Mechanism of RNA-protein interactions in tobacco mosaic virus: analysis of the pH stability of virus protein complexes with synthetic polynucleotides

R.K.Ledneva, T.P.Lanina, G.V.Terganova and A.A.Bogdanov

A.N.Belozersky Laboratory of Molecular Biology and Bioorganic Chemistry, Moscow State University, Moscow 117234, USSR

Received 25 March 1980

ABSTRACT

TMV-like RHP complexes were reconstituted from TMV protein and synthetic polynucleotides. Analysis of the pH stability of RHP with polynucleotides containing U, G, or their analogues reveals a correlation between the stability of their structure and the pK values of the bases, and indicates that the -NH-CO-groups of U and G are involved in hydrogen bonding with protein. It is suggested that TMV protein has two U- and one G-specific binding sites which, according to the phase position of the protein subunits relative to the origin of TMV assembly (D. Zimmern (1977), Cell 11, 463) are likely to be organized as UGU. The binding of the A and C residues of RNA with TMV protein is nonspecific. TMV protein groups with pK 6.3, 7.5 and 9.7 were found to be essential in the protein-protein interactions in RHP. A group of the protein with pK 8.2 is also involved in RHP stabilization. Both protein-protein interactions and interactions of protein with RNA phosphate groups were shown to be mediated by a conformational change in the protein induced by base binding. The effect of bases on both types of interactions changes in the order G=U>A, and incorporation of G in RHP proceeds in a compulsory way at the expense of interaction of the neighbouring nucleotide residues in polynucleotides with protein. The data obtained are used to discuss the principles of the cooperativity of the interactions between TMV components and the mechanism of initiation and elongation in TMV self-assembly.

INTRODUCTION

Tobacco mosaic virus (TMV) is a helically organized nucleoprotein, in which the RNA chain is coated by identical protein subunits. General features of TMV arrangement have been elucidated by X-ray diffraction analysis\(^1\). However, the structure of the binding sites for RNA bases is unknown. These sites should be able to bind any base since the protein contacts with RNA throughout its length and, at the same time, to specifica-
ly recognize a nucleotide sequence in the unique region of RNA where virus assembly is initiated (origin of assembly, 0-2). The ability of TMV protein to form virus-like ribonucleoproteins (RNP) with synthetic polynucleotides3,4 allows one to simplify substantially the studies on nucleic acid-protein interactions in TMV. In this work we have studied the pH dependence of the stability of RNP reconstituted from TMV protein and polynucleotides. This approach is based on the well-known property of TMV to dissociate in alkaline media with formation of stable intermediates the existence of which has been attributed to the variability of interaction between nucleotide residues of the RNA and the protein along the rod5. The conclusion is made about the existence in TMV protein of U- and G-specific binding sites and about their role in the maintenance of virus stability.

MATERIALS AND METHODS

TMV protein-polynucleotide complexes. TMV protein was prepared by the acetate method6. The polynucleotides were purchased from Sigma; poly (s4U73, C27) was a gift of Dr. M. Karpeisky (Institute of Molecular Biology, Moscow). RNP were reconstituted from TMV protein and polynucleotides (20:1 w/w) in 0.1 M phosphate buffer, pH 6.6, for 18 hr at 24°C. The reaction mixtures were dialyzed at 4°C against 0.05M NaCl-0.01 M Tris-HCl buffer, pH 7.2, and RNP were sedimented by centrifugation for 2 hr at 40,000 r.p.m. Then RNP were purified by 2 or 3 sedimentations under the same conditions and suspended in the same buffer. Purification of poly(AC)-containing RNP was carried out in 0.05 M NaCl - 0.01 M acetate buffer, pH 6.6, and the dialysis step was omitted. The polynucleotide and protein contents in RNP were determined by the phosphate assay and the Lowry method, respectively.

