RNA binding properties of the coat protein from bacteriophage GA

Jonatha M.Gott⁺, Larry J.Wilhelm[§] and Olke C.Uhlenbeck^{*}

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA

Received August 20, 1991; Accepted November 6, 1991

ABSTRACT

The coat protein of bacteriophage GA, a group II RNA phage, binds to a small RNA hairpin corresponding to its replicase operator. Binding is specific, with a K_d of $71 \mu\text{M}^{-1}$. This interaction differs kinetically from the analogous coat protein-RNA hairpin interactions of other RNA phage and also deviates somewhat in its pH and salt dependence. Despite 46 of 129 amino acid differences between the GA and group I phage R17 coat proteins, the binding sites are fairly similar. The essential features of the GA coat protein binding site are a based-paired stem with an unpaired purine and a four nucleotide loop having an A at position -4 and a purine at -7. Unlike the group I phage proteins, the GA coat protein does not distinguish between two alternate positions for the unpaired purine and does not show high specificity for a pyrimidine at position -5 of the loop.

INTRODUCTION

RNA phage coat proteins act as translational repressors of phage replicase synthesis by binding to a RNA hairpin that includes the ribosome binding site of the replicase gene (1,2). This RNA-protein interaction has been studied extensively in two group I phage, R17 and fr, and the group III phage, Q β (reviewed in 3). The two group I phage coat proteins recognize virtually identical hairpins despite 17 amino acid differences out of 129 residues. In contrast, the group I and III phage coat proteins differ considerably in their RNA recognition elements and are unable to bind each other's operator sequences. The existence of other natural variants which fall between these extremes should make it possible to extend these observations. In addition, the recent determination of the crystal structure of the group I phage MS2 (4) gives us a framework in which to interpret these results.

The group II bacteriophage, typified by phage GA, are closely related to the group I phage based on genome size and organization and antigenic cross-reactivities (5). The sequence and proposed secondary structure of the region of the GA phage genome that includes the replicase ribosome binding site are

strikingly similar to the R17 translational operator (6, Fig. 1). However, the GA hairpin differs from its R17 counterpart by having A's at positions -5 and -6 and possibly by the position of the unpaired A in the stem, since either A₋₁₀ or A₋₁₁ could pair with U₋₁ in the GA sequence. These differences are particularly interesting since the R17 coat protein has a strong preference for pyrimidines at loop residue -5 and requires the precise location of an unpaired purine at -10 for binding. The GA coat protein sequence differs from that of R17 coat protein at 46 of 129 positions (6). Despite the large number of amino acid changes, which are scattered throughout the protein, these proteins are clearly homologous and can be aligned along their entire length (3,6). Thus we expected the GA coat protein to bind this region of its genome. In this paper we show that the GA coat protein is a specific RNA binding protein, describe many of the properties of this RNA-protein interaction, and define essential features of its RNA binding site.

MATERIALS AND METHODS

Phage and Coat Protein

GA phage were kindly provided by Dr. Akikazu Hirashima and were propagated as described by Inokuchi et al. (7). GA coat protein was isolated from purified phage by the acetic acid extraction procedure developed for the R17 coat protein (8) and the same molar extinction coefficient (1.1×10^4 at 280 nm) was used for both proteins. The purity of the coat protein was confirmed by electrophoresis through formic acid gels (9).

RNA Synthesis

RNA fragments were prepared by *in vitro* transcription from synthetic DNA templates by T7 RNA polymerase (10). Transcription reactions contained 40 mM Tris-HCl (pH 8.1 at 37°C), 1 mM spermidine, 5 mM dithiothreitol, 50 $\mu\text{g}/\text{mL}$ bovine serum albumin (BSA), 0.1% (v/v) Triton X-100, and 80 mg/mL poly(ethylene glycol) (M_r 8000). Labeled RNAs were prepared in a 40 μL transcription reaction with 1 mM each of three NTPs, 0.25 mM [α - ^{32}P] NTP (5 μCi), 6 mM MgCl_2 , 100–400 nM template, and 0.05–0.1 mg/mL T7 RNA polymerase.

