EXPRESSION AND CHARACTERISATION OF THE RYEGRASS MOTTLE VIRUS NON-STRUCTURAL PROTEINS

Ina Baļķe, Gunta Resēviča, Dace Skrastiņa, and Andris Zeltiņš

Latvian Biomedical Research and Study Centre, Ratsupītes iela 1, Rīga, LV-1067, LATVIA
E-mail: anze@biomed.lu.lv

Communicated by Pauls Pumpēns

The Ryegrass mottle virus (RGMoV) single-stranded RNA genome is organised into four open reading frames (ORF) which encode several proteins: ORF1 encodes protein P1, ORF2a contains the membrane-associated 3C-like serine protease, genome-linked protein VPg and a P16 protein gene. ORF2b encodes replicase RdRP and the only structural protein, coat protein, is synthesised from ORF3. To obtain the non-structural proteins in preparative quantities and to characterise them, the corresponding RGMoV gene cDNAs were cloned in pET- and pColdI-derived expression vectors and overexpressed in several E. coli host cells. For protease and RdRP, the best expression system containing pColdI vector and E. coli WK6 strain was determined. VPg and P16 proteins were obtained from the pET- or pACYC- vectors and E. coli BL21 (DE3) host cells and purified using Ni-Sepharose affinity chromatography. Attempts to crystallize VPg and P16 were unsuccessful, possibly due to non-structured amino acid sequences in both protein structures. Methods based on bioinformatic analysis indicated that the entire VPg domain and the C-terminal part of the P16 contain unstructured amino acid stretches, which possibly prevented the formation of crystals.

Key words: plant, Ryegrass mottle virus, VPg, P16 protein.

INTRODUCTION

Plant virus genomes encode several types of proteins involved in various activities of the virus life cycle. Virus gene expression leads to formation of structural or coat proteins that protect the virus nucleic acid from environmental damage and several types of unstructured proteins, including polymerases that replicate the genomes, and movement proteins necessary for virus spread from one cell to the next. Additionally, plant virus genomes can code proteases, helicases, methyltransferases, suppressors of posttranscriptional gene silencing (PTGS), and other proteins (Mandahar, 2006).

Sobemoviruses contain single-stranded, positive-sense RNA in size varying between 4.0 and 4.5 kb. Their monopartite genomes contain four open reading frames (ORF). To maximise the coding capacity, sobemoviruses exploit polyprotein synthesis and translational frameshifting gene expression strategies (Tamm and Truve, 2000b). The coding part of the *Sobemovirus* genomes starts with a short ORF1 coding for 11.7–24.3 kDa proteins P1 with low similarity to each other (Tamm and Truve, 2000b; Callaway *et al.*, 2004). The functions of the sobemoviral P1 proteins is still under discussion. Sobemovirus P1 was reported to possess suppressor activity against PTGS, which is a major host defense response against plant viruses. Additionally, cysteine

residues in putative Zn-finger structures were found to be essential for cell-to-cell movement of the P1 protein (Sire *et al.*, 2008). Another possible function of P1 protein is a role in systemic infection, as demonstrated for the *Sobemovirus* member *Cocksfoot mottle virus* (CfMV) and other viruses. P1-deficient CfMV mutant was able to replicate in oat suspension culture but could not infect oat plants systemically (Meier *et al.*, 2006). Interesting results were obtained in experiments with P1 protein of Sobemovirus RYMV with C-terminally fused green fluorescence protein (GFP). P1-GFP fusion protein was shown to possess autoproteolytic activity in *E. coli* and plant cells, and the P1 N-terminus to be involved in the cleavage process. The biological role of the autoproteolytic activity remains to be elucidated (Weinheimer *et al.*, 2010).

All sobemoviruses exploit the -1 ribosomal frameshifting mechanism to produce the characteristic polyproteins (Balke *et al.*, 2007; Meier and Truve, 2007). As a result of the frameshifting, overlapping ORF2a and ORF2b direct the expression of two different polyproteins with a molecular mass of around 100 and 60 kDa, and N-terminal parts of both proteins contain 3C-like serine protease (Pro) and genome-linked virus protein (VPg). The typical genome organisation of sobemoviruses is described in our recent article (Balke *et al.*, 2007). ORF2a-encoded serine proteases are well characterised *Sobemovirus* non-structural proteins.

The main function of *Somebovirus* protease is autoproteolytic polyprotein processing into several mature proteins through E/T, E/S or E/N cleavage sites (Makinen *et al.*, 2000; Satheshkumar *et al.*, 2004). However, unusual proteolytic cleavage between amino acids A and V was also observed in some experiments (Gayathri *et al.*, 2006). The best studied sobemoviral protease from Sesbania mosaic virus (SeMV) can be purified from recombinant *E. coli* cultures, if the coding sequence of N-terminal transmembrane region is removed from the protease gene cDNA (Satheshkumar *et al.*, 2004). Accordingly, crystal structure analysis at resolution of 2.4 Å showed that N-terminally truncated recombinant SeMV serine protease exhibited the characteristic trypsin fold. Several surface-exposed aromatic amino acids (AA) were identified in the protease structure, probably interacting with VPg (Gayathri *et al.*, 2006).

VPg, another functional domain of sobemoviral polyproteins, contains a conserved AA motif (WAD or WGD). It is covalently attached to the genomic RNA at the 5'-end and it is thought to serve as a protein primer for replicase (Makinen *et al.*, 2000). A recently published report suggests that VPg is not necessary for viral RNA synthesis, at least *in vitro*. However, it is possible that the RNA *de novo* synthesis is still VPg-dependent, and requires other viral/host factors in infected plants (Govind and Savithri, 2010). Additionally, SeMV VPg serves as an activator for serine protease *in cis* and *trans,* if prepared in the form of fusion protein Pro-VPg from recombinant *E. coli* (Satheshkumar *et al.*, 2005).

