Molecular Cloning and Sequence Determination of the Genomic Regions Encoding Protease and Genome-Linked Protein of Three Picornaviruses

GUDRUN WERNER, BRIGITTE ROSENWIRTH,* EVA BAUER, JAN-MARCUS SEIFERT, FRED-JOCHEN WERNER, AND JUERGEN BESEMER

Sandoz Forschungsinstitut, A-1235 Vienna, Austria

To investigate the degree of similarity between picornavirus proteases, we cloned the genomic cDNAs of an enterovirus, echovirus 9 (strain Barty), and two rhinoviruses, serotypes 1A and 14LP, and determined the nucleotide sequence of the region which, by analogy to poliovirus, encodes the protease. The nucleotide sequence of the region encoding the genome-linked protein VPg, immediately adjacent to the protease, was also determined. Comparison of nucleotide and deduced amino acid sequences with other available picornavirus sequences showed remarkable homology in proteases and among VPgs. Three highly conserved peptide regions were identified in the protease; one of these is specific for human picornaviruses and has no obvious counterpart in encephalomyocarditis virus, foot-and-mouth disease virus, or cowpea mosaic virus proteases. Within the other two peptide regions two conserved amino acids, Cys 147 and His 161, could be the reactive residues of the active site. We used a statistical method to predict certain features of the secondary structures, such as α helices, β sheets, and turns, and found many of these conformations to be conserved. The hydropathy profiles of the compared proteases were also strikingly similar. Thus, the proteases of human picornaviruses very probably have a similar three-dimensional structure.

A common characteristic feature of picornaviruses besides the structure of the virion is their genome organization and translation strategy. Picornavirus particles contain an infectious, single-stranded RNA molecule of approximately 7,500 nucleotides (2) which is polyadenylated at its 3' terminus (48) and covalently linked to a small viral protein, termed VPg, at its 5' end (24). Upon infection, ribosomes initiate translation at a single site of the viral genome, resulting in the synthesis of a large polyprotein of a molecular weight of approximately 250,000. This polyprotein is cleaved as a nascent chain into three or four primary precursors which are further processed proteolytically to produce all of the viral structural and nonstructural proteins (12, 13, 33, 36, 43).

The family of picornaviruses is subdivided into four genera: the enteroviruses (poliovirus, echovirus, coxsackievirus), cardiovirus (encephalomyocarditis [EMC] virus, Mengo virus), aphthovirus (foot-and-mouth disease virus), and rhinoviruses. Two of these genera are clinically important; human rhinoviruses are the major causative agents of upper respiratory tract infections collectively known as the "common cold" (42), and enteroviruses cause not only dysfunction of the gastrointestinal tract but also aseptic meningitis and other more severe diseases, such as poliomyelitis (23). As many as 115 serotypes of rhinovirus and more than 70 different enteroviruses have been isolated. Because of this high antigenic diversity, it appears unlikely that it would be possible to develop a vaccine protecting against a broad spectrum of rhinovirus or enterovirus infections. However, since picornaviruses share a common genome strategy, the enzymes involved in virus replication may be conserved and thus may represent suitable targets for chemotherapy.

The elucidation of the complete nucleotide sequence of

the poliovirus genome (17, 31) and radiochemical sequencing of poliovirus-specific proteins (19, 38, 39) have confirmed that the RNA is translated into one polyprotein which is processed proteolytically to give rise to all of the known virus-specific polypeptides (25). The majority of these peptides are produced by proteolytic cleavages between a glutamine and a glycine residue. These Gln-Gly cleavages are carried out by a virus-encoded proteinase (11) of a molecular weight of approximately 20,000 originally termed 7c (44) and recently renamed 3C according to the convention on systemic nomenclature of picornavirus proteins (34). Similarly, for EMC virus a polypeptide, originally p22 and now 3C, has been described as being responsible for most capsid maturation cleavages (9, 27) and self-cleavage reactions within the protease precursor molecules (28). This protease obviously plays a crucial role in the processing of viral proteins and thus is indispensable for picornavirus replication.

In addition to being essential for virus growth, a target enzyme for antiviral chemotherapy, ideally, should be conserved among the members of a virus family. To investigate the degree of similarity between picornavirus proteases, we cloned the genome RNAs of echovirus 9, an enterovirus, and two rhinoviruses, serotypes 1A and 14, and determined the nucleotide sequence of the region which, by analogy to poliovirus, encodes the protease. The nucleotide sequence of the region encoding the genome-linked protein VPg (24) or 3B (34), being immediately adjacent, was also determined. The comparison of the nucleotide and deduced amino acid sequences with other available picornavirus sequences revealed remarkable conservation of picornavirus proteases and VPgs.

MATERIALS AND METHODS

Chemicals and enzymes. Nucleoside triphosphates, dideoxynucleoside triphosphates, [α-^{32}P]dCTP, [α-^{35}S]thio-
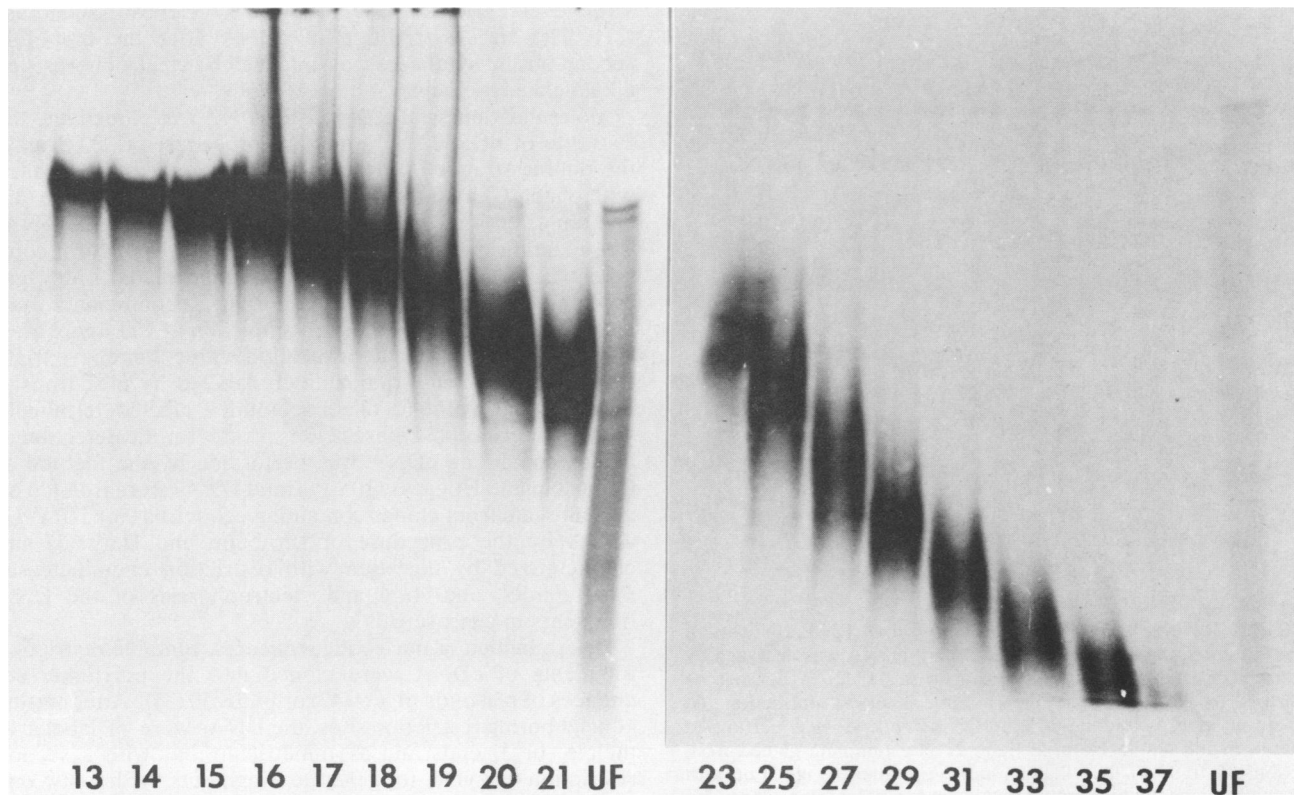
---

* Corresponding author.

