



## Complete Genome Sequence of Transmissible Gastroenteritis Coronavirus PUR46-MAD Clone and Evolution of the Purdue Virus Cluster

ZOLTAN PENZES, JOSE M. GONZÁLEZ, ENRIQUE CALVO, ANDER IZETA, CRISTIAN SMERDOU,  
ANA MÉNDEZ, CARLOS M. SÁNCHEZ, ISABEL SOLA, FERNANDO ALMAZÁN & LUIS ENJUANES\*

*Centro Nacional de Biotecnología, CSIC, Department of Molecular and Cell Biology, Campus Universidad Autónoma,  
Cantoblanco, 28049 Madrid, Spain*

Received January 14, 2001; Accepted March 29, 2001

**Abstract.** The complete sequence (28580 nt) of the PUR46-MAD clone of the Purdue cluster of transmissible gastroenteritis coronavirus (TGEV) has been determined and compared with members of this cluster and other coronaviruses. The computing distances among their S gene sequences resulted in the grouping of these coronaviruses into four clusters, one of them exclusively formed by the Purdue viruses. Three new potential sequence motifs with homology to the  $\alpha$ -subunit of the polymerase-associated nucleocapsid phosphoprotein of rinderpest virus, the Bowman–Birk type of proteinase inhibitors, and the metallothionein superfamily of cysteine rich chelating proteins have been identified. Comparison of the TGEV polymerase sequence with that of other RNA viruses revealed high sequence homology with the A–E domains of the palm subdomain of nucleic acid polymerases.

**Key words:** coronavirus, TGEV, genome, sequence, RNA virus

### Introduction

Transmissible gastroenteritis coronavirus (TGEV) belongs to the *Coronaviridae* family of the *Nidovirales* order [15,17]. TGEV is the prototype of group 1 coronaviruses that includes porcine, canine, feline, and human viruses. TGEV is enveloped and spherical in shape, with an internal core and a helical nucleocapsid [18].

Coronaviruses contain a 27.6–31.3 kb single-stranded positive-sense genomic RNA [15]. The virion RNA functions as a mRNA and is infectious [9]. It contains 7–8 functional genes, 4 or 6 of which (the spike S, membrane M, envelope E, nucleoprotein N, and in some strains an internal (I) open reading frame (ORF) of N gene and the hemagglutinin-esterase (HE)) encode structural proteins [15,35]. In addition, several non-structural proteins are encoded by the coronavirus

genome. The number and location of the non-structural genes vary within coronaviruses of different species. In TGEV the genes are arranged in the order 5'-rep-S-3a-3b-E-M-N-7-3'. Four of them, rep, 3a, 3b, and 7, encode non-structural proteins.

To study the molecular biology of coronaviruses, the recent construction of a cDNA encoding an infectious TGEV RNA [1], the assembly of TGEV genome from six cDNA fragments [72], and the construction of an infectious cDNA clone for human coronavirus (HCoV-229E) [58] will be of great help.

Coronavirus RNA synthesis occurs via an RNA-dependent RNA synthesis process in which mRNAs are transcribed from negative-stranded templates [34,52]. Coronaviruses have transcription regulatory sequences (TRSs) that include a highly conserved core sequence (CS, previously named intergenic sequence [IS]) 5'-CUAAAC-3', or a related sequence, depending on the coronavirus, at sites immediately upstream of most of the genes. Since genes often overlap in the *Nidovirales*, the acronym IS does not seem appropriate

\* Author for all correspondence: Tel.: 34-91-585 4555;  
Fax: 34-91-585 4915; E-mail: L.Enjuanes@cnb.uam.es

in these cases and the acronym CS could reflect the nature of the highly conserved sequence contained within the TRS. These sequences represent signals for the transcription of subgenomic mRNAs [34,52]. Both genome-size and subgenomic negative-strand RNAs, which correspond in number of species and size to those of the virus-specific mRNAs have been detected [54,55]. The two models compatible with most of the experimental data are leader-primed transcription [34] and discontinuous transcription during negative-strand RNA synthesis [53]. Recently, strong experimental evidence supporting the discontinuous transcription during negative-strand RNA synthesis has been reported [3,62]. Also the leader-primed transcription has received additional support [41].

The complete sequence of a coronavirus genomic RNA has been first determined for the avian coronavirus infectious bronchitis virus (IBV) [8]. Since then, several other members of the *Coronavirus* genus have been fully sequenced, including mouse hepatitis virus (MHV) strains A59 [44] and JHM [37], HCoV-229E [26], the TGEV PUR46-PAR strain [13,46], and the bovine coronavirus (BCoV) [71].

TGEV infects both the epithelial cells of the small intestine and the lung cells of newborn piglets, resulting in a mortality of nearly 100%. The Purdue strain of TGEV was isolated for the first time around 1946 by Haelterman's group in the University of Purdue (Lafayette, Indiana) [23,38]. The original virus (PUR46-SW11) was passed exclusively in swine. This virus was adapted to grow in swine testis (ST) cells [6,7] and after 115 passages on this cell line it was cloned and distributed to many laboratories including ours.

During the characterization of one of the oldest *in vivo* passages of the Purdue strain of TGEV (PUR46-SW11) [7,23], we observed that this virulent Purdue strain of TGEV was a mixture of at least two TGEV isolates, with remarkable differences in their *in vivo* and *in vitro* growth [51]. One of them, clone C11, replicated with high titers in the enteric tract and was virulent, while the other one (clone C8) produced low virus titers in enteric tissues and was attenuated.

We report the complete sequence (28,580 nt) of the TGEV PUR46-MAD clone\*, a close relative of PUR46-PAR. The evolution of the Purdue cluster of TGEV, from a highly enteric and virulent strain, to a

\*The nucleotide sequence reported in this paper has been submitted to the GenBank nucleotide sequence database and has been assigned the accession number AJ271965.

clone that does not replicate in the enteric tract of conventional piglets and became attenuated is described. In addition, the sequence identity with other TGEV isolates and potential new sequence motifs identified within the replicase domain are reported.

## Materials and Methods

### *Cells and Viruses*

Viruses were grown in ST cells [39]. The PUR46-SW11 virus is a historical sample of the Purdue strain of TGEV isolated by Haelterman's group [23,38]. It was obtained by passing the first TGEV field isolate 11 times in swine intestine; this virus was kindly provided as a 20% suspension of small intestine cells by Dr. M. Pensaert (Gent, Belgium) [23,38]. From the uncloned virus passaged once in ST cells (PUR46-SW11-ST1), the PUR46-SW11-ST2-C8 (abbreviated PUR46-C8) and PUR46-SW11-ST2-C11 (abbreviated PUR46-C11) clones were plaque-purified [51]. The PUR46-SW11-ST115 was obtained from the PUR46-SW11 by 115 passages in ST cells and was distributed by L. Saif (Ohio State University) to other laboratories, leading to strains PUR46-MAD [31,50] and PUR46-PAR [13,46]. The PUR46-MAD strain was derived from the PUR46-SW11-ST115 strain by five cloning steps in ST cells. The selected clone was named PUR46-MAD in reference to the name of the strain (first three letters), year of isolation (two digits) and the specific clone (last three letters). We have used a similar nomenclature to name other strains derived in different laboratories. The Purdue virus strain NEB72 [50], was renamed PTV (Purdue-type virus) because of its sequence similarity with the PUR46 strain [2]. The PTV clone was probably derived by the passage of a Purdue strain of TGEV in gnotobiotic pigs by the pulmonary route followed by passage in gnotobiotic pig lung cell cultures, and in diploid swine testicular cells with exposure to an acidic (pH 3) environment and incubation with trypsin (M. Welter, Dallas Center, IA).

The original TGEV strains that do not belong to the Purdue cluster have been reported [50].

### *RNA Isolation*

Genomic RNA was extracted from partially purified virus as described [40]. Briefly, ST cells cultivated in roller bottles (500 cm<sup>2</sup>) were infected at MOI 5.

Medium was harvested at 22 h post-infection (hpi) and virions were partially purified as described [31]. The viral pellet was dissociated in 500  $\mu$ l of TNE buffer (0.04 M Tris-hydrochloride pH 7.6, 0.24 M NaCl, 15 mM EDTA) containing 2% SDS, and digested with 50 ng of proteinase K (Boehringer Mannheim) for 30 min at room temperature. RNA was extracted twice with phenol–chloroform and precipitated with ethanol. Cytoplasmic RNA from TGEV infected cells was extracted using a buffer containing urea-SDS and phenol–chloroform [51].