Measurement of the pH stability of RNP. Titration of RNP was performed at 20-24°C by 50 to 100-fold dilution of RNP suspensions with appropriate mixtures of 0.1 M NaH2PO4-Na2HPO4 (to attain pH values higher than 9.2 conc. NaOH was added to 0.1 M Na2HPO4). Measurements were performed at intervals of
0.25 pH unit. The pH values of samples were constant during measurements. The dissociation of RNP was monitored by the decrease in CD at $\lambda_{\text{max}}$ of the difference spectrum (Table 1), the spectra being measured 2-3 min after dilution of the RNP suspension. The percentage of CD decrease was taken as percentage of dissociation. The difference CD spectra were calculated by subtracting the sum of the CD spectra of the components from the CD spectra of RNP as described previously. Titration of TMV protein (in 0.1 M phosphate buffer, pH 7.2) and of poly(U)-containing RNP in the pH range 5.0-9.0 (Fig. 1) was performed as described above but, before CD measurement, the diluted samples had been left to stand for 18 hr at 20-24°C; in this case measurements were performed at intervals of 0.1 pH unit. The CD spectra were recorded using a Roussel-Jouan II dichrograph. Alterations in the RNP structure were also registered by measuring RNP sedimentation coefficients at certain pH values in a scanning Beckman Model E ultracentrifuge.

RESULTS AND DISCUSSION

The simplest models for studying the interaction of TMV protein with RNA bases are RNP containing homopolynucleotides. However, out of the four types of homopolynucleotides, only poly(A) and poly(U) form complexes with TMV protein. The mechanism of G and C binding was investigated with the use of their copolymers with A and U; poly(I) was also used as an analogue of poly(G) for the same purpose. Most important characteristics of these RNP are listed in Table 1. The pH dependence of RNP stability was studied by sedimentation analysis and circular dichroism; the latter is highly sensitive to conformational changes in TMV and in A- and U-containing RNP.

The study of the pH stability has demonstrated that all the complexes obtained by us, except RNP with poly(AC), are stable at pH lower than 7.2. Increase in the pH of the medium induces dissociation of RNP, which entails a decrease in the sedimentation coefficients ($S_{20,w}$) and CD. In a more acidic medium (pH 5.5) all RNP have somewhat higher $S_{20,w}$, and RNP with $S_{20,w} \leq 100$ display slightly higher CD compared to that at pH 7.2 (Table 1).
TABLE I
Characteristics of TLV protein complexes with synthetic polynucleotides

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly(U)</td>
<td>3:1.2</td>
<td>269.1 6.9 6.5 100 75 75 75 (8.1)</td>
<td>270</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pH 5.5, pH 6.6, pH 7.2</td>
<td>pH 8.0</td>
<td></td>
</tr>
<tr>
<td>poly(A)</td>
<td>3:1</td>
<td>7.2 170 125 125 125(10.5)</td>
<td>30(10.5); 115(8.5); 110(8.5); 275 (9.5)</td>
<td>9.6</td>
</tr>
<tr>
<td>poly(I)</td>
<td>3:1</td>
<td>6.6 150 120 90(9.5)</td>
<td>275 (10.3)</td>
<td></td>
</tr>
<tr>
<td>poly(A&lt;sub&gt;46&lt;/sub&gt;,C&lt;sub&gt;54&lt;/sub&gt;)</td>
<td>3:1.4</td>
<td>275 1230 105 105(8.0); 90(9.75)</td>
<td>252</td>
<td>10.3</td>
</tr>
<tr>
<td>poly(A&lt;sub&gt;90&lt;/sub&gt;,U&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>3:1.1</td>
<td>271 6.1 5.7 100</td>
<td>275</td>
<td>10.1</td>
</tr>
<tr>
<td>poly(A&lt;sub&gt;63&lt;/sub&gt;,G&lt;sub&gt;37&lt;/sub&gt;)</td>
<td>3:1.2</td>
<td>273 6.2 5.7 130</td>
<td>275</td>
<td>10.4</td>
</tr>
<tr>
<td>poly(U&lt;sub&gt;73&lt;/sub&gt;,C&lt;sub&gt;27&lt;/sub&gt;)</td>
<td>3:1.1</td>
<td>347 4.4 4.3 160 100(8.25); 20(10.25)</td>
<td>360</td>
<td>9.9</td>
</tr>
<tr>
<td>poly(U&lt;sub&gt;45&lt;/sub&gt;, C&lt;sub&gt;55&lt;/sub&gt;)</td>
<td>3:1.1</td>
<td>6.7 150</td>
<td>272</td>
<td>10.1</td>
</tr>
<tr>
<td>poly(U&lt;sub&gt;60&lt;/sub&gt;,G&lt;sub&gt;40&lt;/sub&gt;)</td>
<td>3:1.4</td>
<td>270 3.9 150</td>
<td>275</td>
<td>10.4</td>
</tr>
<tr>
<td>poly(A&lt;sub&gt;24&lt;/sub&gt;, G&lt;sub&gt;39&lt;/sub&gt;, C&lt;sub&gt;12&lt;/sub&gt;)</td>
<td>3:1.6</td>
<td>272 4.6 3.1 125</td>
<td>273</td>
<td>10.6</td>
</tr>
<tr>
<td>TLV RNA</td>
<td>3:1</td>
<td>271 5.1 5.1 275</td>
<td>10.9</td>
<td></td>
</tr>
</tbody>
</table>

