

An insect picornavirus may have genome organization similar to that of caliciviruses

Eugene V. Koonin^a and Alexander E. Gorbalenya^b

^a*Institute of Microbiology, USSR Academy of Sciences, 117811 Moscow, USSR* and ^b*Institute of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, 142782 Moscow Region, USSR*

Received 29 November 1991

Computer-assisted analysis of the amino acid sequence of the product encoded by the sequenced 3' portion of the cricket paralysis virus (CrPV), an insect picornavirus, genome showed that this protein is homologous not to the RNA-directed RNA polymerases, as originally suggested, but to the capsid proteins of mammalian picornaviruses. Alignment of the CrPV protein sequence with those of picornavirus and calicivirus capsid proteins demonstrated that the sequenced portion of the insect picornavirus genome encodes the C-terminal part of VP3 and the entire VP1. Thus CrPV seems to have a genome organization distinct from that of other picornaviruses but closely resembling that of caliciviruses, with the capsid proteins encoded in the 3' part of the genome. On the other hand, the tentative phylogenetic trees generated from the VP3 alignment revealed grouping of CrPV with hepatitis A virus, a true picornavirus, not with caliciviruses. Thus CrPV may be a picornavirus with a calicivirus-like genome organization. Different options for CrPV genome expression are discussed.

Picornavirus; Calicivirus; Capsid protein; RNA-polymerase; Viral genome organization

1. INTRODUCTION

CrPV is classified as an insect picornavirus (reviewed by Moore et al. [1]). CrPV has a number of picornavirus-like characteristics, i.e. icosahedral virions of about 27 nm in diameter, S_w^{20} of 167 S and buoyant density of 1.34 in CsCl, 3 species of capsid proteins with M_r of about 30 kDa, and single-stranded (ss) virion RNA which has approximately 8,000 nucleotides in length. The RNA is polyadenylated at its 3' end and apparently contains a covalently linked protein (VPg) at its 5' end [2]. Also, it has been shown that the three capsid proteins of CrPV are generated by proteolytic processing of a precursor polypeptide, in line with the picornavirus expression strategy [3].

The sequence of 1600 3' terminal nucleotides of CrPV genomic RNA has been reported [2]. It has been speculated that, similarly to mammalian picornaviruses, this portion of CrPV genome might encode the RNA-dependent RNA polymerase, and counterparts to the

three most conserved sequence motifs of positive strand RNA virus polymerases [4,5] have been tentatively identified [2]. Surprisingly, however, no significant overall sequence similarity could be detected between the hypothetical CrPV polymerase and those of mammalian picornaviruses (or any other positive strand RNA viruses).

Here we report results of detailed amino acid sequence comparisons indicating that the polypeptide product of the 3'-terminal part of the CrPV genome is in fact homologous to the capsid proteins of picornaviruses, not to the RNA polymerases.

2. MATERIALS AND METHODS

The sequences were from Swissprot database, except for FeCV [6] and RHDV [7]. Comparison of query amino acid sequences with the Swissprot data bank (Release 18) was performed using the program Quick of the Genebee program package for biopolymer sequence analysis [8]. Local similarity plots for pairs of sequences were generated using the program Dothelix of the Genebee package [9]. Alignments produced using the program OPTAL as previously described, with the adjusted alignment scores computed as the number of standard deviations (SD) above the mean of 25 random simulations [10]. Tentative phylogenetic trees were generated either by a clustering procedure (UPGMA algorithm; [11]), or by the so-called maximum topological similarity algorithm [12].

3. RESULTS AND DISCUSSION

3.1. The 3' part of the CrPV genome encodes a protein homologous to the picornaviruses capsid proteins, not to their RNA-polymerases

In the course of systematic comparative analysis of positive strand RNA virus RNA polymerases [13], the

Abbreviations: CrPV, cricket paralysis virus; FMDV01, foot-and-mouth disease virus, serotype O1; EMCR, encephalomyocarditis virus, Rueckert strain; MENGO, Mengo virus; TMEBEAN, Theiler murine encephalomyelitis virus, strain BeAn 8386; HAV, hepatitis A virus, Los Angeles (Chiron) strain; BEV, bovine enterovirus; POLIO1M, poliomyelitis virus, serotype 1, strain Mahoney; RHINO2, 14 human rhinovirus, respective serotypes (all – picornaviruses); FeCV, feline calicivirus; RHDV, rabbit hemorrhagic disease virus (both – caliciviruses).

