1346–1353 Nucleic Acids Research, 1994, Vol. 22, No. 8

The 3′-untranslated region of alfalfa mosaic virus RNA 3 contains at least two independent binding sites for viral coat protein

Chantal B.E.M.Reusken, Lyda Neeleman and John F.Bol*
Institute of Molecular Plant Sciences, Gorlaeus Laboratories, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands

Received February 17, 1994; Revised and Accepted March 22, 1994

ABSTRACT

The 3′-termini of the three genomic RNAs of alfalfa mosaic virus contain a common sequence of 145 nucleotides (nt) with a specific binding site for coat protein (CP). This sequence consists of several stem/loop structures interspersed with single-stranded AUGC-motifs; in RNA 3 this folding pattern is extended to a region upstream of the homologous sequence. By band-shift assays a minimum of two specific binding sites for CP were identified near the 3′-end of RNA 3. Site 1 consists of the region between nt 11 and 127 from the 3′-end and contains two AUGC-motifs. Site 2 is located between nt 133 and 208 from the 3′-end in a sequence that is largely unique to RNA 3 and contains also two AUGC-motifs. Deletion studies revealed that the two sites could bind CP independently of each other and permitted the identification of sequence elements that are essential for the activity of each site. By site-directed mutagenesis it was shown that the AUGC-motifs are important for binding of CP to both sites. These binding sites may play a role in the phenomenon that each genomic RNA has to be complexed with a few CP molecules to initiate infection. Later in the replication cycle they may act as origins for the assembly of virus particles.

INTRODUCTION

Although the overall structure of the tripartite RNA genomes of bromoviruses, cucumoviruses, ilarviruses and alfalfa mosaic virus (AlMV) is similar, there are two major differences leading to a subdivision of these virus groups. First, the RNAs of bromo- and cucumoviruses contain a 3′-terminal tRNA-like structure that is absent in the RNAs of AlMV and ilarviruses. Second, a mixture of the three genomic RNAs of bromo- and cucumoviruses is infectious, whereas a mixture of AlMV or ilarvirus RNAs 1, 2 and 3 is infectious only after addition to the inoculum of a few molecules of coat protein (CP) per RNA molecule or of RNA 4, the subgenomic messenger for CP. This early function of CP of AlMV and ilarviruses has been termed 'genome activation' (for a review see 1).

RNAs 1 and 2 of AlMV encode the replicase subunits P1 and P2, respectively, whereas RNA 3 encodes both the movement protein P3 and CP. The sequence of the subgenomic RNA 4 is identical to that of the 3′-terminal 881 nucleotides of RNA 3. The four RNAs are separately encapsidated into bacilliform particles that are predominantly stabilized by protein-RNA interactions (2). The observation that AlMV RNAs are able to withdraw CP subunits from virus particles *in vitro* indicated that the RNAs contain specific binding sites with a high affinity for CP (3). In RNA 4 these high-affinity binding sites were predominantly located in the 3′-noncoding region (4). The 3′-terminal 145 nucleotides (nt) of RNAs 1–4 show a sequence similarity of 80% and, despite the nucleotide differences in these sequences, all termini can be folded into a similar secondary structure (5). The proposed secondary structure consists of a number of RNA hairpins flanked by single-stranded AUGC-motifs and is supported by enzymatic structure mapping (6). Similar structures have been observed at the 3′-termini of the ilarviruses tobacco streak virus (TSV) (7) and lilac ring mottle virus (8). Studies on the degradation by ribonucleases of complexes between AlMV or TSV RNAs and CP revealed that protected fragments were located near the 3′-termini of the genomic RNAs. In addition a few internal sequences were found to be protected in these assays and most could also be folded into stem/loop structures flanked by AUGC-boxes (9, 10, 11). TSV CP binds to the 3′-termini of AlMV RNAs and activates the AlMV genome and also the reverse is true although there is no significant sequence similarity between the two CP's. Further evidence that binding of CP to specific binding sites near the 3′-termini of viral RNA is related to the early function of CP in the viral replication cycle was provided by the finding that each genomic RNA in an AlMV inoculum has to be complexed with a few dimers of CP molecules to initiate infection (12). Removal of the N-terminal 25 amino acids of CP interferes with RNA binding and genome activation (13). Recently, it was confirmed that the basic N-terminal peptide of AlMV CP is involved in RNA binding and genome activation (14).

When plants are infected by transient expression of an inoculum consisting of AlMV cDNAs fused to the 35S promoter from cauliflower mosaic virus DNA, the requirement for CP in the

*To whom correspondence should be addressed

inoculum is greatly reduced (15). When P1 and P2 are constitutively expressed from nuclear integrated cDNAs in transgenic tobacco plants (P12 plants), CP is no longer required for the infection of these plants with AlMV RNA 3 (16). These results have led to the hypothesis that binding of CP to the 3′-termini of AlMV and ilarvirus RNAs is required to protect these RNAs from 3′-terminal nucleolytic degradation during primary translation until a replicase enzyme has been assembled (15). With bromo- and cucumo-viruses the 3′-terminal tRNA-like structure could provide such a stabilizing function (17).

In addition to its early function and its structural role, AlMV CP has been proposed to play a role in AlMV RNA replication. When cowpea protoplasts were infected with RNAs 1 and 2 and CP the level of viral plus-strand RNA accumulation was 100-fold reduced whereas the accumulation of minus-strand RNA was increased compared to an infection with the complete genome (18). Deletions or frameshifts in the P3 gene did not affect RNA 3 accumulation in protoplasts from transgenic P12 plants but similar mutations in the CP gene resulted in a 100-fold reduction in RNA 3 accumulation (19, 20). It was proposed that CP has a regulatory role in switching the viral replication complex from symmetric synthesis of viral plus- and minus-strand RNAs to asymmetric plus-strand RNA synthesis although the possibility could not be ruled out that in the absence of functional CP an enhanced degradation of plus-strand RNAs occurs. More specific evidence for a role of CP in RNA synthesis was provided by the observation that AlMV replicase activity could be trapped with antibodies to CP that were bound to agarose beads (21). Recently, Houwing and Jaspars (22) reported that addition of CP increases the accumulation of plus-strand RNA in an *in vitro* RNA synthesizing system and these authors put forward the hypothesis that the early function of CP is related to its putative role in viral RNA synthesis.