* To whom correspondence should be addressed

Present addresses: ⁺ Department of Molecular Biology and Microbiology, School of Medicine, Case Western Reserve, Cleveland, OH 44106 and [§]Synergen, Inc., Boulder, CO, USA

Nucleotides were purchased from Sigma. RNA fragments were gel purified by electrophoresis on 20% denaturing polyacrylamide gels run in 90 mM Tris-borate/2mM EDTA, cut out, eluted overnight in 0.1 M Tris-HCl (pH 8)/1 mM EDTA, and ethanol precipitated.

Filter binding assays

The association constant between coat protein and each RNA fragment was determined with a nitrocellulose filter retention assay described in detail by Carey et al. (8). A constant, low concentration of ^{32}P -labeled RNA was mixed with a series of coat protein concentrations between 0.1 nM and 1 μM in 10 mM magnesium acetate, 80 mM KCl, 80 $\mu\text{g}/\text{mL}$ serum albumin, and 100 mM Tris-HCl (pH 8.5 at 4°C) (TMK). After incubation at 4°C for 20 min, the mixture was filtered through a nitrocellulose filter and the amount of complex retained on the filter determined by liquid scintillation counting. For each experiment the data points were fit to a retention efficiency and a K_d value, assuming a bimolecular equilibrium (8).

RNA excess assays were used to determine the fraction of active protein. Experiments were carried out as above except that in these assays coat protein was held constant at 0.3 μM while the RNA concentration was varied between 0.12 nM and 0.45 μM .

Kinetic measurements on the GA and R17 coat proteins were carried out using RNA variant 9 under the conditions described above. For dissociation rate measurements complexes were formed by incubating RNA and coat proteins for 30–60 minutes on ice. Coat protein concentrations giving approximately 90% bound RNA were used for complex formation. Dissociation was initiated by either the addition of saturating amounts of unlabeled RNA hairpin or by twenty five-fold dilution in the same buffer at 4°C. Association rates were measured by adding saturating amounts of coat protein to RNA in TMK, mixing, and filtering at the times indicated.

RESULTS AND DISCUSSION

RNA binding by GA coat protein

To test whether GA coat protein could bind its replicase operator, we synthesized an RNA containing the appropriate portion of the GA genome with three G residues added at the 5' end to increase transcriptional efficiency (variant 1, Fig. 2A). As expected, GA coat protein binds to this synthetic RNA hairpin

Figure 1. Comparison of potential alternative secondary structures for the region of the GA phage genome that includes the replicase start site (left and center) with the R17 translational operator (right). Sequence differences between the two are shaded on the R17 hairpin. Numbers refer to nucleotide position relative to the A of the initiator AUG of the replicase gene, which is boxed.

with high affinity (Fig. 2A). The K_a for this interaction, 71 μM^{-1} , is quite similar to the value (100 μM^{-1}) obtained for the equivalent interaction with R17 components in the same buffer. RNA excess experiments (Fig. 2B) indicate that the protein is 100% active in RNA binding, assuming a stoichiometry like that of R17 of two protein monomers per RNA molecule (11). It is therefore likely that, like the other phage, GA regulates replicase translation with coat protein.

Although their affinities for their operators are similar, the GA and R17 coat proteins differ in their kinetics of binding RNA. The dissociation rate is seven-fold faster for GA coat protein (0.95 min^{-1}) than for the R17 coat protein (0.13 min^{-1} , Fig. 3A). These k_{-1} values are quite slow, indicating that once formed, both RNA-protein complexes are relatively stable. Interestingly, GA coat protein binds its RNA target approximately five-fold faster ($2.6 \times 10^8 \text{ M}^{-1} \text{ min}^{-1}$) than does R17 coat protein ($4.7 \times 10^7 \text{ M}^{-1} \text{ min}^{-1}$). This association rate for GA coat protein is only an estimate since the RNA is almost completely bound by the first time point (Fig. 3B). However, when k_1 is calculated from the values of K_a and k_{-1} for GA, a value of $1.6 \times 10^8 \text{ M}^{-1} \text{ min}^{-1}$ is obtained, which is still more than three fold faster than R17. Such a difference in the rate of association of the same substrate by similar proteins is not uncommon. For instance, k_1 values for NADH binding to lactate dehydrogenases from different sources can differ by over fifteen-fold (12). The