#### *Cloning and Sequencing Analysis*

The complete sequence of the clone PUR46-MAD was assembled starting from the sequence of a 9.7 kb defective minigenome (DI-C) derived from the virus [40]. This defective TGEV genome has three deletions of about 10, 1.1, and 7.7 kb in ORFs 1a, 1b, and after initiation of the S gene, respectively. The sequence of minigenome DI-C, the homologous sequence within the virus genome, and that of the 7.7 kb deletion were obtained using RNAs that were amplified by RT-PCR [40]. The resulting PCR products were cloned into pBluescript (Stratagene), pGEM-T (Promega), pCR2.1 (Invitrogen), or pSL1190 (Pharmacia) using standard procedures [49]. cDNA clones covering most of the genome were sequenced with Sequenase 2.0 (USB) or an ABI 373A automated sequencing machine (Applied Biosystems Inc.).

The TGEV PUR46-MAD 5'- and 3'-end sequences were determined by primer extension using the 5'/3' RACE (Boehringer Mannheim) starting from 0.5  $\mu$ g of cytoplasmic RNA from virus infected cells. The RT-PCR amplification was performed using the primer 801 rs with a reverse sequence from nt 782 to 801 (see complete TGEV sequence). The primer used to sequence the 5'-end was 364 rs (including nt 365–385). The 3'-end sequence was determined using the primer X3.311vs with virus sense sequence from nt 28,381 to 28,400. The presence of two consecutive 'C' at position 20,347 was assessed by digestion of the cDNA with the BstII restriction endonuclease.

The core sequence was obtained by characterizing at least three clones of independent origin. Sequence data were compiled using the Wisconsin Package software Version 9.1 – UNIX, Genetics Computer Group (GCG) (Madison, Wisconsin). Sequences obtained were compared to those of previously published TGEV strains [13,32,40,46,50]. Sequence differences

were confirmed by sequencing three independently derived RT-PCR clones or by direct viral RNA sequencing [19].

#### *Sequence Comparison and Motif Identification*

Sequence comparison was made by using the Wisconsin Package software version 9.1 – UNIX. The pairwise distances within the group of aligned sequences were obtained using the Jukes–Cantor program of the GCG. The identification of sequence motifs was done with the Psi-Blast program using the Swiss-Prot database available through the European Bioinformatics Institute. Sequences were aligned using the Clustal W sequence alignment program for DNA and proteins [27,59].

## **Results**

#### *Complete Sequence of the TGEV PUR46-MAD Strain*

The complete sequence of the PUR46-MAD genome has been determined and it was comprised of 28,580 nt without the poly(A) tail. The 5' two-thirds of this RNA genome (20,368 nt) encode the viral RNA-dependent RNA replicase, while the structural genes are located at the 3'-end of the genome (8,214 nt). It is assumed that the PUR46-MAD RNA has a 5' terminal cap by analogy with other coronavirus genomes [34]. The viral RNA starts with the sequence 5'-ACUUUUAAAG-3', as determined by 5' extension. At the 3'-end the TGEV genome has a poly(A) tail of unknown length.

#### *Evolution of the Purdue Virus Cluster*

The Purdue virus cluster (Table 1) is defined as a set of viruses closely related in sequence, that are derived from the original PUR46-SW11 strain of TGEV. The sequence differences among these viruses are shown (Fig. 1) in relation to the sequence of the PUR46-MAD, the prototype strain of our laboratory. The Purdue virus cluster includes two clones that were isolated from the original *in vivo* virus stock (virulent PUR46-C11 and attenuated PUR46-C8), clone PUR46-MAD (passaged 120 times in ST cells) with reduced replication in the enteric tract and partially attenuated, and clone PTV that does not replicate within the gut of conventional piglets and is fully attenuated (Table 1).

Table 1. Characteristics of the TGEV Purdue virus cluster

Virus	Simplified names	Tropism <sup>a</sup>		Virulence
		Growth in respiratory tract (Pfu/g tissue)	Growth in enteric tract (Pfu/g tissue)	
PUR46-SW11-ST2-C11	PUR46-C11	10 <sup>6</sup>	10 <sup>7</sup>	Virulent
PUR46-SW11-ST2-C8	PUR46-C8	10 <sup>6</sup>	10 <sup>3</sup>	Part. attenuated
PUR46-MAD-ST120	PUR46-MAD	10 <sup>6</sup>	10 <sup>3</sup>	Part. attenuated
PUR46-PTV-ATT	PUR46-PTV	10 <sup>6</sup>	0	Fully attenuated

<sup>a</sup>Growth of TGEV in conventional, colostrum fed swine.

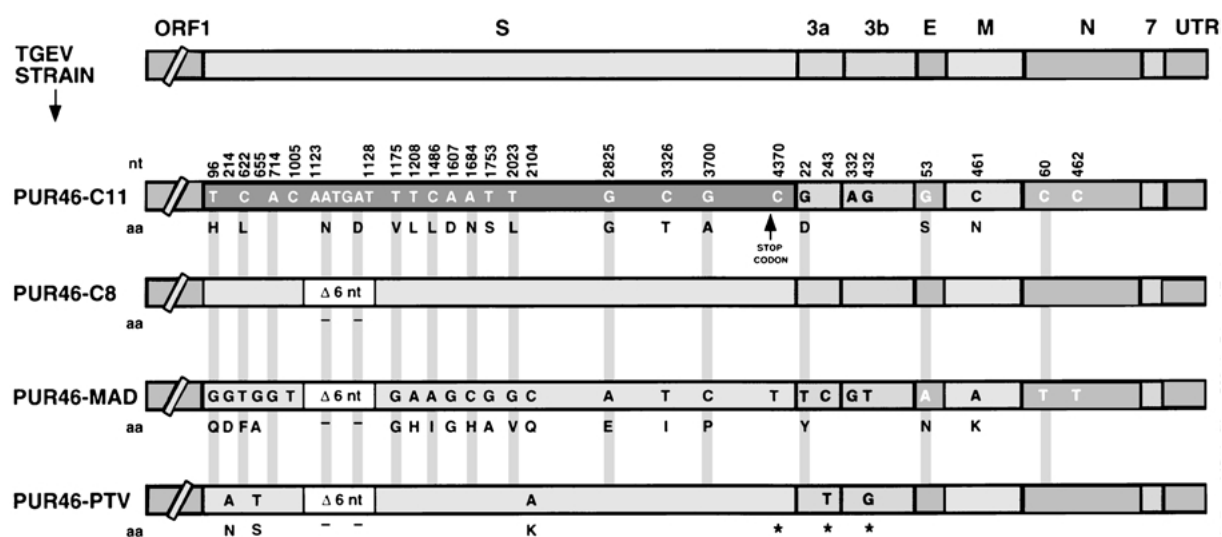


Fig. 1. Nucleotide sequence comparison between members of the Purdue virus cluster. The nucleotide (inside bars) and amino acid (below bars) substitutions at the 3'-end 8.2 kb of four members of the Purdue cluster are indicated in relation to the PUR46-MAD clone (only the differences are highlighted). The viruses are organized from low to high passage number. The approximate location of the different genes (top bar) and the location of the nucleotide substitutions (above second bar) are indicated. Residue numbers are provided in relation to the 'A' of the initiation codon of each gene except ORFs 3a and 3b nucleotide numbers that both refer to the initiation of ORF 3a. S gene numbers refer to the sequence of the PUR46-C11 clone which has an insertion of six nucleotides in relation to the sequence of the PUR46-MAD clone. The origin of the sequences used is indicated in the Material and Methods section. \* denotes nucleotide changes in non-coding regions. Vertical shading is provided to facilitate alignment, nt, nucleotide position, aa, amino acid. Stop codon, end of S gene.

The link between these cluster members is their passage history [51] or their sequence identity within the 3'-end 8,214 nt (Fig. 1). PTV only has 5 nt changes within the 3'-end 8.2 kb in comparison to the PUR46-MAD clone (Fig. 1). This accumulation of nucleotide substitutions represents 0.57 nt changes per one thousand nucleotides, much lower than the 2.5 per one thousand nucleotides accumulated between the PUR-C8 and PUR-C11 clones.