Correspondence address: E.V. Koonin, National Center for Biotechnology Information, National Library of Medicine, Bldg. 38A, N.I.H., 8600 Rockville Pike, Bethesda, MD 20894, USA.

ALIGNMENT OF VP3 (1C)

```

1  FMDVO1 88  tf1AgLaQYYTqYsGTInlhFmFtgptdakARyMVaYaPpg---meppkTPeaaahcihaewDT
2   EMCR  91  tf1AaLSrnFaqYRGS�VYTFvFtgTammkGKfLIaYtPpg---agkptSrdQamqatyaIwDL
3   MENGO 91  tf1AaLSrnFaqYRGS�VYTFvFtgTammkGKfLIaYtPpg---agkptSrdQamqatyaIwDL
4  TMEBEAN 92  sm1AaVarnFnqYRGS�nF1FvFtgaamvkGKfLIaYtPpg---agkptTrdQamqStyaIwDL
5  POLIO1M 97  tmlGeI1NYYThWaGS�kFTF1FcgfmmaTGKLLVSYaPpg---adppkkrkEamlgthVIWDI
6   BEV   96  sl1GtLakhYtqWsGSVeITcmFtgTfntTGKVLLaYtPpg---gdmpnrreEamlgthVIWdf
7   RHINO2 96  t11GeISsYFThWtGS�rFSFmFcgTantTvKLLLaYtPpg---iaeptTrkDamlgthVIWdV
8   RHINO14 94  t11GeIvQYYThWsGS�rFSImYtgpalsSAKLILaYtPpg---argpqrreEamlgthVVWdI
STRUCTURE  aaaaA2aaaabbbbbDbbbbbbb      bbbbEbbbb      aaBaaaabbbbFbbb

```

```

9   HAV   98  talAsIcQmFcFWRGdLVFdfQvfpTkYHSGRLLfcFVPgnelidVTgiTlkQaTtapcaVMDI
          : : : * : *** : * : * * * * * * * * * * * * * * * * * * * * * * : : * : : * : * :
10  CRPV  ?   hthGyVTNaFTYWRGSIVYTFkFvkTqYHSGRLrISFIPyyynttISstGTP-DVSrTqkIVVDL
          : : : : * * * : * : * * * * * * * * * * * * * * * * * * * * * * : : * :
11  FeCV  ?   pylthLak1YvaWsGSvdvrFsisgSgvfgGKLaaivVPpg----IdpvqstsMlqyphVLFda
12  RHDV  ?   pftAvLSQmYagWaGmQFrFivagSgvfgGRLvravIPpg----IeiGpgIEVrqfphVVIda

```

```

1  -glnSkfTFSIPYLSaadYtYtasd---vaettNvQ-GwVclfqIThgk-----aDgdAlvvLaSagk-Dfe
2  -glnSsySFTVPFIspthFrmvgtD---qVnitNaD-GwVtvwqLTpLtypp--gcptsakiltMvSagk-Dfs
3  -glnSsySFTVPFIspthFrmvgtD---qanitNvD-GwVtvwqLTpLtypp--gcptsakiltMvSagk-Dfs
4  -glnSsfntaPFIsphYrqtSyT---spTitsvD-GwVtvwkLTpLtypp--gtptnsdiltLvSagd-Dft
5  -glqSscTmvVPWIsnttYrqtI-----ddSftEg--GyIsvfyqTrIVvpl--stpreMdilgfvSacn-Dfs
6  -glqSsITlvIPWIsashFrgvsndDvlnyqyyaa--GhVtiwyqTnMVipp--gfpNtaGimMiaaqp-Nfs
7  -glqStISMvVPWIsashYrnts-----pgrsts---GyItcwyqTrLVipp--qtpptarllcfvSgck-Dfc
8  -glqStIvmTIPWtsgvqFrYtd-----pdTytsa--GfLscwyqTsLilpp--ettgqVyllsfiSapc-Dfk
Str  bbG1bbb      bG2      bbbbbbbHbbbbbb      bbbbbbbblbbbbb

