Molecular cloning, sequence analysis and expression of genome segment 7 (S7) of *Antheraea mylitta* cypovirus (AmCPV) that encodes a viral structural protein

Venkata Ramana Murthy Chavali · Ananta K. Ghosh

Received: 2 May 2006/Accepted: 7 August 2006/Published online: 26 January 2007 © Springer Science+Business Media, LLC 2007

Abstract The Genome segment 7 (S7) of the 11 double stranded RNA genomes from Antheraea mylitta cypovirus (AmCPV) was converted to cDNA, cloned and sequenced. The nucleotide sequence showed that segment 7 consisted of 1789 nucleotides with an ORF of 530 amino acids and could encode a protein of approximately 61 kDa, termed P61. The 5' terminal sequence, AGTAAT and the 3' terminal sequence, AGAGC of the plus strand was found to be the same as genome segment 10 of AmCPV encoding polyhedrin. No sequence similarity was found by searching nucleic acid and protein sequence databases using BLAST. The secondary structure prediction showed the presence of 17α -helices, 18 extended β -sheets along the entire length of P61. The ORF of segment 7 was expressed in E. coli as His-tagged fusion protein, purified through Ni-NTA chromatography, and polyclonal antibody was raised in rabbit indicating that P61 is immunogenic. Immunoblot analysis using this antibody on viral infected cells as well as purified polyhedra showed that P61 is a viral structural protein. Motif scan search showed some similarity of P61 with Inosine monophosphate dehydrogenase (IMPDH) cystathionine- β -synthase (CBS) domain at the C-terminus and it was hypothesized that by binding to single stranded viral RNA through its CBS domain P61 may help in virus replication or transcription.

Keywords Cypovirus · Antheraea mylitta · P61 · Cloning · Sequencing · Recombinant protein expression · CBS domain

V. R. M. Chavali · A. K. Ghosh (⊠) Department of Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur 721302, India e-mail: aghosh@hijli.iitkgp.ernet.in

Introduction

Cypoviruses (CPVs) belong to the genus Cypovirus in the family *Reoviridae* [1, 2]. These viruses infect midgut cells of a wide range of insects belonging to the order Diptera, Hymenoptera and Lepidoptera [3-7]. Viral infection is characterized by the production of large numbers of occlusion bodies called polyhedra in the cytoplasm of infected cells. The genomes of CPVs like those of other members of the Reoviridae are usually composed of 10 double stranded (ds) RNA segments (S1–S10) [1] although, in some cases, a small, eleventh segment (S11) has been reported [8]. Each dsRNA is composed of a plus-stranded mRNA and its complementary minus strand in an end to end base paired configuration except for a protruding 5' cap on the plus strand. On the basis of electrophoretic migration patterns of the dsRNA segments in agarose or acrylamide gels, they have been classified into 14 different types [1, 5, 9]. Among the family *Reoviridae*, complete sequences of dsRNA genomes have been reported for members of the genera Orthoreovirus, Rotavirus, Orbivirus, Phytoreovirus, Coltivirus, Seadornavirus, Dinovernavirus, and putative members of *Fijivirus* and *Cypovirus*, [10–17].

From the *Cypovirus*, complete nucleotide sequence of type 14 *Lymnatria dispar* CPV (LdCPV-14) [18], type 1 *Dendrolimus punctatus* CPV (DpCPV-1) [19, 20], type 15 *Trichoplusia ni* CPV (TnCPV-15) [21], and type 1 *Bombyx mori* CPV (BmCPV-1) [22–27] have been reported and deposited in the Genbank. The elucidation of these cypovirus sequences has led to a better understanding of the possible functions that each dsRNA segment may play a role in viral pathogenesis. In case of BmCPV, segment 1, 3, 4, 6, 7, 10 encodes structural proteins VP1, VP2, VP3, VP4, VP5, polyhedrin, and segment 2, 5, 8, 9 encodes viral non-structural proteins RNA dependent RNA polymerase, p101, p44, NS5, respectively [22–27]. Segments 1, 3, 4, 6, 7, and 10 of DpCPV encode viral structural proteins while segments 2, 5, 8, 9 and 10 encode non-structural proteins [19, 20]. Besides these, polyhedrin genes of several other cypoviruses such as *Euxoa scandens* CPV (EsCPV) [28], *Orgyia pseudotsugata* CPV (OpCPV), *Heliothis armigera* CPV (HaCPV) [29], *Choristoneura fumiferana* CPV (CfCPV) [30], and *Uranotaenia sapphirina* CPV (UsCPV) [31] have been molecularly characterized but no sequence homology has been found among them.

The Indian non-mulberry saturniidae silkworm, Antheraea mylitta, produces an exotic variety of silk called tasar silk. Being wild in nature, a major population of these silkworms is getting destroyed by viral infection [32] and we have reported earlier that a type IV CPV called AmCPV containing 11 dsRNA segments is the causative agent [33]. AmCPV belongs to genus Cypovirus of the family Reoviridae [33]. Molecular cloning and characterization of AmCPV genome segment 9 and 10 of this CPV show that segment 9 codes for viral non-structural protein NSP38 having RNA binding property [34], and segment 10 codes for polyhedrin [35], respectively but no sequence similarity has been found with any other sequences in the public databases. To understand the role of each genome segment of AmCPV in virus replication and pathogenesis, complete characterization of all its genome segments is necessary but the structure and functions of no other genome segments except 9 and 10 have been reported. Here we report molecular cloning, sequencing and characterization of genome segment 7 of AmCPV, and show by immunoblot analysis that it encodes a 61 kDa viral structural protein.

