Volume 319, number 1,2, 95–99 FEBS 12155 March 1993

TGMV replication protein AL1 preferentially binds to single-stranded DNA from the common region

P. Thömmes, T.A.M. Osman, R.J. Hayes and K.W. Buck

Department of Biology, Imperial College of Science, Technology and Medicine, London SW7 2BB, UK

Received 22 December 1992; revised version received 18 January 1993

The AL1 protein of tomato golden mosaic virus (TGMV) is encoded by the viral DNA and has been shown to be essential for viral DNA replication. We have over-expressed the AL1 open reading frame in *E. coli* and purified the protein from bacterial extracts to near homogeneity. Using various different techniques we have studied the interaction of the AL1 protein with DNA. The AL1 protein is able to bind to DNA containing the common region of the viral genome, which can be demonstrated by photochemical cross-linking. Binding is 4-fold stronger to single-stranded than to double-stranded DNA. Antibodies against the AL1 protein can be used to precipitate the protein–DNA complex. The binding to single- and double-stranded DNA is specifically to the common region since a DNA fragment unrelated to TGMV is not shifted in a gel retardation assay.

Geminivirus; Replication; DNA binding protein; Common region; Origin of rolling circle replication

1. INTRODUCTION

Geminiviruses are a group of plant viruses with small circular single-stranded DNA genomes [1] which replicate in the nucleus of the plant cell by a rolling circle mechanism [2,3]. Tomato golden mosaic virus (TGMV) belongs to a subgroup of the geminiviruses which have genomes with two components, DNA A and DNA B, which encode several proteins involved in virus DNA replication, cell-to-cell movement, transcriptional regulation or particle structure [4]. Genes on both DNAs, which are encoded by either the virion strand or the complementary DNA strand, are transcribed bidirectionally [5] and are separated by an intergenic region. The nucleotide sequences of the two DNA components of TGMV have little sequence homology except for ~ 200 nucleotides within the intergenic region, which are almost identical in the two molecules [6]. This so-called 'common region' contains at least one origin of DNA synthesis [7], as well as transcriptional promoter and regulatory sequences [4].

TGMV protein AL1, which is an absolute requirement for DNA replication [8–10], has recently been shown to bind specifically within the common region [11]. In these experiments, the AL1 protein was expressed in insect cells using a baculovirus expression vector or transgenic AL1 plants. The formation of AL1–DNA complexes was demonstrated by immunoprecipitation from cell extracts using AL1 monoclonal antibodies. Here we report the expression of the AL1 protein in *E. coli* and its purification to near homogeneity. Using the purified protein and a variety of binding methods, we demonstrate that it binds with much higher affinity to single-stranded DNA of the common region than to the double-stranded DNA form. The significance of this preferential binding to single-stranded DNA is discussed in relation to possible functions of the AL1 protein in viral replication and transcriptional control.

Correspondence address: P. Thömmes, Department of Biology, Imperial College of Science, Technology and Medicine, London SW7 2BB, UK.

Abbreviations: BSA, bovine serum albumin; DTT, dithiothreitol; ORF, open reading frame; PCR, polymerase chain reaction; PMSF, phenylmethylsulfonyl fluoride; RCNMV, red clover necrotic mosaic virus; RCR, rolling circle replication initiator protein; TGMV, tomato golden mosaic virus.

2. MATERIALS AND METHODS

2.1. *Expression and purification of the AL1 protein*

The open reading frame (ORF) of the AL1 protein was expressed in *E. coli* and purified from the bacterial lysate. By in vitro mutagenesis *Nde*I and *Bgl*II sites were introduced at the start and just after the end of the AL1 ORF in a clone of TGMV DNA A in pEMBL9X [12]. This *Nde*I/*Bgl*II fragment of DNA A was cloned between the *Nde*I and *Bam*HI sites in pGEMEX-2*174, a modified version of the vector pGEMEX-2 (Promega) in which a second *Nde*I site had been destroyed [13]. By substituting the T7 gene 10 with the AL1 gene, the latter can be transcribed under the control of the T7 promoter by the bacteriophage T7 RNA polymerase provided by BL21 (pLysS) cells. Expression and purification of the AL1 protein was carried out as described for the Tobacco mosaic virus movement protein [14]. Cells were broken by sonication (14 μ, 4 min) and insoluble material was spun down. The AL1 protein was purified from inclusion bodies by washing the pellet successively with buffer L (10 mM Tris-HCL, pH 8, 10% glycerol, 1 mM EDTA, 200 mM NaCl, 1 mM DTT, 1 mM PMSF) containing 200 mM NaCl, 1 M NaCl, 1 M urea, 4 M urea, and 0.5% SDS. The fraction that had been solubilized with 4M urea, which contained most of the AL1 protein, was dialysed extensively