1) Since TLV and virus-like RITP are helical structures comprising $N$ turns of the polynucleotide and $(N + 1)$ turns of the protein moiety, the RNA:protein molar ratio would decrease (from 3:1 characteristic of TLV and long RITP) with decreasing rod length. The ratio 3:1.09 observed for poly(A)- RITP is due to partial stripping of protein subunits from poly(A) during RITP purification at pH 7.2. 2) $\lambda_{max}$ of the difference CD spectra of RITP at which the RITP dissociation was monitored. 3) pH values at which the dissociation of RITP is 50%. 4) In brackets, pH values at which $S_{20, w}$ were measured.
Structural rearrangements of RNP in the pH range 5.0-9.0 were studied in detail for RNP with poly(U). The relative change in CD at pH values higher and lower 7.0 (Fig. 1) was found to be the more pronounced the shorter was RNP; hence it should reflect the change in the RNP conformation on the terminal turns of the helix. As one can see from Fig. 1 there are two transitions with $pK_1 \sim 6.3$ and $pK_2 \sim 7.5$. The free protein seems to undergo similar conformational changes; in this case, however, higher proton concentrations are required, i.e. $pK_1 \sim 5.7$ and $pK_2 \sim 7.1$. The transition in the free protein with $pK \sim 7.0$ at $\mu = 0.1$ that is due to the uptake of one proton per protein subunit was previously recorded by the CD technique. Recently the $pK_1$ value was estimated at $6.4 \pm 0.5$, which is somewhat higher than the $pK \sim 5.7$ found in this work. These $pK$ values were assigned to two carboxylates which were believed by Caspar to form hydrogen-bonded carboxyl-carboxylate pairs with $pK$ values being higher in the virus. The parallelism in the shifts of the curves in Fig. 1 allows one to suggest that the pH-dependent change in the CD of RNP reflects the titration of these carboxyl-carboxylate pairs in the protein and that the $pK$ values of these pairs in RNP rise as a result of the polynucleotide-protein interaction. Our estimate of $pK_2 (\sim 7.5)$ coincides with that in TMV but the $pK_1$ value ($\sim 6.3$) obtained in this work differs from that in the mentioned report ($pK 5.8$). As shown below, the change
Fig. 2. pH dependence of dissociation of RNP.

a. Dissociation curves of RNP with poly(A)-1, with poly(I)-2, with poly(U)-3 and of TMV - 4. Note that the dissociation curves of the native and in vitro reconstituted preparations of the virus are identical.


The final concentrations of poly(N) in samples were ~5 x 10^-5 M.

In the protonation state of these carboxyl-carboxylate pairs affects the stability of various RNP to a different degree.