The role of the ORF2a C-terminal part beyond VPg has been under discussion for a long time. It is not clear whether it has a functional role, or if it is only the byproduct of proteolytic cleavage. Recent studies using purified ORF2a encoded C-terminal fragments from SeMV, demonstrated ATPase and nucleic acid binding activity for the corresponding recombinant proteins (Nair and Savithri, 2010a). Such properties are characteristic, for example, for plant virus-encoded helicases. As demonstrated for recombinant TGBp1 proteins from potex- and hordeiviruses, helicase RNA unwinding activity is dependent on Mg ions and ATP (Kalinina *et al.*, 2002).

Sobemovirus replicases, or RdRP, are expressed as C-terminal parts of polyproteins after ORF2a frameshifting into ORF2b, resulting in a polyprotein with a domain arrangement of Pro-VPg-RdRP. To characterise the SeMV RdRP, the enzyme was expressed in recombinant *E. coli* as a thioredoxin-fusion protein. Purified SeMV RdRP was shown to be able to synthesise RNA from specific genomic and subgenomic RNA, with and without the protein primer VPg. The stem-loop structure at the 3'-end of the genome is important for the enzyme activity (Govind and Savithri, 2010).

Ryegrass mottle virus (RGMoV) is a plant virus, which belongs to the genus *Sobemovirus* and infects several species of *Gramineae* plants (Zhang *et al.*, 2001a). According to three-dimensional (3D) crystal structure analysis data,

RGMoV particles have icosahedral T=3 quasisymmetry and contain 180 coat protein subunits with the canonical jellyroll β–sandwich structure (Plevka *et al.*, 2007). Complete genome nucleotide sequence alignments demonstrated phylogenetic grouping of Sobemoviruses in two groups of viruses infecting, monocotyledonous and dicotyledonous plants, respectively. Interestingly, RGMoV, which infects monocotyledonous plants but groups with dicotyledon-infecting viruses is an exception to this rule (Sereme *et al.*, 2008). Recently, we resequenced the complete genome of RGMoV and found that, unlike that described previously (Zhang *et al.*, 2001a), RGMoV RdRP is translated via an -1 ribosome frameshifting mechanism, similar to other members of the *Sobemovirus* genus. Also, according to sequence analysis, ORF2a of RGMoV encodes typical sobemovirus 3C-like serine protease with a strongly conserved triad and substrate-binding amino acids. Several putative protease cleavage sites were identified in RGMoV polyproteins (for the localisation of the sites see Balke *et al.*, 2007).

While protease and coat protein 3D structures have been determined, very little is known about the structure of other *Sobemovirus* nonstructural proteins. CD spectral analysis suggests that *Sobemovirus* VPg's have a "natively unfolded" structure (Satheshkumar *et al.*, 2005a). No information is available about the structure of P1, RdRP and other sobemoviral polyprotein functional domains.

To characterise the structure and function of proteins, it is necessary to purify them in quantitative amounts. In contrast to viral coat proteins, which in many cases are actively synthesized, the expression level of virus nonstructural proteins is very low in infected plants. For example, the output of plant virus RdRP, if purified from natural sources, can be as low as 0.002 mg/kg plant material (Bates *et al.*, 1995). Therefore, the corresponding protein genes have to be cloned in heterologous expression vectors and purified from recombinant organisms.

The goal of the present study was to overexpress, purify and characterize the RGMoV unstructural proteins, both experimentally and using computer-assisted programmes.

MATERIALS AND METHODS

Construction of RGMoV nonstructural protein expression plasmids. The plasmid containing full length cDNA copy of RGMoV was used as the source of coding sequences for nonstructural protein genes (P1, Δ50Pro, VPg, P16 and RdRP) (Balke *et al.*, 2007). Nucleotide and amino acid sequences were obtained from Genbank file No. EF091714. For each protein construction, specific oligonucleotides containing restriction sites for rapid cloning in the corresponding expression vector were designed. Oligonucleotides used for PCR-mediated cloning are summarized in Table 1. For production of recombinant proteins in *E. coli* the following commercially available expression vectors were used: pET-28a(+), pET-Duet-1, pACYC-Duet-1 (all Novagen, USA), pCold-I (Takara, Ja-

216

Proc. Latvian Acad. Sci., Section B, Vol. **64** (2010), No. 5/6.

OLIGONUCLEOTIDES USED IN CLONING OF RGMOV NONSTRUCTURAL PROTEINS

Oligonucleotide	Sequence	Introduced restriction site
RG2-VPg-BamHI-F	GGATCCAAACGGAGAGCAGGGAGCGCGCGAGA	BamHI
RG2-VPg-HindIII-R	AAGCTTTCATTCTTCACGGGCGTCCCAATCA	HindIII
RG2-SerP-HindIII-R	AAGCTTTCATTCACTGGATTCACAGTTTGCATGGAA AATGGAGA	HindIII
RG-SerP-BglII-R	AGATCTTCATTCACTGGATTCACAGTTTGCATGGAAAATGGAGA	BglII
RG-SerP-RAW-6H-F	CCATATGCGCGCTTGGCTAGCAACCTCTCTGGA	NdeI
RG-P16-F	CCATATGTCCACAGGAAATGATATTCCTTTAAACTGCCAGCA	NdeI
RG-P16-BglII-R	AGATCTTCAAGCTGAGGGGGACCCCTGGACT	BglII
RG-P1-F-NdeI	CATATGCCTTCAGTGGTTATCGAGGTTTGCTCAT	NdeI
RG2-P1-HindIII-R	AAGCTTTCAATGATGTCTAGTCCAAGACTGCCCT	HindIII
RG-P1-BglII-R	AGATCTTCAATGATGTCTAGTCCAAGACTGCCCT	BglII
RG-Pol-BglII-R	AGATCTTCATTCTTCCGGGATTTCTCCTTCA	BglII
RG-Pol-E/T1983-NdeI-F	CATATGACTGCTAGAACGAGTACTAACGAGCTCT	NdeI
RG2-Pol- E/T3369-indIII-R	AAGCTTTCATTCTTCCGGGATTTCTCCTTCA	HindIII

pan). pET-derived vectors contain a six-histidine coding sequence upstream of the recombinant protein insertion site, allowing to produce His-tagged proteins and to purify them using affinity chromatography columns. pET-Duet-1 and pACYC-Duet-1 are designed for coexpression of two target genes in the same vector. pCold-I vector is suitable for efficient protein expression at extremely low temperatures from cold-shock gene *csp*A. This vector also contains the N-terminal His-tag sequence.