against buffer L and used for binding experiments. Generally, most of the protein remained soluble after dialysis; occasionally up to 20% of the total protein precipitated during freeze–thaw cycles. The fraction solubilized by 0.5% SDS, which also contained AL1 protein, was used for antibody production.

2.2. *DNA*

DNA of the common region of TGMV DNA A was amplified by 30 cycles of PCR between primers starting at nucleotide 5 and 235, respectively [6]. The PCR reaction contained 10 mM Tris-HCl, pH 8.8, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 0.2 mM dNTPs, 500 pmol/ml of each primer and 200 ng/ml template DNA (TGMV DNA A *Eco*RI/*Xho*I fragment) and 2.5 U *Taq* DNA polymerase. Cycles were 1 min at 95°C, 1 min at 40°C and 20 s at 72°C. A fragment of a cDNA clone of RCNMV RNA 2 extending from position 126 to 347 [15] was amplified in the same way. When the DNA was required for photochemical cross-linking experiments, dTTP was replaced by 0.2 mM BrdUTP. For radiolabeling DNA, dCTP was reduced to 20 μM and 5–10 μCi [α-^{32}P]dCTP (3,000 Ci/mmol) were added. The DNA was purified by Magic PCR prep (Promega) according to the manufacturer's protocol, and used directly for the experiments. For producing single-stranded DNA the PCR product was heated to 100°C followed by immediate shock cooling.

2.3. *Antibodies*

A rabbit was injected intramuscularly with AL1 protein (500 μg in 0.5 ml) mixed with an equal volume of Imject Alum (Pierce) five times at two-weekly intervals. The IgG fraction was purified from the serum on protein A agarose using the ImmunoPure gentle Ag/Ab buffer system (Pierce) according to the manufacturer's directions.

2.3. *Immunoprecipitation*

18 μl AL1 protein (various amounts) in buffer L were mixed with 2 μl ^{32}P-labeled DNA (50,000 cpm) in water and 2.2 μl buffer B (300 mM HEPES, pH 7.5, 10 mM DTT, 1 mg/ml BSA, 50 mM $MgCl_2$) for 1 h at 0°C. Antibodies (30 μg) were added followed by further incubation for 3 h. Finally formaldehyde-fixed *Staphylococcus aureus* cells (30 μl of a 10% solution; BRL) were incubated with the immunocomplex with shaking for 30 min. Washing of the immunocomplex was as described [16]. The final pellet was resuspended in 5% TCA containing 25 mM sodium pyrophosphate and filtered through Gf/C filters (Whatman). The filters were washed with the same TCA solution, dried, and the radioactivity determined by liquid scintillation counting.

2.5. *Photochemical crosslinking*

Incubation of AL1 protein and labelled DNA was as described in section 2.4 except that buffer B contained 0.1 M KCl. Then the samples were irradiated in open tubes with 2J in a Stratalinker (Stratagene). Excess DNA was removed by digestion with 10 U of DNaseI and incubation for 1 h at 37°C. After addition of gel loading buffer, samples were electrophoresed through 12% SDS-polyacrylamide gels [17]. Gels were stained with Coomassie blue, dried, and autoradiographed.

2.6. *Gel retardation assay*

The incubation of DNA and AL1 protein was as described in section 2.4 but with the omission of buffer B. The samples were then mixed with gel loading buffer and electrophoresed on a 4% native polyacrylamide gel in Tris-acetate-EDTA buffer [18] until the Bromophenol blue dye had reached the end. The gels were then dried and autoradiographed.

