Molecular cloning and complete sequence determination of RNA genome of human rhinovirus type 14

(recombinant DNA/translation frame/sequence comparisons/proteolytic cleavage sites)

PIA L. CALLAHAN, SATOSHI MIZUTANI, AND RICHARD J. COLONNO

Department of Virus and Cell Biology Research, Merck Sharp & Dohme Research Laboratories, West Point, PA 19486

Communicated by Edward M. Scolnick, October 3, 1984

ABSTRACT The genomic RNA of human rhinovirus type 14 was cloned in *Escherichia coli* and the complete nucleotide sequence was determined. The RNA genome is 7212 nucleotides long. A single large open reading frame of 6536 nucleotides was identified, which starts at nucleotide 678 and ends 47 nucleotides from the 3' end of the RNA genome. Comparisons of the specified proteins with those of other picornaviruses showed a striking homology (44–65%) between rhinovirus and poliovirus. The rhinovirus genomic RNA is rich in adenosine (32.1%) and strongly favors an adenosine or uridine in the third position of codons. The predicted map locations of all the rhinovirus structural and non-structural proteins and their proposed proteolytic cleavage sites are described.

Rhinoviruses are the most important common cold viruses known. The name "rhinovirus" reflects the prominent nasal involvement seen in infections with these viruses. Human rhinoviruses (HRVs) are members of the picornavirus family, which also contains the entero- (polio, coxsackie), cardio- [encephalomyocarditis (EMC), Mengo], and aphthoviruses [foot-and-mouth disease virus (FMDV)] (1). As many as 115 HRV serotypes have been identified (2). Similar to other picornaviruses, the rhinoviruses contain a single-stranded RNA genome that serves as a monocistronic mRNA for the synthesis of all viral structural and non-structural proteins. Studies with poliovirus have indicated that a protein (VPg) is attached by a tyrosine- O^4 -phosphodiester bond to the 5' pUp of the RNA genome. In addition, all the picornaviruses contain a short poly(A) stretch at the 3' end of their genomic RNA (3). A variety of experimental results have suggested that the genomic RNA of picornaviruses is translated into a single polypeptide from which functional viral proteins are derived by proteolysis.

The fact that the RNA genome of picornaviruses is infectious, illustrates the importance of knowing the sequence of the RNA genome. The complete nucleotide sequences of the genomic RNAs of polio, EMC, and FMD viruses recently have been determined (4–7). We now report the sequence of the RNA genome of a single HRV serotype.

MATERIALS AND METHODS

Virus Growth and Purification. HRV type 14 (HRV-14) was obtained from the American Type Culture Collection and plaque-purified by standard techniques. The growth and purification of HRVs have been described recently (8).

Genomic RNA Isolation. Purified HRV-14 was digested with proteinase K (0.5 mg/ml) in 10 mM Tris Cl/1 mM EDTA/0.3% NaDodSO₄, pH 7.5, for 30 min at 37°C. Viral genomic RNA was separated from digested protein by oligo(dT)-cellulose chromatography (9). After precipitation with ethanol, the RNA was suspended in 0.3 ml of 10 mM Tris Cl/1 mM EDTA/0.5% NaDodSO₄, pH 7.5, and layered on a 12-ml preformed 15–30% (wt/wt) sucrose gradient in the same buffer but containing 0.1 M NaCl. RNA was sedimented at 23,000 rpm for 17 hr at 4°C in a Beckman SW40 rotor. Gradients were fractionated (0.4 ml) with an ISCO gradient fractionator and the A_{260} of each fraction was measured. A major peak of viral RNA sedimenting at 35 S, relative to the position of marker 18S and 28S RNAs in a parallel gradient, was pooled and precipitated with ethanol.

Construction of Double-Stranded cDNA. The procedures used for making cDNA to the genomic RNA and converting the cDNA to oligo(dC)-tailed, double-stranded cDNA were as described by Maniatis *et al.* (10). To insert the oligo(dC)-tailed cDNA into pBR322, 25 ng of pBR322-oligo(dG) cloning vector was annealed with 13, 26, or 39 ng of oligo(dC)-tailed cDNA and used to transform calcium-treated *Escherichia coli* strain RRI (10). The transformants were selected on tetracycline-containing agar plates.