FIG. 1. Reverse transcription of echovirus 9 RNA: fractionation of first-strand cDNA on an alkaline sucrose gradient. ^{32}P-labeled first-strand cDNA was prepared by reverse transcription and subjected to centrifugation through an alkaline sucrose gradient (15 to 30% [wt/vol] sucrose in 0.9 M NaCl–0.1 M NaOH) in the SW65 Ti Beckman rotor at 39,000 rpm for 20 h at 1°C. cDNA was collected in 40 fractions, and size distribution was analyzed by electrophoresis through alkaline agarose gels (1.4% agarose in 50 mM NaCl–1 mM EDTA; running buffer, 30 mM NaOH–2 mM EDTA). The numbers at the bottom denote the respective fractions. Lane UF shows unfractionated first-strand cDNA.

dATP, and [γ-^{32}P]ATP were purchased from Amersham International, and oligo(dT) was from Boehringer GmbH, Mannheim, Federal Republic of Germany. M13 vectors mp8 and mp9 were obtained from Amersham International, oligo(dG)-tailed pBR322 was from New England Nuclear Corp., Boston, Mass., and cloning vector pUC9 was from P-L Biochemicals, Inc., Milwaukee, Wis. Cloning vector pUR250 was a gift from U. Ruether, Cologne, Federal Republic of Germany. RNase inhibitor was from Biotec Inc. Avian myeloblastosis virus reverse transcriptase was obtained from J. Beard, Life Sciences, Inc., St. Petersburg, Fla.; T4 DNA ligase, terminal deoxynucleotidyl transferase, and T4 polynucleotide kinase were from P-L Biochemicals; nuclease S1 was from Sigma Chemical Co., St. Louis, Mo.; proteinase K and DNA polymerase I Klenow fragment were from Boehringer GmbH; and restriction enzymes were from Boehringer GmbH and Bethesda Research Laboratories, Inc., Gaithersburg, Md.

Viruses. Echovirus type 9, strain Barty, was originally supplied by A. B. Sabin and kindly provided by H. J. Eggers, Cologne, Federal Republic of Germany. Human rhinovirus type 1A (HRV-1A), strain 2060, was obtained from the American Type Culture Collection (ATCC). Human rhinovirus type 14LP (HRV-14LP) is a large-plaque variant isolated in our laboratory from a serially passaged stock of rhinovirus type 14, strain 1059, which was originally from ATCC. The viruses were plaque purified by standard techniques, and their identities were verified by neutraliza-tion assays with type-specific antiserum purchased from ATCC.

Cells. The GMK cell line, a continuous line derived from African green monkey kidney cells by H. Lennartz, Hamburg, Federal Republic of Germany, was given to us by H. J. Eggers, Cologne, Federal Republic of Germany. HeLa Ohio cells were from ATCC. Both cell lines were propagated as monolayers.

Bacteria. *Escherichia coli* K-12-5K was kindly provided by W. Goebel, Wuerzburg, Federal Republic of Germany.

Virus growth and purification. Confluent GMK cell monolayers were infected with echovirus 9 at a multiplicity of infection of 10 to 20 PFU per cell; when 80% confluent, HeLa cell monolayers were infected with 2 to 5 PFU of HRV-1A or HRV-14LP per cell. Infected cells were incubated at 37°C (for echovirus) or 34°C (for rhinovirus) in Eagle minimum essential medium (6) without serum but containing in addition 25 mM HEPES (N-2-hydroxyethylpiperazine-N'-2- ethanesulfonic acid), pH 7.0 (for echovirus 9), or 25 mM HEPES and 30 mM MgCl$_2$ (for rhinoviruses) until the cytopathic effect was complete. Cells were disrupted by freezing and thawing, debris was removed by centrifugation, and the virus particles were precipitated twice by addition of dry polyethylene glycol 20,000 (to 6% [wt/vol] for echovirus 9 and to 4% [wt/vol] for the rhinoviruses) in the presence of 0.15 M NaCl. The pellets were dissolved in 0.01 M Tris chloride (pH 7.4)–0.15 M NaCl, and the virus particles were purified by zonal centrifugation through 15 to 30% sucrose

—23500
— 9700
— 6600
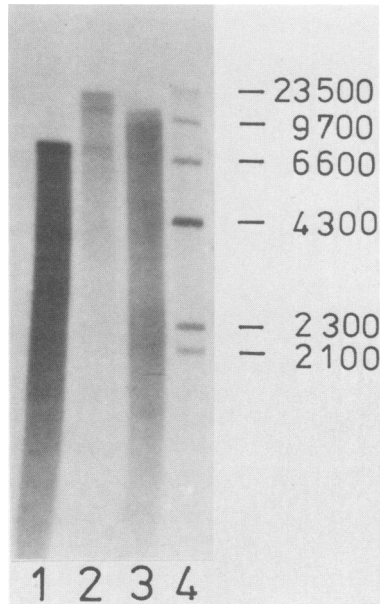— 4300
— 2300
— 2100

1 2 3 4

FIG. 2. Reverse transcription of echovirus 9 RNA: analysis of cDNA after synthesis of the second strand. cDNA was analyzed by electrophoresis through an alkaline agarose gel (1.4% agarose as described in the legend to Fig. 1). Lane 1 shows unfractionated first-strand cDNA. Second-strand cDNA synthesis was performed either with full-length material (lane 2) as a template gained from fractions 11 to 13 of the alkaline sucrose gradient (Fig. 1) or with more heterogeneous first-strand cDNA as a template (lane 3) pooled from fractions 14 to 17. Size marker DNA was run in parallel (lane 4). Marker DNA is shown in base pairs.

gradients in the same buffer. Fractions containing virus were identified by measuring the optical density at 260 nm. Virus purified by this procedure was >90% pure as judged from sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis of capsid proteins, followed by Coomassie blue staining.

Isolation of virus RNA. Virus RNA was extracted directly from the appropriate sucrose gradient fractions by proteinase K digestion for 1 h at 37°C in a reaction mixture containing 250 μg of proteinase K per ml, 0.01 M Tris chloride (pH 7.5), 0.5% SDS, and 1 mM EDTA. RNA was precipitated with 3 volumes of ethanol after addition of sodium acetate (pH 5.0) to 0.3 M, dissolved in 0.5 to 1.0 ml of H_2O, and extracted three times with an equal volume of phenol (neutralized and saturated with 0.01 M Tris chloride [pH 7.5]–1 mM EDTA)-chloroform-isoamyl alcohol (100:50:4). The RNA was collected by ethanol precipitation and further purified by zonal centrifugation through 15 to 30% sucrose gradients in 0.01 M Tris chloride (pH 7.5)–0.1 M NaCl–0.5% SDS–1 mM EDTA. Fractions containing RNA were identified by measuring the OD_{260}. The RNA was precipitated with ethanol and washed several times with 70% ethanol in H_2O to remove SDS.

Synthesis of double-stranded cDNAs. ^{32}P-labeled first-strand cDNA was prepared by reverse transcription with avian myeloblastosis virus reverse transcriptase and oligo(dT) as a primer as described by Maniatis et al. (21). Full-length cDNA was separated from smaller fragments by zonal centrifugation through alkaline sucrose gradients (14) and used as a template for the synthesis of double-stranded cDNA. Self-primed synthesis of the second DNA strand was achieved by sequential incubation with avian myeloblastosis

virus reverse transcriptase, followed by Klenow polymerase (21). The size distribution of cDNA after the first- and second-strand syntheses was analyzed by electrophoresis on alkaline agarose gels.

Molecular cloning of cDNA. The cDNA was cloned into the PstI site of pBR322 by strandard procedures (21). Briefly, the double-stranded, nuclease S1-treated cDNA was tailed with oligo(dC) by terminal deoxynucleotidyl transferase (32) and annealed with PstI-cut, oligo(dG)-tailed pBR322 DNA. The reaction mixture was taken directly to transform competent E. coli 5K cells which were then plated onto agar containing 10 μg of tetracycline per ml. Recombinants were identified by their sensitivity to ampicillin at 100 μg/ml after replica plating. Colony hybridization with either appropriate restriction fragments radioactively labeled by nick translation (21) or synthetic oligonucleotides labeled terminally with polynucleotide kinase (21) as probes for the detection of clones containing cDNA was performed by the method of Grunstein and Hogness (10). Plasmid DNAs were isolated on a small scale from clones containing echovirus 9 or HRV-1A cDNA by the procedure of Birnboim and Doly (3) and characterized by digestion with restriction endonucleases PstI, EcoRI, and AccI and electrophoresis of the DNA fragments in agarose gels.