3. RESULTS

The AL1 ORF of TGMV DNA A was cloned into pGEMEX-2*174 under the control of the T7 RNA polymerase promoter. AL1 was expressed in large amounts, representing > 50% of the cellular protein. Most of this protein is located in inclusion bodies and is insoluble in 200 mM salt. This enabled the protein to be purified by washing with increasing salt concentrations and finally be solubilized in 4 M urea. After dialysis against low salt buffer generally most of the protein remained soluble. The final preparation (Fig. 1A, left panel) consisted of two nearly homogenous polypeptides of 41 and 29 kDa. The latter is a preparational degradation product of the former, because (i) its amount increases during the purification process and (ii) it reacted in Western blots with polyclonal antibodies against AL1 which had been affinity purified against the 41 kDa polypeptide (data not shown). Neither the 41 kDa nor the 29 kDa proteins were detected in bacteria containing the vector lacking the AL1 ORF.

To compare the ability of the AL1 protein to bind to single-stranded and double-stranded forms of the TGMV common region, the common region of DNA A from nucleotides 5 to 235 was amplified by PCR with the inclusion of BrdUTP and [α-^{32}P]dCTP. Increasing amounts of protein were incubated with either single- or double-stranded DNA and binding was assayed by photochemical cross-linking (Fig. 1A, right panel). Both the full-length AL1 protein (41 kDa) and 29 kDa degradation product were found to bind to DNA approximately equally when compared to the amount of each polypeptide in the preparation. The proteins covalently linked to radiolabeled nucleotides showed a slightly slower electrophoretic migration than the free protein, as found previously in cross-linking experiments with other proteins [14,15]. The bands detected by autoradiography were diffuse, particularly when high amounts of protein bound to single-stranded DNA, suggesting decreased accessibility of the DNA to DNAse digestion.

There was a pronounced difference between the binding of the AL1 protein to single-stranded compared to double-stranded DNA. Whilst 200 ng of protein bound little double-stranded DNA, resulting in a weak signal, as little as 25 ng protein bound enough single-stranded DNA to show two clearly visible bands. Quantification of the intensities of the bands by scanning densitometry indicated that the binding of the AL1 protein to single-stranded DNA was about 4-fold stronger than to double-stranded DNA.

A similar conclusion was reached using anti-AL1 antibodies for immunoprecipitation experiments (Fig. 2). Neither the protein itself, nor the preimmune serum, is able to precipitate DNA, indicating that non-specific complexes were not formed under the conditions used. However, the anti-AL1 antibodies were able to precipitate the protein–DNA complex. Again, the binding of AL1 to single-stranded DNA was much stronger than to double-stranded DNA, differing by a factor of six to ten.

To investigate, whether the specificity of binding of the AL1 protein to the common region of TGMV is