The 3'-end of the PUR46-MAD genome has complete sequence identity with clone C8. Comparison of

the 3'-end 8.2 kb sequences of clones PUR46-C11 and PUR46-C8 revealed 22 nt differences, 14 of them in the S gene (Fig. 1). Three of these nucleotide substitutions were in non-coding regions, one downstream the S gene stop codon (nt S-4370) and upstream the 3a gene, and two on the 3b gene (nts 3b-332 and 3b-432). The other nucleotide substitutions were scattered through the other 3'-end genes. In addition, there was a 6 nt deletion in the PUR46-C8 clone. This deletion has been considered a trade mark of all TGEV Purdue strains since it is present in all Purdue

	PUR46 -C11	PUR46 -C8	PUR46 -MAD	PUR46 -PAR	TOY56	MIL65	BRI70	TAI83	FRA86	ENG86	HOL87
PUR46-C11	0.00	0.35	0.35	0.46	1.42	1.39	2.07	2.74	3.17	3.08	2.98
PUR46-C8		0.00	0.00	0.16	1.49	1.61	2.20	2.91	3.20	3.17	3.06
PUR46-MAD			0.00	0.16	1.49	1.61	2.20	2.91	3.26	3.17	3.06
PUR46-PAR				0.00	1.56	1.68	2.27	2.98	3.31	3.23	3.11
TOY56					0.00	1.51	2.10	2.64	3.28	3.19	3.19
MIL65						0.00	0.95	1.79	2.29	2.21	2.32
BRI70							0.00	1.74	2.29	2.24	2.35
TAI83								0.00	2.83	2.80	2.94
FRA86									0.00	0.68	0.96
ENG86										0.00	0.82
HOL87											0.00

Fig. 2. Computing distances among the S genes of TGEVs and PRCoVs. The pairwise distances within the group of aligned sequences were calculated by the Jukes–Cantor methods of the GCG. The complete sequence of the S gene was used to compute the distances except for the MIL65 strain. In that case, the first 2,230 nt were used. Viruses with S genes with close computing distance values have been grouped and enclosed within the same box. The origin of the sequences used is indicated above. The name of the viruses is composed of three letters related to their geographical origin or classical name, followed by two numbers indicating the year of isolation, and a code that refers to the particular clone. PUR46-C11, PUR46-C8, PUR46-MAD, and PUR46-PAR are different clones of the Purdue cluster of TGEVs. TOY56, MIL65, BRI70, and TAI83 are other strains of TGEV. FRA86, ENG86, and HOL87 are different PRCoV strains.

isolates sequenced except the parental PUR46-C11 clone [10,11,46,47,50,67].

The sequences of the S genes from PUR46-C8 and PUR46-C11 clones were compared with those of the S genes from other nine TGEV strains, by computing the distances among their S genes using the Jukes–Cantor method. The results indicated that the 11 virus isolates could be grouped into four clusters according to their sequence homology (Fig. 2). These clusters had increasing computing distances with viruses of the PUR46 cluster and with the TOY56, ranging between 0.0–0.5, 1.3–1.7, 2.0–2.98, and 2.98–3.4, and were formed by the isolates: (i) Purdue-type viruses (PUR46-C11, PUR46-C8, PUR46-MAD, and PUR46-PAR); (ii) TOY56 and MIL65-AME; (iii) BRI70 and TAI83, and (iv) Porcine respiratory coronavirus (PRCoV) strains FRA86-RM4, ENG86-II, and HOL87, respectively. This organization of TGEVs into clusters matches the previously reported evolutionary tree [50].

The PUR46-MAD and the PUR46-PAR have similar virulence. Both clones are attenuated in colostrum-fed swine and virulent in colostrum-deprived animals [2,4,21,51]. PUR46-MAD replicates to a limited extent within the enteric tract (between  $10^2$  and  $10^3$  pfu/gram

of tissue), and causes the death of two-day-old newborn piglets ( $LD_{50} = 1 \times 10^4$  pfu/animal). The PUR46-PAR clone was the first TGEV strain completely sequenced [13]. The 29 nt substitutions detected between PUR46-MAD and PUR46-PAR clones are responsible for 14 amino acid (aa) changes (Table 2). On some occasions, these changes represented insertions or deletions. One of these changes was a nucleotide (nt 20,347) deletion in the PUR46-PAR that led to a frame shift located in a region close to the end of ORF 1b and two nucleotide differences (one insertion and one deletion in the PUR46-MAD) in the non-coding region at the 3'-end of the genome (nt 28,331 and 28,440), respectively (Table 2). Within the region that encodes the structural proteins at the 3'-end of the genome (nts 20,365–28,580), 12 nt differences were found, five of which resulted in amino acid changes (Table 2).

#### TGEV Genome Organization

The nine ORFs identified in the TGEV genome (PUR46-MAD clone) are summarized (Table 3). The first 93 nt of the TGEV sequence correspond to the leader, defined as the motif preceding the first CS

Table 2. Sequence differences between PUR46-MAD and PUR46-PAR RNAs

Position nt	PUR46-MAD	PUR46-PAR	Amino acid change	Amino acid position
2,029	T	C	Ser→Phe	1a-572
2,609	T	C	Asn	1a-765
3,437	A	C	Asp→Glu	1a-1041
6,926	C	T	Tyr	1a-2207
7,437	C	A	Thr→Pro	1a-2375
7,455	G	C	Gln→Glu	1a-2381
7,478	T	C	Gly	1a-2388
11,501	C	T	Val	1a-3729
13,549	G	A	Lys	1b-404
14,812	G	A	Leu	1b-825
16,139	C	G	Ala→Pro	1b-1268
18,473	G	A	Ile→Val	1b-2046
19,575–76	AT	TA	Val→Asp	1b-2413
19,591	T	G	Lys→Asn	1b-2418
19,592	G	T	Phe→Val	1b-2419
20,347	CC	C	frame shift	1b-2670
20,578	G	A	Asn→Asp	S-72
22,480	C	A	Ile→Leu	S-705
22,551	C	T	Ile	S-729
23,244	G	A	Glu	S-960
25,138	G	T	—	—
25,258	T	G	—	—
26,699	G	A	Asp→Gly	M-195
26,704	A	G	Val→Met	M-197
28,043–44	TA	AT	Asn→Ile	N-376
28,331	T	—	—	—
28,440	—	A	—	—

5'-CUAAAC-3'. The CS is afterwards repeated along the genome at different nucleotide distances (3–37 nt) from the first codon (AUG) of each gene (Fig. 3A). In addition, there is another 5'-CUAAAC-3' sequence 120 nt after the first initiation codon of the S gene. In principle, this CS could be responsible for the synthesis of a mRNA that has not been detected, although its size similarity with that of the S gene could have prevented its identification (S. Alonso, I. Sola, and L. Enjuanes, unpublished data).

Transcription in coronavirus requires the discontinuous synthesis of the mRNAs in order to link the leader to the coding sequences of each mRNA. This process requires a complementarity between the sequences downstream of the 3'-end of the leader and the sequences flanking the complement of the CS (cCS) in the negative strand [34,52,62]. The extent of this complementarity could regulate transcription and was calculated for the TGEV PUR46-MAD strain using

Table 3. PUR46-MAD sequence features

Feature	Start nt	Stop nt	Start aa	Stop aa
<i>Open reading frame</i>				
ORF1a	315	12,368		
ORF1b	12,338	20,368		
ORF2, S	20,365	24,708		
ORF3a	24,827	25,042		
ORF3b	25,136	25,870		
ORF4, E	25,857	26,105		
ORF5, M	26,116	26,904		
ORF6, N	26,917	28,065		
ORF7	28,071	28,307		
<i>Consensus sequence</i>				
cs <sup>a</sup>	94	99		
CS, S	20,333	20,338		
cs, S <sup>b</sup>	20,485	20,490		
CS, 3a	24,798	24,803		
CS, 3b <sup>c</sup>	25,119	25,124		
CS, E	25,814	25,819		
CS, M	26,107	26,112		
CS, N	26,905	26,910		
CS, 7	28,062	28,067		
<i>Replicase domain</i>				
ORF1a, RVPh	3,123	3,551	937	1,079
ORF1a, Papain proteinase	3,552	4,133	1,080	1,273
ORF1a, Papain proteinase	5,037	5,624	1,575	1,770
ORF1a, BBPI	6,594	6,782	2,094	2,156
ORF1a, 3C-like proteinase	8,943	9,851	2,877	3,179
ORF1a, GFL	11,898	12,329	3,862	4,005
ORF1a, Mth	12,117	12,311	3,935	3,999
Ribosomal slip site (RSS)	12,332	12,338		
Pseudoknot	12,342	12,409		
ORF1b, Pol	13,925	14,833	4,538	4,840
ORF1b, MIB	15,095	15,322	4,928	5,003
ORF1b, Hel	15,929	16,228	5,206	5,305
ORF1b, VD	18,827	19,006	6,172	6,231
ORF1b, CD	19,136	20,080	6,275	6,589

<sup>a</sup>CS, consensus sequence 'CUAAAC'.

<sup>b</sup>There is no experimental evidence that this canonical CS is used.