Materials and methods

Silkworm and virus

CPV-infected Indian non-mulberry silkworm *A. mylitta* was collected from the tasar silk farms of West Bengal and Jharkhand states of India.

Purification of polyhedral bodies, isolation of total genomic RNA, and extraction of genome segment 7 RNA

Polyhedra were purified from the midguts of infected silkworm larvae by sucrose density gradient centrifugation according to the method of Hayashi and Bird [36] with some modification [33]. Genomic RNA was extracted from the purified polyhedra by the standard guanidinium isothiocyanate method [37] and fractionated in 1% agarose gel. The genome segment 7 was excised from ethidium bromide-stained gel and eluted using Qiaquick gel extraction kit (Qiagen).

Molecular cloning and sequencing of genome segment 7 RNA

Segment 7 genomic RNA of AmCPV was converted to cDNA as described by a sequence independent RT method [38] by using two primers (AG1 and AG2). The 3'-end of 5'-phosphorylated primer, AG1, (5' PO₄-CCCGGATCCGTCGACGAATTCTTT-NH₂ 3') was blocked by NH₂ to prevent its concatenation in subsequent dsRNA/DNA ligation reactions. Approximately 200 ng of purified segment 7 RNA was taken and primer AG1 was ligated to both 3'ends of dsRNA by using T4 RNA ligase for 1 h at 37°C. The tailed RNA was denatured by heating, annealed to primer AG2 (5'AAAGAATTCGTCGACGGATCCGGG 3'), which is complementary to AG1, and reverse transcribed at 55°C for 50 min by using Thermoscript reverse transcriptase (Invitrogen). The template RNA from RNA/cDNA hybrid was removed by digestion with RNaseH, and reannealing of the cDNA strands was done by incubating at 65°C for 2 h. The cDNA ends were repaired by incubation with Taq DNA polymerase (Bioline) at 72°C for 20 min and cloned into pCR2.1-TOPO vector (Invitrogen) to make plasmid pCR2.1 TOPO/AmCPV7. After transforming in E. coli TOP 10 cells, plasmids were isolated and characterized by EcoRI digestion. Recombinant plasmids containing proper size insert were then sequenced using Bigdye Terminator (ABI) in an ABI 3100 automated DNA sequencer with M13 forward and reverse primers as well as internal primers designed from deduced sequences.

Sequence analysis

Genome Sequence of AmCPV segment 7 was analyzed by Sequencher program and homology searches were done using BLAST [39]. Conserved motifs were identified using MotifScan program (http://myhits.isbsib.ch/cgi-bin/motif_scan). The molecular weight of the deduced protein, isoelectric point, and composition were calculated using protein calculator program (http://www.scripps.edu/~cdputnam/protcalc. html) [40]. Secondary structure was predicted using PHD program [41] and GOR4 [42].

Northern hybridization

In order to verify cloning of the segment 7 cDNA from the corresponding RNA of AmCPV, all of the genomic dsRNA segments were separated in a 10% polyacrylamide gel and observed by staining with ethidium bromide. The RNA segments in the gel were then denatured by brief treatment with 0.1 M NaOH, neutralized in TAE buffer and electroblotted onto nitrocellulose membrane. The membrane was then hybridized with ³²P-labelled cloned segment 7 cDNA from AmCPV, washed and autoradiographed [34, 43].

Expression of AmCPV segment 7 in E. coli

The entire protein-coding region (530 amino acids) of AmCPV segment 7 cDNA, from nucleotide 18 to 1610, was amplified by PCR from plasmid pCR2.1 TOPO/ AmCPV7 by using Accuzyme DNA polymerase (Bioline) and two synthetic primers, AGCPV54F (5' AGTAATCGGGATCCGAAATG 3'; forward primer) and AGCPV62R (5' TGTAGTCGTCGACCT TCATTA 3'; reverse primer), complementary to bases 1-20 and bases 1600-1620, respectively, and containing BamHI (in the forward primer) and SalI (in the reverse primer) restriction enzyme sites (underlined). The amplified PCR product was digested with BamHI and SalI, separated on a 1% agarose gel and purified from the gel by using a gel extraction kit (Qiagen). Purified DNA was ligated to BamHI/SalI-digested pQE-30 vector (Qiagen) in-frame with a sequence encoding six histidine residues at the N-terminus. The resulting recombinant plasmid, pQE-30/AmCPV7, was then transformed into E. coli M15 and colonies were screened following BamHI and SalI digestion.