Fig. 2 demonstrates that at pH above 7.2 the RNP with poly(A) is the least stable of the homopolyribonucleotide RNP, whereas RNP with poly(I) and poly(U) are much more stable. Introduction of U or G residues into poly(A) produces a considerable stabilizing effect on RNP structure, while inclusion of C works to the contrary.

It is obvious that the presence of a G or U lends stability to the structure of RNP in an alkaline medium that is comparable with the stability of the virus itself, although the most stable RNP complexes are somewhat less stable than TMV (Table 1). The correlation of the pH stability of RNP and the changes in the pK values of acidic dissociation of the bases in the order hypoxanthine (9.62)< guanine (10.00)=uracil (10.06) indicates the participation of the imino ring proton of the base in the formation of the hydrogen bond with the protein. The same relationship has been revealed in the interaction of the pro-
tein with analogues of poly(U). Increase in the acidity of the proton in 4-thiouracyl ($pK \sim 8.2^{14}$) is responsible for dissociation of RNP with poly($^4$UC) at pH values lower than those for RNP with poly(UC) (Table 1). On the other hand, the inability of poly($\text{h}_2\text{U}$) to form RNP with TMV protein found out in this work seems to result from the lower tendency of dihydrouracyl to the formation of the hydrogen bond due to both the decrease in the positive character of H at N(3) ($pK \sim 12^{13}$) and the flexibility of the structure of the base. It is noteworthy that guanine and hypoxanthine which, unlike adenine, form stable bonds with protein, have the same structural element, i.e. an -NH-CO-group capable of forming two hydrogen bonds. And although it cannot be unequivocally deduced from our data that the bases form two rather than one hydrogen bond, such a double-point binding is energetically much more favourable.

With uracyl, the most probable candidate for the formation of the second hydrogen bond is $\text{O}(4)$; this follows from the great differences in the binding of U and C by TMV protein in RNP. The presence of the invariant -NH-CO-group in U and G may at first seem to indicate the existence of a common site for the binding of these bases. However, since the hydrogen bonds are highly sensitive to the distance between the interacting groups, the relative spatial localization of the -NH-CO-fragment of the bases and the respective groups of the protein in the binding site should be the same. With the fixed conformation of nucleotide residues of intraviral RNA $^1$, this is obviously not the case for U and G. Hence one is compelled to accept the existence of different sites for specific recognition of these RNA bases.

Dissociation of RNP with poly(A) displays a complex pattern with a smooth phase in the pH range 7.0 to 9.0 and an abrupt one at pH 9.0-10.2. Introduction of C in poly(A) causes dissociation of the complex in the pH range where the poly(A)-protein complex is stable. Since poly(A) and poly(AC) have no groups that ionize in the pH range 6.7 to 10.2, the observed structural rearrangements should be ascribed to the changes in the ionization of the functional groups of the protein.

Most critical for the stability of RNP with poly(A) is the
ionization of the group with pK \sim 9.7. Since none of the amino acid residues that ionize around this pH, i.e. Cys, Tyr, Lys, is localized in the region of RNA binding, the dissociation of RNP is due to the disruption of the protein-protein interactions and indicates that Tyr or Lys (but not Cys, as can be inferred from the chemical modification data) is involved in the stabilization of the helical packing of the protein subunits in TMV. Dissociation of RNP in the pH range 7.0 to 9.0 seems to depend on the deprotonation of the carboxyl-carboxylate pair (pK \sim 7.5) and another group with pK \sim 8.2 in the protein. A group with pK 8.2 in TMV was detected previously and assumed to be a carboxyl group localized in the proximity of the phosphate group of RNA. Deprotonation of these groups, like of that with pK \sim 9.7, does not affect the structure of the G- and U-containing complexes, but induces dissociation of RNP with poly(A) which is subsequently enhanced by ionization of the group with pK \sim 9.7. In the case of RNP with poly(AC), in the pH range 6.7 to 7.0 the structure is stabilized at the expense of A residues, and the dissociation of the complex in the sites of C localization is likely to result from the ionization of the anomalously titrated protein carboxyl group in RNP (pK \sim 6.3).