Determination of nucleotide sequences. Suitable restriction fragments of cDNA were cloned into the polylinker sequences of plasmids pUC9 (47) or pUR250 (35). After cutting at neighboring restriction sites, the DNAs were ^{32}P labeled at their 5' or 3' ends and asymmetrically cut with a second restriction enzyme to generate fragments with only one radioactively labeled end. Since one of these fragments was only a few nucleotides long, the other could be sequenced without further purification by using the chemical methods described by Maxam and Gilbert (22). Alternatively, restriction fragments were subcloned into the mp8 and mp9 derivatives of bacteriophage M13 and sequenced by the dideoxynucleotide method of Sanger et al. (37). Both strands of cDNA were sequenced completely in all cases. The derived nucleotide sequences were entered into a Hewlett-Packard HP1000 computer, where the information was stored and processed with a computer program developed in our laboratories for this purpose.

RESULTS

Viral RNA was extracted from virion particles and purified as detailed in Materials and Methods. Single-stranded cDNA was synthesized from this RNA with oligo(dT) as a primer for avian myeloblastosis virus reverse transcriptase. The cDNA was rendered double stranded with reverse transcriptase, followed by Klenow polymerase. To obtain significant amounts of full-length double-stranded cDNA, it proved essential to separate full-length first-strand transcripts from smaller fragments by centrifugation through alkaline sucrose gradients before they could serve as templates for self-primed synthesis in the second-strand reaction. Figure 1 shows an alkaline agarose gel illustrating the size range of the products of first-strand cDNA synthesis of echovirus 9 RNA after centrifugation through an alkaline sucrose gradient. Sucrose gradient fractions 11 to 13 or 14 to 17 were the templates for the second-strand reaction, and the resulting double-stranded cDNA was again analyzed by electrophoresis on an alkaline agarose gel (Fig. 2).

The double-stranded cDNA obtained from the two reaction mixtures differed considerably. When cDNA from fractions 11 to 13 containing only full-length DNA was used as a template (Fig. 2, slot 2), a clearly discernible band of
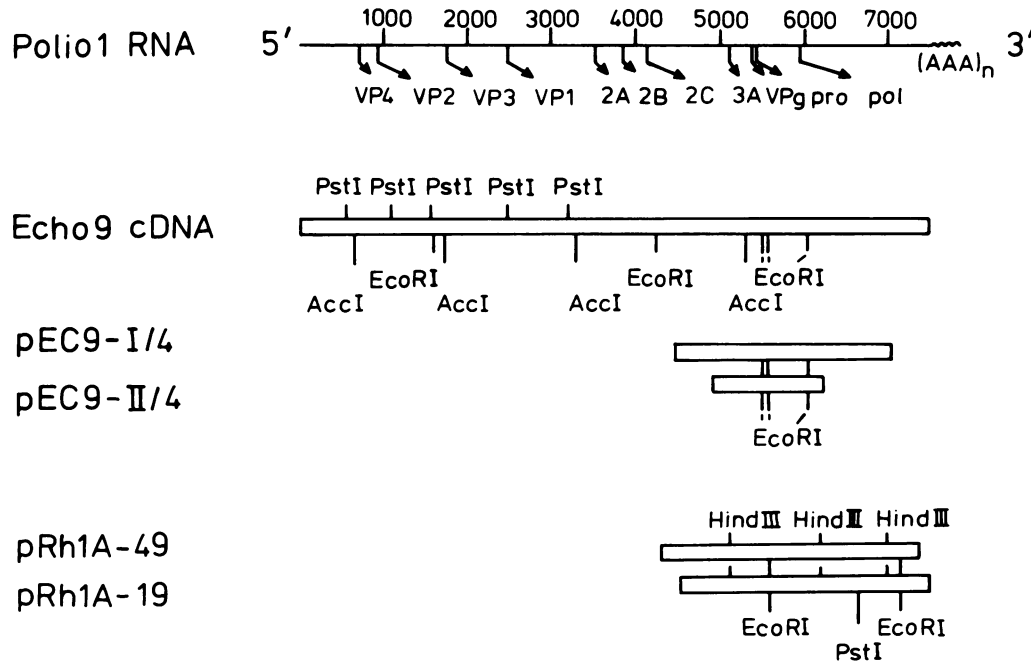
FIG. 3. Alignment of the echovirus 9 cDNA restriction map with the biochemical map of poliovirus 1. Echovirus 9 cDNA was digested with combinations of two of the following restriction enzymes: PstI, EcoRI, and AccI. Fragments were separated by electrophoresis through alkaline agarose gels (1.4% agarose in 50 mM NaCl–1 mM EDTA; running buffer, 30 mM NaOH–2 mM EDTA). Those fragments representing the 5' end of the RNA were identified by their migration behavior in alkaline gels as compared with that in neutral gels; they appeared twice as long under denaturing conditions as under nondenaturing conditions because of the fold-back loop produced by reverse transcriptase. The genomic position of viral cDNA from two clones containing echovirus 9 sequences pEC9-I/4 and pEC9-II/4 and HRV-1A sequences pRh1A-49 and pRh1A-19, which were isolated as described in Materials and Methods, are depicted schematically in the lower part of this figure.

full-length cDNA (approximately 15,000 nucleotides long) was detected; in addition, we found one slightly shorter species (13,500 nucleotides) probably due to a strong stop signal for reverse transcriptase. When more heterogeneous cDNA (fractions 14 to 17) served as a template (Fig. 2, slot 3), the products of the second-strand reaction were a mixture of cDNAs of approximately 13,000 nucleotides long to less than 2,000 nucleotides long. Thus, by taking only full-length first-strand transcripts as templates, we could synthesize full-length double-stranded cDNA from echovirus 9 and HRV-14LP RNA. For HRV-1A, the amount of starting material, i.e., purified viral RNA, was too low to obtain sufficient cDNA of full length in the first-strand reaction. In this case, second-strand synthesis was performed with cDNA as a template corresponding to the fragment sizes of sucrose gradient fractions 12 to 16.

For echovirus 9, the production of full-length double-stranded cDNA enabled us to establish a restriction map of the viral genome directly from the cDNA (Fig. 3). The orientation of the end fragments relative to the RNA genome was deduced from their migration behavior in alkaline agarose gels; fragments representing the 5' end of the RNA appeared twice as long in alkaline gels as in neutral gels because of the fold-back loop produced by the self-primed reverse transcription. The restriction map made it possible to identify the fragment in which we were interested. Alignment of the echovirus 9 cDNA restriction map with the known biochemical map of poliovirus 1 (25) suggested that the EcoRI 490-base pair restriction fragment produced by the two 3' proximal cuts should contain protease sequences. This fragment was inserted into the EcoRI site of pUR250 and sequenced chemically (22). Translation of the DNA sequences obtained in the six possible reading frames and

comparison of the amino acid sequences with those of poliovirus indeed revealed homology with the protease.

Since the EcoRI 490-base pair fragment lacked the 5'-terminal part of the protease gene, it was labeled radioactively by nick translation and used as a hybridization probe to screen our cDNA library, obtained by random cloning of double-stranded echovirus 9 cDNA into pBR322. Two clones thus identified and used for further sequence analysis of the protease and VPg genes of echovirus 9 are depicted schematically in Fig. 3.

The restriction enzymes EcoRI, BamHI, and HindIII cut HRV-14LP cDNA too rarely to allow the construction of a restriction map. In this case, we used the endonucleases AluI, HaeIII, TaqI, Sau3A, and EcoRI* which cut DNA frequently and cloned the fragments into the polylinker site of M13 vectors mp8 and mp9. Clones overlapping most of the genome were sequenced by the dideoxynucleotide method (37). All cDNA clones could be localized at a specific position of the genome by comparing the deduced amino acid sequences with those of poliovirus with the aid of a computer program. In this way, the protease and VPg regions of HRV-14LP were easily identified.