Detailed knowledge on the origin of assembly of plus-strand RNA plant viruses is limited to studies on tobacco mosaic virus (TMV) (23). Possibly, binding of CP to the 3′-termini of AlMV RNAs has an early function in the initiation of infection and a late function as the first step in the assembly of virions. Binding sites for BMV CP appeared to map internally in BMV RNA 1 and not in the 3′-terminal sequence of this RNA that contains the tRNA-like structure (24). Also for turnip crinkle virus CP, binding sites mapped internally in the replicase and CP genes (25, 26). In the present work we used band-shift assays with mutant RNA transcripts to further characterize the CP binding site near the 3′-terminus of AlMV RNA 3. The 3′-untranslated region of this RNA was found to contain at least one CP binding site in the region that is homologous to RNAs 1 and 2, and one binding site which involves sequences that are unique to RNA 3.

MATERIALS AND METHODS

Construction of mutant cDNAs

The cloning of plasmids pT72-42 , pTE6 and pTE7 with inserts that can be transcribed with T7 RNA polymerase has been described previously (27, 28). pT72-42 contains a full-length copy of RNA 4, pTE6 contains the 5′-terminal 681 nucleotides (nt) of RNA 4 and pTE7 contains the 3′-terminal 208 nt downstream of a pBR322 sequence of 783 nt (Fig. 2, upper panel). When pT72-42 and pTE7 are linearized by *Sma*I digestion, T7 transcripts are obtained with two nonviral C-residues at the 3′-terminus. pTE7 was used to make the derivative pBRWT with a *Pst*I site at the 3′-end of the cDNA. To this goal an internal *Pst*I site at the junction of the pBR and viral sequence in pTE7 was removed by digestion of this plasmid with *Pst*I and religation after the ends were made blunt with T4 DNA polymerase. The *Bst*X I/*Sma*I fragment of the modified pTE7, which contains the 3′-terminal 164 nt of RNA 4, was replaced by the corresponding fragment of pAL3. Plasmid pAL3 contains a full-length DNA copy of AlMV RNA 3 with a *Pst*I site at the 3′-end, just upstream of the *Sma*I site (29). When pBRWT is linearized with *Pst*I, T7 transcripts will be obtained with correct 3′-termini.

Plasmids pTE2 to pTE5 contain various 3′-terminal deletions of the RNA 4 sequence, followed by a *Pst*I restriction site (28). These plasmids were cleaved with *Bst*X I at a site 164 upstream of the 3′-end of RNA 4 and *Sca*I at a site 983 bp downstream of the viral insert, and the resulting fragments were used to replace the corresponding *Bst*X I/*Sca*I fragment of pBRWT. This yielded constructs pBRTE2 to pBRTE5 with 3′-terminal deletions ranging from 11 to 133 nt (Fig. 3, upper panel). Nucleotide numbers from the 3′-end of RNAs 3 and 4 are given a minus-sign as prefix.

In addition to a *Bst*X I site at position −164, the 3′-noncoding region of RNAs 3 and 4 contains a *Dra*III site at position −127. Deletion of the *Bst*X I/*Dra*III fragments from constructs pBRWT, pBRTE2, pBRTE3 and pBRTE4 resulted in constructs pBRWTΔBD, pBRTE2ΔBD, pBRTE3ΔBD and pBRTE4ΔBD, respectively (Fig. 4, upper panel). As the restriction sites were made blunt with T4 DNA polymerase before religation of the gaps, the sequence from −169 to −127 is deleted in these mutants.

The stem/loop structures in the 3′-noncoding region of RNA 3 are separated by five single-stranded AUGC-boxes which are numbered 1 to 5 from the 3′-terminus. By site-directed mutagenesis the sequence of boxes 1 to 5 has been changed into AGGC, yielding plasmids p3AC1 to p3AC5, respectively (van der Vossen *et al.*, submitted for publication). To introduce these mutations in pBRWT the 178 bp *Bst*X I/*Sma*I fragment of pBRWT, which contains the 3′-terminal 164 nt of RNA 3, was replaced by the corresponding fragments of p3AC1 to p3AC5, yielding pBRAC1 to pBRAC5, respectively (Fig. 5, upper panel). Like pBRWT, these constructs have a *Pst*I site at the 3′-end of the cDNA. Deletion of the *Bst*X I/*Dra*III restriction fragment (nt −169/−127) from mutants pBRAC1, pBRAC2 and pBRAC3 resulted in mutants pBRAC1ΔBD, pBRAC2ΔBD and pBRAC3ΔBD, respectively (Fig. 6, upper panel).

In vitro transcription

Transcripts are indicated by the name of the plasmid without the prefix 'p'; e.g. plasmid pTE7 yields transcript TE7. Prior to transcription, pT72-42 and pTE7 were linearized with *Sma*I. All other mutants were linearized with *Pst*I and made blunt ended with T4 DNA polymerase to yield the corresponding transcripts. Linearization of pTE7 with *Pst*I yielded transcript TE11 (Fig. 2). In addition, mutants pBRAC4, pBRAC5 and pBRWT were linearized with *Dra*III and made blunt ended with T4 DNA polymerase to yield transcripts BRAC4xD, BRAC5xD and BRWTxD, respectively (Fig. 6, upper panel). BRWTxB (Fig. 3) was transcribed from pBRWT after linearization with *Bst*X I and making the end blunt with T4 DNA polymerase.

Transcription mixtures were incubated for 2 h at 37°C. For the synthesis of unlabeled transcripts the mixture (100 μl) consisted of 10 μl of 10× transcription buffer (400 mM Tris.HCl, pH 7.6, 60 mM $MgCl_2$, 20 mM spermidine, 100 mM NaCl), 10 μl 0.1 M dithiothreitol, 5 μl of a solution of 10 mM of each

of the four nucleoside triphosphates, 1 μl RNasin (Promega, 40 u/μl), 5 μl T7 RNA polymerase (BRL, 50 u/μl) and 2 μg linearized template DNA. For the synthesis of ^{32}P-labeled transcripts, the mixture (50 μl) contained 5 μl of 10× transcription buffer, 5 μl of 0.1 M dithiothreitol, 2.5 μl of a solution containing 20 mM ATP, CTP and GTP, 2.5 μl 2 mM UTP, 5 μl [α-^{32}P]UTP (Pharmacia, specific activity 15 TBq/mM), 1 μl RNasin, 5 μl T7 RNA polymerase and 2 μg linearized template DNA.

Isolation of virus particles and RNAs

AlMV particles and native RNA 4 were purified as described previously (3). RNA of turnip yellow mosaic virus was isolated as described by Stols and Veldstra (30).