Characterization of Selected Clones. Colonies (354) were picked and each was used to inoculate 2 ml of L broth containing 12 mg of Na₂HPO₄, 6 mg of KH₂PO₄, 1 mg of NaCl, 2 mg of NH₄Cl, and 4 mg of glucose. The cultures were shaken at 37°C for 16 hr, then 1.5 ml of each culture was pelleted by centrifugation. Plasmid DNA was isolated by using the miniprep protocol of Holmes and Quigley (11). The resulting plasmid DNA preparations were each suspended in 50 μ l of H₂O, 10 μ l of each was digested with the restriction endonuclease *Pst* I, and the size of its cDNA insert was determined by electrophoresis in 1% agarose gel. Clones containing inserts of at least 1500 base pairs (bp) were mapped by colonyhybridization studies (12).

Construction of Deletion Subclones. Deletion clones were constructed as described by Hong (13). Briefly, 1 μ g of isolated cDNA insert from either clone 7 or clone 186 was ligated to 2.8 μ g of linearized, *Pst* I-digested pUC9 DNA (10). *E. coli* strain HB101 was transformed with each of the ligation mixtures and transformants were selected on agar plates containing ampicillin at 100 μ g/ml. pUC9 plasmids containing clone-7 or clone-186 inserts were linearized with DNase I and digested with *Sal* I, and the resulting overhanging ends were filled in with DNA polymerase (10). After blunt-end ligation, the DNA was used to transform *E. coli* strain HB101 cells. Transformants were screened for size of inserts, and clones were selected that represented a distribution of sizes.

Nucleic Acid Sequencing of cDNA Inserts. Digestion of clones 7, 57, and 186 and their subsets of deletion clones with various commercially available restriction enzymes generated a set of overlapping DNA fragments. These fragments were ³²P-labeled as described (10). Labeled DNA fragments were asymmetrically cut with a second restriction enzyme to generate fragments with only one radioactively

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HRV, human rhinovirus; EMCV, encephalomyocarditis virus; FMDV, foot-and-mouth disease virus; PV-1, poliovirus type 1; bp, base pair(s).

labeled end. After isolation by electrophoresis in 5% acrylamide/bisacrylylcystamine gels (14), labeled DNA fragments were sequenced by using the chemical methods described by Maxam and Gilbert (15). Alternatively, synthetic deoxynucleotide primers and Sanger dideoxy sequencing techniques (16) were used to complete the determination of the nucleotide sequence. Both strands of cDNA were sequenced in most instances, and all sequence reactions were done at least twice. A continuous DNA sequence of 7227 nucleotides was obtained using the Intelligenetics software package. All ³²Plabeled triphosphates were purchased from Amersham. DNA-modifying enzymes and restriction endonucleases were purchased from Boehringer Mannheim.

RESULTS

Cloning of Rhinovirus Double-Stranded cDNA. Genomic RNA from plaque-purified HRV-14 was extracted and then purified by oligo(dT)-cellulose chromatography and sucrose gradient centrifugation. The 35S genomic RNA was used as a template for the synthesis of cDNA, using reverse transcriptase and an oligo(dT) primer. Following second-strand synthesis with DNA polymerase I, the cDNA was trimmed with nuclease S1 and tailed with oligo(dC) prior to insertion into the Pst I site of oligo(dG)-tailed pBR322. The newly constructed plasmids were then used to transform competent E. coli RRI cells. Each resulting cDNA clone was characterized for the size of its cDNA insert. Clones containing inserts of >1500 bp were analyzed by restriction endonuclease mapping and colony hybridization to determine their relationship to each other. Clones 7 (3230 bp), 57 (1910 bp), and 186 (3223 bp) were found to overlap and represent a 6784-nucleotide segment of the genomic RNA. Clone 186, which contained a 3' poly(A) of 15 nucleotides hybridized to total cellular RNA from infected but not uninfected cells, and thus contained viral-encoded sequences (data not shown). The precise map location of each clone was determined from restriction digests and is shown in Fig. 1. Since clone 7 did not appear to extend far enough to represent the 5'-terminal region, a synthetic primer complementary to viral genome nucleotides 864-878 was used to generate a new set of cDNA clones as above. A new cDNA clone designated 198 was isolated, which overlapped cDNA clone 7 and extended 430 nucleotides toward the 5' end of the genome