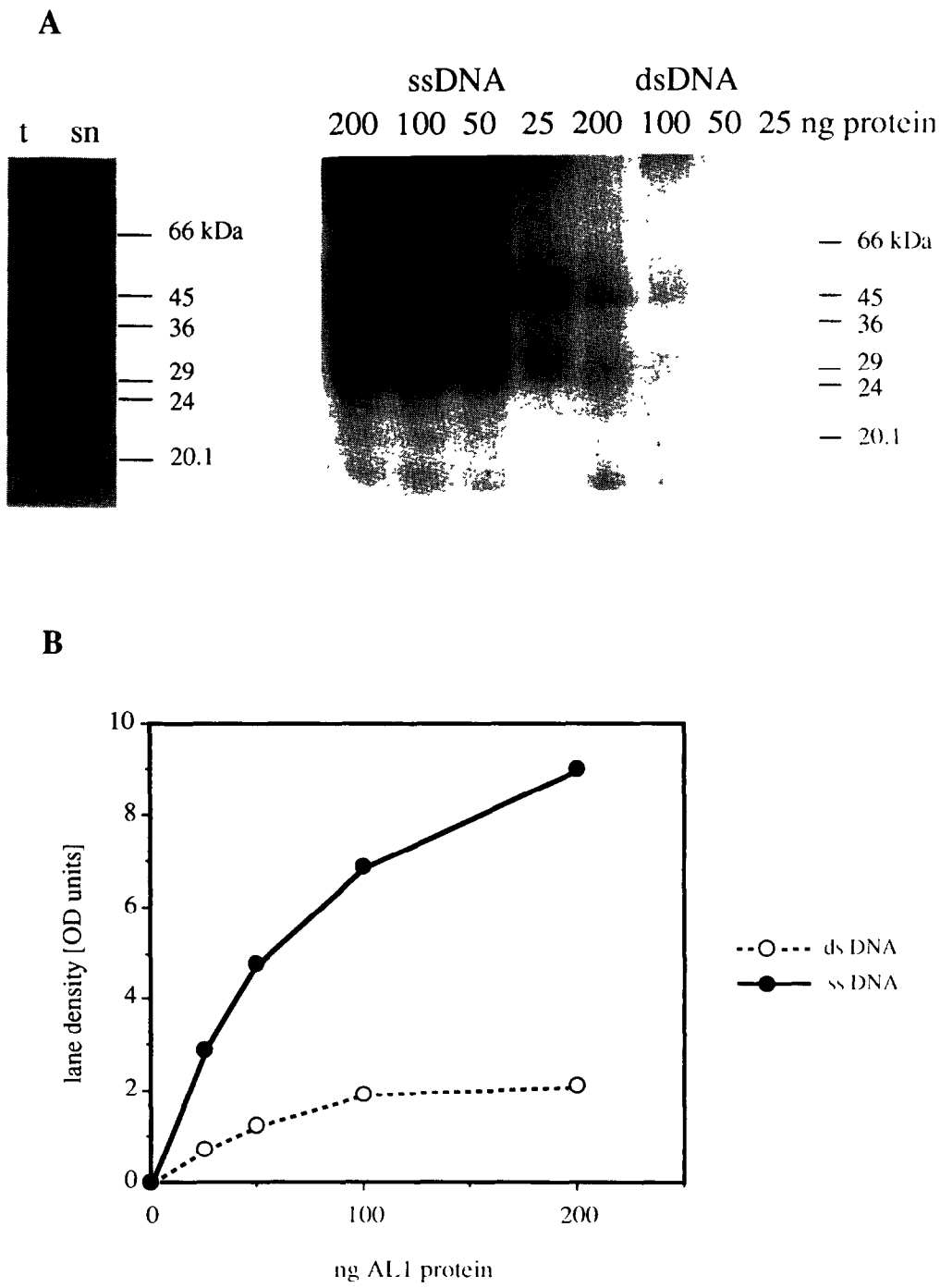

Fig. 1. The 41 and 29 kDa polypeptides of AL1 preferentially bind to single-stranded DNA of the common region. (A) Left panel: AL1 protein was purified from the bacterial lysate as described. After dialysis against low salt buffer the total protein (t) or the supernatant after centrifugation representing the soluble fraction (sn) were applied to a SDS-polyacrylamide gel and stained with Coomassie; right panel: increasing amounts of AL1 protein, as indicated above each lane, were incubated with labelled common region DNA. Cross-linking between the protein and the DNA was achieved by irradiation with UV light. After digesting the excess DNA the proteins were separated by SDS-PAGE and autoradiographed. (B) Quantification of the autoradiography shown in A (right panel) by laser scanning densitometry.

maintained also to the single-stranded DNA we created a probe of similar length from the completely unrelated RNA virus, red clover necrotic mosaic virus (RCNMV). Single and double-stranded DNA of the common region and RCNMV were incubated with the AL1 protein and electrophoresed through a native polyacrylamide gel in a gel retardation assay (Fig. 3). The AL1 protein was used in saturating amounts so that both single- and

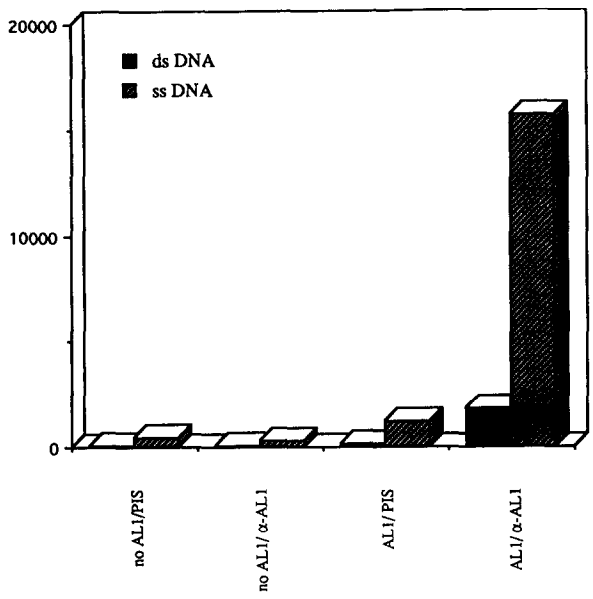

Fig. 2. Immunoprecipitation of the AL1–DNA complex. AL1 protein was incubated with labelled common region DNA either double-stranded or denatured by boiling. To show the specificity of the binding the protein–DNA complex was immunoprecipitated by antibodies against the AL1 protein and *Staphylococcus aureus* cells. The amount of DNA precipitated was quantified by liquid scintillation counting. Average input radioactivity was 50,000 cpm. PIS, pre-immune serum; α-AL1, antibody to the AL1 protein.