Figure 2. A) Coat protein excess binding curve for ^{32}P -labeled RNA variant 1 (inset) at 4°C in TMK buffer. B) RNA excess assay under the same conditions.

difference does suggest that an additional step may be limiting in RNA binding by R17 coat protein. One possibility is that the R17 coat protein exists in two or more forms in equilibrium with one another, only some of which are active in RNA binding. Alternatively, the R17 coat protein may undergo a conformational change upon binding, with only one form of the complex retained by nitrocellulose filters in a filter binding assay. Both association rates are considerably below the expected rate of diffusional

collision. Since the coat protein and RNA hairpin are both reasonably large, structured molecules, this most likely reflects the infrequency of productive collisions due to orientation effects.

RNA binding by GA coat protein is less dependent on ionic strength than that of other RNA phage coat proteins. Figure 4A shows that the K_a is a function of monovalent cation concentration. The K_a of the GA interaction decreases with increasing ionic strength, suggesting a significant contribution to binding by ionic contacts. However, the slope of the line is less steep than the ionic strength dependence of other phage coat proteins, yielding a maximum estimate of only three ion pairs for the GA interaction (see discussions in 13, 14). This may reflect differences in the charged residues of the RNA-binding face of the coat protein. By superimposing the aligned GA sequence on the MS2 coat protein tertiary structure, seven positively charged and four negatively charged residues are predicted to fall on the viral interior. In contrast, MS2 (or R17) coat protein has six positively charged and three negatively charged residues in this region. The five charged amino acid changes between GA and R17 (R43K, D54Q, K55N, K83R, and D89E) do not suggest a simple explanation for the salt

Figure 3. A) Dissociation of pre-formed complexes between RNA variant 9 and either GA coat protein (filled circles) or R17 coat protein (open circles) at 4°C in 100 mM Tris-HCl (pH 8.5), 10 mM magnesium acetate, 80 mM KCl, and 80 μ g/ml serum albumin. B) Association kinetics between RNA variant 9 and the GA coat protein (filled circles) or R17 coat protein (open circles) under the same conditions.

Figure 4. A) Ionic strength dependence of K_a . Protein excess binding curves with variant 1 were obtained at 4°C in 20 mM MOPS (pH 7.0), 10 mM magnesium acetate, and 80 μ g/ml serum albumin with KCl added to give the indicated $[M^+]$. B) pH dependence of K_a . Protein excess binding curves with RNA variant 1 were carried out at 4°C in 10 mM magnesium acetate, 80 mM KCl, and 80 μ g/ml serum albumin with 100 mM Na acetate (pH 5.0), MES (pH 6–6.5), MOPS (pH 7–7.5), or Tris-HCl (pH 8–8.5).

Figure 5. Sequence and potential secondary structures of RNA variants 2–18.

dependence. In fact, the reason is likely to be fairly complex since the ionic strength dependence of the association and dissociation rates differs between the two coat proteins (J.M.G., unpublished results).

The pH optimum for GA coat protein binding was obtained by determining the K_a at different pHs from protein excess binding curves. A plot of $\log K_a$ vs pH (Fig. 4B) indicates that the pH optimum is around seven, which differs from the pH optima for both the group I phage coat proteins (pH 8.5 for R17 and fr) (13,15) and the Q β coat protein (pH 6) (14). The pH dependence is not steep enough to reflect protonation of a particular side chain and the GA coat protein has no histidines or cysteines which might ionize in this range. However, since different optima were obtained for the GA and R17 coat proteins using the same RNA hairpin, the altered pH dependence is probably due to subtle effects on protein conformation.

