52.
[12] Romaniuk PJ, Lowary P, Wu HN, Stormo G, Uhlenbeck OC. RNA binding site of R17 coat protein. Biochemistry 1987;26:1563–8.
[13] Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. Crystal structure of an RNA bacteriophage coat protein–operator complex. Nature 1994;371:623–6.
[14] Valegård K, Murray JB, Stonehouse NJ, Van den Worm S, Stockley PG, Liljas L. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein–RNA interactions. J Mol Biol 1997;270:724–38.
[15] van den Worm SH, Stonehouse NJ, Valegård K, Murray JB, Walton C, Fridborg K, et al. Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. Nucleic Acids Res 1998;26:1345–51.
[16] Grahn E, Moss T, Helgstrand C, Fridborg K, Sundaram M, Tars K, et al. Structural basis of pyrimidine specificity in the MS2 RNA hairpin-coat-protein complex. RNA 2001;7:1616–27.
[17] Helgstrand C, Grahn E, Moss T, Stonehouse NJ, Tars K, Stockley PG, et al. Investigating the structural basis of purine specificity in the structures of MS2 coat protein RNA translational operator hairpins. Nucleic Acids Res 2002;30:2678–85.
[18] Persson M, Tars K, Liljas L. PRR1 coat protein binding to its RNA translational operator. Acta Crystallogr Sect D Biol Crystallogr 2013;69:367–72.
[19] Chao JA, Patskovsky Y, Almo SC, Singer RH. Structural basis for the coevolution of a viral RNA–protein complex. Nat Struct Mol Biol 2008;15:103–5.
[20] Witherell GW, Uhlenbeck OC. Specific RNA binding by Qβ coat protein. Biochemistry 1989;28:71–6.
[21] Lim F, Spingola M, Peabody DS. The RNA-binding site of bacteriophage Qβ coat protein. J Biol Chem 1996;271:31839–45.
[22] Golmohammadi R, Fridborg K, Bundule M, Valegård K, Liljas L. The crystal structure of bacteriophage Qβ at 3.5 Å resolution. Structure 1996;4:343–54.
[23] Spingola M, Peabody DS. MS2 coat protein mutants which bind Qβ RNA. Nucleic Acids Res 1997;25:2808–15.
[24] Lim F, Peabody DS. RNA recognition site of PP7 coat protein. Nucleic Acids Res 2002;30:4138–44.
[25] Horn WT, Tars K, Grahn E, Helgstrand C, Baron AJ, Lago H, et al. Structural basis of RNA binding discrimination between bacteriophages Qβ and MS2. Structure 2006;14:487–95.
[26] Schrödinger L. The PyMOL Molecular Graphics System, version 1.5.0.1. at http://www.pymol.org; 2012.
[27] Spingola M, Lim F, Peabody DS. Recognition of diverse RNAs by a single protein structural framework. Arch Biochem Biophys 2002;405:122–9.
[28] Rowsell S, Stonehouse NJ, Convery MA, Adams CJ, Ellington AD, Hirao I, et al. Crystal structures of a series of RNA aptamers complexed to the same protein target. Nat Struct Biol 1998;5:970–5.
[29] Convery MA, Rowsell S, Stonehouse NJ, Ellington AD, Hirao I, Murray JB, et al. Crystal structure of an RNA aptamer–protein complex at 2.8 Å resolution. Nat Struct Biol 1998;5:133–9.
[30] Klovins J, Overbeek GP, Van den Worm SHE, Ackermann H-W, Van Duin J. Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. J Gen Virol 2002;83:1523–33.
[31] Kazaks A, Voronkova T, Rumnieks J, Dishlers A, Tars K. Genome structure of caulobacter phage phiCb5. J Virol 2011;85:4628–31.
[32] Plevka P, Kazaks A, Voronkova T, Kotelovica S, Dishlers A, Liljas L, et al. The structure of bacteriophage phiCb5 reveals a role of the RNA genome and metal ions in particle stability and assembly. J Mol Biol 2009;391:635–47.
[33] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The Proteomics Protocols Handbook. Toyowa, NJ: Humana Press; 2005. p. 571–607.
[34] Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newslett Protein Crystallogr, 26; 1992.
[35] Evans PR. Scala. Joint CCP4 + ESF-EAMCB Newslett Protein Crystallogr, 33; 1997. p. 22–4.
[36] Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr Sect D Biol Crystallogr 1994;50:760–3.
[37] Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Crystallogr 1997;30:1022–5.
[38] Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr Sect D Biol Crystallogr 1997;53:240–55.
[39] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr Sect D Biol Crystallogr 2010;66:486–501.
[40] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr Sect D Biol Crystallogr 2010;66:12–21.
[41] Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol 2007;372:774–97.
[42] Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr Sect D Biol Crystallogr 1996;52:842–57.

Please cite this article as: Rumnieks Janis, Tars Kaspars, Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene, J Mol Biol (2013), http://dx.doi.org/10.1016/j.jmb.2013.08.025