Isolation of AlMV CP

CP was purified from AlMV particles essentially as described previously (31). A virus solution of 25 mg/ml in PE buffer (10 mM phosphate, 1 mM EDTA, pH 7.0) was slowly added under vigorous shaking at 4°C to an equal volume of a solution containing 1 M $MgCl_2$ and 5% 2-mercapto-ethanol. After incubation for 20 min on ice, the RNA precipitate was removed by centrifugation at 9000 *g* for 10 min at 4°C. The supernatant with the protein was dialyzed at 4°C for 48 h against distilled water, 48 h against a buffer containing 0.2 M acetic acid, 5 mM EDTA, adjusted to pH 5.5 with NaOH, and again for 48 h against distilled water. After dialysis the preparation was centrifuged at 9000 *g* for 10 min at 4°C to remove aggregated material and the protein concentration was measured by UV absorption. The protein was stored in small aliquots at −20°C until use.

Gel mobility shift assays

Unless mentioned otherwise, incubation mixtures (15 μl) were made by adding 50 ng CP to PE-buffer containing ^{32}P-labeled transcript (10,000 cpm) and 50 ng of the corresponding unlabeled transcript. Before mixing, the RNAs were heated for 10 min at 65°C and chilled on ice. The resulting mixture with an RNA/CP ratio of 1:10 was incubated for 30 min at room temperature. When indicated, competitor RNAs were added at a 10-fold mass excess (450 ng per incubation mixture) prior to addition of CP. Native RNA 4 and TYMV RNA were used as homologous and nonhomologous competitors, respectively. After incubation, 1.5 μl of sample buffer (0.25% bromophenolblue, 0.25% xylene cyanol and 30% glycerol in water) was added and the samples were loaded onto a nondenaturing 3.5% polyacrylamide gel with 1× Tris-borate-EDTA as the running buffer. After electrophoresis for 3.5 h at 375 V, the gel was dried and exposed to X-ray film.

RESULTS

Specificity of the binding of CP to RNA 4

Transcript T72-42 corresponds to the sequence of RNA 4 and was used to investigate the possibility of studying the binding of CP to AlMV RNAs by band-shift assays. Increasing amounts of CP were added to a mixture of labeled transcript and 50 ng of unlabeled T72-42, resulting in mixtures that contained 0.25 to 130 CP molecules per RNA molecule. Fig. 1A shows that a shift of the labeled transcript becomes detectable at a CP/RNA ratio of 1; at a CP/RNA ratio of 130 the transcript no longer entered the gel. In a virus preparation two RNA 4 molecules are encapsidated in a protein shell of 132 CP subunits (32). Thus,

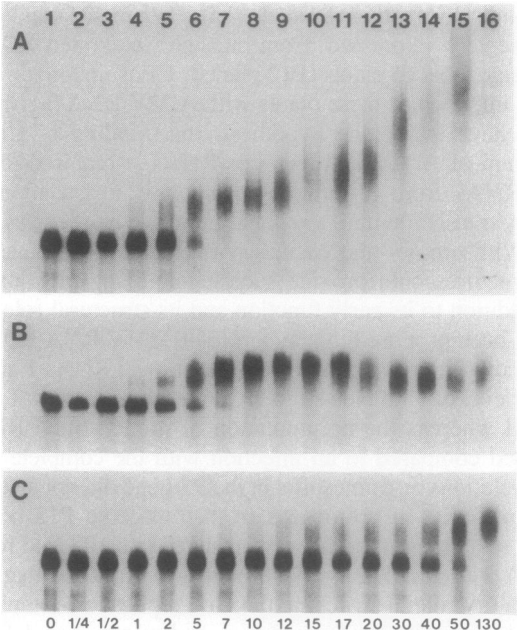

Figure 1. Band-shift analysis of the binding of CP to transcript T72-42, corresponding to RNA 4. The number of CP molecules that were added per transcript molecule is given below panel C. Binding of CP to the transcript was studied in the absence of unlabeled competitor RNA (panel **A**), or in the presence of a 10-fold mass excess of TYMV RNA (panel **B**) or AlMV RNA 4 (panel **C**).

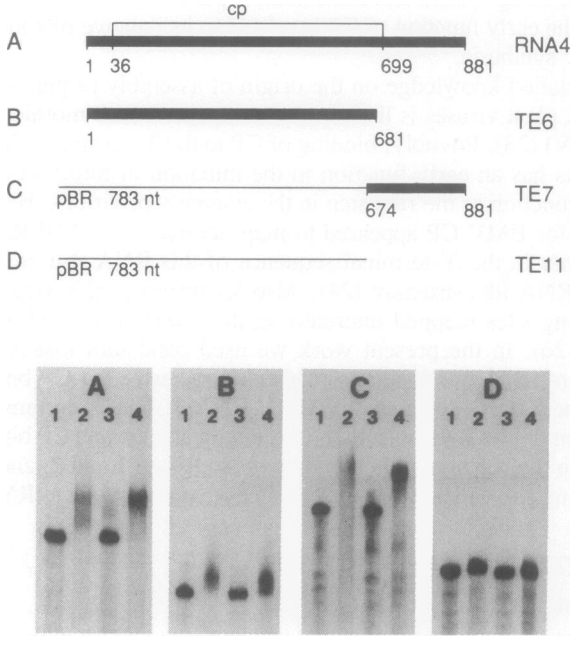

Figure 2. Binding of CP to the 3′-terminus and internal sites of RNA 4. **Upper panel:** Schematic representation of transcripts TE6, TE7 and TE11. Viral sequences are indicated by bold lines; thin lines represent pBR322-derived sequences. Numbering of nucleotides is according to the sequence of RNA 4. **Lower panel:** Band-shift analysis of the binding of CP to transcripts T72-42 (panel **A**), TE6 (panel **B**), TE7 (panel **C**) and TE11 (panel **D**). In lanes **1** no CP was added to the transcripts. Binding of CP to the transcripts was studied in the absence of unlabeled competitor RNA (lanes **2**), and in the presence of a 10-fold mass excess of RNA 4 (lanes **3**) or TYMV RNA (lanes **4**).