```

```

9  tgvqStLrFrVPWisetPYrvnrytkSahqkgEytaiGkLIvyCynrLtsps--nVashVrvNvylSaI--NLe
          : : * * * * * * : * : * : * : * : * : * : * : * : * : * : * : * : * :
10 R-tsTeVSFTVPYIaSrPWlycIrpESswLSkDNkD-GaLMynCVSgIVrte--vLnQlVAaQnVfSeI--DVi
          * : * * * * * : * : * : * : * : * : * : * : * : * : * : * : * : * :
11 R-qvepViFSIPdLrStlyhlms-----dTD-t---tsLVimvyndLInpyandsNssgcivtVeTkpgpDfk
12 R-SlepVTiTMPdLrpnmYhptg-----dpglv---ptLVlsvynnLInpf--ggstsaiqvtVeTrpseDfe

```

ALIGNMENT OF VP1 (1D)

```

1  lrLpvdaraE----->-----
2  lkMpisPapwspq-----
3  lkMpisPapwspq-----
4  lrMpisPtkwvpq-----
5  vrLlrdtthieqk-alaq ---glgqmlEsmIdNTvRetvg-----aatsrdalpnteAsgpthskEIP
6  frIQkdrEdmtqt-ailq ---ndpgkmlkDaIdkqvagalv-----agtttsthsvatdstP
7  lrMardtNLhlqsgaiaq -----npvENyIdEvlinevly-----vpinssnpttsnsaP
8  lrLmkdtQtisqtvalte ---glgdeleEEvIvEktKqt-----vasI-ssgpkhtqkVP
Str  bbb

```

```

9  CfaplyhaMDvttq---- vgddsggFs--tTVsteqnV---pdpQvgittmrdLkgkanRgkmDVsGvQapVg
          * : : * : : * : : * : : * : : * : : * : : * : :
10 CeVNggPDLEfagp---- tcpsmYPagDDTLADTrKLeaertqEysnnednrIttqcsRIvAQVmGeDhEIP
          : *
11 fhLlkpPgsm1th-----
12 fvMiraPssktvd-----

```

Fig. 1.