<sup>c</sup>This CS has the sequence CUAAAU, i.e., it has the sixth nucleotide mutated to 'U' in relationship to the canonical CS.

two procedures: by computing the complementary nucleotides in an uninterrupted segment of sequence around the CS, or by calculating the total number of complementary nucleotides for a sequence segment including the 6 nt of the CS and 12 nt flanking both the 5'- and the 3'-ends of the CS (30 nt total) (Fig. 3B). The amount of each mRNA produced after infection with the PUR46-MAD strain, as determined by Northern blot analysis with a probe specific for the 3'-end of the genome (results not shown) was

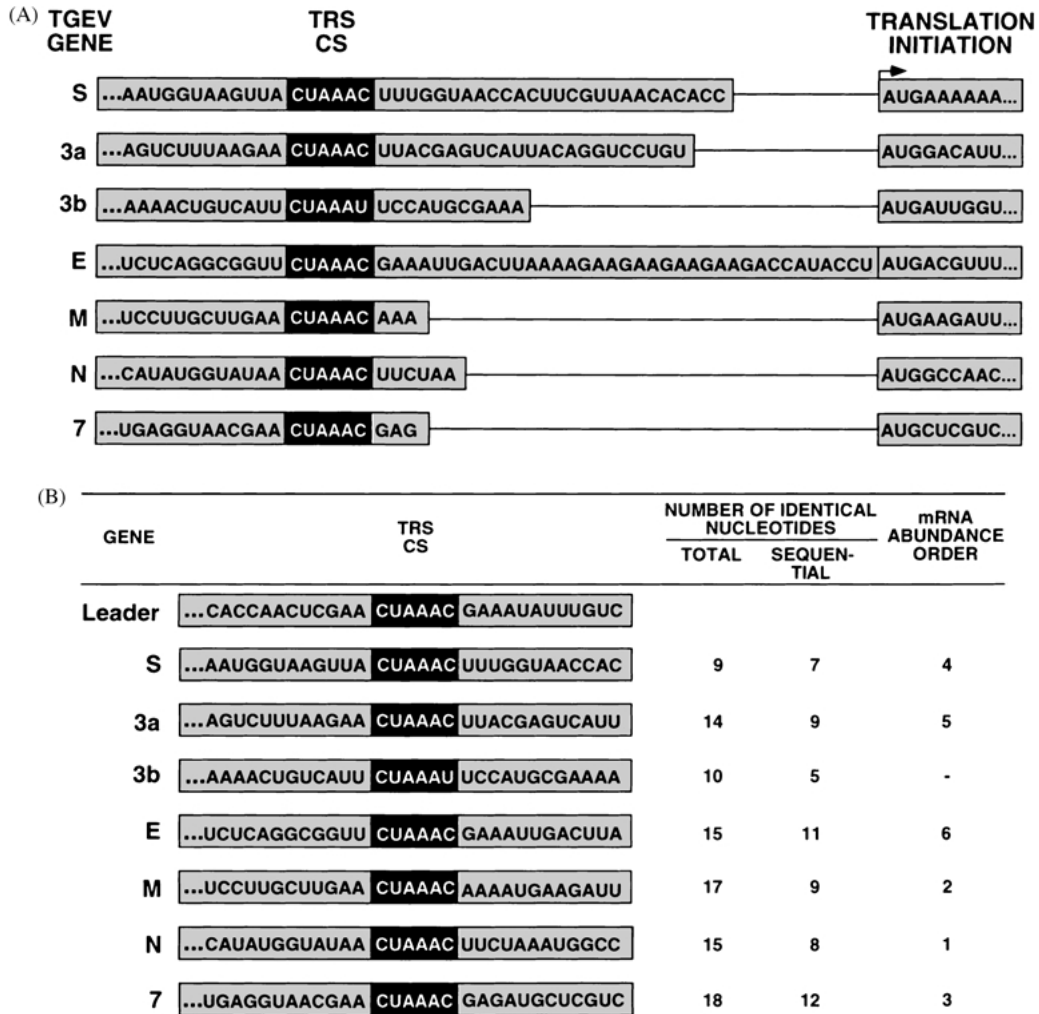


Fig. 3. Sequences flanking the core sequence of each TGEV PUR46-MAD clone gene. (A) Preceding each gene of PUR46-MAD clone, the core sequence (CS) 5'-CUAAAC-3' (black boxes) is present at different distances from the initiation of the translation except in gene 3b, in which the second 'C' has been replaced by a 'U'. The CS is a domain of the TRS that has a weakly defined size. The name of the corresponding virus gene is indicated to the left of each bar. (B) Sequences of 30 nt including the CS plus 12 nt flanking 5' upstream and 12 nt downstream of the CS, present at the 5'-end of each PUR46-MAD virus gene, were aligned with the 3'-end of the leader. The number of identical nucleotides in an uninterrupted sequence segment, or within all the 30 nt compared, is indicated in the columns under the headings sequential or total, respectively. Numbers in the third column indicate the abundance order of the corresponding mRNA (numbers 1 and 6 representing the most and the least abundant mRNA, respectively), determined by integrating the mRNA bands observed in a Northern blot analysis, using a <sup>32</sup>P-labeled probe specific for the 3'-end of the genome (data not shown). Numbers and letters to the left of each bar indicate gene name.

not related to the extent of the potential basepairing (Fig. 3B).

The largest mRNA is the genomic RNA that also serves as the mRNA for ORF 1a and 1b. The remainder are subgenomic mRNAs designated mRNA 2-7 (with the exception of the mRNA 3-1 corresponding to ORF 3b), in the order of decreasing size, encoding ORFs 2 (S), 3a, 3b, 4 (E), 5 (M), 6 (N), and 7 (Table 3).

In the PUR46-MAD clone of TGEV, and in the other Purdue strains, the CS corresponding to the ORF3b has the sequence 5'-CUAAAU-3' where the 'C' in the last position of the CS is replaced by a 'U'. Consequently, mRNA 3-1 encoding gene 3b was not observed [30]. In contrast, this RNA has been detected in cells infected with the MIL65 strain of TGEV which has a standard CS in the homologous position [67].

A potential internal ORF starting at amino acid 77 is observed within the N gene. This ORF is within the same frame as the full-length N protein (383 aa) and could lead to a potential truncated N protein of 306 aa with an estimated molecular mass of 35 kDa. A truncated N protein with an estimated molecular mass of around 41 kDa, instead of 44 kDa of the full-length protein, has been regularly observed by Western blot analysis in TGEV infected ST cells using N specific monoclonal antibodies (results not shown). This band is larger than the one expected for the truncated protein associated to a potential internal initiation of translation and possibly corresponds to a protease cleaved product (see below).

#### Predicted Domains in TGEV ORF 1a–1b

The precise location of PUR46-MAD ORF 1a–1b predicted motifs (Table 3) and their distribution along the genome is indicated (Fig. 4). These include already described motifs such as two papain-like proteinase domains (PL1 and PL2), a 3C-like (3CL) protease domain, a growth factor-like (GFL) domain, the ribosomal slippage site 5'-UUUAAAC-3' (RSS), the pseudoknot (PKnt), the polymerase (Pol), metal ion binding domain (MIB), helicase (Hel), ORF 1b variable domain (VD), and a conserved domain (CD) [13].

In addition, we have identified three potential new domains (Figs. 4 and 5) showing variable sequence homology with other sequences: (i) 28% (41/148) amino acid identity with a phosphoprotein of rinderpest virus (RVPh). This protein has 507 aa and is probably a component of the active RNA-directed RNA polymerase alpha-subunit that may function in template binding [69] (Fig. 5A); (ii) 30% (15/49) amino acid identity with the invariant active site (core region) of the W1P1 Bowman–Birk serine proteinase inhibitor (BBPI) described in plants, and significant identity with other BBPI proteinases [42,48]. These proteins have 102 aa including seven highly conserved cysteine residues. Interestingly, four of these residues are also conserved within the TGEV replicase sequence (Fig. 5B); and (iii) 25% (18/72) amino acid identity with LeMTA metallothionein-like protein (MTh) of plants and of significant identity with other MTh [68]. Of the 72 aa that represent the full-length of this metallothionein, 14 are cysteines and 7 of them are also conserved in the TGEV motif (Fig. 5C). Further work needs to be done to determine whether TGEV would have the activities potentially encoded by the identified domains.