For protein expression, recombinant bacteria was grown in 5 ml LB medium containing ampicillin (100 μ g/ml) and kanamycin (25 μ g/ml) for 3 h at 37°C and then induced with 1 mM IPTG for an additional 4 h at the same temperature. Bacteria were harvested by centrifugation, lysed by boiling with sample loading buffer (60 mM Tris-HCl, pH 6-8; 10% glycerol; 2% SDS; 5% β -mercaptoethanol and 1 µg/ml bromophenol blue) for 3 min and then loaded onto a 5% stacking gel cast above a 10% resolving SDS-polyacrylamide gel. After electrophoresis, the protein bands in the gel were stained with Coomassie brilliant blue [37]. The molecular mass of the expressed recombinant protein was determined by comparison with standard protein molecular mass markers and by using Quantity One software in Gel-Doc 2000 (Bio-Rad).

Purification of His-tagged protein

Recombinant bacteria containing pQE-30/AmCPV7 were grown in 1 l LB medium and induced with IPTG as described above. The insoluble His-tagged fusion protein (P61) was first purified as inclusion bodies [44]. After solubilizing the inclusion bodies in 6 M guanidine hydrochloride, further purification of protein was carried out using a Ni-NTA agarose kit (Qiagen) according to the manufacturer's protocol. The total amount of purified protein was quantitated by Bradford method [45] using BSA as the standard and purity was checked by SDS-10% PAGE [46].

Rabbit immunization and production of polyclonal antibodies

One rabbit was immunized with bacterially expressed, purified, recombinant His-tagged protein (P61) by standard methods [34, 47]. In brief, purified protein (550 mg) was mixed with Freund's complete adjuvant, and injected subcutaneously at multiple sites. Three booster doses with Freund's incomplete adjuvant and the same amount of protein were administered via the same route at 4-week intervals. Twelve days after the final booster, blood was collected, serum prepared and the antibody titer was determined by ELISA [47]. Specific antibody was purified by antigen (p61-Sepharose) affinity chromatography [47, 48].

Western blot analysis

For detecting the expression of P61 in infected cells, protein samples from dissected midgut of AmCPV infected and uninfected 5th instar larvae were prepared by homogenizing the tissue in PBS followed by centrifugation at 10,000g for 10 min [34]. Each protein sample in the supernatant was boiled in sample loading buffer and run on SDS-10% PAGE under reducing conditions. After electrophoresis, proteins from the gel were transferred onto a duralose membrane (Stratagene) using a transblot cell (Bio-Rad) according to manufacturer's protocol. The membrane was then blocked in PBS containing 3% BSA for 1 h at room temperature followed by incubation with 500fold diluted affinity purified anti-p61 polyclonal antibody for 1 h. After washing with PBS, the membrane was incubated with 200-fold diluted Protein Aconjugated horse raddish peroxidase for 1 h and then washed, and color development was done by using HPO color development kit (Bio-Rad).

The nucleotide sequence data reported here for segment 7 of AmCPV have been deposited in Genbank database and have been assigned the Accession No: DQ512389.

Results

Cloning of AmCPV genome segment 7

AmCPV genome segment 7 was purified from the genomic RNA of AmCPV after separating by 1% agarose gel electrophoresis, converted to cDNA and cloned into pCR 2.1 TOPO vector to create plasmid pCR 2.1 TOPO/AmCPV7. Digestion of isolated plasmids from transformed *E. coli* with *Eco*RI and their analysis by agarose gel electrophoresis showed an insert of 1.8 kb (data not shown) indicating cloning of full length segment 7 cDNA.

Northern analysis

Among the 11 dsRNA segments present in AmCPV the cDNA of segment 7 specifically hybridized with segment 7 genomic RNA, confirming the cloning of segment 7 genomic RNA from AmCPV (Fig. 1).



Fig. 1 Northern blot analysis of AmCPV genome RNA. AmCPV genome segments were resolved by 10% PAGE as discrete bands (Lane 1) (segment 11 was not visible as it ran out off the gel under these running conditions) and hybridized with ³²P-labelled cloned segment 7 cDNA (Lane 2). Arrow indicates hybridization of segment 7 cDNA to segment 7 genomic RNA. The number and sizes of different RNA segments of AmCPV are shown in the left

Molecular analysis of genome segment 7

Sequencing of AmCPV segment 7 showed that it consisted of 1789 nucleotides with a long single ORF of 530 amino acids. The ORF started with an ATG codon at nucleotide 18 and terminated in a TAG codon at nucleotide 1608 (Fig. 2a). Seventeen nucleotides upstream of the start codon and 179 nucleotides downstream of stop codon were present as untranslated sequences. The molecular weight of the encoded protein was deduced as 61 kDa and we termed the protein as P61. The 5' terminal sequences AGTAAT, and 3' terminal sequence AGAGC were found to be same as the 5' and 3' terminal sequences of polyhedrin gene encoded by AmCPV genome segment 10. No significant homology was detected with any nucleotide or protein sequences available in the public databases using BLAST. The deduced amino acid composition resulted in an isoelectric point of 6.07 and showed that P61 is rich in isoleucine (8.5%), leucine (8.1%), aspartate (7.9%), lysine (7.7%), threenine (7.5%) and glutamate (6.7%) residues with an isoelectric point of 6.07. Four potential N-linked glycosylation sites (at positions 68-70, 124-126, 175-177 and 309-311) and several phosphorylation sites were found within the protein coding region.