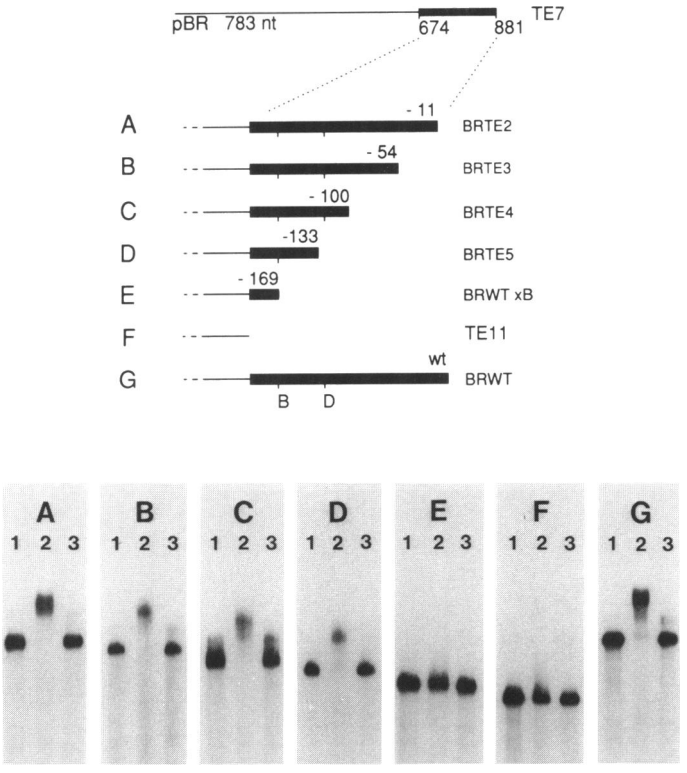

Figure 3. 3′-Deletion analysis of CP binding sites in the 3′-terminal 208 nt of RNA 3. **Upper panel:** Schematic representation of transcripts. Viral sequences are indicated by bold lines; thin lines represent pBR322-derived sequences. Transcript BRWT contains the 3′-terminal 208 nt of RNA 3. 3′-Deletions of 11, 54, 100, 133 and 169 nt are indicated. **Lower panel:** Band-shift analysis of the binding of CP to transcripts BRTE2 (panel **A**), BRTE3 (panel **B**), BRTE4 (panel **C**), BRTE5 (panel **D**), BRWTxB (panel **E**), TE11 (panel **F**) and BRWT (panel **G**). In lanes **1** no CP was added to the transcripts. Binding of CP to the transcripts was studied in the presence of a 10-fold mass excess of TYMV RNA (lanes **2**) or RNA 4 (lanes **3**).

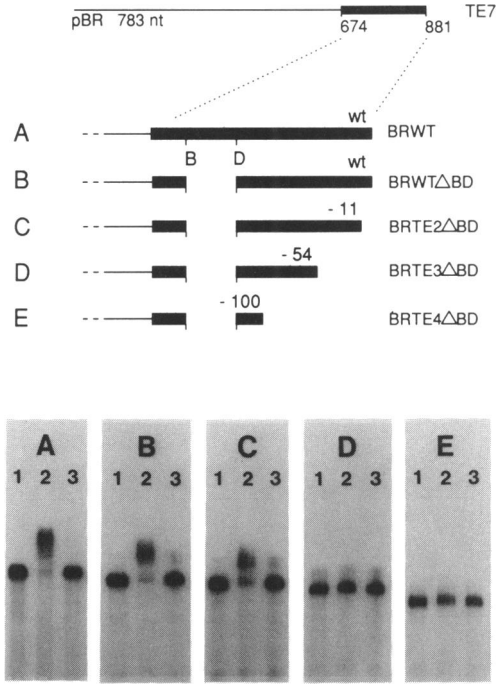

Figure 4. 3′-Deletion analysis of CP binding sites in the 3′-terminal 127 nt of RNA 3. **Upper panel:** Schematic representation of transcripts. Viral sequences are indicated by bold lines; thin lines represent pBR322-derived sequences. Transcript BRWT contains the 3′-terminal 208 nt of RNA 3. In the other transcripts the BD-fragment (nt −169/−127) is deleted in addition to 3′-terminal deletions of 11, 54 and 100 nt. **Lower panel:** Band-shift analysis of the binding of CP to transcripts BRWT (panel **A**), BRWTΔBD (panel **B**), BRTE2ΔBD (panel **C**), BRTE3ΔBD (panel **D**) and BRTE4ΔBD (panel **E**). In lanes **1** no CP was added to the transcripts. Binding of CP to the transcripts was studied in the presence of a 10-fold mass excess of TYMV RNA (lanes **2**) or RNA 4 (lanes 3).

a CP/RNA ratio of 66 reflects the ratio found in virions. Addition of a 10-fold mass excess of nonhomologous competitor RNA (TYMV RNA) reduced the shift of labeled complexes formed at CP/RNA ratios higher than 10 indicating that the competitor was able to withdraw CP molecules from these complexes that were nonspecifically bound to the transcript. However, in the presence of this nonhomologous competitor, the formation of a specific complex between transcript and CP could be clearly observed (Fig. 1B). When a 10-fold excess of the homologous competitor (RNA 4) was added most of the transcript did not shift at CP/RNA ratios up to 30 (Fig. 1C). In subsequent experiments a CP/transcript ratio of 10 was used to study CP binding and addition of a 10-fold mass excess of TYMV RNA and RNA 4 was used to demonstrate the specificity of the binding (Fig. 1, lanes 8).

Binding of CP to the 3′-terminus and internal sites of RNA 4

Transcript TE6 contains nucleotides 1/681 of RNA 4 and transcript TE7 contains nucleotides 674/881 (i.e. the 3′-terminal 208 nt of RNA 4) fused to a pBR322-derived sequence of 783 nt (Fig. 2, upper panel). These transcripts were used to roughly map the CP binding sites in RNA 4. In panel B of the lower panel of Fig. 2, lane 1 shows the migration of the free transcript TE6. Addition of CP results in a complete shift of the transcript (lane 2) which is completely abolished by addition of the homologous competitor (lane 3). After addition of the heterologous competitor only part of the transcript was retarded. When a similar experiment was done with transcript TE7, all labeled material remained bound in a specific complex after addition of the heterologous competitor (panel C in the lower panel of Fig. 2). As controls in the experiment of Fig. 2, transcript T72-42 (full-length RNA 4, panel A) and transcript TE11 (pBR322 sequence only, panel D) were included. The observation that TE11 does not form a detectable complex with CP demonstrates that the shift of TE7 is due to binding of CP to the RNA 4 specific sequence in this transcript. The relatively weak complex formation of TE6 may be due to binding of CP to the region of nucleotides 425/474 which has been identified as an internal binding site in RNA 4 (33). In our studies we decided to focus on a further characterization of the 3′-terminal binding site present in transcript TE7. As all cDNA clones used in this study were derived from a full-length DNA copy of RNA 3, we will discuss this 3′-binding site by referring to its location in RNA 3 instead of RNA 4.