Five motifs (A–E) have been defined in the palm subdomain of nucleic acid polymerases [24]. The amino acid sequence of the TGEV RNA polymerase

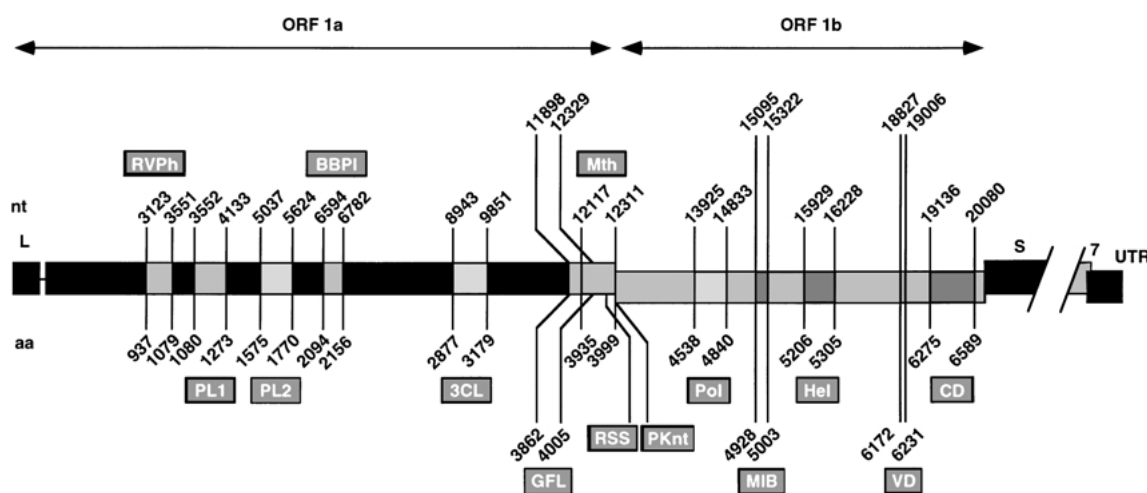


Fig. 4. Schematic representation of sequence domains identified along the PUR46-MAD sequence. These domains include: PL1 and PL2, 3CL protease domain, GFL domain, RSS 5'-UUUAAAC-3', PKnt, Pol, MIB domain, Hel, ORF 1b VD, and a CD [13]. In addition, new domains showing sequence homology (Fig. 5) with a RVPh, a Bowman–Birk type serine proteinase inhibitor (BBPI), and a metallothionein-like protein (MTh) are also indicated. The predicted biological activity has not been experimentally proven. The position of the first and last nt or aa of each domain within the virus sequence is shown.



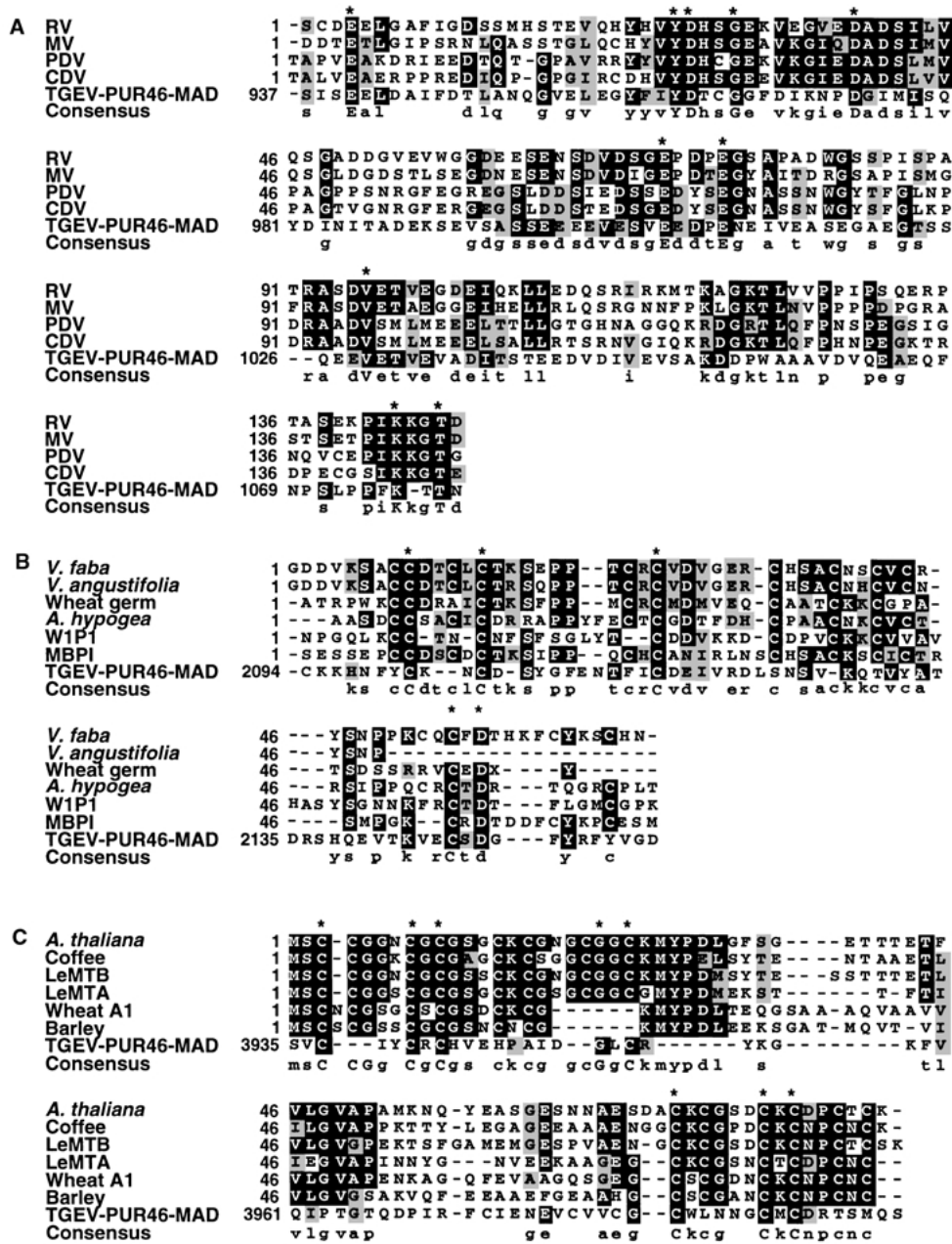


Fig. 5. Predicted similarity of PUR46-MAD sequence with other functional proteins. Alignment with an alphavirus phosphoprotein (A), with Bowman-Birk type proteinase inhibitors (B), and with a protein belonging to the metallothionein family (C) are shown. Alignment with a fragment of rinderpestvirus (RV) phosphoprotein results in a 28% aligned score (A), in contrast with other phosphoproteins of phocid distemper virus (PDV), canine distemper virus (CDV), and measles virus (MV), where the aligned scores are 12%, 10%, and 16%, respectively. The sequences in (A) were previously reported [69]. The sequences in (B) for *Vicia faba*, *Vicia angustifolia*, wheat germ, *Arachis hypogea*, wound-induced protein from maize (W1P1), and mung bean proteinase inhibitor (MBPI) were previously reported [42,48]. Sequences (C) for *Arabidopsis thaliana*, coffee, *Lycopersicon esculentum* L. metallothionein (LeMT), wheat A1, and barley were reported [68]. Black and gray boxes indicate identity or similarity, respectively, with the corresponding residue in other sequences. Complete residue identity in all included sequences is denoted with an asterisk. Domain prediction was performed using the Psi-Blast program and the sequences were aligned using the ClustalW program. The number to the left of each sequence indicates the amino acid aligned or the amino acid within the replicase polyprotein (TGEV-PUR46-MAD).