All of the double-stranded HRV-1A cDNA was cloned into pBR322 without previous separation of size classes. To identify clones containing HRV-1A cDNA encoding the protease region, a synthetic oligonucleotide was used as a hybridization probe. Comparison of the cDNA sequences of poliovirus 1, echovirus 9, and HRV-14LP revealed a region of relatively conserved nucleotide sequence at positions 5667 to 5687 (numbering according to poliovirus 1) of the protease gene. An oligonucleotide representing the corresponding sequence of 21 bases of HRV-14LP was synthesized, labeled terminally with ^{32}P, and served as a hybrid-

A

```
         9            18            27            36        45            54            63            72            81
GGC CCA GCG TTT GAA TTC GCC GTG GCG ATG ATG AAA AGA AAC GCC AGT ACA GTG AAA ACC GAG TAT GGT GAA TTC ACC ATG
Gly Pro Ala Phe Glu Phe Ala Val Ala MET MET Lys Arg Asn Ala Ser Thr Val Lys Thr Glu Tyr Gly Glu Phe Thr MET

        90            99           108           117       126           135           144           153           162
CTT GGT ATT TAT GAC AGA TGG GCG GTG TTA CCA CGC CAC GCC AAA CCT GGT CCC AGC ATC TTG ATG AAT GAT CAG GAA GTT
Leu Gly Ile Tyr Asp Arg Trp Ala Val Leu Pro Arg His Ala Lys Pro Gly Pro Ser Ile Leu MET Asn Asp Gln Glu Val

       171           180           189           198       207           216           225           234           243
GGC GTG TTG GAT GCC AAG GAA CTG GTC GAT AAA GAT GGG ATA AAC CTA GAA CTG ACA CTC CTG AAG CTC AAC CGT AAT GAA
Gly Val Leu Asp Ala Lys Glu Leu Val Asp Lys Asp Gly Ile Asn Leu Glu Leu Thr Leu Leu Lys Leu Asn Arg Asn Glu

       252           261           270           279       288           297           306           315           324
AAG TTC AGA GAC ATT AGA GGG TTT CTA GCG AGA GAA GAA GTT GAA GTG AAT GAA GCT GTC CTA GCA ATA AAC ACA AGT AAA
Lys Phe Arg Asp Ile Arg Gly Phe Leu Ala Arg Glu Glu Val Glu Val Asn Glu Ala Val Leu Ala Ile Asn Thr Ser Lys

       333           342           351           360       369           378           387           396           405
TTC CCT AAC ATG TAT ATA CCC GTG GGC CAG GTA ACT GAC TAC GGG TTC TTG AAC CTG GGT GGG ACC CCC ACG AAG AGA ATG
Phe Pro Asn MET Tyr Ile Pro Val Gly Gln Val Thr Asp Tyr Gly Phe Leu Asn Leu Gly Gly Thr Pro Thr Lys Arg MET

       414           423           432           441       450           459           468           477           486
CTC ATG TAT AAC TTC CCA ACC AGA GCA GGT CAG TGT GGA GGT GTC CTC ATG TCA ACA GGG AAA GTC CTC GGA ATA CAC GTA
Leu MET Tyr Asn Phe Pro Thr Arg Ala Gly Gln Cys Gly Gly Val Leu MET Ser Thr Gly Lys Val Leu Gly Ile His Val

       495           504           513           522       531           540           549
GGA GGA AAT GGA CAC CAC GGG TTC TCA GCC GCA CTC CTC AGG CAT TAC TTC AAC GAG GAG CAG
Gly Gly Asn Gly His His Gly Phe Ser Ala Ala Leu Leu Arg His Tyr Phe Asn Glu Glu Gln
```

B

```
         9            18            27            36        45            54            63            72            81
GGA CCA AAC ACA GAA TTT GCA CTA TCC CTG TTA AGG AAA AAC ATA ATG ACT ATA ACA ACC TCA AAG GGA GAG TTC ACA GGG
Gly Pro Asn Thr Glu Phe Ala Leu Ser Leu Leu Arg Lys Asn Ile MET Thr Ile Thr Thr Ser Lys Gly Glu Phe Thr Gly

        90            99           108           117       126           135           144           153           162
TTA GGC ATA CAT GAT CGT GTC TGT GTG ATA CCC ACA CAC GCA CAG CCT GGT GAT GAT GTA CTA GTG AAT GGT CAG AAA ATT
Leu Gly Ile His Asp Arg Val Cys Val Ile Pro Thr His Ala Gln Pro Gly Asp Asp Val Leu Val Asn Gly Gln Lys Ile

       171           180           189           198       207           216           225           234           243
AGA GTT AAG GAT AAG TAC AAA TTA GTA GAT CCA GAG AAC ATT AAT CTA GAG CTT ACA GTG TTG ACT TTA GAT AGA AAT GAA
Arg Val Lys Asp Lys Tyr Lys Leu Val Asp Pro Glu Asn Ile Asn Leu Glu Leu Thr Val Leu Thr Leu Asp Arg Asn Glu

       252           261           270           279       288           297           306           315           324
AAA TTC AGA GAT ATC AGG GGA TTT ATA TCA GAA GAT CTA GAA GGT GTG GAT GCC ACT TTG GTA GTA CAT TCA AAT AAC TTT
Lys Phe Arg Asp Ile Arg Gly Phe Ile Ser Glu Asp Leu Glu Gly Val Asp Ala Thr Leu Val Val His Ser Asn Asn Phe

       333           342           351           360       369           378           387           396           405
ACC AAC ACT ATC TTA GAA GTT GGC CCT GTA ACA ATG GCA GGA CTT ATT AAT TTG AGT AGC ACC CCC ACT AAC AGA ATG ATT
Thr Asn Thr Ile Leu Glu Val Gly Pro Val Thr MET Ala Gly Leu Ile Asn Leu Ser Ser Thr Pro Thr Asn Arg MET Ile

       414           423           432           441       450           459           468           477           486
CGT TAT GAT TAT GCA ACA AAA ACT GGG CAG TGT GGA GGT GTG CTG TGT GCT ACT GGT AAG ATC TTT GGT ATT CAT GTT GGC
Arg Tyr Asp Tyr Ala Thr Lys Thr Gly Gln Cys Gly Gly Val Leu Cys Ala Thr Gly Lys Ile Phe Gly Ile His Val Gly

       495           504           513           522       531           540
GGT AAT GGA AGA CAA GGA TTT TCA GCT CAA CTT AAA AAA CAA TAT TTT GTA GAG AAA CAA
Gly Asn Gly Arg Gln Gly Phe Ser Ala Gln Leu Lys Lys Gln Tyr Phe Val Glu Lys Gln
```

ization probe to screen the HRV-1A clones. From several clones giving a positive hybridization signal, two were selected which were assumed to contain inserts including the protease gene (Fig. 3). As a control, a set of echovirus 9 clones was also tested against this oligonucleotide, and only those clones containing the protease gene gave a positive result (data not shown). Sequencing of the HRV-1A protease gene was performed by subcloning EcoRI, HindIII, AluI, and HaeIII restriction fragments in M13 vectors mp8 and mp9, followed by application of the dideoxynucleotide method (37).

The nucleotide and deduced amino acid sequences of the regions encoding the proteases and VPgs of echovirus 9, HRV-14LP, and HRV-1A are given in Fig. 4A through C and 5, respectively.