Specificity of RNA binding

Sequence differences in the GA and R17 genomes suggest that the two coat proteins may recognize different elements of their respective binding sites. To investigate the possible effects of substitutions at these and other positions, we have measured the binding of GA coat protein to a number of RNA hairpins. K_a values for the RNA variants shown in Fig. 5 are given in Table 1. The GA coat protein binds tightly to the operator regions of the closely related group I phage fr (variant 2) and R17 (variant 4) and poorly to the equivalent region from the more distantly related group III phage Q β (variant 3). This is not surprising since variant 3 differs in both loop size and in the position of its unpaired nucleotide from the GA binding site. This result, along with data from other variants which show no binding, indicate that GA coat protein binds RNA with considerable specificity.

Only one of the loop nucleotides is essential for GA coat protein binding. While replacement of A₋₆ with U has little or no effect (variants 2 vs 4) and substitution of a G for A₋₇ reduces the K_a only three to four-fold (variant 5), a G at position -4 (variants 6 and 7) eliminates RNA binding. This result is similar to that with R17 coat protein (16) and may reflect either a site-specific contact to the base at -4 or a more general effect on the hairpin structure. Changes at position -5 (variants 8-11), an important

recognition element for both R17 and fr coat proteins, have only modest effects on GA coat protein affinity. Thus GA coat protein is less sensitive to changes in the loop than coat proteins from group I phage.

GA coat protein binding to the 'bulge' variants (variants 12-18) reflects the conformational ambiguity of its genomic binding site. Like R17 and fr (15-17), GA coat protein requires an unpaired purine within the stem (variants 12-16). Substitution of a U at this position eliminates binding (variant 18), as does the addition of a paired base opposite this residue in the helix (variant 17). However, unlike the R17 coat protein which only binds RNAs containing a 'bulge' at position -10 (3), the GA coat protein can bind hairpins having an unpaired A at either position -10 or -11 (variants 12-14 and 16). Since solution NMR experiments indicate that this base is intercalated into the RNA helix (18), it is likely that the major role of the unpaired A is structural, allowing various parts of the RNA to be brought into juxtaposition with appropriate protein side chains. Given its relative insensitivity to changes in the loop, such precise positioning of loop residues may be less critical for the GA interaction.

The essential features of the GA coat protein binding site are summarized in Fig. 6. Like the R17 coat protein, GA requires a base-paired stem with an unpaired purine and a four nucleotide loop containing an A at -4 and a purine at -7. Unlike the R17 binding site, the unpaired purine can be at either of two positions, -10 or -11. In addition, the GA coat protein does not show a preference for particular nucleotides at position -5 of the loop. Thus, despite their similar roles during phage infection, the details of these phage coat protein-RNA hairpin interactions are idiosyncratic.

Although the GA and R17 coat proteins differ at 46 positions, the residues likely to be involved in the recognition of unique features of their RNA binding sites can be narrowed to a few candidates. In the MS2 crystal structure (4), the inner surface of the phage is formed by approximately 29 amino acids from each monomer. Of these, only nine residues differ between the GA and R17 coat proteins: positions 43, 54, 55, 83, and 89, mentioned above, and 33 (L/I), 52 (G/S), 59 (A/T), and 87 (S/N, GA/R17 for each). These amino acids cluster in three distinct regions of the protein structure. One group, consisting of residues 52, 54, and 55, lies within the loop between β strands E and

Table 1.

	RNA Variant	K_a (μM^{-1})
GA operator	1	71
fr operator	2	91
Q β operator	3	0.6
R17 operator	4	91
loop variants	5	20
	6	<0.1
	7	<0.1
	8	56
	9	166
	10	17
	11	357
bulge variants	12	166
	13	50
	14	59
	15	91
	16	14
	17	<0.1
	18	<0.1

Figure 6. Essential features of the GA coat protein binding site. Dots represent positions that can be changed without altering the K_a significantly, the size of the dot corresponding to the number of mutations that were made at that position. R represents positions that must be purines. The arrows indicate that the unpaired purine in the stem can be in either of two positions.