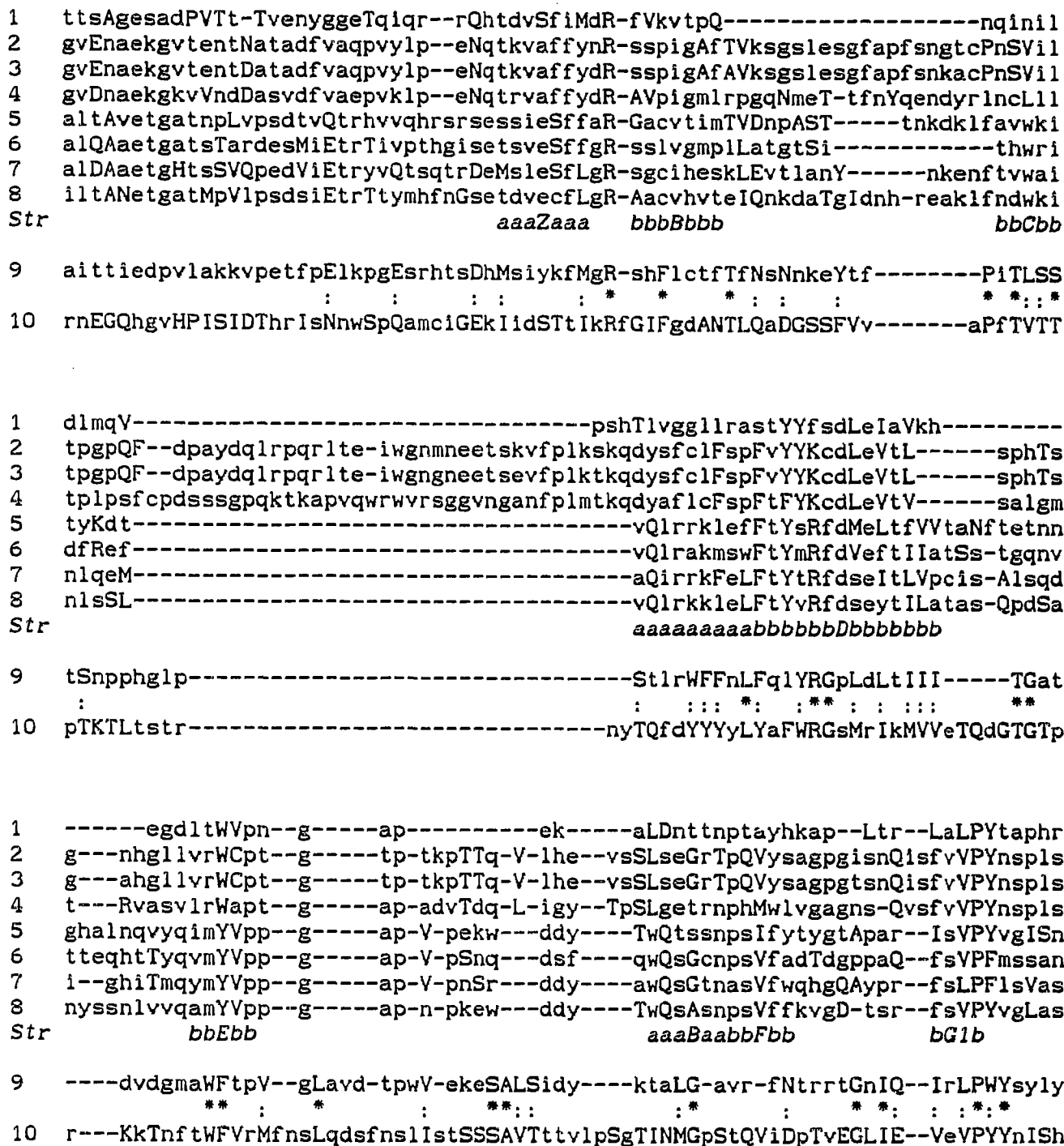


Fig. 1.

product of the 3' part of the CrPV genome was compared to all sequences in the Swissprot bank. As perhaps could have been anticipated, the sequence most similar to that of CrPV belonged to the polyprotein of a picornavirus, HAV. Surprisingly, however, this sequence was not from the polymerase region of the HAV polyprotein, but rather comprised a 30 residue segment

of the capsid protein VP3. The next highest scores were with the homologous sequences of other picornaviruses. To further explore the nature of the product encoded by the 3' part of CrPV genome, we generated the plots of local similarity between its sequence and picornavirus capsid proteins. This led to delineation of additional regions of significant similarity in picornavirus VP3 and

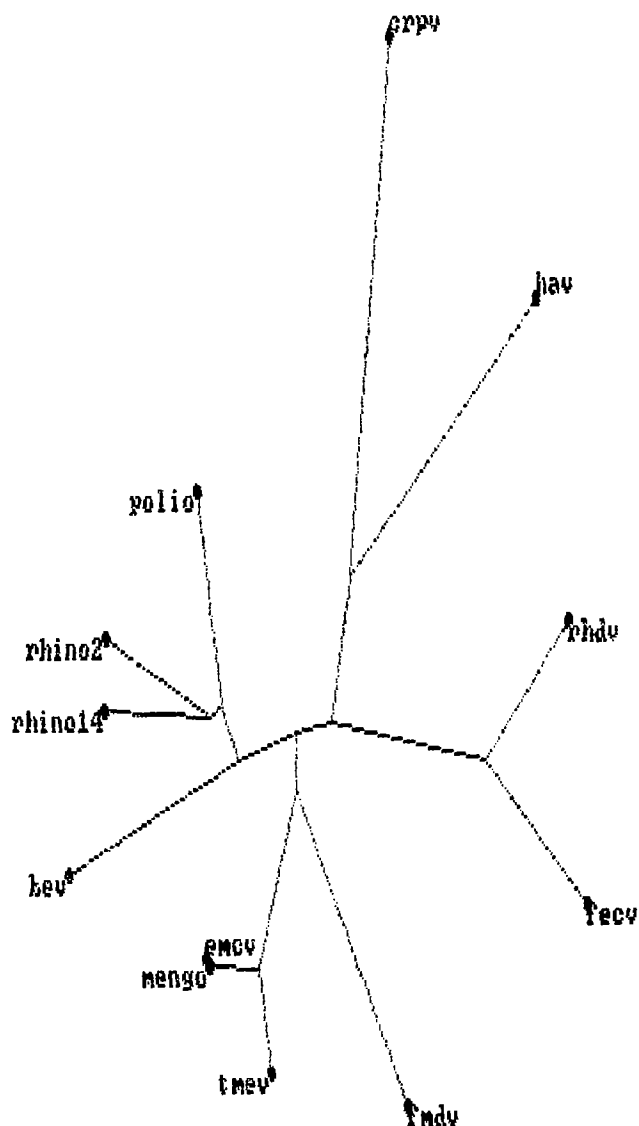


Fig. 2. Tentative phylogenetic tree for VP3 proteins. The rootless tree generated by the maximum topological similarity algorithm is shown. The branch lengths are proportional to evolutionary rates of the respective species.