DISCUSSION

A comparison of the deduced amino acid sequences of proteases and VPgs of echovirus 9, HRV-14LP, and HRV-1A with those of poliovirus 1 and HRV-2 is presented in Fig. 5 and 6. While this manuscript was in preparation, the nucleotide sequences of HRV-14 (4, 41) and HRV-2 (40) were published. The HRV-14LP variant we have sequenced does not differ in nucleotide sequence from the published

C

```
        9              18             27             36             45             54             63             72             81
GGT CCA GAA GAA GAA TTT GGA AGG TCA ATT CTC AAA AAC AAT ACT TGT GTG ATT ACT ACA GGT AAT GGA AAA TTT ACA GGT
Gly Pro Glu Glu Glu Phe Gly Arg Ser Ile Leu Lys Asn Asn Thr Cys Val Ile Thr Thr Gly Asn Gly Lys Phe Thr Gly

        90             99            108            117            126            135            144            153            162
CTT GGT ATA CAT GAC AGA ATT CTA ATC ATC CCA ACA CAT GCT GAT CCA GGT AGA GAG GTC CAA GTT AAT GGT GTC CAC ACT
Leu Gly Ile His Asp Arg Ile Leu Ile Ile Pro Thr His Ala Asp Pro Gly Arg Glu Val Gln Val Asn Gly Val His Thr

       171            180            189            198            207            216            225            234            243
AAG GTT CTA GAC TCA TAT GAT CTT TAT AAT AGA GAT GGA GTT AAA CTT GAA ATA ACG GTC ATA CAA TTA GAT AGA AAT GAA
Lys Val Leu Asp Ser Tyr Asp Leu Tyr Asn Arg Asp Gly Val Lys Leu Glu Ile Thr Val Ile Gln Leu Asp Arg Asn Glu

       252            261            270            279            288            297            306            315            324
AAA TTT AGG GAC ATT AGA AAG TAT ATA CCT GAA ACA GAA GAC GAT TAT CCA GAA TGC AAT TTG GCA CTT TCA GCT AAT CAA
Lys Phe Arg Asp Ile Arg Lys Tyr Ile Pro Glu Thr Glu Asp Asp Tyr Pro Glu Cys Asn Leu Ala Leu Ser Ala Asn Gln

       333            342            351            360            369            378            387            396            405
GAT GAA CCA ACT ATA ATT AAA GTA GGA GAT GTA GTG TCC TAT GGC AAT ATT TTG CTT AGT GGA AAT CAA ACA GCC AGA ATG
Asp Glu Pro Thr Ile Ile Lys Val Gly Asp Val Val Ser Tyr Gly Asn Ile Leu Leu Ser Gly Asn Gln Thr Ala Arg MET

       414            423            432            441            450            459            468            477            486
CTT AAA TAT AAT TAC CCC ACA AAA TCA GGG TAT TGT GGA GGG GTA CTA TAT AAA ATT GGT CAA ATT CTA GGT ATT CAT GTG
Leu Lys Tyr Asn Tyr Pro Thr Lys Ser Gly Tyr Cys Gly Gly Val Leu Tyr Lys Ile Gly Gln Ile Leu Gly Ile His Val

       495            504            513            522            531            540            549
GGT GGA AAT GGA AGG GAT GGT TTT TCA GCT ATG TTA CTT AGA TCA TAC TTT ACA GAT ATT CAG
Gly Gly Asn Gly Arg Asp Gly Phe Ser Ala MET Leu Leu Arg Ser Tyr Phe Thr Asp Ile Gln
```

FIG. 4. Nucleotide and deduced amino acid sequences of the echovirus 9 genomic region encoding the protease (A), the HRV-14 LP genomic region encoding the protease (B), and the HRV-1A genomic region encoding the protease (C).

HRV-14 strains in the regions coding for protease and VPg. HRV-2 shows close homology in the protease and VPg genomic regions to HRV-1A (see below), and the corresponding sequences are included for comparison. Tables 1 and 2 summarize nucleotide and amino acid sequence homologies of proteases and VPgs, respectively. The percentage of identical amino acid residues between pairs of proteases compared varies from 43.7 to 80.9%, and that between VPgs varies from 36.4 to 90.5%. The closest homology is found between HRV-1A and HRV-2 and between the two enteroviruses poliovirus 1 and echovirus 9. HRV-14LP, interestingly, seems to take an intermediate position; it is as closely related to the enteroviruses as to the other two rhinoviruses. The comparison of the percentages of identical

nucleotides between pairs of viruses leads to the same conclusion: HRV-1A and HRV-2 are most closely related, followed by the enteroviruses. The high degree of homology between the protease and VPg genes of HRV-1A and HRV-2 suggests a close evolutionary relationship of these two viruses.

In three regions of the human picornavirus proteases, stretches of amino acid residues are highly conserved (Fig. 6): Arg 79 to Arg 87, Gly 145 to Ile 151, and His 161 to Gly 166. Quite strikingly, the conserved amino acid residues Gly 145-X-Cys 147 and His 161 are identical to the amino acids known to be present in the active center of plant sulfhydryl endopeptidases, such as papain, stem bromelain, or actinidin (15), of streptococcal protease (46), and of mammalian

```
           9              18             27             36             45             54             63
1    GGA GCA TAC ACT GGT TTA CCA AAC AAA --- AAA CCC AAC GTG CCC ACC ATT AGG ACA GCA TTG GTA CAA
     Gly Ala Tyr Thr Gly Leu Pro Asn Lys     Lys Pro Asn Val Pro Thr Ile Arg Thr Ala Leu Val Gln

           9              18             27             36             45             54             63
2    GGC GCC TAC ACA GGG ATG CCC AAC AAG --- AAA CCT AAG GTG CCC ACC CTA AGA CAG GCC AAA GTG CAA
     Gly Ala Tyr Thr Gly MET Pro Asn Lys     Lys Pro Lys Val Pro Thr Leu Arg Gln Ala Lys Val Gln

           9              18             27             36             45             54             63
3    GGA CCA TAT TCT GGT AAC CCG CCT CAC AAT AAA CTA AAA GCC CCA ACT TTA CGC CCA GTT GTT GTG CAA
     Gly Pro Tyr Ser Gly Asn Pro Pro His Asn Lys Leu Lys Ala Pro Thr Leu Arg Pro Val Val Val Gln

           9              18             27             36             45             54             63
4    GGA CCA TAC TCA GGT GAA CCT AAA CCT --- AAA ACC AAA GTA CCT --- GAA AGA AGA GTA GTT GCT CAA
     Gly Pro Tyr Ser Gly Glu Pro Lys Pro     Lys Thr Lys Val Pro     Glu Arg Arg Val Val Ala Gln

           9              18             27             36             45             54             63
5    GGA CCA TAT TCA GGA GAA CCA AAG CCC --- AAG ACT AAA ATC CCA --- GAA AGG CGT GTA GTA ACA CAG
     Gly Pro Tyr Ser Gly Glu Pro Lys Pro     Lys Thr Lys Ile Pro     Glu Arg Arg Val Val Thr Gln
```