Secondary structure prediction using PHD and GOR4 programs showed that 45.8% of the residues are likely to form random coils, 34.2% would form α -helices and 20% would form extended sheets. A total of 17 α -helices and 18 extended β -sheets were found along the entire length of protein (Fig. 2b). MotifScan search showed significant similarity of P61 with Inosine monophosphate dehydrogenase (IMPDH) enzyme from 201 to 526 amino acid residues containing one Cystathionine Beta Synthase (CBS) domain at C-terminus (440–497 amino acid residues) with the characteristic of sheet/helix/sheet/sheet/helix topology (Fig. 3).

Expression and purification of genome segment 7 encoded protein

To express segment 7 encoded protein in *E. coli*, the ORF was cloned in pQE30 vector, expressed in *E. coli* and purified as 6X His-tag fusion protein through Ni-NTA affinity chromatography. Analysis of induced bacterial lysate and purified protein through SDS 10%-PAGE showed the production of ~63 kDa protein (Fig. 4, lanes 2 and 3) but, no such protein band was found in uninduced bacterial lysate (Fig. 4, lane 1).

(a)					1	8 1	init	iat	ior		ndor								
lnt	AGTAA	rcga	CGA	GCG	AAA	TGT	ATC	GTG	CCA	GGT	TTT	TCG	AAT	CCG	CCG	ААА	GGC.	ATC	TTCCC
laa						М	Y	R	Α	R	F	F	Е	s	Α	Е	R	н	L P
63nt	ATCGT	ICTC	CATC	ACT	AAC	ACA	CTT	TTC	AAA	.GAG	AAA	ATA	GCA	CCA	ATT	TTT.	AAG.	AAT	ATT
16aa		L	I	T	N	T	L	F	K	E	K	I	A	P	I	F	K	N	I
123nt	AAATCI	AACC	TAC	'GGA	GAT	GAA	ATT	TTT	GAT	CAG	GTT	AAA	TCA	ACT	ACT	GCA	GTC.	AGA	TAC
36aa	K S	T	Y	G	D	E	I	F	D	Q	V	K	S	T	T	A	V	R	¥
183nt	TTAATO	GAAC	AAT	'GAT	GCT	'ATG	GTC	ACC	GAC	TGG	AAA	AAT	GAG	ACT	GAA	GTC	CGA	AAA	TGG
56aa	L M	K	N	D	A	M	V	T	D	W	K	<u>N</u>	E	T	E	V	R	K	W
243nt	CTGAA	rgg1	TTC	AAT	GCT	GTC	CCA	CAA	ATA	.GAC	GCC	CTA	AAC	GAT	AAG	TAT	TAT	CCA	ATC
76aa	LN	G	F	N	A	V	P	Q	I	D	A	L	N	D	K	Y	Y	P	I
303nt	TTTAC	CTCI	ATG	TCT	AGG	ATA	ATA	AAG	TCG	CCI	'GAG	TTA	GCA	AAG	ACG	GTT	TTT.	AAT	AAG
96aa	F T	S	M	S	R	I	I	K	S	P	E	L	A	K	T	V	F	N	K
363nt	ATATCO	GTTC	TAT	GTG	ATT	CCA	GAT	'AAT	GCA	ACT	CAA	CCA	TCT	ACG	ATC	GAG	GAC.	AAT	TTC
116aa	I S	F	I	V	I	P	D	<u>N</u>	A	T	Q	P	S	T	I	E	D	N	F
423nt	CACTA	TATO	AAA	LCTA	GTT	CCT	'AAT	CCT	CAT	ATC	AAG	GGT	CAG	CAA	ATT	GAC	GAG	TTT	GTA
136aa	HY	I	K		V	P	N	P	H	I	K	G	Q	Q	I	D	E	F	V
483nt	AGGTG	CATC	GAG	AAT	ATT	GCT	ACC	'AAG	ACA	.GCC	AAT	CGA	CTT	AGG	AAC	CCA	CAG	CCC	AAT
156aa	R C	I	E	N	I	A	T	K	T	A	N	R	L	R	N	P	Q	P	N
543nt	AACAC	CACA	CAA	ATG	GAA	GCT	'AAA	CCT	GAA	CCA	GTA	ATA	TCC	CAA	GAC	GAT	CTA	GAT	ATA
176aa		T	Q	M	E	A	K	P	E	P	V	I	S	Q	D	D	L	D	I