F. This loop forms an arm-like extension into the interior of the phage and could, for example, contact the RNA stem. Residues 33, 43, and 83 fall at the other end of the β -sheet within the adjacent β_D , β_E , and β_F strands. Two of these, the arginine and lysine residues at positions 43 and 83, make up part of a positively charged patch which could interact with the phosphate backbone. The third region is made up of amino acids 59, 87, and 89, which cluster about the two-fold axis of symmetry at the dimer interface. By virtue of its position, this region is also probably involved in RNA binding.

It is quite possible that these nine amino acids are responsible for the differences in RNA binding by the GA and R17 coat proteins. The small nuclear ribonucleoproteins U1A and U2B'' recognize related RNA hairpins within the U1 and U2 RNAs, respectively. By exchanging only eight amino acids in the amino-terminal domains, Mattaj and colleagues have largely reversed the binding specificity of these two proteins (19). Similar experiments can now be done on the GA and R17 proteins to help define the residues which contact particular regions of the RNA hairpin.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Lars Liljas for providing the MS2 crystal structure coordinates and to Dr. Gary Witherell for helpful discussions. This work was supported by a National Institutes of Health Postdoctoral Fellowship Grant (GM-12452) to J. M. G. and by a National Institutes of Health Grant (GM-36944) to O. C. U. We also thank the W. M. Keck Foundation for their generous support of RNA science on the Boulder Campus.

REFERENCES

1. Lodish, H.F. and Zinder, N.D. (1966) *J. Mol. Biol.* 19, 333–348.
2. Bernardi, A. and Spahr, P.F. (1972) *Proc. Natl. Acad. Sci.* 69, 3033–3037.
3. Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C. (1991) *Prog. Nucl. Acid. Res. Mol. Biol.* 40, 185–220.
4. Valegård, K., Liljas, L., Fridborg, K. and Unge, T. (1990) *Nature* 345, 36–41.
5. van Duin, J. (1988) In Calendar, R. (ed.), *The Bacteriophages I*. Plenum Press, New York, pp. 117–167.
6. Inokuchi, Y., Takahashi, R., Hirose, T., Inayama, S., Jacobsen, A.B. and Hirashima, A. (1986) *J. Biochem.* 99, 1169–1180.
7. Inokuchi, Y., Hirashima, A. and Watanabe, I. (1982) *J. Mol. Biol.* 158, 711–730.
8. Carey, J., Cameron, V., de Haseth, P.L. and Uhlenbeck, O.C. (1983) *Biochemistry* 22, 2601–2609.
9. Strauss, E.G. and Kaesberg, P. (1970) *Virology* 42, 437–452.
10. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) *Nucleic Acids Res.* 15, 8783–8798.
11. Beckett, D. and Uhlenbeck, O.C. (1988) *J. Mol. Biol.* 204, 927–938.
12. Hammes, G.G. and Schimmel, P.R. (1970) In Boyer, P.D. (ed.), *The Enzymes*, Vol. 2. Academic Press, New York, p. 67.
13. Carey, J. and Uhlenbeck, O.C. (1983) *Biochemistry* 22, 2610–2615.
14. Witherell, G.W. and Uhlenbeck, O.C. (1989) *Biochemistry* 28, 71–76.
15. Wu, H.-N., Kastelic, K.A., and Uhlenbeck, O.C. (1988) *Nucleic Acids Res.* 16, 5055–5066.
16. Romaniuk, P.J., Lowary, P., Wu, H.-N., Stormo, G. and Uhlenbeck, O.C. (1987) *Biochemistry* 26, 1563–1568.
17. Wu, H.-N. and Uhlenbeck, O.C. (1987) *Biochemistry* 26, 8221–8227.
18. Borer, P.N., Bishop, K.D., Gott, J.M. and Uhlenbeck, O.C. (in preparation).
19. Scherly, D., Boelens, W., Dathan, N.A., van Venrooij, W.J. and Mattaj, J.W. (1990) *Nature* 345, 502–506.