sites could be found within approximately 30 residues around the probable boundary between the two proteins. Cleavage at each of them would result in predicted VP1 size of about 240–270 amino acid residues, in good agreement with both the typical values for mammalian picornaviruses (274–302 residues, with the exception of aphthoviruses having large deletions in VP1), and with the electrophoretic estimates for CrPV. In common with HAV, CrPV probably bears an extension at the N-terminus of the putative VP1. On the other hand, it appears to lack the C-terminal disordered region found in VP1s of all mammalian picornaviruses (Fig. 1). In the latter, this region appears to constitute the hinge between the structural and non-structural

domains of the polyprotein. Thus its absence in CrPV may be related to the alternative genome organization of this virus.

Serological cross-reaction between CrPV and encephalomyocarditis virus has been reported [17]. Although the present observations do not allow one to delineate the common antigenic determinant, they show that conserved segments do exist in the (putative) capsid proteins of these viruses, one of them probably conferring the cross-reactivity.

The resulting alignment of VP3, which was the best conserved portion of the picornavirus capsid protein alignment, and the only sequence showing significant conservation between calicivirus and picornavirus capsid proteins, was used to generate tentative phylogenetic trees using two different methods. Fig. 2 shows the rootless tree produced by the maximum topological similarity algorithm which has been designed specifically to overcome the unequal evolutionary rate effects [12]. Obviously, the tree split in three major subdivisions, one of which included the caliciviruses, the other HAV and CrPV, and the third one all the remaining mammalian picornaviruses. The dendrogram generated by the clustering procedure separated HAV and CrPV as an outgroup, while grouping together caliciviruses and other picornaviruses (not shown). Phylogenetic analysis of viral RNA-dependent RNA polymerases and putative RNA helicases clearly indicated that caliciviruses are related to but distinct from picornaviruses (A.E.G., unpublished observations). Thus we believe that the result of cluster analysis of the VP3s was probably due to an anomalously high rate of evolution of HAV and CrPV.

3.2. Possible genome organization of CrPV

We showed here that the sequenced portion of an insect picornavirus genome encodes putative viral capsid proteins. As the assignment of this sequence to the 3' part of virus RNA by King et al. [2] seems to be quite solid, confirmed by direct sequencing of a part of the poly(A) tail, this observation implies a genomic organization for CrPV that is basically different from the picornavirus one. Apparently, CrPV encodes the proteins mediating genome replication and expression in its 5' part, and capsid proteins in the 3' part. This gene arrangement obviously resembled that found for the caliciviruses [7,16,18], as well as for several other groups of positive-strand RNA viruses (reviewed by Strauss and Strauss [19], and Lai [20]). On the other hand, the results of phylogenetic analysis suggested an evolutionary relationship between CrPV and HAV, which is obviously a 'true' picornavirus. Also, the classification of CrPV as a picornavirus, not a calicivirus, is supported by the absence of substantial similarity between the VP1-related sequence of CrPV and the calicivirus capsid protein, and by the fact that CrPV produces three capsid protein species, like picor-

naviruses, and not one, like caliciviruses. This discrepancy raises the important 'tempo and mode' problem in virus evolution, suggesting that similarities in genome organization do not necessarily mirror the phylogenetic grouping of viruses [21]. Determination of the portion of CrPV genome encoding non-structural proteins will allow a better assessment of the phylogenetic position of this interesting virus.

3.3. *The options for CrPV genome expression strategy*

There seem to exist three options for CrPV genome expression: (i) capsid protein precursor is generated by translation of a subgenomic RNA; this expression strategy would be analogous to that exploited by caliciviruses [6,7]; (ii) the capsid protein precursor constitutes the C-terminal portion of the single polyprotein product of the virus genome, resembling the expression strategy of potyviruses [22]; (iii) translation of the capsid protein precursor is initiated independently.