double-stranded forms of the common region DNA were completely bound. In both cases, soluble complexes which entered the gel were formed. However, no binding of either the single-stranded or the double-stranded DNA of the RCNMV probe could be detected even when the gel was over-exposed. When retardation of the labeled common region probe was competed with up to a 1,000-fold excess of unlabeled homologous or heterologous probes, only the homologous probe competed for binding (data not shown).

4. DISCUSSION

Previous attempts to use the AL1 protein in *E. coli* for binding studies were unsuccessful because of its insolubility [11]. However, we have taken advantage of this initial insolubility to purify the protein to near homogeneity. Furthermore we have shown that the protein can be solubilized in 4 M urea and that most of it remains soluble after removal of the urea. This has enabled binding studies to be carried out with the purified protein and, in addition to an immunological assay, to employ photochemical cross-linking and gel retardation analysis. The results confirm the conclusion, based on results using an immunological approach with unpurified extracts of insect or plant cells containing the AL1 protein, which found that the AL1 protein binds specifically to double-stranded DNA of the common region [11]. The finding that both the 41 kDa and 29 kDa proteins bind equally well indicates that the full-length AL1 protein is not needed for binding.

In addition to a specific binding to the common region we find that the AL1 protein binds with a much higher affinity to single-stranded than to double-stranded DNA without losing its specificity. Binding to single-stranded DNA is consistent with the proposed functions of the AL1 protein in DNA replication. The TGMV AL1 protein, and equivalent proteins in other geminiviruses, have a C-terminal domain containing sequence motifs characteristic of helicases [19]. Several eukaryotic DNA helicases have been found to bind specifically or preferentially to single-stranded DNA [20]. Geminivirus AL1 proteins also have an N-terminal domain with sequence motifs characteristic of rolling circle replication initiator proteins (RCRs) of bacterial plasmids [21]. The most widely studied RCR, the gene A protein of bacteriophage ΦX174, has been shown to be a sequence-specific endonuclease. Cleavage only takes place on single-stranded DNA or supercoiled double-stranded DNA containing AT-rich single-stranded regions [22]. Hence the binding of the AL1 protein to single-stranded DNA of the common region is consistent with its postulated functions both as a DNA helicase and as an RCR.

The following model can now be proposed for the role of the AL1 protein in geminivirus DNA replication. (i) The AL1 protein binds initially to a previously identified AT-rich site [11] in supercoiled DNA of the common region, which encompasses the promoter and transcriptional start site of the AL1 ORF. This causes

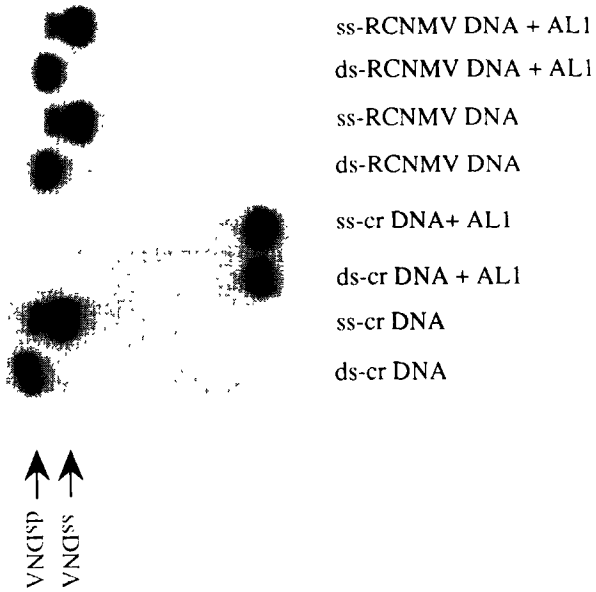

Fig. 3. Specificity of DNA binding by the purified AL1 protein. AL1 protein was incubated with labelled DNA of the common region (cr) or unrelated RCNMV DNA, as indicated above each lane. The reaction products were separated by native-PAGE and the position of the DNA fragments visualized by autoradiography.