It can be suggested that dissociation of a hypothetical RNP with poly(C) should have been very easy, and, as will be clear from the following discussion, it can be regarded as a dissociation of a repolymerized form of the protein. However, since repolymerization of the protein seems to preclude the building of any polynucleotide into RNP, reconstitution of RNP with poly(C) is impossible in principle.

The different stability of poly(A)- and poly(AC)-containing RNP can be explained by high adenine hydrophobicity. Increase in the hydrophobicity as a result of the introduction of the second ring to the cytosine base makes possible the formation of a TMV protein complex with poly(3,N\textsuperscript{4}-ethenocytidylic acid). The data on the interaction of TMV protein with poly(εC) and the inference about the inability of pyrimidine O(2) group to bind with protein add up to the conclusion that C does not form any specific bonds in a ribonucleo-
protein. This conclusion seems to hold for A as well and hence accounts for the common property of the poly(A)-and poly(AC)-containing complexes, i.e. for the greater, compared to U- and G-containing RNP, dependence of their stability on protein-protein interactions.

The greater stability of the protein-protein linkages in the virus compared to repolymerized protein is ascribed to the presence of RNA and its interaction with the protein that compensates for the adverse effect of the ionization of the two carboxyl-carboxylate pairs on the helical packing of the subunits. It is obvious then that the stabilization of the protein-protein interactions by RNA is due to the binding of U and G and, to a lesser degree, of A. Moreover, the data obtained in this work show that not only the protein-protein, but also the electrostatic nucleic acid-protein interactions in RNP are determined by the nature of the bases in the polynucleotide. This follows from the fact that the interaction of the phosphate groups of the polynucleotides with the protein is masked in the course of RNP dissociation, which depends only on the nature of the bases. This means that the electrostatic nucleic acid-protein interaction in RNP are possible as long as the bases of the polynucleotide remain in the binding sites of the protein.

It is known from X-ray diffraction data that three arginine residues (90, 92, and probably 113) are involved in the interaction with phosphate groups of RNA. In the free protein (20S disk) these arginine residues are located in the flexible loop of residues 89-113. This region of TMV protein has a much more ordered conformation in TMV. Thus the dependence of electrostatic nucleic acid-protein interactions on RNA base-protein binding observed in this work may be interpreted as being the result of a rearrangement of the protein structure in the region of phosphate-binding sites induced by binding of RNA bases.

In accordance with the recent model of the RNA organization in TMV, the binding sites of the bases and phosphate groups are separated; the surface of the protein subunit that faces the 5'-end of the particle carries three binding sites.
for the bases and one for the phosphate (Arg 113), and the opposite surface bears two binding sites for the phosphate, Arg 90 and Arg 92 \(^1\). With such localization of the binding sites, protein subunits of the adjacent layers of the RNP helix should be involved in the fixation of each turn of the polynucleotide. In this connection the bases can affect the electrostatic nucleic acid-protein interactions in two ways: (1) directly, e.g. by stabilizing the position of the three arginine residues on the protein subunit which binds the bases, the electrostatic binding of the two adjacent polynucleotide turns being affected thereby; (2) indirectly, via the protein of the proximal (in the 5'-direction) layer that has Arg 90 and Arg 92 available for phosphate binding (the effect of this type is unidirectional and operative for the binding of just one turn of the polynucleotide). As the protein, which always mediates the effect of bases on the electrostatic nucleic acid-protein interactions, also participates in the formation of specific protein-protein contacts, all types of the intermolecular bonds in a nucleoprotein prove to be affected by the base-protein interactions.

The effect of the bases on the stability of RNP structure depends on the mode of their binding and is displayed in the pattern of the dissociation curves (Fig. 2). It seems that the initial part of the dissociation curve of RNP with poly(A) reflects the consecutive removal of protein subunits from the end of the particle owing to the ionization of the carboxyl protein group with \(pK \approx 7.5\). This process is promoted by the ionization of the second group of the protein with \(pK \approx 8.2\). But as the group with \(pK \approx 9.7\) is ionized, the dissociation of the complex either greatly accelerates or becomes cooperative. In the case of U- and G-containing RNP, the dissociation is always cooperative. One can suggest that the interaction of A and the protein can secure only one turn of the nucleoprotein helix, mostly via the stabilization of protein-protein interaction, whereas the formation of specific hydrogen bonds between the protein and U and G residues affects also the electrostatic nucleic acid-protein interaction of the adjacent (in the 3'-end direction) turn of the polynucleotide, supporting thereby the structure of two turns of the RNP helix.