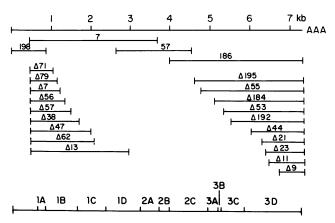


FIG. 1. Mapping of HRV-14 cDNA clones to the viral RNA genome. The topmost line represents the HRV-14 RNA genome; numbers above the line indicate sequence positions in kilobases (kb), and AAA, the poly(A) tail. The positions of cDNA clones 198, 7, 57, and 186 were determined by restriction endonuclease mapping. Deletion clones (designated by Δ) were constructed from cDNA clones 7 and 186 as described in *Materials and Methods*. The HRV-14 protein map is shown at the bottom and was determined by analogy to poliovirus. Protein nomenclature is that of Rueckert and Wimmer (17).

RNA. Sequence data (see below) indicated that clone 198 represents a sequence starting at genomic RNA nucleotide 13.

Construction of Deletion Clones. To aid in the rapid sequencing of cDNA clones 7 and 186, a subset of deletion clones representing various deletions from the 5' and 3' ends of clones 186 and 7, respectively, were constructed. Inserts of clones 7 and 186 were put into pUC9 plasmids that had been linearized with DNase I and cut with the restriction endonuclease *Sal* I. No *Sal* I or *Pst* I sites were found in the HRV-14 sequence. Following ligation and transformation of *E. coli*, clones were selected to give a set of clones that had inserts corresponding to various lengths of the original clones 7 and 186 (Fig. 1).

Nucleotide Sequence of the Rhinovirus Genome. The sequence of cDNA clone inserts were determined predominantly (>95%) by using the Maxam-Gilbert chemical sequencing method (15). The 5' end of the RNA genome and regions lacking convenient restriction sites were sequenced by using dideoxy sequencing methods (16) in which synthetic primers were synthesized from sequenced regions and used to prime cDNA synthesis on the genomic RNA template. Deletion clones were labeled at the EcoRI or BamHI sites of the pUC9 polylinker and sequenced by using the Maxam-Gilbert method. The first 6 nucleotides at the 5' end were confirmed by sequential single-nucleotide extension of a ³²P-labeled synthetic oligodeoxynucleotide representing genome nucleotides 7-27. The results of this experiment confirmed the 5' sequence U-U-A-A-A for the HRV-14 RNA genome, since addition of only dTTP extended the primer by 4 nucleotides and addition of dATP and dTTP extended the primer by 6 nucleotides, whereas addition of dCTP and dGTP had no effect (data not shown). The first 10 nucleotides of poliovirus type 1 (PV-1) and HRV-14 are identical.

The Intelligenetics software package was used to merge sequence gel readings into a single DNA sequence containing 7227 nucleotides including the poly(A) tail (Fig. 2). Computer-generated translation of the DNA sequence in all three reading frames (Fig. 3) revealed a single long open reading frame 6537 nucleotides long. The open reading frame initiates at nucleotide 629, following four unused AUG codons in the same reading frame at nucleotide positions 59, 332, 395, and 419. In addition, there are nine AUG codons in the other two reading frames prior to the AUG at position 629.

Base composition of the HRV-14 genomic RNA, minus the poly(A), showed a high adenine content (32.1%), followed by uracil (27.3%), guanine (20.5%) and cytosine (20.1%). Codon frequency analysis supported the high adenine content by showing a strong preference for adenosine and uridine in the third position of codons. The triplet CGG does not occur at all, and GCG and UCG occur only two and three times, respectively. The predicted polyprotein contains only 1.1% tryptophan, 2.0% cysteine, 2.4% histidine, and 2.4% methionine, but it is rich in leucine (8.9%) and threonine (8.0%). The computer-generated lengths and molecular weights for all the viral proteins are listed in Table 1. This assignment of protein map positions was done solely by analogy to the PV-1 sequence. The postulated proteolytic cleavage sites (Table 2) must be considered tentative until protein sequences are obtained.