The 3′-terminal sequence of RNA 3 contains at least two binding sites for CP

To localize CP binding sites in the 3′-terminal 208 nt of RNA 3, mutants BRTE2 to BRTE5 with 3′-deletions of 11, 54, 100 and 133 nt were constructed (Fig. 3, upper panel). The lower panel of Fig. 3 shows the migration of the free transcripts (lanes

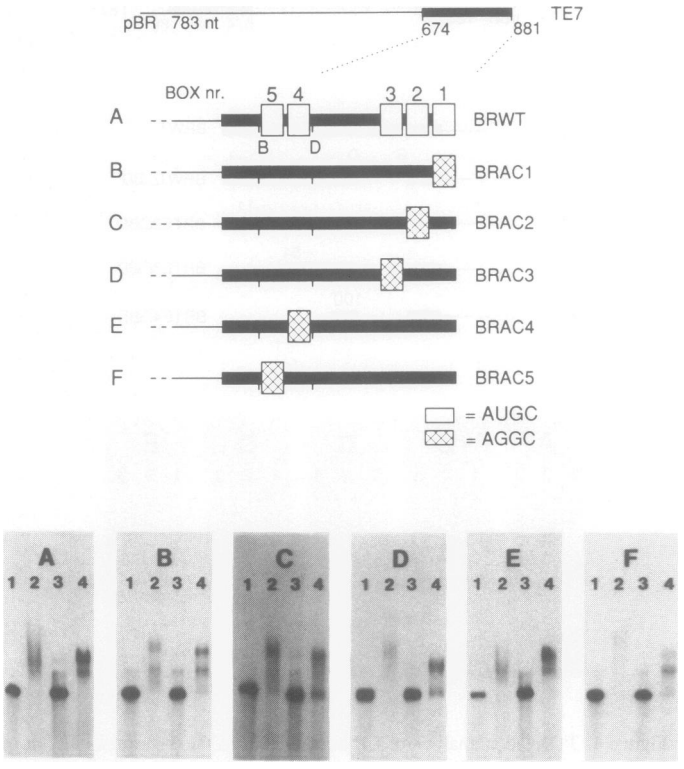

Figure 5. CP binding to 3′-terminal transcripts of 208 nt with mutations in the AUGC-motifs. **Upper panel:** Schematic representation of transcripts. Viral sequences are indicated by bold lines; thin lines represent pBR322-derived sequences. In transcript BRWT the open boxes represent AUGC-motifs 1 to 5 in the 3′−208 nt of RNA 3. Mutations of these motifs to the sequence AGGC are indicated by hatched boxes. **Lower panel:** Band-shift analysis of the binding of CP to transcripts BRWT (panel **A**), BRAC1 (panel **B**), BRAC2 (panel **C**), BRAC3 (panel **D**), BRAC4 (panel **E**) and BRAC5 (panel **F**). In lanes **1** no CP was added to the transcript. Binding of CP to the transcripts was studied in the absence of unlabeled competitor RNA (lanes **2**) and in the presence of a 10-fold mass excess of RNA 4 (lanes **3**) or TYMV RNA (lanes **4**).

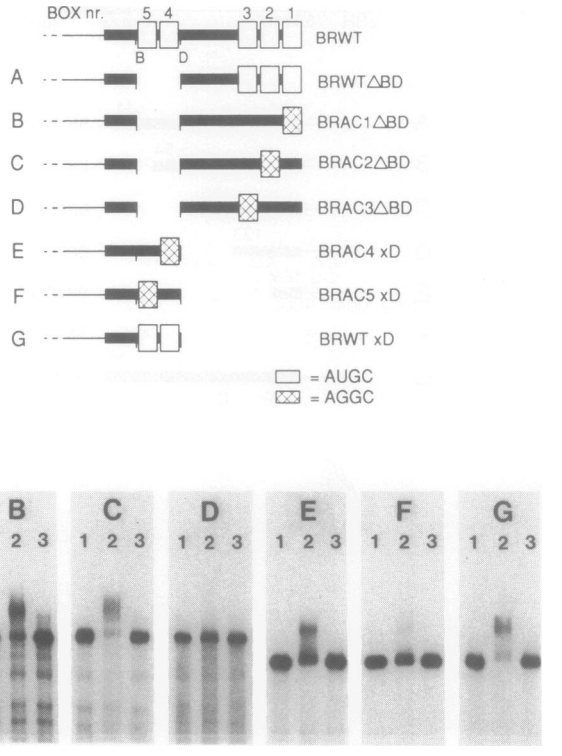

Figure 6. CP binding to truncated 3′-terminal transcripts with mutations in the AUGC-motifs. **Upper panel:** Schematic representation of transcripts. Viral sequences are indicated by bold lines; thin lines represent pBR322-derived sequences. Transcript BRWT contains the 3′-terminal 208 nt of RNA 3. Open boxes represent AUGC-motifs 1 to 5; hatched boxes represent mutant sequences AGGC. Deletions of the BD-fragment (nt −169/−127, boxes 4 and 5) or the 3′-terminal 127 nt of RNA 3 (boxes 1, 2 and 3) are indicated. **Lower panel:** Band-shift analysis of the binding of CP to transcripts BRWTΔBD (panel **A**), BRAC1ΔBD (panel **B**), BRAC2ΔBD (panel **C**), BRAC3ΔBD (panel **D**), BRAC4xD (panel **E**), BRAC5xD (panel **F**) and BRWTxD (panel **G**). In lanes **1** no CP was added to the transcripts. Binding of CP to the transcripts was studied in the presence of a 10-fold mass excess of TYMV RNA (lanes **2**) or RNA 4 (lanes **3**).

1), and the complexes formed with CP in the presence of a 10-fold mass excess of TYMV RNA (lanes 2) or RNA 4 (lanes 3). The 3′-deletions of 11 to 133 nt did not affect specific complex formation of the transcripts with CP (panels A to D). As controls, transcripts TE11 (panel F) and BRWT (panel G) were included in the experiment of Fig. 3. Transcript BRWT is similar to TE7 but is derived from a template that is linearized with *Pst*I instead of *Sma*I. When plasmid pBRWT was linearized with *Bst*X I, transcript BRWTxB was obtained. This transcript lacks the 3′-terminal 169 nt of RNA 3 and does not give a detectable complex formation with CP (panel E in the lower panel of Fig. 3). From these results we conclude that at least one specific binding site for CP is present in the region between nt 133 and 208 from the 3′-end of RNA 3 (region −208/−133) and that the region −169/−133 is an essential part of this binding site.