The first hypothesis, though attractive because of the analogy with caliciviruses, seems to contradict the results of radiolabelling of virus-specific RNA in CrPV-infected cells that has not revealed subgenomic species [3]. The second hypothesis, on the other hand, is not readily compatible with the pactamycin mapping data suggesting that the capsid proteins of CrPV are encoded immediately downstream of the ribosome entry site [23]. The third possibility is apparently consistent with all the available experimental data. Clearly, internal initiation of translation is not a typical expression strategy of eukaryotic positive strand RNA viruses. On the other hand, growing evidence shows that initiation of translation of picornavirus mRNAs occurs internally, and their 5'-non-coding sequences retain the ability to direct initiation when placed downstream of a 5' gene in a genetically engineered mRNA (for review see [24]). Moreover, artificial dicistronic construct has been recently described, in which EMC virus 5'-non-coding sequence was introduced between the poliovirus genes encoding capsid and nonstructural proteins; strikingly, such RNA produced infectious virus [25]. Thus it cannot be ruled out that CrPV is a natural 'monster' with a similar, 'mixed' procaryotic/eucaryotic expression strategy. Further experiments on CrPV genome structure and expression will show whether this rather bold speculation will hold true, or this virus actually

exploits one of the more conventional strategies, whereas some of the available data require reevaluation.

REFERENCES

- [1] Moore, N.F., Reavy, B. and King, L.A. (1985) *J. Gen. Virol.* 66, 647-659.
- [2] King, L.A., Pullin, J.S.K., Stanway, G., Almond, J. and Moore, N.F. (1987) *Virus Res.* 6, 331-344.
- [3] Reavy, B. and Moore, N.F. (1983) *Virology* 131, 551-554.
- [4] Kamer, G. and Argos, P. (1984) *Nucleic Acids Res.* 12, 1194-2007.
- [5] Koonin, E.V., Gorbalenya, A.E., Chumakov, K.M., Donchenko, A.P. and Blinov, V.M. (1987) *Mol. Genetika* 7, 27-39.
- [6] Neill, J.D., Reardon, I.M. and Heinrichson, R.L. (1991) *Virus Res.* 11, 59-72.
- [7] Meyers, G., Wirblich, C. and Thiel, H.-J. (1991) *Virology* 184, 664-676.
- [8] Brodsky, L.I., Drachev, A.L., Tatuzov, R.L. and Chumakov, K.M. (1991) *Biopolimery i Kletka* 7, 10-14.
- [9] Leontovich, A.M., Brodsky, L.I. and Gorbalenya, A.E. (1990) *Biopolimery i Kletka* 6, 14-21.
- [10] Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. and Koonin, E.V. (1989) *J. Mol. Evol.* 28, 256-268.
- [11] Sneath, P. and Sokal, R. (1973) *Principles of Numerical Taxonomy*, San Francisco.
- [12] Chumakov, K.M. and Yushmanov, S.Yu. (1988) *Mol. Genetika* n. 3, 3-9.
- [13] Koonin, E.V. (1991) *J. Gen. Virol.* 72, 2197-2206.
- [14] Palmenberg, A.C. (1989) in: *Molecular Aspects of Picornavirus Infection and Detection* (B.L. Semler and E. Ehrenfeld eds.) pp. 211-241, Am. Soc. Microbiol., Washington.
- [15] Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. and Brown, F. (1989) *Nature* 327, 709-716.
- [16] Tohya, Y., Taniguchi, Y., Takahashi, E., Utagawa, E., Takeda, N., Miyamura, K., Yamazaki, S. and Mikami, T. (1991) *Virology* 183, 810-814.
- [17] Tinsley, T.W., MacCallum, F.O., Robertson, J.S. and Brown, F. (1984) *Intervirology* 21, 181-186.
- [18] Neill, J.D. (1990) *Virus Res.* 17, 145-160.
- [19] Strauss, E.G. and Strauss, J.H. (1986) in: *The Togaviridae and Flaviviridae* (S. Schlesinger and M.J. Schlesinger eds.) pp. 35-90, Plenum, New York.
- [20] Lai, M.M.C. (1990) *Ann. Rev. Microbiol.* 44, 303-333.
- [21] Koonin, E.V. (1991) *Virus Genes* 5, 273-282.
- [22] Dougherty, W.G. and Carrington, J.C. (1988) *Annu. Rev. Phytopathol.* 26, 123-143.
- [23] Reavy, B. and Moore, N.F. (1981) *J. Gen. Virol.* 55, 429-438.
- [24] Jackson, R.J., Howell, M.T. and Kaminski, A. (1990) *Trends Biol. Chem.* 15, 477-483.
- [25] Wimmer, E., Hellen, C., Jang, S.K., Litterst, M., Molla, A., Paul, A., Pestova, T. and Witherel, G. (1991) in: *Workshop on the Regulation of Translation in Animal Virus-Infected Cells* (Sonnenberg, N. and Carrasco, L. eds.) pp. 18-19, Fundacion Juan March, Madrid.