603nt	GGGAC	GGAC	CAT	'ACA	CCT	GGG	CTC	GAC	TTA	GAT	GAC	GAA	CCC	TGT	AGC	GAG	GTT.	AGT	GAT
196aa	G T	D	H	T	P	G	L	D	L	D	D	E	P	C	S	E	V	S	D
663nt	CAAAG	rgac	CATC	TTG	CCA	GTA	GAA	ATA	CAC	ACI	'GAA	GCT	ATC	ATT	GAG	CCT	GAA	CCA	AAG
216aa	Q S	D	I	L	P	V	E	I	H	T	E	A	I	I	E	P	E	P	K
723nt	CCTCA	GCGI	TACA	TTG	AAA	GTT	'AAA	ACT	CCC	CTA	ACC	GTT	CAA	GAG	AGT	AAA	GTA	GAT	GTG
236aa	PQ	R		L	K	V	K	T	P	L	T	V	Q	E	S	K	V	D	V
783nt 256aa	CACGC	AGTA V	ACCG P	GTA V	GAT D	ATA I	GCG A	TCT S	ATC I	CCA P	LCTT	ATT I	CCG P	GAA E	CAT H	CAA Q	AAC N	GTT V	GAG E
843nt 276aa	CCAGC	ACCA P	AGCG	CCA P	TCC S	CAT H	CAC H	GCT A	AGT S	TCC S	GAA E	TCC S	ATT I	TGT C	AAA K	TGG W	CTC L	GAT D	CAA Q
903nt	TGTGAO	GAGO	TCT	'AAA	TCA	TAT.	'GAA	CAG	AAA	GCT	TCC	ATA	AAC	GTG	ACT	GGC	TTA	CGA	GTT
296aa	C E	R	S	K	S	Y	E	Q	K	A	S	I	<u>N</u>	V	T	G	L	R	V
963nt	GGCGA	rcga	NTCA	GTC	TTT	AAA	GTC	ACT	TTG	TCT	ЪАА	GAA	CGA	GGA	ATT	TTA	CGA	CAA	TTA
316aa	G D	R	S	V	F	K	V	T	L	S	К	E	R	G	I	L	R	Q	L
1023nt	CGAGA(CTTG	AAA	CTT	AGC	GAC	CAT	AAG	ATA	CAT	ATC	TTT	TAC	AAC	GAT.	ATA	CTA	CTA	ACT
336aa	R D	L	K	L	S	D	H	K	I	H	I	F	Y	N	D	I	L	L	T
1083nt	GATCGO	GGAI	'ACA	TTA	ATA	TCA	ATT	AAG	GCA	.GGC	GAA	GAT	GAT	ACT	TTC.	AAC.	ATT	GTT	ACT
356aa	DR	D	T	L	I	S	I	K	A	G	E	D	D	T	F	N	I	V	T
1143nt	GTGAAA	AGAC	GAG	AGC	AGC	GAG	AAG	CCC	CTT	GTG	GTA	ACA	TTG	AAA	ACG	GAC.	AAT	GTA	CCT
376aa	VK	D	E	S	S	E	K	P	L	V	V	T	L	K	T	D	N	V	P
1203nt	TTTAT(GCAI	GGT	GTT	CAC	CCA	Q	CAC	AGC	ATA	ACA	TCT	GAG	CTT	GGC	GAT	TGG	GCC	TTT
396aa	F M	H	G	V	H	P	Q	H	S	I	T	S	E	L	G	D	W	A	F
1263nt	ACTGC	TATO	GTG	GAC	TCA	TTA	GCT	TAT	AAT	CAC	CTC	AGA	TTC	TTC.	AAG.	ACA	GCA.	AAT	TTG
416aa	T A	M	V	D	S	L	A	Y	N	H	L	R	F	F	K	T	A	N	L
1323nt	GATGA	CGAI	'AAG	TTC	CTG	ATT	ACT	ATT	GGT	TCG	AAA	TAC	GCC	ATA	GAA	GAT	TTC.	ACC	TCA
436aa	DD	D	K	F	L	I	T	I	G	S	K	Y	A	I	E	D	F	T	S
1383nt	ATTGA	AGGC	TTG	AAG	ACG	AAT	CAT	AAA	CGT	CAT	AAG	ACA	TCT	AAC	TTA	CTT	CTA	TTC	CAA
456aa		G	L	K	T	N	H	K	R	H	K	T	S	N	L	L	L	F	Q
1443nt	TCCAC	ATCO	CTT	GAA	GGA	TTA	TTG	GCT	GCT	GCC	GGT	GAT	GAT	CGA	TTG.	AAA	AGA.	ACC	TAC
476aa		S	L	E	G	L	L	A	A	A	G	D	D	R	L	K	R	T	Y
1503nt	ATATT	rgat	ACC	ATT	GAG	AAA	AGT	TTT	AAT	AAA	TAT	TTA	GAA	GAA	GAT	GGA	AGT	GAT	TAC
496aa		D	T	I	E	K	S	F	N	K	Y	L	E	E	D	G	S	D	¥
1563nt 516aa	TCAATO S M	GTTI F	CAG Q	CAG Q	AGA R	ACG T	AGT S	GGA G	GTA V	ATT I	GGT G	TTA L	ATG M	AAG K	TAG 53	ACC 0aa	GAC'	TAC	AAT
1623nt	ATGCC	AGAI	TAA	CCA	ATT	AAG	TAT	AAC	ACT	TTA	GGA	CAT.	ATA	TAT	CCA	ACA	TTA	GTA	TAG
1683nt	TTTGG	rggī	GGT	ACA	ATG	GCG	GTT	AGT	GTT	TGA	GAA	ACT	GGG	TCG	TGA	TTT	TGG	GTA	GGG
1743nt	TTTGA	CTGG	GCT	CGA	GAG	TTT	CAA	AAC	AGT	AGC	CGT	CGG	TAT	AGA	GC	178	9nt		