Earlier studies have demonstrated a specific binding site for CP in the sequence of the 3′-terminal 80 nt of RNA 3/4 (33). To confirm the existence of two independent CP binding sites in the 3′−208 nt of RNA 3, the *Bst*X I/*Dra*III restriction fragment (BD-fragment) was deleted from various cDNA constructs, resulting in a deletion of nt −169/−127 in the corresponding transcripts. The structure of these transcripts is shown in the upper panel of Fig. 4 whereas the lower panel of Fig. 4 shows the results of band-shift assays with these transcripts. Transcript BRWT was included as a control (panel A). Deletion of the BD-fragment from BRWT had little effect on the specific binding of CP to the transcript (panel B). An additional deletion of the 3′−11 nt of RNA 3 had a minor effect on the efficiency of CP binding (panel C) whereas additional 3′-deletions of 54 and 100 nt abolished the specific binding of CP (panels D and E, respectively). The results shown in Fig. 4 confirm the presence of at least one specific CP binding site in the 3′-terminal 133 nt of RNA 3 and show that the −11/−54 region is an essential part of the binding site. The results obtained with transcripts BRTE5 (Fig. 3, panel D) and BRTE2ΔBD (Fig. 4, panel C) demonstrate that the regions −208/−133 and −127/−11 can bind CP independently of each other.

Role of AUGC-boxes in CP binding

The 3′-noncoding sequence of 182 nt of RNA 3 contains a number of stem/loop structures that are interspersed with five single-stranded AUGC-sequences which have been numbered 1 to 5 starting from the 3′-terminus (Fig. 7). To study the role of these AUGC-boxes in CP binding, mutant plasmids pBRAC1 to pBRAC5 were constructed (Fig. 5, upper panel). Transcripts BRAC1 to BRAC5 consist of a pBR322 sequence fused to the

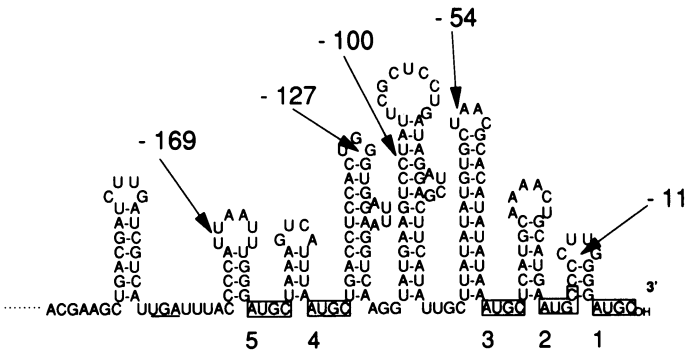

Figure 7. Proposed secondary structure of the 3′-terminus of RNA 3. AUGC-motifs 1 to 5 are boxed. The UGA-termination codon of the CP gene is underlined. The positions of nt 11, 54, 100, 127 and 169 from the 3′-end are indicated by arrows.

3′-terminal 208 nt of RNA 3 with AUGC-box 1 to 5 changed into AGGC, respectively. The lower panel of Fig. 5 shows the migration of the free transcripts (lanes 1), and the complexes formed in the absence of competitor (lanes 2) or in the presence of a 10-fold excess of RNA 4 (lanes 3) or TYMV RNA (lanes 4). Compared to the complex formation obtained with the wild type transcript BRWT in the presence of heterologous competitor (panel A, lane 4), a minor effect on complex formation is observed for transcripts with a mutation in box 1 (panel B, lane 4), box 2 (panel C, lane 4), box 3 (panel D, lane 4) or box 5 (panel F, lane 4). In contrast to the wild type transcript, part of these mutant transcripts are not complexed to CP under the conditions used. No effect of the mutation in box 4 (panel E, lane 4) was detectable.

In view of the finding that the 3′-208 nt of RNA 3 contain a minimum of two independent CP binding sites, it could well be that a possible inactivation of one site by mutation of an AUGC-box is masked by the presence of the second wild type binding site. Therefore, the effect of mutations in the AUGC-boxes was studied with transcripts in which either the 3′-distal binding site was inactivated by deletion of the BD-fragment (region −169/−127) or the 3′-proximal binding site was inactivated by deleting the 3′-terminal 127 nt of RNA 3 (Fig. 6, upper panel). Deletion of the 3′-terminal 127 nt was done by linearizing plasmids pBRAC4 and pBRAC5 with *Dra*III to obtain transcripts BRAC4xD and BRAC5xD, respectively. The results of band-shift experiments with these transcripts are shown in the lower panel of Fig. 6. The migration of the free transcript is shown in lanes 1, and complexes with CP formed in the presence of a 10-fold excess of TYMV RNA and RNA 4 are shown in lanes 2 and 3, respectively. Wild type transcripts with a deletion of the upstream CP binding site (BRWTΔBD) or the downstream CP binding site (BRWTxD) are shown in panels A and G, respectively. When the upstream CP binding site is deleted, mutation of the U-residue in AUGC-boxes 1 and 2 slightly reduces the fraction of the transcript that is bound to CP (lane 2 of panels B and C, respectively) whereas the mutation in box 3 abolishes CP binding (panel D, lane 2). When the downstream CP binding site is deleted from the transcripts, the mutation in box 4 substantially reduces the fraction of the transcript that is bound to CP (panel E, lane 2) and the mutation in box 5 almost completely inhibits binding of CP to the transcript (panel F, lane 2). The results of Fig. 6 demonstrate that AUGC-boxes 4 and 5 are important elements of the upstream binding site(s) whereas particularly box 3 is essential for the activity of the downstream binding site(s). Apparently, in the experiment shown in Fig. 5, inactivation of the upstream binding site by mutations in boxes 4 or 5 was masked by the presence of the wild type downstream binding site and inactivation of the downstream binding site by the mutation in box 3 was largely obscured by the activity of the upstream binding site.