To amplify the corresponding RGMoV genes or coding sequences of polyprotein domains, a high-fidelity DNA polymerase iProof (BioRad, USA) was used in PCR reactions according to the protocol of the manufacturer. PCR reactions were carried out in Verity 96 Well Thermal Cycler (Applied Biosystems, USA). The obtained PCR products were first ligated in the pTZ57 vector (Fermentas, Lithuania). After selection of insert containing clones, several clones were sequenced using a BigDye cycle-sequencing kit and ABI Prism 3100 Genetic analyser (Applied Biosystems, USA). PCR product clones with expected cDNA sequences were excised from pTZ57 and subcloned in the corresponding expression vector. P16, Δ50Pro, and RdRP were cloned in the pET-Duet-1 vector using NdeI and BglII restriction sites. VPg was cloned using BamHI and HindIII restriction sites in the pACYC-Duet-1 vector. The P1-coding sequence was inserted in NdeI and BamHI restriction sites in to the pCold-I vector. Additionally, P16 from the pET-Duet-1 was recloned in to pET-28a(+) vector using restriction sites NdeI and XhoI. Similarly, genes of Δ50Pro and RdRP were recloned in pCold-I vector using the same restriction sites NdeI/BglII. All restriction enzymes were from Fermentas (Lithuania).

Nonstructural protein expression and purification. After isolation of the corresponding expression vector DNA, *E. coli* BL21(DE3) (New England Biolabs, USA) and WK-6 (Zell and Fritz, 1987) were transformed with the corresponding plasmid to obtain RGMoV protein expression clones. In all cases, at least three expression clones were prepared from single colonies. Expression cultures in 20 ml of 2TY-medium with corresponding antibiotics (kanamicin, chloramphenicol or ampicillin) were cultivated according to the expression vector protocol, provided by Novagen or Takara, respectively. pET-vector containing cultures were grown at +30 ^{o}C and 200 rpm until optical density (OD) reached 0.8 measured at 600 nm. After induction with 0.5 mM IPTG the cultivation was continued at +20 ^{o}C for an additional 16 h. For pColdI-containing cultures, cultivation was started at +30 ^{o}C and, after induction with IPTG at OD_{600}=0.3, continued overnight at +15 ^{o}C. For preparative cultures, the volumes were scaled up to 200 ml.

Cell biomass were harvested by low speed centrifugation, suspended in ProFound lysis buffer (Thermo Scientific, USA) and sonicated on ice using a cell disintegrator Soniprep 150 (MSE, UK). The soluble protein fraction was separated by centrifugation for 30 min at 13000 rpm and +4 ^{o}C. Proteins containing the N-terminal His-tag were purified using N-sepharose affinity chromatography columns and a B-PER 6xHis Fusion Protein Purification Kit (Thermo Scientific, USA) according to the protocol for native protein purification provided by the manufacturer. In cases of VPg and P16 proteins, the final steps of the purification were ammonium sulfate precipitation and dialysis against 100 volumes of 10 mM Tris, pH 7.0.

Sample analysis in denaturing polyacrylamide gels (SDS-PAGE). All samples were analysed in 12% SDS-PAGE. The samples for gel analyses were diluted with a buffer containing 50 mM Tris pH 7.0, 1% SDS, 50% glycerin; 5% mercaptoethanol; 0.004% Bromephenol blue and heated for 10 min at +95 ^{o}C. SDS/PAGE gels were visualised by Coomassie blue R-250 staining.

Sample analysis by mass spectrometer MALDI TOF. A MALDI-suitable matrix solution was prepared from 7.6 mg of 2.5-dihydroxyacetophenone (2.5-DHAP; Bruker Daltonics, Germany) suspended in 375 µl ethanol. To the

2.5-DHAP solution, 125 μl of an aqueous diammonium hydrogen citrate solution (80 mM) was added.

Protein-containing solution (1 mg/ml) was diluted with aqueous 2% trifluoroacetic acid (TFA) solution. Then, 2 μl of 2,5-DHAP matrix solution was added, mixed and applied onto a Bruker AnchorChip 600 (Bruker Daltonics, Germany) target and left to dry. After drying, the 600 μm anchors were well-covered with protein/matrix mixture. The analysis was performed using a MALDI TOF mass spectrometer (Bruker Daltonics, Germany).

Protein crystallisation. JCSG-plus HT-96 and Structure screen I+II HT-96 protein crystallization solution kits (Molecular Dimensions, UK) were used to determine the optimum conditions for VPg and P16 crystal formation. Ni-Sepharose-purified VPg and P16 (6 mg/ml) were mixed 1:1 with a corresponding solution from the kit and spotted on the cover of a 96-well Microbatch plate. The wells of the plate were filled with 40 μl of the same buffer. The crystal formation was monitored under light microscope every 3–7 days for an eight-week period.

Additionally, to identify crystallisation conditions for VPg, a sample was sent to a high-throughput screening facility located at the Hauptman-Woodward Medical Research Institute, USA. The screening facility ensures microbatch-under-oil crystallisation experiments under 1536 different conditions (Luft *et al.*, 2003).

Computer-assisted protein sequence analysis. To identify folded and unfolded regions of proteins, the software FoldIndex was used (Prilusky *et al.*, 2005), which is based on the average amino acid residue hydrophobicity and net charge of the sequence. The predictions were performed at window size 51 and step 1. Alternatively, to identify disordered regions in RGMoV proteins, the software PONDR was used (http://www.pondr.com; Romero *et al.*, 1997; Dunker *et al.*, 2002). The prediction programme of transmembrane helices in proteins (http://www.cbs.dtu.dk/; TMHMM version 2.0) was used for RGMoV 3C-serine protease and RdRP domain analysis.