FIG. 5. Nucleotide and deduced amino acid sequence of picornavirus genomic regions encoding VPg. Rows: 1, poliovirus 1; 2, echovirus 9; 3, HRV-14LP; 4, HRV-1A; 5, HRV-2.

```
                5                        10              15          20          25
1  Gly Pro Gly Phe Asp Tyr Ala Val Ala Met Ala Lys Arg Asn Ile Val Thr Ala Thr Thr Ser Lys Gly Glu Phe Thr Met Leu
2      Ala     Glu Phe                     Met     Arg Lys     Ala Ser         Glu Tyr
3      Asn Thr Glu Phe         Leu Ser Leu Leu Arg Lys         Met     Ile                     Gly
4      Glu Glu Glu Phe Gly Arg Ser Ile Leu Asn         Thr Cys Val Ile         Gly Asn     Lys         Gly
5      Glu Glu Glu Phe Gly Met Ser Leu Ile Ile His     Ser Cys Val Ile         Glu Asn     Lys         Gly

         30      35          40              45          50              55
1  Gly Val His Asp Asn Val Ala Ile Leu Pro Thr His Ala Ser Pro Gly Glu Ser Ile Val Ile Asp Gly Lys Glu Val Glu Ile
2      Ile Tyr     Arg Trp     Val             Arg Lys     Pro     Leu Met Asn Asp Gln     Gly Val
3      Ile         Arg     Cys Val Ile                 Gln     Asp Leu Val Asn     Gln Lys Ile Arg Val
4      Ile         Arg Ile Leu     Ile                 Asp     Arg Glu Val Gln Val Asn     Val His Thr Lys Val
5      Val Tyr         Phe Val Val Val                 Asp     Lys Glu Gln Val         Ile Thr Thr Lys Val

      60          65              70          75          80
1  Leu Asp Ala Lys Ala Leu Glu Asp Gln Ala Gly Thr Asn Leu Glu Ile Thr Ile Ile Thr Leu Lys Arg Asn Glu Lys Phe Arg
2                  Glu     Val     Lys Asp     Ile                     Leu Leu Leu     Asn
3  Lys     Lys Tyr Lys     Val     Pro Glu Asn Ile             Leu     Val Leu     Gln     Asp
4          Ser Tyr Asp     Tyr Asn Arg Asp     Val Lys             Val         Val         Asp
5  Ile     Ser Tyr Asp     Tyr Ser Lys Asn Ile Lys             Val Leu Lys         Asp

   85          90          95              100             105             110
1  Asp Ile Arg Pro His Ile Pro Thr Gln Ile Thr Glu Thr Asn Asp Gly Val Leu Ile Val Asn Thr Ser Lys Tyr Pro Asn Met
2      Gly Phe Leu     Glu Glu Asp Val     Val     Glu Ala         Ala Ile             His Ser Asn Asn Phe
3      Gly Phe     Ser Glu Asp Leu ---     Gly Val     Ala Thr     Val         Leu Ser Ala Asn Gln Phe Thr     Thr
4      Lys Tyr         Glu Thr Glu Asp Asp Tyr Pro Glu Cys Asn     Ala Leu Ser Ala Asn Asn Asp Glu Pro Thr
5      Arg Tyr         Asn Asn Glu Asp Asp Tyr Pro Asn Cys Asn     Ala Leu Leu Ala Asn Gln Pro Glu Pro Thr

      115             120             125             130             135             140
1  Tyr Val Pro Val Gly Ala Val Thr Glu Gln Gly Tyr Leu Asn Leu Gly Gly Arg Gln Thr Ala Arg Thr Leu Met Tyr Asn Phe
2  Ile Ile         Gln         Asp Tyr Phe         Thr Pro     Lys Met             Arg     Tyr
3  Ile Leu Glu     Pro     Met Ala Leu Ile Ser Ser Thr Pro Asn Met Ile Arg         Asp Tyr
4  Ile Ile Lys     Asp Val Ser Tyr Asn Ile Leu Ser Asn     Met     Met Lys         Tyr
5  Ile Ile Asn     Asp Val Ser Tyr Asn Ile Leu Ser Asn         Met Lys         Ser Tyr

         145             150             155             160             165
1  Pro Thr Arg Ala Gly Gln Cys Gly Gly Val Ile Thr Cys Thr Gly Lys Val Ile Gly Met His Val Gly Gly Asn Gly Ser His
2  Ala         Leu Met Ser                     Leu     His
3      Lys Thr             Leu Cys Ala     Ile Leu Phe Ile             Arg Gln
4      Lys Ser Tyr         Leu Tyr Lys Ile Gln Ile Leu Ile             Arg Asp
5      Lys Ser Tyr         Leu Tyr Lys Ile Gln Leu Leu Ile             Arg Asp

      170             175             180
1  Gly Phe Ala Ala Ala Leu Lys Arg Ser Tyr Phe Thr Gln Ser Gln Gln
2      Ser         Leu     His         Asn Glu Glu
3      Ser Gln Lys Gln     Val Glu Lys
4      Ser Met Leu         Asp Ile
5      Ser Met Leu         Asp Val
```
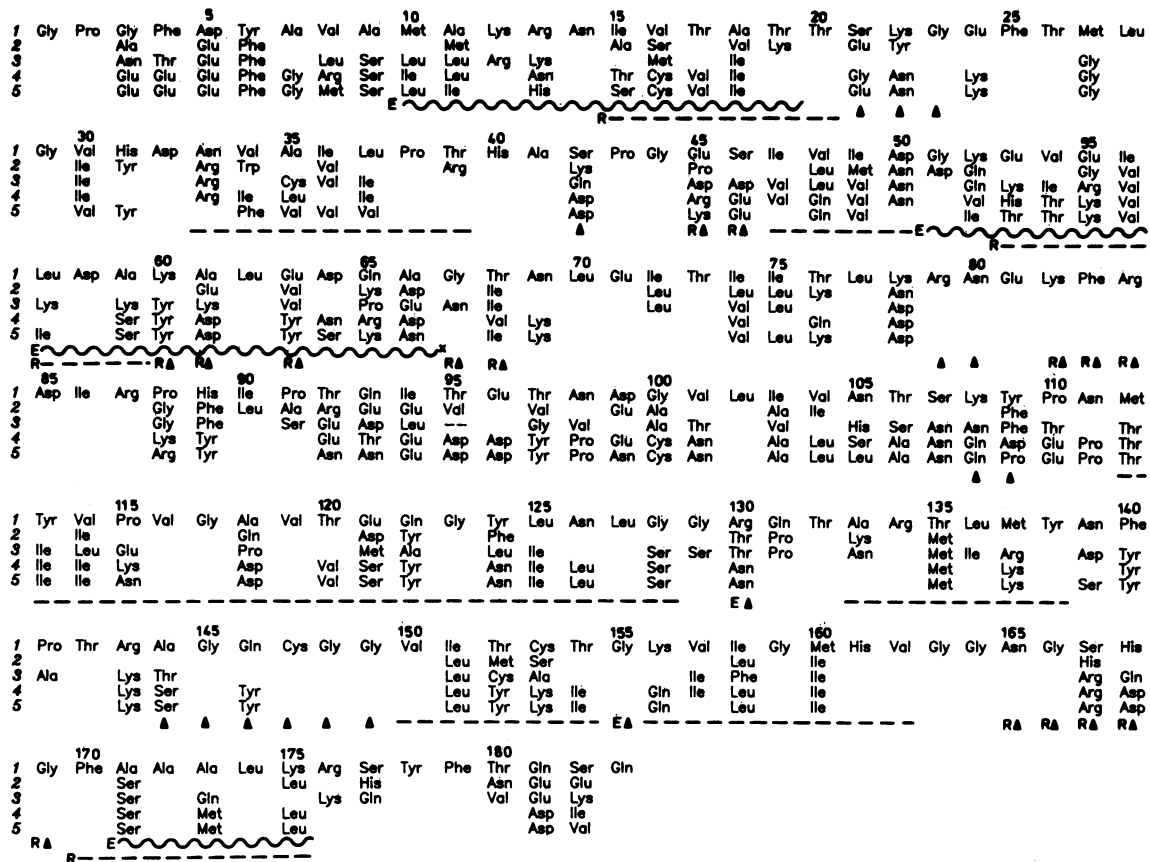
FIG. 6. Comparison of picornavirus protease amino acid sequences and secondary structures. Rows: 1, poliovirus 1; 2, echovirus 9; 3, HRV-14 LP; 4, HRV-1A; 5, HRV-2. Amino acid residues which are conserved as compared with the poliovirus 1 protease sequence are not depicted. Hypothetical secondary structures were calculated by the statistical method of Garnier et al. (8) and are symbolized as follows: ∿∿, α-helical domains; ———, β sheets; ▲, turns. E and R represent structures common only to enteroviruses or rhinoviruses, respectively. X marks a domain which is common for enteroviruses and HRV-14 but not for the two other rhinoviruses.

proteases, such as the rat liver cathepsins (15). Mechanistic (30) and crystallographic (15) studies have elucidated the mode of action of this type of proteases and established the essential role of Cys and His in the catalytic process. Therefore, one is led to the conclusion that the human picornavirus proteinases probably belong to the cysteine class of proteolytic enzymes.

From the comparison of the protease amino acid sequences of animal picornaviruses, namely EMC virus (26) and foot-and-mouth disease virus (5), and cowpea mosaic virus (20), a plant comovirus, to those of poliovirus, Argos et al. (1) came to conclusions similar to ours. These three viruses contain stretches of amino acids similar to those of poliovirus in the regions around Cys 147 and His 161. The homologies are not as striking as for the human picornaviruses but are clearly significant. The use of specific inhibitors also led to results which indicated that the 3C protein of EMC virus is a cysteine proteinase (9, 29). However, apart from these conserved Gly-X-Cys and His residues, there is no further homology detectable with papain or other members of the papain family. Thus, the picornavirus proteases represent a different class of enzymes not homologous but possibly mechanistically similar to papain. The sequence Arg 79 to Arg 87 has no obvious counterparts in the EMC virus (26), foot-and-mouth disease virus (5), or cowpea mosaic virus (20) proteases and therefore seems to be a unique feature of human picornaviruses.