DISCUSSION

Previous studies on the interaction of CP with the 3′-termini of native AlMV RNAs revealed that RNA 4 could form specific complexes with one or three dimers of CP (33). In the complex with one dimer a 3′-terminal sequence of up to 83 nucleotides was protected from nuclease degradation, whereas in the complex with three dimers additional sequences of nt −94/−122 and −128/−214 were protected. Subsequently, it was shown that in RNAs 1 and 2 a 3′-terminal sequence of 81 nt and subfragments thereof were protected against nuclease degradation while in RNA 3 sequences of nt −1/−79 and −152/−180 were protected (9, 10). To permit a more detailed analysis of these binding sites, we developed a technique to investigate the interaction between CP and RNA transcripts of AlMV cDNAs by band-shift assays. Our results showed that the 3′-terminal sequence of RNA 3 contains two regions that are able specifically to bind CP independently of each other. Although we cannot rule out the possibility that each of these two regions contain multiple binding sites for CP, we refer to these regions as 'site I' and 'site II'. Site I is represented by the sequence of nt −11/−127 from the 3′-end of RNA 3 and is completely localized in the 3′-terminal sequence of 145 nt that is homologous to RNAs 1 and 2. Site II is found between nt −133/−208 in a region that is largely unique to RNA 3. Inactivation of one binding site by deletions or point mutations could be observed only when the second binding site was deleted from the viral transcript. Under the standard conditions used, all labeled transcripts containing either site I or site II were fully complexed with CP in the incubation mixture whereas some of the transcripts with modified sites showed only partial binding. However, we have not yet performed the kinetic studies necessary to measure differences in binding activity quantitatively. Recently, an apparent K_d of 500 nM was determined for the binding of AlMV CP to a transcript containing the 3′-terminal 162 nt of RNA 3 (14).

Fig. 7 shows that the sequence of the 3′-terminal 127 nt of RNA 3 is predicted to contain four stem/loop structures, three of which are flanked by the downstream sequence AUGC (boxes 1, 2 and 3). Deletion of the 3′-terminal 11 nt had little effect on the activity of site I but deletion of nt −11/−54 abolished the binding activity (Fig. 4). Apparently, the latter sequence is essential for the activity of site I. A study of the binding of CP to 3′-terminal fragments of RNA 1 by a nitrocellulose filter retention assay revealed that fragments containing sequences −1/−36 or −28/−81 were both able to bind CP (10). This indicated the presence of at least two CP binding sites in the 3′-terminal 81 nt of RNA 1. Recently, evidence has been provided that a 3′-terminal 39 nt fragment of RNA 3 is also able to bind CP (14). Like RNA 1, the 3′-terminal 127 nt of RNA 3 may contain two CP binding sites and inactivation of one of these sites by deletion of the 3′-11 nt could be masked by the presence of a second binding site in this region. The observation

that a mutation in AUGC-box 1 reduced the efficiency of the binding of the 3'-terminal 127 nt sequence of RNA 3 supports this notion. However, a single mutation in AUGC-box 3 abolished the binding activity of the 3'−127 nt sequence (Fig. 6). This would point to the presence of a single binding site in this region. Although the binding site(s) have to be characterized in more detail by site-directed mutagenesis, our results thus far demonstrate the importance of the AUGC-motifs in CP binding to site I.

Similar to site I, site II (nt −133/−208) contains several stem/loop structures two of which are flanked by downstream AUGC-motifs (boxes 4 and 5, Fig. 7). Deletion analysis revealed that the sequence of nt −133/−169, which contains these two boxes, is essential for the activity of site II (Fig. 3). The effect of a deletion of nt −169/−208 on the activity of site II has not yet been analyzed. As such a deletion would effect the stem/loop structure upstream of box 5 it is quite possible that the binding site(s) in site II extend upstream of nt −169. Although the region of site II overlaps 12 nt with the 3'-terminal sequence of RNA 3 that is homolgous to RNAs 1 and 2, the binding activity of site II appears to be unique to RNA 3. AUGC-boxes 4 and 5 are not present in RNAs 1 and 2, and mutations in these boxes reduced and abolished the activity of site II, respectively (Fig. 6). Again, these data underline the importance of the AUGC-motifs in CP binding. The observation that binding of CP to RNAs 1 and 2 did not protect any detectable region of the 3'-noncoding sequences of these RNAs except for the 3'-81 nt, further indicates that site II is unique to RNA 3 (10).

We have put forward the hypothesis that binding of CP to the 3'-termini of RNAs 1, 2 and 3 in an inoculum mixture is required to protect these termini from nucleolytic degradation during the early phase of the replication cycle (15). A function of this binding in the regulation of viral RNA synthesis has been proposed by others (4, 22). In addition to CP binding sites, the 3'-termini of the AlMV RNAs contain *cis*-acting sequences involved in recognition of the viral replicase. In an *in vitro* assay, the sequence of nt −133/−163 of RNA 3 was found to be sufficient to act as a promoter for minus-strand RNA synthesis by the purified AlMV replicase whereas the sequence of the 3'-terminal 120 nt was not (28). It could be that early in the infection cycle binding of CP to site I promotes binding of the replicase to the minus-strand promoter while binding of additional CP molecules to site II would shut-off minus-strand RNA synthesis at a later stage. However, the absence of site II in RNAs 1 and 2 makes such a regulatory function unlikely. All 3'-terminal *cis*-acting sequences required for replication of RNAs 1 and 2 in protoplasts were found to be located in the 3'−127 nt of these RNAs (van Rossum *et al.*, unpublished) whereas replication of RNA 3 required a 3'-terminal sequence of at least 169 nt (van der Vossen *et al.*, submitted for publication). This indicates that due to dissimilarities in the nucleotide sequence of the homologous region, the 3'-terminal sequences of the three genomic RNAs are functionally different. We favour the hypothesis that the affinity for CP and/or replicase of sites I in RNAs 1 and 2 is higher than that of site I in RNA 3. In this view, the reduced affinity of site I in RNA 3 could be compensated by its duplication into site II which could result e.g. in an enhanced cooperativity of CP binding. Evidence for such a cooperativity is provided by the detection of complexes of RNA 4 with one or three dimers of CP bound to the 3'-end (33). Recently, AUGC-boxes 1 to 5 were mutated into corresponding AGGC-boxes in an infectious cDNA clone of RNA 3 (van der Vossen *et al.*, submitted for publication). Mutation of boxes 1, 4, or 5 had little effect on the infectivity of the RNA 3 transcript towards P12 plants or protoplasts but a mutation of box 2 strongly reduced infectivity whereas mutation of box 3 abolished infectivity. This would suggest that binding of CP to site I is critical for the infectivity of RNA 3 whereas binding to site II is dispensable. However, it is possible that *in vivo* a single mutation in site II has little effect on the putative cooperative binding of CP to sites I and II. The effect of AUGC-mutations in site II on *in vitro* binding of CP could be due to the fact that in our studies sites I and II were disconnected. To test our hypothesis, we will quantitatively compare the affinity of the 3'-terminal sequences of RNAs 1, 2 and 3 for CP and replicase in *in vitro* assays.