RESULTS

Cloning, expression and purification of RGMoV unstructural proteins. Sequence analysis is needed to be able to separately express domains of RGMoV unstructural proteins. As demonstrated for other sobemoviruses, the formation of functionally active domains is dependent on polyprotein cleavage by 3C-serine protease at amino acid pair sites. E/T, E/S or E/N (Makinen *et al.*, 2000; Satheshkumar *et al.*, 2004). As the substrate specificity for RGMoV protease is not known, it was assumed that the enzyme cleaves the polyprotein at the same sites. The relative localisation of putative protease cleavage sites was shown in our previous paper (Balke *et al.*, 2007). Based on that study, localization of the ORF2a (551 AA) protease domain was set between amino acids 1 and 317, VPg domain from

AA 318 to 382, and P16 domain from AA 383 to 551. The order of domains in ORF2a polyprotein and FoldIndex analysis is shown in Figure 1A. For sobemoviral proteases, it is known that the 5'-terminal part of the gene codes a membrane anchor domain (Satheshkumar *et al.*, 2004). As shown in Figure 1B, prediction using TMHMM programme revealed the putative transmembrane anchor sequence (An) to be between AA 1 and 50 in the Pro sequence, which can significantly reduce the solubility of the recombinant protein. Therefore, to stimulate the formation of soluble protein in *E. coli* cells, the coding sequence of first 50 AA was removed from the Pro expression clone.
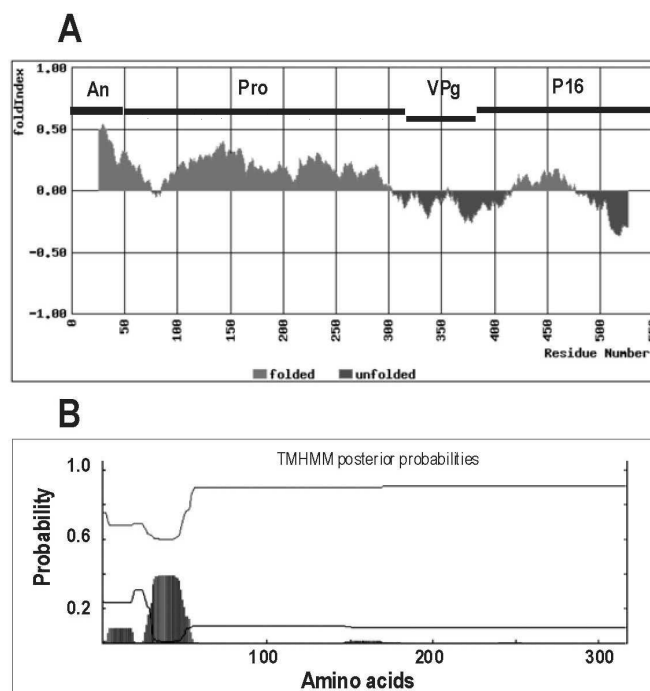

Fig. 1. Prediction of RGMoV ORF2a domain structure. A, analysis of ORF2a using the programm FoldIndex (window size 51, step 1). Folded amino acid regions have positive FoldIndex values, unfolded — negative, respectively. Localisation of putative functional domains in ORF2a is indicated in respect to amino acid sequence: An, membrane anchor sequence of protease, Pro, 3C-like serine protease, VPg, genome-linked virus protein, P16, ORF2a C-terminal domain. B, prediction of transmembrane helices in RGMoV protease using TMHMM programme. Putative transmembrane region is indicated as a shaded curve.

All ORF2a fragments (Δ50Pro, VPg, P16) and coding sequences for P1 and RdRP were amplified by PCR, and cloned in expression vectors as described in Materials and Methods. We tested the formation of serine protease in different expression systems. Protease production was undetectable after expression of Δ50Pro gene from the pET-derived vector in BL21(DE3) cells (Fig. 2A). A higher expression level was observed if the same cells contained the protease gene of the pColdI vector (Fig. 2B). The best production level of Δ50Pro was achieved in the case of expression of host *E. coli* WK6 and pColdI-derived vector. SDS/PAGE showed that the larger Δ50Pro was found in soluble protein fraction (Fig. 2C, lane 3). To purify the protein, the Δ50Pro-containing fraction was loaded on Ni-Sepharose column and eluted with imidazole buffer.
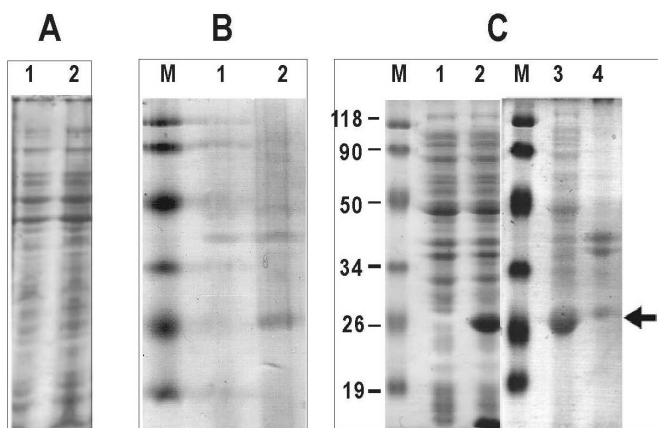
Fig. 2. SDS/PAGE analysis of Δ50Pro expression. A, pET-Du-Δ50Pro expression in *E. coli* BL21(DE3); B, pCold-Δ50Pro expression in *E. coli* BL21(DE3); C, pCold-Δ50Pro expression in *E. coli* WK-6. M, protein weight marker (Fermentas, Lithuania); 1, non induced total cell lysates; 2, induced total cell lysates; 3, soluble protein fraction; 4, cell debris and insoluble proteins.

However, the protein was not able to bind to the column material and was found in unbound protein fraction (data not shown).