down-regulation of transcription of the AL1 ORF. Indeed such down-regulation has recently been demonstrated [23]. (ii) The putative helicase activity of the AL1 protein can then unwind the region containing the origin of replication. This would allow the same or an additional AL1 molecule to bind to, and cleave, the single-stranded DNA at a specific site to create a 3'-OH primer to initiate rolling circle replication. The origin of rolling circle replication in one geminivirus [3] has been located to an inverted repeat sequence, potentially capable of forming a stable stem–loop structure and present in the common region or its equivalent in all geminiviruses that have been sequenced. In TGMV, this sequence is adjacent to the AL1 binding site in the common region double-stranded DNA [11]. Further work will be required to confirm these putative biochemical activities of the AL1 protein.

Acknowledgements: P.T. thanks the European Molecular Biology Organisation for the award of a long-term fellowship.

REFERENCES

[1] Hull, R., Stanley, J. and Briddon, R.W. (1991) Arch. Virol. Suppl. 2, 173–177.
[2] Saunders, K., Lucy, A. and Stanley, J. (1991) Nucleic Acids Res. 19, 2325–2330.
[3] Stenger, D.C., Revington, G.N., Stevenson, M.C. and Bisaro, D.M. (1991) Proc. Natl. Acad. Sci. USA 88, 8029–8033.
[4] Buck, K.W. (1993) in: Encyclopedia of Virology (Webster, R.G. and Granoff, A., eds.) Academic Press, London, in press.
[5] Sunter, G. and Bisaro, D.M. (1989) Virology 173, 647–655.
[6] Hamilton, W.D.O., Stein, V.E., Coutts, R.H.A. and Buck, K.W. (1984) EMBO J. 3, 2197–2205.
[7] Lazarowitz, S.G., Wu, L.C., Rogers, S.G. and Elmer, J.S. (1992) Plant Cell 4, 799–809.
[8] Elmer, J.S., Brand, L., Sunter, G., Gardiner, W.E., Bisaro, D.M. and Rogers, S.G. (1988) Nucleic Acids Res. 16, 7043–7060.
[9] Hayes, R.J. and Buck, K.W. (1989) Nucleic Acids Res. 17, 10213–10222.
[10] Hanley-Bowdoin, L., Elmer, J.S. and Rogers, S.G. (1990) Proc. Natl. Acad. Sci. USA 87, 1446–1450.
[11] Fontes, E.P.B., Luckow, V.A. and Hanley-Bowdoin, L. (1992) Plant Cell 4, 597–608.
[12] Brough, C.L., Hayes, R.J., Morgan, A.J., Coutts, R.H.A. and Buck, K.W. (1988) J. Gen. Virol. 69, 503–514.
[13] Pauls, T.L. and Berchtold, M.W. (1993) Methods Enzymol., in press.
[14] Citovsky, V., Knorr, D., Schuster, G. and Zambryski, P. (1990) Cell 60, 637–647.
[15] Osman, T.A.M., Miller, S.J., Marriott, A.C. and Buck, K.W. (1991) J. Gen. Virol. 72, 213–216.
[16] Thömmes, P., Reiter, T. and Knippers, R. (1986) Biochemistry 25, 1308–1314.
[17] Laemmli, U.K. (1970) Nature 277, 680–685.
[18] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
[19] Gorbalenya, A.E., Koonin, E.V. and Wolf, Y.I. (1990) FEBS Lett. 262, 145–148.
[20] Thömmes, P., Ferrari, E., Jessberger, R. and Hübscher, U. (1992) J. Biol. Chem. 267, 6063–6073.
[21] Koonin, E.V. and Ilyina, T.V. (1992) J. Gen. Virol. 73, 2763–2766.
[22] Mansfield, A.D.M., Langeveld, S.A., Baas, P.D., Jansz, H.S., van der Marel, G.A., Veeneman, G.H. and van Boom, J.H. (1980) Nature 288, 561–566.
[23] Haley, A., Zhan, X., Richardson, K., Head, K. and Morris, B. (1992) Virology 188, 905–909.