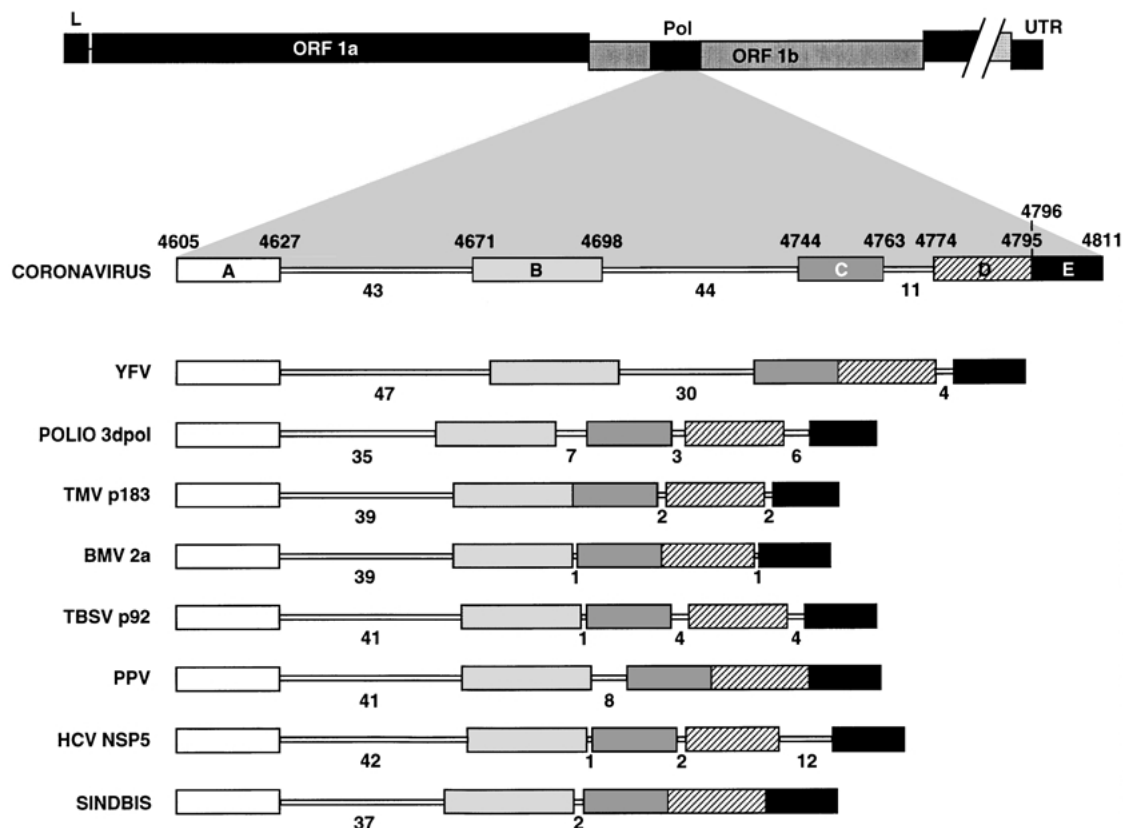


Fig. 6. Comparison of the coronavirus polymerase sequence with that of other RNA viruses. The general organization of the different palm subdomain polymerase motifs is shown, indicating the beginning and termination of the previously defined A, B, C, D, and E motifs [24,43]. Yellow fever virus (YFV), tobacco mosaic virus (TMV), brome mosaic virus (BMV), tomato bushy stunt virus (TBSV), plum pox virus (PPV), human hepatitis C virus (HCV), and Sindbis, Sindbis virus. L, leader. Pol, polymerase. UTR, 3' untranslated region. Large boxes, palm subdomain polymerase motifs. Numbers below thin bars between large boxes indicate the length in amino acids of the sequences linking the motifs. The first and last amino acids of each motif are indicated above the second bar. Motifs A, B, C, D, and E can be identified in the different viruses by the box shading.

was compared to that of other coronaviruses and positive strand RNA viruses and similar domains have been identified in the coronavirus polymerases (Figs. 6 and 7). An interesting difference between the TGEV and other coronaviruses, in relation to polymerases of other RNA viruses, is the presence of a 44 aa linker sequence between B and C motifs in coronaviruses. This is in contrast to a 1–8 aa linker present in other RNA virus polymerases analyzed, except in the yellow fever virus (YFV) with a linker of 30 aa (Fig. 6).

Motif A of TGEV polymerase shows significant homology with the A motif of other positive RNA viruses (Fig. 7). All of these viruses maintain the conserved amino acids D4613 and D4618 of the catalytic site. TGEV motif B has the highest homology with other positive strand RNA viruses with identical amino

acids in the highly conserved positions S4677, G4678, T4682, and N4686 (Fig. 7 Motif B). The coronavirus motif C, relevant in copy fidelity, includes the SDD (aa 4,754–4,756) sequence in substitution to the classic GDD conserved in all positive strand RNA viruses that have been studied. Motifs D and E are less conserved between coronaviruses and other positive strand RNA viruses.

## Discussion

The complete sequence of the PUR46-MAD clone has been determined and its relation with other members of the Purdue cluster of viruses and with other coronaviruses has been defined. In addition, the role of

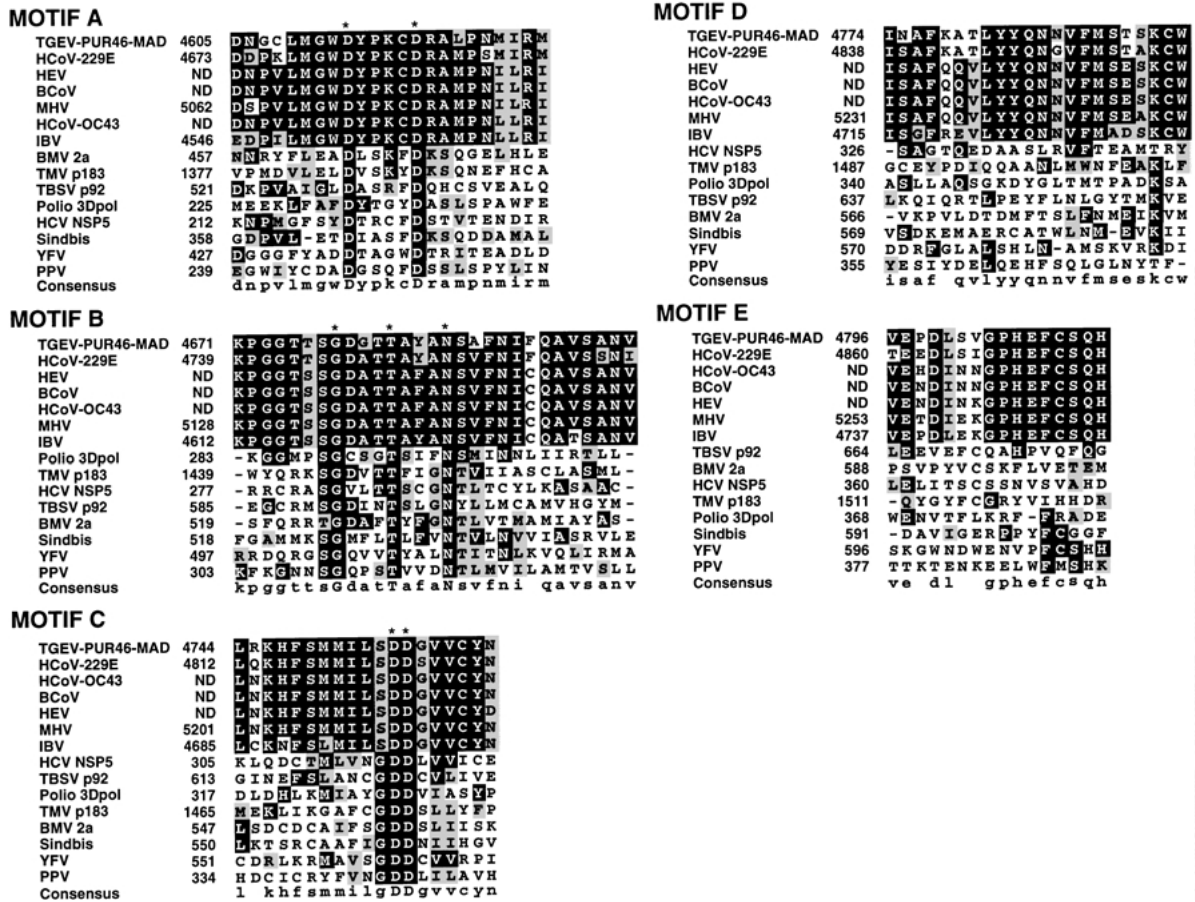


Fig. 7. Alignment of the coronavirus polymerase palm subdomain motifs with the corresponding motifs of other RNA viruses. Five motifs (A–E) have been defined in nucleic acid polymerases [24,43]. The amino acid sequence of the TGEV RNA polymerase motifs is shown in comparison with those of other coronaviruses and of positive strand RNA viruses. The organization of motifs was generated to obtain maximum alignment of highly conserved amino acids in the case of A, B, and C motifs, and it was based on motif length and position of highly conserved residues in the case of D and E motifs with limited homology. Multiple sequence alignments were performed using the ClustalW program [60]. Black and gray boxes indicate identity or similarity, respectively, with the corresponding residue of the other viruses. The sequences included within the alignments are from: (i) TGEV, this manuscript; (ii) HCoV-229E, HEV, BCoV, MHV, HCoV-OC43, IBV [57]; (iii) Polio 3Dpol, TMV p183, HCV NSP5, TBSV p92, BMV 2a [43], and YFV and PPV [22], or from references cited within these publications. HCoV, human coronavirus; HEV, porcine hemagglutinating encephalomyelitis virus; BCoV, bovine coronavirus; MHV, mouse coronavirus; IBV, infectious bronchitis virus; other acronyms as in Fig. 6. Amino acid positions are provided in relationship to the first amino acid of the viral replicase in the case of the coronavirus sequences. Complete residue identity is denoted with an asterisk. ND indicates that the number of the first amino acid is not known because the complete sequence of the virus is not available.

the complementarity between the 3'-end of the leader and the CS has been analyzed, and three new potential sequence motifs have been identified along the replicase gene.