A similar expression strategy was applied in the case of RGMoV P1 protein and RdRP. The P1 protein expression level was undetectable in all tested vectors and host strain combinations. To construct the RdRP expression clone, the part of ORF2a-ORF2b region was chosen, which codes AA 448-947 after the first putative E/T cleavage site in ORF2b protein (see Balke *et al.*, 2007 and Genbank file No. EF091714). Expression level analysis of RdRP demonstrated enhanced protein synthesis in the case of pCold-derived vector and *E.coli* WK6 cells, when compared with
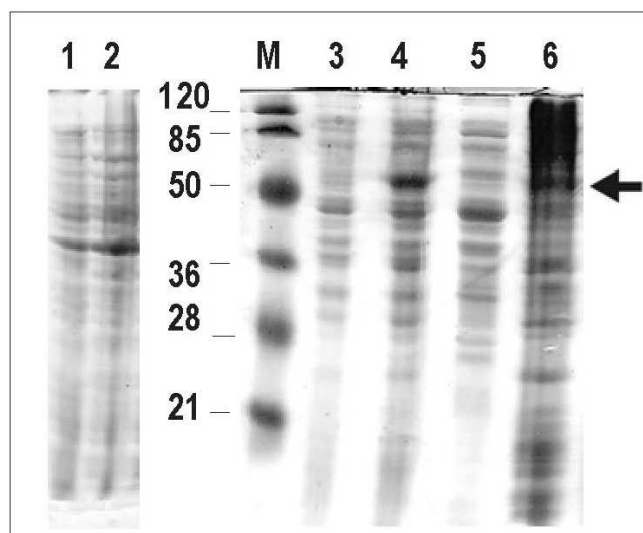

Fig. 3. SDS/PAGE analysis of pET-RdRP expression. 1, 2, pET-Du-RdRP-E/T1983 expression in BL21(DE3); 3–6, pCold-RdRP-E/T1983 expression in WK-6; 1, 3, non induced total cell lysates; 2, 4, induced total cell lysates; 5, soluble fraction; 6, insoluble proteins and cell debris; M, protein marker (Fermentas, Lithuania).

pET-vectors and BL21(DE3) host (Fig. 3). However, RdRP was found in the insoluble protein fraction and purification on a Ni-Sepharose column at native conditions was impossible despite the fact that the expressed RdRP domain did not contain hydrophobic regions in the AA sequence, as predicted by the TMHMM programme.

For expression of the VPg domain, the corresponding cDNA was cloned in the pACYC-Duet-1 vector and *E.coli* BL21(DE3) was chosen as a host strain. As the expression level was sufficient for the preparative production of VPg (Fig. 4; lane 2), other vectors and host strains were not fur-
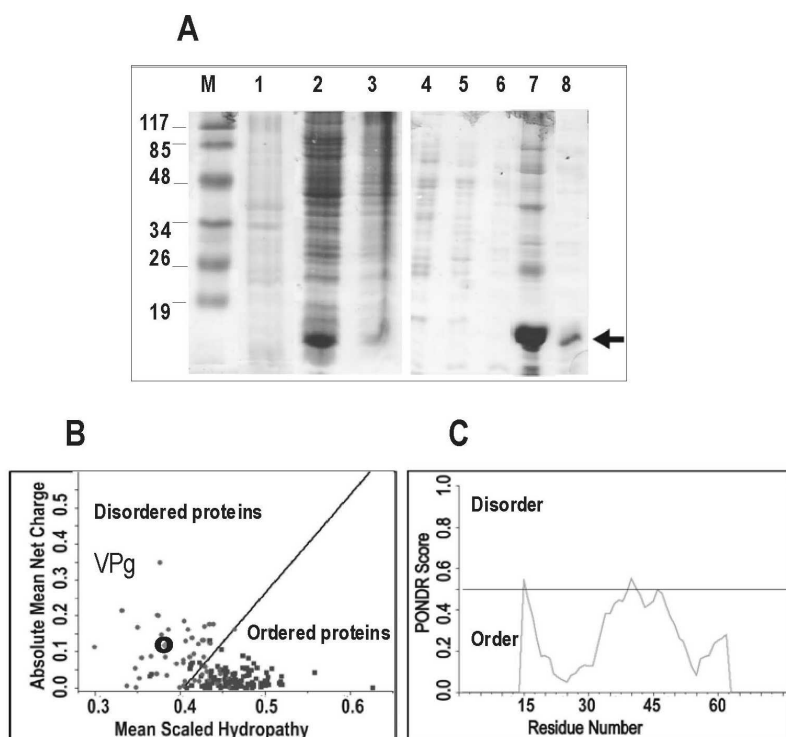

Fig. 4. VPg expression, purification and sequence analysis. A, SDS/PAGE analysis: M, protein weight marker (Fermentas, Lithuania); 1, noninduced total cell lysates; 2, soluble protein fraction; 3, insoluble proteins and cell debris; 4, proteins not bound to Ni-Sepharose column; 5, 6, Ni-Sepharose column washing fraction; 7, proteins eluted from Ni-Sepharose column; 8, VPg after ammonium sulfate precipitation. B, prediction of folding nature of VPg in comparison with other ordered and disordered proteins; the position of VPg is labeled with "O"; C, ordered and disordered regions of VPg depending on amino acid sequence.

Proc. Latvian Acad. Sci., Section B, Vol. **64** (2010), No. 5/6.

219

ther tested. SDS/PAGE analysis revealed that VPg was soluble and could be purified using Ni-Sepharose affinity chromatography columns. As shown in Fig. 4, VPg was bound efficiently to the column material and could be eluted from the column with the imidazole-containing elution buffer (lane 7).

To obtain P16 in preparative amounts for subsequent characterization experiments, the PCR product containing P16 cDNA was cloned in pET-28a+ downstream of the His-tag coding sequence using restriction sites NdeI/BglII(BamHI). The production level of soluble P16 in BL21(DE3) cells was acceptable and other expression systems were not tested. Similar to VPg P16 protein could be efficiently purified using Ni-Sepharose columns (Fig. 5). Surprisingly, in SDS/PAGE, denatured P16 protein (calculated MW= 18928 Da) appeared on gels as a protein with MW approximately 34 kDa (Fig. 5), and the abnormal behavior of P16 in SDS/PAGE did not change even after 30 min boiling in 0.1M DTT /1% SDS buffer. To confirm the P16 origin of the purified protein, the eluted protein (Fig. 5, lane 8) after precipitating with ammonium sulfate and dialysis was tested using mass spectrometry MALDI TOF. Mass spectrometry analysis revealed two major peaks (Fig. 5, B). One peak of 18603 Da approximately corresponded to the P16 monomer with 6 His-tag and without methionine (from AA sequence calculated value MW_{P16}=18908 Da). The second major peak of 37473 Da might be a signal of the P16 dimer. These data demonstrate comparably high stability of the putative P16 dimer, even at MALDI TOF experimental conditions.