The data obtained by us allow interesting conclusions to
be made as to the mechanism of assembly of TMV and to its structure in toto. The primary step in the initiation of virus assembly is the binding of the bases in the RNA Oa region with protein subunits which form the underside of the A-layer of the 20S disk. The binding of the bases induces changes in the structure of the protein that permit the B-layer to approach the A-layer (at R<60 Å). In other words, the bases are the effectors that remove the hindrances for axial protein-protein interactions from the surface of the protein that faces the 5'-end of the forming particle. As a consequence of this rearrangement, the phosphate groups of the Oa region of RNA bind with the B-layer. The formation of the initiation complex is a result of a cooperative interaction of the protein with A, U, and G bases of RNA and is accompanied or terminated by transformation of the structure of the planar disk into a fragment of the nucleoprotein helix. The complex is mostly stabilized by insertion of U and G in their specific binding sites. The subsequent coating of the 5'-tail of RNA is facilitated by polymerization of the protein along RNA, which is elicited by the binding of RNA bases with the surface of the nucleoprotein that faces the 5'-end. The assembly in the 3'-end direction is promoted by the electrostatic nucleic acid-protein interactions that are enhanced by efficient formation of the phosphate-binding sites on the opposite surface of the nucleoprotein owing mostly to the specific binding of the U and G from the preceding RNA turn.

Stabilization of the virus structure requires a certain regularity in the distribution of the U and G residues along the whole RNA chain governed by the location of the U- and G-specific sites in the protein. However, the presence of U or G in each trinucleotide unit is not obligatory for the stabilization of the structure of the RNP helix. As shown with poly (A90, U10) (Fig. 2), the specific binding of, on the average, one U residue per three protein subunits produces a considerable stabilizing effect; this means that U (as well as G) serves as a pin that fastens the RNP helix in its different sites allowing A and C residues (as well as U and G located in nonspecific centers) to be held in place even when the interaction
of these units with the protein is very weak. As this effect is based on the influence of U and G on the binding of the adjacent RNA turn, the distance between the structure-forming U or G residues should not exceed the length of one turn, i.e. at most 49 nucleotide residues.

To provide recognition of the assembly initiation site, the periodicity of U and G in the sequence of $O_a$ should be most pronounced. Analysis of the structure of the $O_a$ region of RNA (line 0 in $O_a$, Fig. 6 in 2) shows that 11 out of 16 G fall into one of the three positions, and 12 out of 13 U are almost equally divided between the two remaining positions of the trinucleotides in RNA. The latter may indicate the existence of two rather than one U-binding sites. Hence the specific base-binding sites in TMV protein are likely to be arranged as UGU, or else as UUG/GUU on the shift of the phase position of the protein subunits relative to $O_a$. Uniform distribution of A and C over the three positions in the same region of $O_a$ agrees with the absence of specific binding sites for these RNA bases in TMV protein. Finally, U and G, when localized in nonspecific sites, probably behave as C and A, respectively, their binding depending first and foremost on their hydrophobicity. As the hydrophobicity of U is too low to ensure its binding, its occurrence in nonspecific sites should be very unfavourable for the stability of the structure compared to a similar situation with G. Hence the existence of two binding sites for U as a means to compensate for the lack of hydrophobicity seems to be highly plausible.

ACKNOWLEDGEMENTS

We are indebted to Professor J.G. Atabekov for a kind gift to TMV, to Drs E.N. Dobrov and Shie Medon for critical consideration of the manuscript and useful discussion.

REFERENCES

6 Praenkel-Conrat, H. (1957) Virology 4, 1-4