*Evolution of the Purdue Cluster of TGEV*

The first nucleotide of PUR46-MAD clone was an A, coinciding with the 5' sequence of the PUR46-PAR

clone [13]. Interestingly, when synthetic TGEV minigenomes were cloned behind T7 bacteriophage promoter [30] or after cytomegalovirus (CMV) promoters [45], where the first engineered viral nucleotides were a 'C' or an 'A', respectively, the synthetic minigenomes were replicated by the helper virus, indicating that the nature of the first nucleotide, at least for minigenome rescue, was not absolutely critical.

The length of the poly(A) tail at the 3'-end of the TGEV genome is not accurately known. Nevertheless, minigenomes or a full length infectious RNA with a poly(A) of 24 residues have been constructed which are efficiently replicated, indicating that 24 residues are enough for TGEV RNA replication [1,30]. In fact, MHV minigenomes with 5-, 10-, and 68-nt poly(A) tails were replicated during BCoV infection [56]. Poly(A) tails of larger length (100–130 nt) have also been detected in coronaviruses [28,33,70]. Coronavirus poly(A) tail is essential for virus replication [56], but the summarized data suggest that there is a high flexibility with the length of this poly(A).

The different members of the Purdue virus cluster (Table 1) are closely related. Two of them, PUR46-C8 and PUR46-C11, were isolated from the same animal. These clones seem to have evolved through the accumulation of nucleotide substitutions and a small deletion. A comparison of the S gene sequences among eleven TGEV isolates (Fig. 2) showed that clone PUR46-C11 had the lowest computing distance (0.35) with clone PUR46-C8, while the computing distances with other TGEVs such as the MIL65 strain and with the PRCoVs were higher than 2.0 and 3.0, respectively. These data strongly suggest that clone C8 is derived from C11, and not from other viruses circulating at the same time and geographical area, such as the MIL65 strain isolated in Fredericksburg, Ohio [5] (R.D. Wesley, Personal communication). The PUR46-C11 clone could be a recent ancestor of the MIL65 strains of TGEV according to the epidemiological tree previously described [50].

The computing distances between the PRCoV isolates (FRA86, ENG86, and HOL87) were higher with the members of the Purdue cluster of viruses than with MIL65 and BRI87, suggesting that the PRCoVs were more likely derived from strains related to MIL65 or BRI87 TGEVs.

A surprising observation was the high conservation of the RNA sequence of the PUR46-MAD virus upon passage on ST cells, since almost one-third of its genome (8,221 nt), that encodes all the structural and three small non-structural proteins, has a complete sequence identity with the PUR46-C8 clone, with only two passages on the same ST cell line. This sequence identity may indicate that the selected virus has a highly favored sequence to grow in ST cells. In contrast, within the full-length PUR46-PAR genome that was passaged in a different cell line (PD-5 cells) [46],

the estimated number ( $1 \times 10^{-4}$ ) of nucleotide substitutions per nucleotide and replication cycle, in relation to the PUR46-MAD was higher and within the expected range for a RNA virus genome [12].

#### *Organization of TGEV Genome*

The amount of each mRNA produced after infection with the PUR46-MAD strain was not proportional to the extent of the potential basepairing, indicating that in TGEV mRNA abundance is not exclusively regulated by the complementarity between the sequences at the 3'-end of the leader and the sequences complementary to the TRSs in the negative RNA strand, in agreement with previous observations in MHV [34,61]. In addition, although the mRNAs closer to the 3'-end of the genome are in general more abundant, the relative amount of each mRNA did not precisely correlate with the proximity of each mRNA leader to the 3'-end of the virus genome, in contrast to what has been described in other positive-strand RNA viruses [20,61] and also in the negative-strand ones [25,29,64]. These results suggest that, in addition to basepairing between the 3'-end of the leader and the TRS complementary sequences [cTRSs], transcription in coronaviruses may be controlled by other viral and cellular factors, including TRS primary and secondary structure. In fact, it has been suggested that the discontinuous transcription that takes place during mRNA synthesis, is probably mediated through the interaction of proteins with both the 3'-end of the leader and with the cTRS, and then, by the binding between these proteins [34,62]. This protein-RNA interaction most likely requires the recognition of a RNA-TRS primary and secondary structure larger than the CS.

The presence of ORFs 3a or 3b in different TGEV strains is variable [2,10,16,36,63,67]. TGEV strains, such as the MIL65, express both ORFs 3a and 3b [65]. In contrast, other strains such as a small plaque (SP) mutant of the MIL65 strain, express none of these ORFs [66]. All these strains infect swine implying that ORFs 3a and 3b are non-essential for virus growth in tissue culture or *in vivo*, facilitating the loss of ORF 3b during the passage of the TGEV Purdue strains or the PRCoV isolates.

The truncated N protein, with an estimated molecular mass of around 41 kDa instead of 44 kDa of the full-length protein, regularly observed by Western blot analysis in TGEV infected ST cells, most likely

corresponds to a caspase-mediated cleavage induced during the apoptosis of TGEV infected cells as previously reported [14]. It has been shown that the N protein sequence VVPD359 located 23 aa residues upstream of the carboxy-terminal end of the N protein is cleaved leading to the apparition of a shorter form of N protein in infected cells. The observed sequence is also present in other coronavirus N proteins, including the PRCoV. This protein is not found in the purified virions [14].

### Sequence Motifs

Polymerase motif C showed that coronaviruses had the DD sequence conserved as in the RNA-dependent DNA polymerases of retroviruses and in RNA-dependent RNA polymerases of double-stranded RNA and segmented (–) strand viruses. But, in contrast to these viruses, coronaviruses have the SDD motif instead of the more common GDD one [24,43,57]. The high conservation among the Groups 1, 2, and 3 coronavirus polymerase domains in relation to other positive strand RNA viruses, and the conservation of additional replicase domains, for example, the carboxy-terminal ORF 1b domain for which no homologue can be found in the other viral replicases, clearly indicates that the *Nidovirales* replicases are more related to each other than to any other group of positive-stranded RNA viruses [17]. The longer linker (44 aa) identified between the polymerase palm subdomain motifs B and C will also support the grouping of coronavirus polymerases as a subset within the positive-stranded RNA viruses. Motif B of the coronavirus polymerase sequence is also more closely related to the poliovirus polymerase than to the homologous domain of the other viruses analyzed.

Three new potential domains have been identified in the TGEV replicase showing limited amino acid homology with the  $\alpha$ -subunit of the polymerase-associated nucleocapsid phosphoprotein of rinderpest virus, the Bowman–Birk type of proteinase inhibitors, and the metallothionein superfamily of cysteine rich chelating proteins [48,68,69]. We think that the sequence identities observed are possibly significant because of the number of conserved residues and, at least in the cases of the BBIP and metallothioneins, due to the highly conserved cysteine residues, generally relevant to protein structure and function. Nevertheless, the limited sequence homology observed does not imply that these domains will provide the virus with

the corresponding activities. The role of these domains in TGEV replication is being investigated.

### Acknowledgements

We would like to thank the helpful comments of Dr. Alexander Gorbalenya on sequence motifs. This research was supported by grants from the Comisión Interministerial de Ciencia y Tecnología (CICYT, Spain), the Dirección General de Investigación (Community of Madrid), the European Union (Key Action 2. Control of Infectious Diseases Programs), and Fort-Dodge Veterinaria S.A. (Spain). Z.P., C.S., F.A., and I.S. received fellowships or contracts from European Union Control of Infectious Diseases Programs. J.M.G. received fellowships from the Spanish Department of Science and Technology. A.I. received fellowships from the Education Department of the Basque Government. E.C. received a fellowship from the Fundación Ramón Areces, Madrid.

### References

- Almazan F., González J.M., Péñzes Z., Izeta A., Calvo E., Plana-Durán J., and Enjuanes L., *Proc Natl Acad Sci USA* 97, 5516–5521, 2000.
- Ballesteros M.L., Sánchez C.M., and Enjuanes L., *Virology* 227, 378–388, 1997.
- Baric R.S. and Yount B., *J Virol* 74, 4039–4046, 2000.
- Bernard S. and Laude H., *J Gen Virol* 76, 2235–2241, 1995.
- Bohl E.H. and Cross R.F., *Ann NY Acad Sci* 176, 150–161, 1971.
- Bohl E.H., Gupta P., Olquin F., and Saif L.J., *Infect Immun* 6, 289–301, 1972.
- Bohl E.H. and Kumagai T., *Proceedings United States Livestock Sanitary Association* 69, 343–350, 1965.
- Bourns M.E.G., Brown T.D.K., Foulds I.J., Green P.F., Tomley F.M., and Binns M.M., *J Gen Virol* 68, 57–77, 1987.
- Brian D.A., Dennis D.E., and Guy J.S., *J Virol* 34, 410–415, 1980.
- Britton P., Mawditt K.L., and Page K.W., *Virus Res* 21, 181–198, 1991.
- Chen C.-M., Cavanagh D., and Britton P., *Vir Res* 38, 83–89, 1995.
- Domingo E. and Holland J.J., *The Evolutionary Biology of Viruses-Mutation Rates and Rapid Evolution of RNA Viruses*. Raven Press, New York, 1994.
- Eleouet J.F., Rasschaert D., Lambert P., Levy L., Vende P., and Laude H., *Virology* 206, 817–822, 1995.
- Eleouet J.F., Slee E.A., Saurini F., Castagne N., Poncet D., Garrido C., Solary E., and Martin S.J., *J Virol* 74, 3975–3983, 2000.