Crystallization experiments. To obtain crystals for protein three-dimensional structure studies, purified VPg and P16 proteins were prepared for crystal formation using the hanging drop method. Two solution kits were used to test the crystal formation depending on factors such as combinations of precipitant (PEG4000, ammonium sulfate, 2-propanol and others) at different pH and presence of specific cations and anions. Altogether 192 different solutions were tested, but none of them promoted the crystal formation even after eight-week incubation. Then, a VPg protein sample was sent to the Hauptman-Woodward Medical Research Institute to determine conditions for crystallization, from 1536 different solutions. Also, in this case, the conditions for VPg crystal formation were not found. These data indicate that recombinant P16 and VPg proteins in the form of separate domains can not be crystallised, making impossible the subsequent crystallographic 3D-structure analysis.

Bioinformatic protein sequence analysis. Based on numerous protein crystallization experiments, it is well known that unstructured regions in protein structure can inhibit crystal formation (Pantazatos *et al.*, 2004). The reason for unsuccessful crystallisation experiment with VPg and P16 may be disordered regions in the structure of these proteins. Initial FoldIndex-supported analysis of complete RGMoV ORF2a revealed several disordered regions in the ORF2a polyprotein, and a significant part amino acids are in the disordered state (193 from 551). Accordingly to FoldIndex, the complete VPg domain and at least a part of the P16 domain represents a clearly unfolded structure, in contrast to the well-structured protease domain (Fig. 1A). To obtain additional data about VPg and P16, a bioinformatics tool (PONDR) was chosen for specific analysis of both recombinant proteins. Calculations of the VPg net charge and hydropathy profile in clearly placed VPg among the disordered proteins, however, the whole VPg domain, according to PONDR score, could be classified as rather ordered (Fig. 4; B,C).
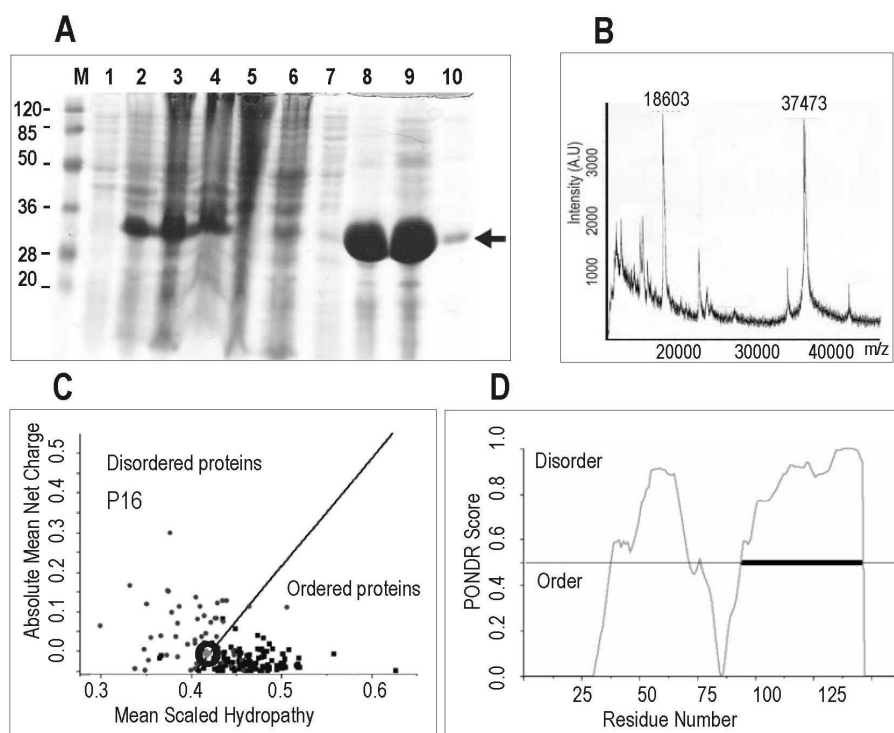

Fig. 5. Expression, purification and analysis of P16. A, SDS/PAGE analysis of pET-P16 expression and purification on Ni-Sepharose column; M, protein weight marker (Fermentas, Lithuania); 1, noninduced total cell lysate; 2, IPTG induced total cell lysate; 3, soluble protein fraction; 4, insoluble proteins and cell debris; 5, proteins not bound to Ni-Sepharose column; 6, 7, proteins eluted with washing buffer; 8, 9, eluted P16 fractions; 10, P16 after ammonium sulfate precipitation. B, P16 analysis by mass spectrometer. C, prediction of folding nature of P16 in comparison with other ordered and disordered proteins; the position of P16 is labeled with "O"; D, ordered and disordered regions of P16 depending on amino acid sequence.

220

Proc. Latvian Acad. Sci., Section B, Vol. **64** (2010), No. 5/6.

PONDR analysis results for P16 revealed the position of the protein as close to the border between ordered and disordered proteins (Fig. 5, C). Also, the C-terminal part of the P16 protein appeared to be clearly disordered, with the PONDR score close to 1.0 (Fig. 5, D).

DISCUSSION

The production of stable, soluble proteins in preparative amounts is one of the most important steps in the determination of protein functional studies and structure. Here we attempt to clone, overexpress and purify plant virus RGMoV nonstructural proteins, with the aim to characterise their function and structure. Several plasmid vector *E. coli* host combinations were tested in order to maximise the output of corresponding proteins after cultivation. In cases of protease and RdRP, the expression level of both proteins could be considerably enhanced, when the pColdI-derived vector was used in combination with *E. coli* WK6 cells, whereas pET-derived vectors and BL21(DE3) were inefficient. However, the system was not suitable for P1 protein expression. Apparently, various systems of expression of P1 with different N-terminal fusion partners have to be tested. In the case of RdRP the larger part of the protein was insoluble in recombinant *E. coli,* despite of the 6His-tag containing a leader sequence of pColdI-origin at the N-termini of RdRP and low cultivation temperature. An efficient alternative has been demonstrated in the case of SeMV RdRP: soluble polymerase can be obtained in the form of thioredoxin fusion protein (Govind and Savithri, 2010). Also, in the case of RGMoV RdRP, different fusion proteins will be tested in the future to attempt to obtain the soluble protein for subsequent characterisation experiments.