Fig. 2 (a) Complete nucleotide (nt) and deduced amino acid (aa) sequences of AmCPV genome segment 7. The CBS motif found between 440 and 497 amino acid residues is represented by **-----** below the amino acid sequences. The initiation and termination codons are shown in bold. Four potential N-linked glycosylation sites are double underlined. (b) Secondary structure prediction was represented above the amino acid residues. The cylinders (**_____**) represent helix, arrows (**_____**) represent the beta strands and the joining lines (—) represent random coil



Fig. 2 continued

Detection of P61 expression in virions and viral infected cells

To demonstrate the expression of P61 in virions and virus infected cells, CPV infected *A. mylitta* midgut cells and purified polyhedral bodies were analyzed through SDS-10% PAGE and subjected to immunoblot analysis using affinity purified anti-P61 polyclonal antibodies. A major immunoreactive protein band of approximately 65 kDa was found in lane containing lysate of infected midgut cell (Fig. 5, lane 2) and polyhedral bodies (Fig. 5, lane 3), but not in lane containing lysate of uninfected cell (Fig. 5, lane 1)

Discussion

AmCPV although infects and destroys an economically important insect species, *Antheraea mylitta*, which produces tasar silk, but except segments 9 and 10, none of its other genome segments have been fully characterized at the molecular level. But full molecular characterization of this CPV is very much needed to understand the viral life cycle, pathogenesis and to find target sites for the development of effective anti-viral compound. Here we describe molecular analysis of another of its genome segment, Segment 7, by cloning, sequencing and expressing in E. coli. The 1789 nucleotides long cloned cDNA consists of an ORF of 530 amino acids and its molecular mass is calculated to be approximately 61 kDa. Presence of conserved terminal sequences at 5' and 3' ends is a characteristic feature found in different genome segments of most of the virions of the *Reoviridae* family [49]. At the 5' and 3' end of AmCPV segment 7, AGTAAT and AGAGC sequences were found as observed at the 5' and 3' ends of AmCPV genome segment 10 encoding polyhedrin indicating the genome structure of this CPV may follow the same pattern as observed in other CPVs, but cloning and sequencing of other genome segments of AmCPV are required to confirm it. Absence of any significant sequence similarity of S7 either at nucleotide



Fig 3. Alignment of CBS domains found in AmCPV segment 7 with that of enzymes, channel and few hypothetical proteins. The sequences are annotated with their SWISS-PROT identifier followed by the start point and end point of each domain. AmCPV_S7/440-497 denotes CBS domain found in AmCPV segment 7, from 440 to 497. IMDH_ARATH, IMDH_METKA, IMDH_DROME, IMD2_MESAU, IMH3_YEAST and IM-H4_YEAST are inosine monophosphate dehydrogenases from *Arabidopsis thaliana, Methanopyrus kandleri, Drosophila melanogaster, Mesocricetus aurateus* and *Saccharomyces cervisiceae*,

respectively. YR33_THEPE is a hypothetical protein from *Thermophilum pendens*. YTFL_HAEIN is a hypothetical protein from *Haemophilus influenza*.CLC3_HUMAN is a chloride channel protein from *Homo sapiens*. AAKG_RAT is a 5' AMP activated protein kinase from *Rattus norvegicus*. KSF1_ECOLI is a polysialic acid capsule expression protein from *E. coli*. CBS_RAT is a cystathionine beta synthase H-450 protein from *Rattus norvegicus*. ACUB_BACSU is acetoin utilization protein B from *Bascillus subtilis*. The shaded boxes indicate the conserved residues in the CBS domains from different species



Fig. 4 Analysis of protein samples (recombinant *E. coli* lysate and purified P61) on SDS-12% PAGE. lane 1, Uninduced *E. coli* lysate; lane 2, IPTG-induced *E. coli* lysate; lane 3, Ni-NTA purified P61. Arrow indicates the position of expressed protein

or amino acid level with the available database indicates that segment 7 of AmCPV encodes a novel protein and further confirms that a new type of CPV infects tasar silk worm, *A. mylitta* which has been reported earlier [33–35].

Cloning of S7 ORF in bacterial expression vector and analysis of bacterially expressed recombinant Histagged fusion protein showed the production of a protein of 63 kDa, the molecular weight of which is \sim 2 kDa higher than the calculated molecular weight of 61 kDa. This reduction in the mobility of the protein might be due to the presence of high proportions of hydrophobic and acidic amino acids residues such as isoleucine (8.5%), leucine (8.1%), aspartic acid (7.9%), and glutamic acid (6.7%) residues. Anomalous electrophoretic behavior of several proteins such as pig heart calpastatin have been found in SDS-PAGE, which are rich in hydrophilic and acidic amino acid residues [50].