DISCUSSION

HRVs are the leading cause of the common cold in humans (18). Although up to 115 antigenically distinct serotypes have been identified, little is known at the molecular level about any of the viruses. A knowledge of the RNA genome sequence is necessary for a full understanding of the functions of the encoded proteins involved in virus structure

20 40 60 80 100 TTAAAACAGC GGATGGGTAT CCCACCATTC GACCCATTG GIGTAGTACT CTGGTACTAT GIACCETTGT ACGCCTGCTT AMANTCCCA
140 160 180 200 220 240 260 280 CCCATGAAAC GTTAGAAGGT GGACATTAAA GTACAATAGG TGGCGCCCATA TCGTAGGGGT TCTATGTACA AGCACITCIG TTTCCCAGGA GGGAGGGTATA GGCGGGGGGGAGGTATA GGCCGACGT AGCCAACTAC GTAGCAGGTA GTACCAACTAC GTAGCAGGTA GTACCAACTAC GTAGCAGGTA GTACCAACTAC GTAGCAGGTA GTACCAACTAC GTAGCAGGTA GTACCAACTAC GTAGCAGGTA GTACGAGTA GTACCAACTAC GTAGCAGGTA GTACGAGTA
GARGETICA TCANGTGAT TITECTICA CTAGTITGGT COATGAGGET AGGATITGCT CAGGGGTGAE GETGTCTAG CETGGGGGG GGCAACCCA GETIATGCTG GGAGGGCCTT TTAGGAGAT GETGGAGA CTGCATGTG CTGGTTGTG AGTCCTGGG CCCCTGAATG
CONCERTANCET TANCETARGE CACATEGA GOLTATGEC ACGATECAGE GOTTETAAGE CONTREGE CANTECEGE ACGEGACEGA CTACTITIGE TETECATTET CITEATATE TETTATET CATATAGATATA TATACATATA CTETEAT 11 A (PPA) 643 763 763 763 763 763 763 763 763 763 76
NET GLY ALS GLT WAL SET THE GLI VAS SET GLY SET HIS GLU ASS GLI ASS GLI ASS GLI ALS ALS GLA
Ser Lys Phe Thr Glu Pro Val Lys Asp Leu MET Leu Lys Gly Ala Pro Ala Leu Asn Ser Pro Asn Val Glu Ala Cy Gly Tyr Ser Asp Arg Val Glu Glu Glu Glu Glu Glu Gly An Leu Gly Asn Ser Thr I Glu Thr Glu Glu Ala Ala Asn Ala 943 103 103 103 103 103 103 103 103 103 10
103
TTA CCA CAT GCA TCC ANG GAT ATG GGT GTG TTG GGG CAA AAC ATG TTT TTC CAC TCA CTA GGA AGA TCA GGT TAC ACA GTA CAC GTA CAC GTA CAC GTA CAC GTA CAC AAA TTC CAT AGC GGT TGT CTA CTT GTA CTT GTA TATA CCA GAA CAC CAA Leu Pro Asp Ala Leu Lys Asp MET GLY Val Phe Gly Gln Asn MET Phe Phe His Ser Leu Gly Arg Ser Gly Tyr Thr Val His Val Gln Cys Asn Ala Thr Lys Phe His Ser Gly Cys Leu Leu Val Val Val Val Val ILe Pro Glu His Gln 1203 1303
LIG GCT TCA CAT CAC GCT GGC AAT GTT TCA GTT AAA TAC ACA TTC AGG CAT CCA GGT GAA CGT GGT ATA GAT TTA TCA TCT GCA AAT GAA GTG GGA GGG CCT GTC AAG GAT GTC ATA TAC AAT ATG GAT GGT ACT TTA TTA GGA AAT CTG Leu Ala Ser His Clu Gly Gly Asn Val Ser Val Lys Tyr Thr Phe Thr His Pro Cly Clu Arg Cly ILe Asp Leu Ser Ser Ala Asn Clu Val Cly Cly Pro Val Lys Asp Val ILe Tyr Asn MET Asn Cly Thr Leu Leu Gly Asn Leu