As shown in our experiments, N-terminally truncated RGMoV 3C-like serine protease containing 6His-tag can be successfully expressed in soluble form. However, the protein was not able to bind to a Ni-Sepahrose column, indicating the absence of an accessible His-tag in the protein structure. One possible reason for the missing His-tag in Δ50Pro was suggested by FoldIndex analysis: the protein contains a short disordered region between AA 76 and 83 (Fig. 1A) including Lys and Arg residues. This region can be cleaved by *E. coli* trypsin-like enzymes (Pantazatos *et al.*, 2004), resulting in a N-terminally truncated protein without the His-tag. SDS/PAGE analysis indirectly confirmed this, as the calculated Δ50Pro MW was in 28.7 kDa, but the observed value was only approximately 26 kDa (Fig. 2C). On the other hand, unusual self-cleavage by Δ50Pro can not be excluded, as discussed in the case of SeMV protease.

VPg, another RGMoV structural protein, was successfully expressed and purified using His-tag affinity chromatography. All attempts to crystallize the VPg were unsuccessful, despite the fact that more than 1500 crystallization solutions tested. The reason for the inability of VPg to form crystals might be unstructured regions in the protein structure. Bioinformatics analysis using PONDR demonstrated disor-

dered properties of VPg when compared with other proteins. The significant part of the VPg amino acid sequence approximately corresponds to a PONDR score of 0.5. Such protein segments are often predicted to be so-called molecular recognition elements or features (MoRFs). MoRFs are short amino acid stretches able to undergo disorder-to-order transitions upon specific binding with functional partners (Vacic *et al.*, 2007). Plant virus VPg, including members of *Sobemovirus* group, are characterized as extended "natively unfolded proteins" lacking a distinct 3D-structure and existing at different conformations at physiological conditions (Hebrard *et al.*, 2009). Interestingly, VPg was demonstrated to modulate sobemoviral protease activity (Gayathri *et al.*, 2006), if prepared in the form of fusion protein Pro-VPg. One might expect that, in the form of the fusion protein, VPg would be more structured and support crystal formation needed for VPg 3D-structural analysis. Interestingly, also immunisation experiments confirmed the disordered nature of VPg. Immune responses in mice immunised with purified VPg were very low and the obtained antibodies were not sensitive enough to detect VPg in Western blots at the dilutions higher than 1:50 (data not shown).

Similar to VPg, also P16 protein was not able to form crystals in various conditions. Bioinformatic analysis revealed a clearly unstructured C-terminal part of the protein (Fig. 5, D), which might be expected to negatively influence crystal formation. Additionally, P16 demonstrated several interesting properties. Firstly, the mobility of P16 in SDS/PAGE was unusual, possibly due to the amino acid content of P16. The C-terminal part of the protein contains 14 basic AA (Arg and Lys), resulting in an isoelectric point of 9.26 for P16. Positively charged proteins typically behave anomalously in SDS/PAGE, as shown also for SeMV P8 protein (Nair and Savithri, 2010b). Secondly, our mass spectrometry analysis data suggest that P16 protein can be a homodimer in physiological conditions. Thirdly, P16 can bind nucleic acids, as observed in preliminary gel-retardation experiments (data not shown). Purified P16, if incubated with specific RGMoV RNA, as well as with unspecific RNA, cause a typical shift in nucleic acid mobility on agarose gels. Additional studies are necessary to demonstrate the substrate specificity for potential binding partners of P16. The identification of a specific binding partner is important not only for functional characterisation of P16. P16 co-crystallisation experiments with corresponding specific nucleic acids may stimulate the crystallisation process of P16 protein.

Proc. Latvian Acad. Sci., Section B, Vol. **64** (2010), No. 5/6.

221

REFERENCES

Balke, I., Resevica, G., Zeltins, A. (2007). The ryegrass mottle virus genome codes for a sobemovirus 3C-like serine protease and RNA-dependent RNA polymerase translated via -1 ribosomal frameshifting. *Virus Genes* **35**(2), 395–398.

Bates, H.J., Farjah, M., Osman, T.A.M., Buck, K.W. (1995). Isolation and characterization of an RNA-dependent RNA polymerase from Nicotiana clevelandii plants infected with red clover necrotic mosaic dianthovirus. *J. Gen. Virol.,* **76**, 1483–1491.

Callaway, A.S., George, C.G., Lommel, S.A. (2004). A Sobemovirus coat protein gene complements long-distance movement of a coat protein-null Dianthovirus. *Virology*, **330**, 186–195.

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., Obradovic, Z. (2002). Intrinsic disorder and protein function. *Biochemistry*, **41**(21), 6573–6582.

Gayathri, P., Satheshkumar, P.S., Prasad, K., Nair, S., Savithri, H.S., Murthy, M.R. (2006). Crystal structure of the serine protease domain of Sesbania mosaic virus polyprotein and mutational analysis of residues forming the S1-binding pocket. *Virology,* **346**(2), 440–451.

Govind, K., Savithri, H.S. (2010). Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase. *Virology,* **401**(2), 280–292.

Hebrard, E., Bessin, Y., Michon, T., Longhi, S., Uversky, V.N., Delalande, F., Van Dorsselaer, A., Romero, P., Walter, J., Declerck, N., Fargette, D. (2009). Intrinsic disorder in Viral Proteins Genome-Linked: Experimental and predictive analyses. *Virol. J.,* **6**, 23.

Kalinina, N.O., Rakitina, D.V., Solovyev, A.G., Schiemann, J., Morozov, S.Y. (2002). RNA helicase activity of the plant virus movement proteins encoded by the first gene of the triple gene block. *Virology,* **296**(2), 321–329.

Luft J.R., Collins, R.J., Fehrman, N.A., Lauricella, A.M., Veatch, C.K., DeTitta, G.T. (2003). A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.* **142**, 170–179.