The VPgs of human picornaviruses can be easily aligned (Fig. 5). HRV-1A and HRV-2 VPgs are one amino acid residue shorter than the VPg of poliovirus; VPg of HRV-14LP is one residue longer. In position 3, the tyrosine residue which links the protein to the RNA (16) is conserved, as are also the basic amino acids Lys 11 (numbering according to HRV-14LP) and Arg 18 and Gly 5, Pro 7, and Pro 15, which are known to behave as "helix breakers." These residues may be important for the function, namely in the interaction with the RNA, or the secondary structure of VPg.

To get an idea about the structural similarities of the compared proteases, we calculated the relative hydropathy values (Fig. 7) and the probabilities of secondary structures like α helices, β sheets, and turns as well as random coil regions (Fig. 6) by the methods of Kyte and Doolittle (18) and Garnier et al. (8), respectively.

The overall homology of the hydropathy profiles is statistically highly significant (Table 1). For the three conserved stretches of amino acid residues, the hydropathy patterns are very similar: regions 145 to 151 and 161 to 166 are partly hydrophilic and partly hydrophobic as may be expected from an active center "pocket"; region 79 to 87 is highly hydrophilic. Interestingly, the amino acid sequence of HRV-14LP located 3' proximal to this region resembles the enteroviruses more closely than it resembles the HRV-1A and HRV-2, which again indicates an intermediate position
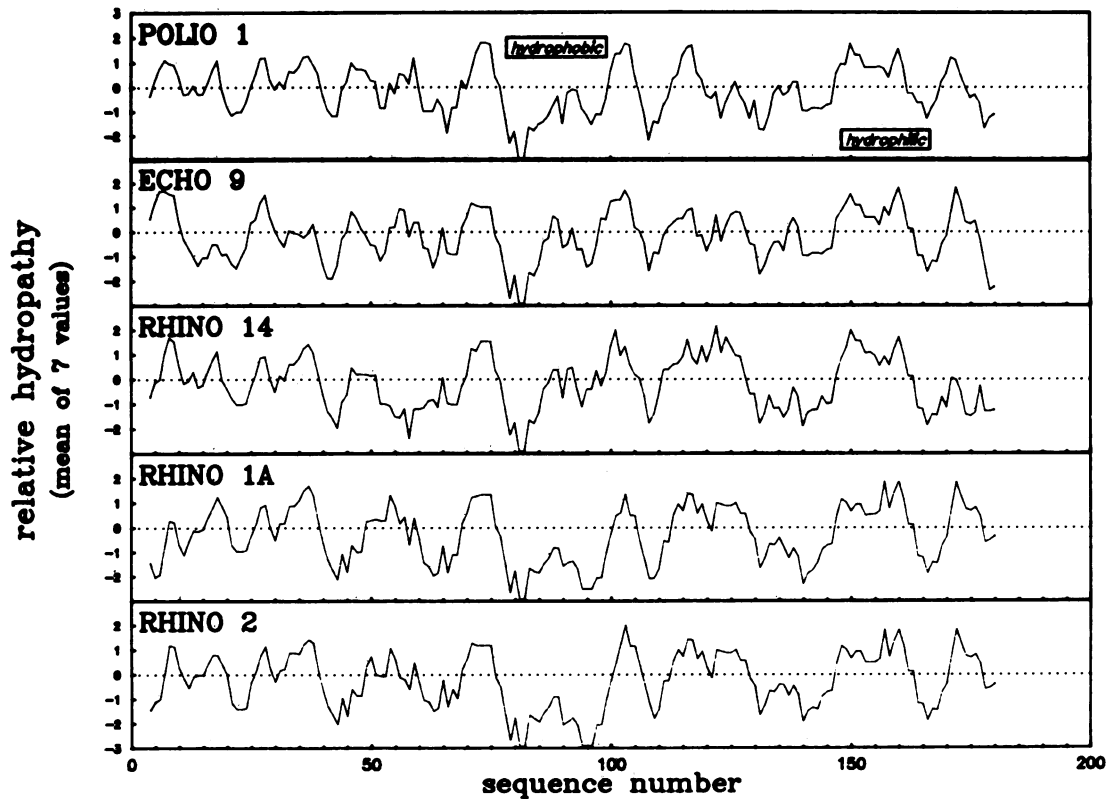
FIG. 7. Hydropathy plots of human picornavius proteases. Relative hydropathy values were calculated as a mean of seven consecutive amino acid residues by the method of Kyte and Doolittle (18) and were plotted against the amino acid sequence number.

for this virus. Though region 79 to 87 is highly hydrophilic, it seems not to be a major antigenic determinant in vivo, which is concluded from the finding that antisera prepared against purified, SDS-denatured proteases of poliovirus 2 and echovirus 9 do not cross-react either with each other or with extracts from rhinovirus-infected cells (B. Rosenwirth and G. Werner, unpublished results). Similarly, antiserum against SDS-denatured poliovirus 1 protease has been shown not to cross-react with echovirus 6- or echovirus 11-induced proteins (7).

The analysis of secondary structures (Fig. 6) clearly reveals structural domains which are common for all human picornaviruses and also some which are only shared by enteroviruses or rhinoviruses. There is a series of turns at

the putative active center at positions 144 to 149 and also in the conserved region 79 to 87. The conserved His 161 is part of a β-sheet structure for all compared viruses. There is also a common extended β-sheet structure at positions 112 to 128; in other regions α helices formed by enterovirus proteases correspond to β sheets formed by rhinovirus proteases. Interestingly, the α-helical domain at positions 51 to 66 is shared by the enteroviruses and HRV-14LP, while HRV-1A and HRV-2 form a β sheet at positions 53 to 59, followed by several turns. In this region, the protease of HRV-14LP obviously differs structurally from that of the other two rhinoviruses.

The statistical methods used for the proteases were not applicable for VPg because of the small size of this polypeptide.

In conclusion, the data presented allow the prediction that the three-dimensional structure and the amino acid residues probably involved in the proteolytic process are highly

TABLE 1. Proteases of picornaviruses

	% Sequence homology[a]				
	Poliovirus 1	Echovirus 9	HRV-14LP	HRV-1A	HRV-2
Poliovirus 1	100	63.4	57.9	50.1	52.9
Echovirus 9	61.7 (0.8115)	100	57.9	52.2	55.1
HRV-14LP	45.9 (0.7346)	52.5 (0.7152)	100	57.0	56.1
HRV-1A	44.3 (0.7533)	44.3 (0.6368)	53.0 (0.6944)	100	75.6
HRV-2	43.7 (0.7775)	48.6 (0.6980)	49.7 (0.7104)	80.9 (0.9632)	100

[a] Upper right half, Nucleotide sequence homologies; bottom left half, amino acid sequence homologies; in parentheses, correlation coefficient of hydropathy plots (18) calculated by the method of Sweet and Eisenberg (45).

TABLE 2. VPgs of picornaviruses

	% Sequence homology[a]				
	Poliovirus 1	Echovirus 9	HRV-14LP	HRV-1A	HRV-2
Poliovirus 1	100	68.2	52.2	57.6	50.0
Echovirus 9	77.3	100	49.3	50.0	45.5
HRV-14LP	43.5	52.2	100	53.6	49.3
HRV-1A	40.9	45.5	56.5	100	73.0
HRV-2	36.4	40.9	56.5	90.5	100

[a] Upper right half, Nucleotide sequence homologies; bottom left half, amino acid sequence homologies.

conserved for the human picornaviruses. This enzyme therefore is a suitable target for antiviral chemotherapy, and it should be possible to find a common inhibitor. Attempts to express the cloned protease gene are in progress, with the aim of preparing a sufficient amount of pure enzyme for crystallization and X-ray analysis of its structure.