Four potential N-linked glycosylation sites have been found in ORF of P61. When expression of this protein was analyzed in AmCPV infected midgut cells or in purified polyhedra by immunoblot analysis using anti-P61 polyclonal antibodies, an immunoreactive band of approximately 65 kDa was observed. This indicates that the protein has undergone glycosylation during expression in eukaryotic cells utilizing some of the potential glycosylation sites found in segment 7. Production of high titer antibodies (10^{-5}) in rabbit as shown by ELISA (data not shown) and western blot indicate that P61 is highly antigenic to induce strong immune response. The antibodies generated in this study could aid in the development of a rapid and simple diagnostic test to detect AmCPV. Further work is required to demonstrate if this antibody possesses any neutralizing activity to this CPV. Detection of P61 protein expression in virus infected midgut cells and in polyhedra by Immunoblot analysis using anti-P61 antibody indicate that P61 codes for a viral structural protein.

Since a significant similarity of P61 was found with IMPDH enzyme containing one CBS domain by motif scan analysis it may be assumed that P61 may possess IMPDH like activity. It has been shown that IMPDH of human, *T. foetus* and *E. coli* can bind to single stranded nucleic acids with nanomolar affinity through

Fig. 5 SDS-PAGE and Western blot analysis of P61. Samples were separated by SDS-10% PAGE and stained with Coomassie brilliant blue (a) or transblotted onto nitrocellulose membrane and reacted with anti- P61 antibody (b) lane M, Molecular weight markers; lane 1, Uninfected and lane 2, CPV-infected midgut cell lysate; lane 3, Purified AmCPV polyhedra. Arrow indicates the position of immunoreactive protein



its CBS domains (not by its catalytic domain) [51]. CBS domains are also found in chloride channels, ABC transporters and protein kinases where their function is unknown [52] but CBS domain mutations have been implicated in a wide array of inherited diseases including myotonia [53] and familial hypertropic cardiomyopathy [54]. Perhaps nucleic acid binding is a general property of CBS domains and P61 encoded by AmCPV segment 7, by binding to single stranded viral RNA through its CBS domain may help in viral replication or transcription.

Acknowledgments The authors thank the Director(s) of Central Tasar Research and Tranining Institute, Ranchi and Jhargram for providing the virus infected *A. mylitta* larvae. Venkata Ramana Murthy Chavali is the recipient of fellowship from Indian Institute of Technology, Kharagpur and Council of Scientific and Industrial Research, New Delhi.

References

- C.C. Payne, P.P.C. Mertens, in *Cytoplasmic Polyhedrosis* Viruses. The Reoviridae, ed. by W.K. Joklik (Plenum, New York, 1983), pp. 425–504
- M.H.V. van Regenmortel, C.M. Fauquet, D.H.L. Bishop, E.B. Carstens, M.K Estes, S.M. Lemon, J. Maniloff, M.A. Mayo, D.J. McGeoch, C.R. Pringle, R.B. Wickner (eds.), Virus Taxonomy, Seventh report of the International Committee on Taxonomy of Viruses (Academic Press, San Diego, 2000)
- S. Belloncik in Cytoplasmic Polyhedrosis Viruses Reoviridae. Advances in Virus Research, vol 36, ed. by K. Maramorsch, F.A. Murphy, A.J. Shatkin (Academic Press, New York, 1989), pp. 173–209.
- S. Belloncik, K. Mori, in *Cypoviruses, The insect viruses*, ed. by L.K. Miller, L.A. Ball (Plenum, New York, 1998), pp. 337–369
- 5. M. Fouillaud, G. Morel, J. Invertebr. Pathol. 64, 89-95 (1994)
- T. Hukuhara, J.R. Bonami, in *Atlas of Invertebrates*, ed. by J.R. Adams, J.R. Bonami (CRC Press, Boca Raton, Florida, 1991), pp. 394–430
- C.C. Payne in: Pathogenesis of Invertebrate Microbial Diseases, E.W. Davidson, (Eds.). Cytoplasmic polyhedrosis viruses. Allanheld, Osmun, Totowa, New Jersey, (1981), pp. 61–100
- M. Arella, C. Lavallee, S. Belloncik, Y. Furuichi, J. Virol. 62, 211–217 (1988)
- S. Belloncik, J. Liu, D. Su, M. Arella, J. Invertebr. Pathol. 67, 41–47 (1996)
- 10. R. Duncan, Virology 260, 316-328 (1999)
- 11. M.K. Estes, J. Cohen, Microbiol. Rev. 53, 410-449 (1989)
- P. Roy, J.J. Marshall, T.J. French, Curr. Top. Microbiol. Immunol. 162, 43–87 (1990)
- 13. N. Suzuki, Semin. Virol. 6, 89–95 (1995)
- H. Attoui, F. Billoir, P. Biagini, J.F. Cantaloube, de R. Chesse, de P. Micco, de X. Lamballerie, Biochem. Biophys. Res. Commun. 273, 1121–1125 (2000)
- H. Attoui, F. Billoir, P. Biagini, de P. Micco, de X. Lamballerie, J. Gen. Virol. 81, 1507–1515 (2000)