Makinen, K., Makelainen, K., Arshava, N., Tamm, T., Merits, A., Truve, E., Zavriev, S., Saarma, M. (2000). Characterization of VPg and the polyprotein processing of cocksfoot mottle virus (genus Sobemovirus). *J. Gen. Virol.,* **81**(11), 2783–2789.

Mandahar, C. (2006). Positive-sense viral RNA. In: *Multiplication of RNA Plant Viruses*. Mandahar, C. (ed.). Springer Verl, Dordrecht, pp. 29–70.

Meier, M., Paves, H., Olspert, A., Tamm, T., Truve, E. (2006). P1 protein of Cocksfoot mottle virus is indispensable for the systemic spread of the virus. *Virus Genes* **32**(3), 321–326.

Meier, M., Truve, E. (2007). Sobemoviruses possess a common CfMV-like genomic organization. *Arch. Virol.,* **152**(3), 635–640.

Nair, S., Savithri, H.S. (2010a). Natively unfolded nucleic acid binding P8 domain of SeMV polyprotein 2a affects the novel ATPase activity of the preceding P10 domain. *FEBS Lett.,* **584**(3), 571–576.

Nair, S., Savithri, H.S. (2010b). Processing of SeMV polyproteins revisited. *Virology* **396**(1), 106–117.

Pantazatos, D., Kim, J.S., Klock, H.E., Stevens, R.C., Wilson, I.A., Lesley, S.A., Woods, V. L., Jr. (2004). Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc. Natl. Acad. Sci. USA,* **101**(3), 751–756.

Plevka, P., Tars, K., Zeltins, A., Balke, I., Truve, E., Liljas, L. (2007). The three-dimensional structure of ryegrass mottle virus at 2.9 A resolution. *Virology,* **369**(2), 364–374.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I., Sussman, J.L. (2005). FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics,* **21**(16), 3435–3438.

Romero, P., Obradovic, Z., Kissinger, C.R., Villafranca, J.E., Dunker, A.K. (1997). Identifying Disordered Regions in Proteins from Amino Acid Sequences. *Proc. I.E.E.E. International Conference on Neural Networks*, pp. 90–95.

Satheshkumar, P.S., Gayathri, P., Prasad, K., Savithri, H.S. (2005). "Natively unfolded" VPg is essential for Sesbania mosaic virus serine protease activity. *J. Biol. Chem.,* **280**(34), 30291–30300.

Satheshkumar, P.S., Lokesh, G.L., Savithri, H.S. (2004). Polyprotein processing: Cis and trans proteolytic activities of Sesbania mosaic virus serine protease. *Virology*, **318**(1), 429–438.

Sereme, D., Lacombe, S., Konate, M., Pinel-Galzi, A., Traore, V.S., Hebrard, E., Traore, O., Brugidou, C., Fargette, D., Konate, G. (2008). Biological and molecular characterization of a putative new sobemovirus infecting Imperata cylindrica and maize in Africa. *Arch. Virol.* **153**(10), 1813–1820.

Sire, C., Bangratz-Reyser, M., Fargette, D., Brugidou, C. (2008). Genetic diversity and silencing suppression effects of Rice yellow mottle virus and the P1 protein. *Virol. J.,* **5**, 55.

Tamm, T., Truve, E. (2000). Sobemoviruses. *J. Virol.,* **74**(14), 6231–6241.

Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N., Dunker, A. K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.*, **6**(6), 2351–2366.

Weinheimer, I., Boonrod, K., Moser, M., Zwiebel, M., Fullgrabe, M., Krczal, G., Wassenegger, M. (2010). Analysis of an autoproteolytic activity of rice yellow mottle virus silencing suppressor P1. *Biol. Chem.,* **391**(2-3), 271–281.

Zell, R., Fritz, H.J. (1987). DNA mismatch-repair in Escherichia coli counteracting the hydrolytic deamination of 5-methyl-cytosine residues. *EMBO J.,* **6**, 1809–1815.

Zhang, F.Y., Toriyama, S., Takahashi, M. (2001). Complete nucleotide sequence of ryegrass mottle virus: A new species of the genus sobemovirus. *J. Gen. Plant Pathol.,* **67**, 63–68.

DAUDZZIEDU AIRENES VĪRUSA NESTUKTURĀLO PROTEĪNU EKSPRESIJA UN RAKSTUROŠANA

Daudzziedu airenes plankumainības vīrusa (RGMoV) vienpavediena RNS genoms ir organizēts četrās atklātajās nolasījuma fāzēs (ORF), kuras kodē vairākus proteīnus: ORF1 kodē P1 proteīnu, ORF2a kodē ar membrānu saistītu 3C serīna proteāzi, ar genomu saistītā vīrusa proteīna VPg un P16 proteīna gēnu. ORF2b kodē replikāzi RdRP, bet vienīgais strukturālais proteīns, apvalka proteīns tiek sintezēts no ORF3. Lai iegūtu nestrukturālos proteīnus preparatīvos daudzumos un raksturotu tos, atbilstošo RGMoV gēnu cDNS tika klonētas pET un pColdI ekspresijas vektoros un ekspresētas vairāku *E. coli* celmu šūnās. Proteāzes un RdRP gadījumā labākā ekspresija izrādījās sistēma, kurā tika izmantots pColdI vektors un *E. coli* WK6 celms. VPg un P16 proteīni tika iegūti no pET vai pACYC vektoriem un *E. coli* BL21(DE3) saimniekšūnām un attīrīti, izmantojot Ni-Sepharose afinitātes hromatogrāfiju. Mēģinājumi kristalizēt VPg un P16 bija neveiksmīgi, iespējams, nestrukturētu aminoskābju sekvenču klātbūtnes dēļ abu proteīnu struktūrā. Uz bioinformātikas metodēm balstīta analīze parādīja, ka viss VPg domēns ir nestrukturēts, bet P16 proteīnam ir nestrukturēta C-terminālā daļa, kas, iespējams, traucēja kristālu veidošanos.

222

Proc. Latvian Acad. Sci., Section B, Vol. **64** (2010), No. 5/6.