ACKNOWLEDGMENTS

LITERATURE CITED

1. Argos, P., G. Kamer, M. J. H. Nicklin, and E. Wimmer. 1984. Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. Nucleic Acids Res. 12:7251–7267.
2. Baltimore, D. 1971. Expression of animal virus genomes. Bacteriol. Rev. 35:235–241.
3. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513–1523.
4. Callahan, P. L., S. Mizutani, and R. J. Colonno. 1985. Molecular cloning and complete sequence determination of RNA genome of human rhinovirus 14. Proc. Natl. Acad. Sci. USA 82:732–736.
5. Carroll, A. R., D. J. Rowlands, and B. E. Clarke. 1984. The complete nucleotide sequence of the RNA coding for the primary translation product of foot and mouth disease virus. Nucleic Acids Res. 12:2461–2472.
6. Eagle, H. 1959. Amino acid metabolism in mammalian cell cultures. Science 130:432–437.
7. Emini, E. A., W. A. Schleif, R. J. Colonno, and E. Wimmer. 1985. Antigenic conservation and divergence between the viral-specific proteins of poliovirus type 1 and various picornaviruses. Virology 140:13–20.
8. Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120:97–120.
9. Gorbalenya, A. E., and Y. V. Svitkin. 1983. Protease of encephalomyocarditis virus: purification and role of sulfhydryl groups in processing of the structural proteins precursor. Biokhimiya 48:442–453.
10. Grunstein, M., and D. S. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA 72:3961–3965.
11. Hanecak, R., B. L. Semler, C. W. Anderson, and E. Wimmer. 1982. Proteolytic processing of poliovirus peptides: antibodies to polypeptide P3-7c inhibit cleavage at glutamine-glycine pairs. Proc. Natl. Acad. Sci. USA 79:3973–3977.
12. Holland, J. J., and E. D. Kiehn. 1968. Specific cleavage of viral proteins as steps in the synthesis and maturation of enteroviruses. Proc. Natl. Acad. Sci. USA 60:1015–1022.
13. Jacobson, M. F., and D. Baltimore. 1968. Polypeptide cleavages in the formation of poliovirus proteins. Proc. Natl. Acad. Sci. USA 61:77–84.
14. Kacian, D. L., and J. C. Myers. 1976. Synthesis of extensive, possibly complete, DNA copies of poliovirus RNA in high yields and at high specific activities. Proc. Natl. Acad. Sci. USA 73:2191–2195.
15. Kamphuis, I. G., J. Drenth, and E. N. Baker. 1985. Thiol proteases. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. J. Mol. Biol. 182:317–329.
16. Kitamura, N., C. J. Adler, P. G. Rothberg, J. Martinko, S. G. Nathenson, and E. Wimmer. 1980. The genome-linked protein of

picornaviruses. VII. Genetic mapping of poliovirus VPg by protein and RNA sequence studies. Cell 21:295–302.
17. Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. van der Werf, C. W. Anderson, and E. Wimmer. 1981. Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature (London) 291:547–553.
18. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105–132.
19. Larsen, G. R., C. W. Anderson, A. J. Dorner, B. L. Semler, and E. Wimmer. 1982. Cleavage sites within the poliovirus capsid protein precursors. J. Virol. 41:340–344.
20. Lomonosoff, G. P., and M. Shanks. 1983. The nucleotide sequence of cowpea mosaic virus B RNA. EMBO J. 2:2253–2258.
21. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.
23. Melnick, J. L., and H. A. Wenner. 1969. The enteroviruses, p. 529–602. In E. H. Lennette and N. J. Schmidt (ed.), Diagnostic procedures for viral and rickettsial disease, 4th ed. American Public Health Association, New York.
24. Nomoto, A., B. Detjen, R. Pozzatti, and E. Wimmer. 1977. The location of the polio genome protein in viral RNAs and its implication for RNA synthesis. Nature (London) 268:208–213.
25. Pallansch, M. A., O. M. Kew, B. L. Semler, D. R. Omilianowski, C. W. Anderson, E. Wimmer, and R. R. Rueckert. 1984. Protein processing map of poliovirus. J. Virol. 49:873–880.
26. Palmenberg, A. C., E. M. Kirby, M. R. Janda, N. L. Drake, G. M. Duke, K. F. Potratz, and M. S. Collett. 1984. The nucleotide and deduced amino acid sequences of the encephalomyocarditis viral polyprotein coding region. Nucleic Acids Res. 12:2969–2985.
27. Palmenberg, A. C., M. A. Pallansch, and R. R. Rueckert. 1979. Protease required for processing picornaviral coat protein resides in the viral replicase gene. J. Virol. 32:770–778.
28. Palmenberg, A. C., and R. R. Rueckert. 1982. Evidence for intramolecular self-cleavage of picornaviral replicase precursors. J. Virol. 41:244–249.
29. Pelham, H. R. B. 1978. Translation of encephalomyocarditis virus RNA in vitro yields an active proteolytic processing enzyme. Eur. J. Biochem. 85:457–462.
30. Polgar, L., and P. Halasz. 1982. Current problems in mechanistic studies of serine and cysteine proteinases. Biochem. J. 207:1–10.
31. Racaniello, V. R., and D. Baltimore. 1981. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. Proc. Natl. Acad. Sci. USA 78:4887–4891.
32. Roychoudhury, R., and R. Wu. 1980. Terminal transferase-catalyzed addition of nucleotides to 3′ termini of DNA. Methods Enzymol. 65:43–62.
33. Rueckert, R. R., T. J. Matthews, O. M. Kew, M. A. Pallansch, C. McLean, and D. R. Omilianowski. 1979. Synthesis and processing of picornaviral polyprotein, p. 113–125. In R. Perez-Bercoff (ed.), The molecular biology of picornaviruses. Plenum Publishing Corp., New York.
34. Rueckert, R. R., and E. Wimmer. 1984. Systematic nomenclature of picornavirus proteins. J. Virol. 50:957–959.
35. Ruether, U. 1982. pUR250 allows rapid chemical sequencing of both DNA strands of its inserts. Nucleic Acids Res. 10:5765–5772.
36. Sangar, D. V. 1979. The replication of picornaviruses. J. Gen. Virol. 26:227–238.
37. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
38. Semler, B. L., C. W. Anderson, N. Kitamura, P. G. Rothberg, W. L. Wishart, and E. Wimmer. 1981. Poliovirus replication

proteins: RNA sequence encoding P3-1b and the sites of proteolytic processing. Proc. Natl. Acad. Sci. USA **78:** 3464–3468.

39. **Semler, B. L., R. Hanecak, C. W. Anderson, and E. Wimmer.** 1981. Cleavage sites in the polypeptide precursors of poliovirus protein P2-X. Virology **114:**589–594.

40. **Skern, T., W. Sommergruber, D. Blaas, P. Gruendler, F. Fraundorfer, C. Pieler, J. Fogy, and E. Kuechler.** 1985. Human rhinovirus 2: complete nucleotide sequence and proteolytic processing signals in the capsid protein region. Nucleic Acids Res. **13:**2111–2126.

41. **Stanway, G., P. J. Hughes, R. C. Mountford, P. D. Minor, and J. W. Almond.** 1984. The complete nucleotide sequence of a common cold virus: human rhinovirus 14. Nucleic Acids Res. **12:**7859–7875.

42. **Stott, E. J., and R. A. Killington.** 1972. Rhinoviruses. Annu. Rev. Microbiol. **26:**503–524.

43. **Summers, D. F., and J. V. Maizel.** 1968. Evidence for large precursor proteins in poliovirus synthesis. Proc. Natl. Acad. Sci. USA **59:**966–971.

44. **Summers, D. F., J. V. Maizel, and J. E. Darnell.** 1965. Evidence for virus-specific noncapsid proteins in poliovirus-infected cells. Proc. Natl. Acad. Sci. USA **54:**505–513.

45. **Sweet, R. M., and D. Eisenberg.** 1983. Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. J. Mol. Biol. **171:**479–488.

46. **Tai, J. Y., A. A. Kortt, T.-Y. Liu, and S. D. Elliott.** 1976. Primary structure of streptococcal proteinase. III. Isolation of cyanogen bromide peptides: complete covalent structure of the polypeptide chain. J. Biol. Chem. **251:**1955–1959.

47. **Vieira, J., and J. G. Messing.** 1982. pUC-plasmids, a M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.

48. **Yogo, Y., and E. Wimmer.** 1972. Polyadenylic acid at the 3'-terminus of poliovirus RNA. Proc. Natl. Acad. Sci. USA **69:**1877–1882.