The 3'-termini of bromo- and cucumovirus RNAs contain a tRNA-like structure. Available evidence indicates that this structure is not involved in CP binding. The internal binding sites for CP that were mapped recently in BMV RNA 1 could play a role in encapsidation of the viral RNA (24). Internal CP binding sites have also been identified in RNAs 1, 2 and 3 of AlMV (9). Infectious cDNA clones of these RNAs are now available to study the possible role of these binding sites in translation, replication or encapsidation of AlMV RNAs (15). In TCV RNA, CP was found to bind to an internal site downstream of the amber readthrough codon in the putative replicase gene, suggesting that CP of this virus could regulate expression of the replicase gene at the translational level (25, 26). Another example of a regulatory role of a viral CP is the activity of CP of small RNA phages as a repressor of translation of the viral replicase gene (34). By a comparison of the effect of mutations on protein/RNA interactions *in vitro* and virus replication *in vivo* we hope to dissect the functions of CP in the AlMV replication cycle.

ACKNOWLEDGEMENTS

Plasmids p3AC1 to p3AC5 were kindly provided by Mr. E.A.G. van der Vossen. TYMV RNA was a generous gift from Mrs. C.J. Houwing. This work was supported in part by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

REFERENCES

1. Jaspars, E.M.J. (1985) In J.W. Davies (ed.), Molecular Plant Virology. CRC Press, Boca Raton, Vol I, pp 155−221.
2. Kaper, J.M. (1973) *Virology*, **55**, 299−304.
3. Van Vloten-Doting, L. and Jaspars, E.M.J. (1972) *Virology*, **48**, 699−708.
4. Houwing, C.J. and Jaspars, E.M.J. (1978) *Biochemistry*, **17**, 2927−2933.
5. Koper-Zwarthoff, E.C., Brederode, F.Th., Walstra, P. and Bol, J.F. (1979) *Nucl. Acids Res.*, **7**, 1887−1900.
6. Quigley, G.J., Gehrke, L., Roth, D.A. and Auron, P.E. (1984) *Nucl. Acids Res.*, **12**, 347−366.
7. Cornelissen, B.J.C., Jansen, H., Zuidema, D. and Bol, J.F. (1984) *Nucl. Acids Res.*, **12**, 2427−2437.
8. Bol, J.F., Cornelissen, B.J.C., Huisman, M.J. and Van Vloten-Doting, L. (1985) In Van Vloten-Doting, L., Groot, G.S.P. and Hall, T.C. (eds), Molecular Form and Function of the Plant Genome. Plenum Press, pp 359−369.
9. Zuidema, D. and Jaspars, E.M.J. (1984) *Virology*, **135**, 43−52
10. Zuidema, D., Cool, R.H. and Jaspars, E.M.J. (1984) *Virology*, **136**, 282−292.
11. Zuidema, D. and Jaspars, E.M.J. (1985) *Virology*, **140**, 342−350
12. Smit, C.H., Roosien, J., Van Vloten-Doting, L. and Jaspars, E.M.J. (1981) *Virology*, **112**, 169−173.
13. Zuidema, D., Bierhuizen, M.F.A. and Jaspars, E.M.J. (1983) *Virology*, **129**, 255−260.

14. Baer, M.L., Houser, F., Loesch-Fries, L.S. and Gehrke, L. (1994) *EMBO J.*, **13**, 727–735.
15. Neeleman, L., Van der Vossen, E.A.G. and Bol, J.F. (1993) *Virology*, **196**, 883–887.
16. Taschner, P.E.M., Van der Kuyl, A.C., Neeleman, L. and Bol, J.F. (1991) *Virology*, **181**, 445–450.
17. Dreher, T.W., Rao, A.L.N. and Hall, T.C. (1989) *J. Mol. Biol.*, **206**, 425–438.
18. Nassuth, A. and Bol, J.F. (1983) *Virology*, **124**, 75–85.
19. Van der Kuyl, A.C., Neeleman, L. and Bol, J.F. (1991) *Virology*, **183**, 687–694.
20. Van der Kuyl, A.C., Neeleman, L. and Bol, J.F. (1991) *Virology*, **185**, 496–499.
21. Quadt, R., Rosdorff, H.J.M., Hunt, T.W. and Jaspars, E.M.J. (1991) *Virology*, **182**, 309–315.
22. Houwing, C.J. and Jaspars, E.M.J. (1993) *Biochimie*, **75**, 617–622.
23. Turner, D.R., Joyce, L.E. and Butler, P.J.G. (1988) *J. Mol. Biol.*, **203**, 531–547.
24. Duggal, R. and Hall, T. (1993) *J. Virol.*, **67**, 6406–6412.
25. Wei, N. and Morris, T.J. (1991) *J. Mol. Biol.*, **222**, 437–443.
26. Wei, N., Hacker, D.L. and Morris, T.J. (1992) *Virology*, **190**, 346–355
27. Langereis, K., Neeleman, L. and Bol, J.F. (1986) *Plant Mol. Biol.*, **6**, 281–288.
28. Van der Kuyl, A.C., Langereis, K., Houwing, C.J., Jaspars, E.M.J and Bol, J.F. (1990) *Virology*, **176**, 346–354.
29. Neeleman, L., Van der Kuyl, A.C. and Bol, J.F. (1991) *Virology*, **181**, 687–693.
30. Stols, A.L.H. and Veldstra, H. (1965) *Virology*, **25**, 508–515.
31. Kruseman, J., Kraal, B., Jaspars, E.M.J., Bol, J.F., Brederode, F.Th. and Veldstra, H. (1971) *Biochemistry*, **10**, 447–455.
32. Bol, J.F. and Jaspars, E.M.J. (1994) In Webster, R.G. and Granoff, A. (eds.), Encyclopedia of Virology. Academic Press, London, Vol I, pp 1442–1447.
33. Houwing, C.J. and Jaspars, E.M.J. (1982) *Biochemistry*, **21**, 3408–3414.
34. Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C. (1991) In Cohn, W.E. and Moldave, K. (eds.), Progress in Nucleic Acid Research and Molecular Biology. Academic Press, London, Vol. 40, pp 185–220.