15. Enjuanes L., Brian D., Cavanagh D., Holmes K., Lai M.M.C., Laude H., Masters P., Rottier P., Siddell S.G., Spaan W.J.M., Taguchi F., and Talbot P., *Virus Taxonomy. Classification and Nomenclature of Viruses-Coronaviridae*. Academic Press, New York, 2000a.
16. Enjuanes L., Siddell S.G., and Spaan W.J., *Coronaviruses and Arteriviruses*. Plenum Press, New York, 1998.
17. Enjuanes L., Spaan W., Snijder E., and Cavanagh D., *Virus Taxonomy. Classification and Nomenclature of Viruses-Nidovirales*. Academic Press, New York, 2000b.
18. Escors D., Ortego J., Laude H., and Enjuanes L., *J Virol* 75, 1312–1324, 2001.
19. Fichot O. and Girard M., *Nucleic Acids Res* 18, 6162, 1990.
20. French R. and Ahlquist P., *J Virol* 62, 2411–2420, 1988.
21. Gebauer F., Posthumus W.A.P., Correa L., Suñé C., Sánchez C.M., Smerdou C., Lenstra J.A., Meloen R., and Enjuanes L., *Virology* 183, 225–238, 1991.
22. Gorbalenya A.E., Koonin E.V., Donchenko A.P., and Blinov V.M., *Nucleic Acids Res.* 17, 4847–4861, 1989.
23. Haelterman E.O. and Pensaert M.B., Pathogenesis of transmissible gastroenteritis of swine, Proc 18th World Vet Congress, Paris 1967.
24. Hansen J.L., Long A.M., and Schultz S.C., *Structure* 5, 1109–1122, 1997.
25. Harmon S.B., Megaw G., and Wertz G.W., *J Virol* 75, 36–44, 2001.
26. Herold J., Raabe T., Schelle-Prinz B., and Siddell S.G., *Virology* 195, 680–691, 1993.
27. Higgins D.G., Thompson J.D., and Gibson T.J., *Meth Enzymol* 266, 383–402, 1996.
28. Hofmann M.A. and Brian D.A., *J Virol* 65, 6331–6333, 1991.
29. Iverson L.E. and Rose J.K., *Cell* 23, 477–484, 1981.
30. Izeta A., Smerdou C., Alonso S., Penzes Z., Méndez A., Plana-Durán J., and Enjuanes L., *J Virol* 73, 1535–1545, 1999.
31. Jiménez G., Correa L., Melgosa M.P., Bullido M.J., and Enjuanes L., *J Virol* 60, 131–139, 1986.
32. Kapke P.A. and Brian D.A., *Virology* 151, 41–49, 1986.
33. Lai M.M.C., Brayton P.R., Armen R.C., Patton C.D., Pugh C., and Stohlman S.A., *J Virol* 39, 823–834, 1981.
34. Lai M.M.C. and Cavanagh D., *Adv Virus Res* 48, 1–100, 1997.
35. Laude H. and Masters P.S., *The Coronaviridae-The Coronavirus Nucleocapsid Protein*. Plenum press, New York, 1995.
36. Laude H., Vanreeth K., and Pensaert M., *Vet Res* 24, 125–150, 1993.
37. Lee H.J., Shieh C.K., Gorbalenya A.E., Koonin E.V., Lamonica N., Tuler J., Bagdzhadzhyan A., and Lai M.M.C., *Virology* 180, 567–582, 1991.
38. Lee K.M., Moro M., and Baker J.A., *Amer J Vet Res* 15, 364–372, 1954.
39. McClurkin A.W. and Norman J.O., *Can J Comp Med Vet Sci* 30, 190–198, 1966.
40. Méndez A., Smerdou C., Izeta A., Gebauer F., and Enjuanes L., *Virology* 217, 495–507, 1996.
41. Mizutani T., Repass J.F., and Makino S., *Virology* 275, 238–243, 2000.
42. Murthy H.M., Judge K., DeLucas L., and Padmanabhan R., *J Mol Biol* 301, 759–767, 2000.
43. O'Reilly E.K. and Kao C.C., *Virology* 252, 287–303, 1998.
44. Pachuk C.J., Bredenbeek P.J., Zoltick P.W., Spaan W.J.M., and Weiss S., *Virology* 171, 141–148, 1989.
45. Penzes Z., González J.M., Izeta A., Muntion M., and Enjuanes L., *Adv Exp Med Biol* 440, 319–327, 1998.
46. Rasschaert D., Gelfi J., and Laude H., *Biochimie* 69, 591–600, 1987.
47. Register K.B. and Wesley R.D., *J Vet Diagn Invest* 6, 16–22, 1994.
48. Rohrmeier T. and Lehle L., *Plant Mol Biol* 22, 783–792, 1993.
49. Sambrook J., Fritsch E.F., and Maniatis T., *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989.
50. Sánchez C.M., Gebauer F., Suñé C., Méndez A., Dopazo J., and Enjuanes L., *Virology* 190, 92–105, 1992.
51. Sánchez C.M., Izeta A., Sánchez-Morgado J.M., Alonso S., Sola L., Balasch M., Plana-Duran J., and Enjuanes L., *J Virol* 73, 7607–7618, 1999.
52. Sawicki S.G. and Sawicki D.L., *J Virol* 64, 1050–1056, 1990.
53. Sawicki S.G. and Sawicki D.L., *Adv Exp Med Biol* 440, 215–220, 1998.
54. Sethna P.B., Hofmann M.A., and Brian D.A., *J Virol* 65, 320–325, 1991.
55. Sethna P.B., Hung S.-L., and Brian D.A., *Proc Natl Acad Sci USA* 86, 5626–5630, 1989.
56. Spagnolo J.F. and Hogue B.G., *J Virol* 74, 5053–5065, 2000.
57. Stephensen C.B., Casebolt D.B., and Gangopadhyay N.N., *Virus Res* 60, 181–189, 1999.
58. Thiel V., Herold J., and Siddell S., *Nidovirus: Coronavirus and Arterivirus-Full-Length Genomic Sequence of Bovine Coronavirus*. Kluwer Academic/Plenum Publishers, New York, 2001.
59. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., and Higgins D.G., *Nucleic Acids Res* 24, 4876–4882, 1997.
60. Thompson J.D., Higgins D.G., and Gibson T.J., *Nucleic Acids Res* 22, 4673–4680, 1994.
61. van der Most R.G. and Spaan W.J.M., *The Coronaviridae-Coronavirus Replication, Transcription, and RNA Recombination*. Plenum Press, New York, 1995.
62. van Marle G., Dobbe J.C., Gultyaev A.P., Luytjes W., Spaan W.J.M., and Snijder E.J., *Proc Nat Acad Sci USA* 96, 12056–12061, 1999.
63. Vaughn R.M., Halbur P.G., and Paul P.S., *J Virol* 69, 3176–3184, 1995.
64. Wertz G.W., Perepelitsa V.P., and Ball L.A., *Proc Natl Acad Sci USA* 95, 3501–3506, 1998.
65. Wesley R.D., Cheung A.K., Michael D.M., and Woods R.D., *Virus Res* 13, 87–100, 1989.
66. Wesley R.D., Woods R.D., and Cheung A.K., *J Virol* 64, 4761–4766, 1990.
67. Wesley R.D., Woods R.D., and Cheung A.K., *J Virol* 65, 3369–3373, 1991.
68. Whitelaw C.A., Le Huquet J.A., Thurman D.A., and Tomsett A.B., *Plant Mol Biol* 33, 503–511, 1997.
69. Yamanaka M., Dale B., Crisp T., Cordell B., Crubman M., and Yilma T., *Virology* 190, 553–556, 1992.
70. Yogo Y., Hirano N., Hino S., Shibuta H., and Matumoto M., *J Biochem* 82, 1103–1108, 1977.
71. Yoo D. and Pei Y., *Nidovirus: Coronavirus and Arterivirus-Full-Length Genomic Sequence of Bovine Coronavirus*. Kluwer Academic/Plenum Publishers, New York, 2001.
72. Yount B., Curtis K.M., and Baric R.S., *J Virol* 74, 10600–10611, 2000.