- H. Attoui, F.M. Jaafar, M. Belhouchet, P. Biagini, J.-F.Cantaloube, de P. Micco, de X. Lamballerie, Virology 343, 212–223 (2005)
- N. Nakashima, M. Koizumi, H. Watanabe, H. Noda, J. Gen. Virol. 77, 139–146 (1996)
- S. Rao, M. Shapiro, D. Lynn, K. Hagiwara, R. Dean, G.R. Carner, GenBank Accession Nos. NC_003006 - NC_003015 (2001)
- S.L. Zhao, C.Y. Liang, J.J. Hong, H.G. Xu, H.Y. Peng, Virus Res. 94, 17–23 (2003)
- S.L. Zhao, C.Y. Liang, J.J. Hong, H.Y. Peng, Arch. Virol. 148, 1357–1368 (2003)
- S. Rao, G. Carner, S. Scott, D. Heckel, GenBank Accession Nos. NC_002557-NC_002559, NC_002567, NC_002560-NC_002566 (2000)
- K. Hagiwara, M. Tomita, K. Nakai, J. Kobayashi, S. Miyajima, T. Yoshimura, J. Virol. 72, 5762–5768 (1998)
- K. Hagiwara, M. Tomita, J. Kobayashi, S. Miyajima, T. Yoshimura, Biochem. Biophys. Res. Commun. 247, 549– 553 (1998)
- 24. K. Hagiwara, T. Matsumoto, J. Gen. Virol. 81, 1143–1147 (2000)
- K. Hagiwara, J. Kobayashi, M. Tomito, T. Yoshimura, Arch. Virol. 146, 181–187 (2001)
- K. Hagiwara, S. Rao, W.S. Scott, R.G. Carner, J. Gen. Virol. 83, 1477–1482 (2002)
- K. Ikeda, H. Nakazawa, R. Alain, S. Belloncik, H. Mori, Arch. Virol. 143(2), 241–248 (1998)
- F. Fossiez, S. Belloncik, M. Arella, Virology 169, 462–465 (1989)
- M.S. Galinski, Y. Yu, B.R. Heminway, G.S. Beaudreau, J. Gen. Virol. 75, 1969–1974 (1994)
- F. Echeverry, J. Bergeron, W. Kaupp, C. Guertin, M. Arella, Gene 198, 399–406 (1997)
- A. Shapiro, T. Green, S. Rao, S. White, G. Carner, P.P. Mertens, J.J. Becnel, J. Virol. 79(15), 9430–9438 (2005)
- 32. M.S. Jolly, S.K. Sen, M.M. Ahsan, Tasar Culture. (Ambika publishers, Bombay, 1974)
- K.R. Qanungo, S.C. Kundu, A.K. Ghosh, Acta Virol. 44, 349–357 (2000)
- K.R. Qanungo, S.C. Kundu, J.I. Mullins, A.K. Ghosh, J. Gen. Virol. 83, 1483–1491 (2002)
- Sinha-U. Datta, C.V.R. Murthy., A.K. Ghosh, Biochem. Biophys. Res. Commun. 332, 710–718 (2005)
- 36. Y. Hayashi, F.T. Bird, Can. J. Microbiol. 6, 695-701 (1970)
- F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl (eds.), *Current Protocols in Molecular Biology*. (Wiley, New York, 1995)
- P.R. Lambden, S.J. Cooke, E.O. Caul, I.N. Clarke, J. Virol. 66(3), 1817–1822 (1992)
- S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25, 3389–3402 (1997)
- P. Zhang, N. Briones, C. Liu, C.K. Brush, T. Powdrill, Y. Belosludtsev, M. Hogan, Nucleosides, Nucleotides Nucleic Acids 20(4–7), 1251–1254 (2001)
- 41. B. Rost, C. Sander, Proteins 19, 55-72 (1994)
- 42. J. Garnier, J.-F. Gibrat, B. Robson, Meth. Enzymol. 266, 540–553 (1996)
- 43. A.P. Feinberg, B. Vogelstein, Anal. Biochem. **132**, 6–13 (1983)
- 44. J.E. Caligan, B.M. Dunn, H.L. Ploegh, D.W. Speicher, P.T. Wingfield, *Current Protocols in Protein Science*. (Wiley, New York, 1995)

- 46. U.K. Laemmli, Nature 227, 680-685 (1970)
- E. Harlow D. Lane, Using Antibodies: A Laboratory Manual. (Cold Spring Harbor Laboratory. Cold Spring Harbor, New York, 1988)
- 48. J. Sambrook E.F. Fritsch T. Maniatis, *Molecular Cloning:* A Laboratory Manual, 2nd edn. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989)
- 49. J.T. Patton, E. Spencer, Virology 277(2), 217-225 (2000)
- M. Maki, E. Takano, T. Osawa, T. Ooi, T. Murachi, M. Hatanaka, J. Biol. Chem. 263(21), 10254–10261 (1988)
- J.E. McLean, N. Hamaguchi, P. Belenky, S.E. Mortimer, M. Stanton, L. Hedstrom, Biochem. J. 379, 243–251 (2004)
- 52. A. Bateman, Trends Biochem. Sci. 22, 12–13 (1997)
- 53. C. Ponting, J. Mol. Med. 75, 160-163 (1997)
- E. Blair, C. Redwood, H. Ashrafian, M. Oliveira, J. Broxholme, B. Kerr, A. Salmon, I. Ostman-Smith, H. Watkins, Hum. Mol. Genet. 10, 1215–1220 (2001)