CTC ATT TTC CCT CAC GAC TTC ATT TAAT CTA AGA ACC AAT AAT ACA GCC GAC ATA GTG CATA CTA CATA AAC TCA GTA CCC ATT CAT TAA TCA CAC GTT CAC ATA GTG TCA CTG ATC CTC ATT GTC CTT ATT GTC CTT AGA CAAT GTT CAC TTG TCA CTA TTG GTC CTT TAAT CTA CAA CAAT GTT CAC TTG TCA CTA TTG CAC ATT GTC CTT AGA CAAT GTT CAC ATT ATT GTC CTT AGA CAAT GTT CTT ATT GTT CTT ATT GTT ATT GTT CTT ATT GTT ATT GTT ATT GTT CTT ATT GTT ATT GTT ATT GTT ATT GTT ATT GTT ATT GTT CTT ATT GTT ATT
ACT GGA GCA ACT CCC TCA CTC CCT ATA ACA GTC ACA ATA GCA CCT ATG TGC ACT GAG TTC TCT GGG ATA AGG TCC AAGT CCA AGT GCA CAAGT TG CCA ACT ACA ACT TTG CCG GCG TCA GGA CAAT TTC TTG ACC ACA GAT GAC AGG Thr Gly Ala Thr Pro Ser Leu Pro Ile Thr Val Thr Ile Ala Pro MET Cys Thr Glu Phe Ser Gly Ile Arg Ser Lys Ser Ile Val Pro Gln Gly Leu Pro Thr Thr Thr Leu Pro Gly Ser Gly Gin Phe Leu Thr Thr Asp Asp Arg
1813 CAA TCC CCC AGT GCA CTG CCA AAT TAT GAG CCA ACT CCA AGA ATA CAC ATA CTA GGG AAA GTT CAT AAC TTG CTA GAA ATT ATA CAG GTA GAT ACA CTC ATT CCT AGA AAT AAC AGG GTA GAG GTT AAC AGT TAC CTC ATA Gin Ser Pro Ser Ais Leu Pro Asm Tyr Glu Pro Thr Pro Arg 11e His 11e Leu Gly Lys Val His Asm Leu Leu Glu 11e 11e Gin Vai Asp Thr Leu 11e Pro MET Asm Asm Thr His Thr Lys Asp Glu Val Asm Ser Tyr Leu 11e
CCA CTA MAT GCA AAC AGG CAA AAT GAG CAG GTT TTT GGG ACA AAC CTG TTT ATT GGT GAT GGG GTG GTG GGG GTC TTC AAA ACT ACT CTT CTG GGT GAA ATT GTT CAG TAC TAT TGG TCG ATG TTT GGT GAT TACT GGT Pro Leu Asn Ale Asn Arg Gln Asn Glu Gln Val Phe Gly Thr Asn Leu Phe 11e Gly Asp Gly Val Phe Lyg Thr Thr Leu Leu Gly Clu Lie Val Gin Try Tyr Thr His Try Ser Gly Ser Leu Asp Ass Ser MET Try Thr Clu
2033 2013 2013 2013 2013 2013 2013 2013
2203 2233 2233 2233 2233
Gly Val Gin Phe Arg Tyr Thr Asp Pro Asp Thr Tyr Thr Ser Ala Gly Phe Leu Ser Cys Try Tyr Gin Thr Ser Leu Ile Leu Pro Pro Glu Thr Thr Gly Gin Val Tyr Leu Leu Ser The Ile Ser The Ile Ser Ala Cys Pro Lasp The Lys Leu 2233 AGG CTG ATG AAA GAT ACT CAA AGT ATC TCA AGA CAT GTG CAA CTG AGTG CAT TA GAA GAA GTG ATT C GTG GAA AGA AGA CAG GG GCC CAA ATG ACT ACT CAT GAA GAA GTC ACT CAT TA GAA GAA GTC ATG GTG GCC CAA TA CAG AGA GTC ATG CAG AAA GTC ATG CAG GAA GTC ATG CAG GAA GTC ATG CAG AGA GTC ATG CAG GAA GTC ATG GAA GTC ATG GAA GAA GTC ATG GTG GTC CAG ATA GTC ATG CAG GAA GAA GTC ATG GAA GTC AT
Arg Leu MET Lys Asp Thr Gin Thr Tie Ser Gin Thr Val Ale Leu Thr Glu Gij Asp Glu Leu Gly Gha And Chu Ch Ch Chu Glu Gly Cha And Chu Ch Ch Cha And Arg Chu Ch Cha And Arg Chu Ch Cha And Arg Chu Cha Cha And Chu Cha And Chu Cha And Chu Cha Cha And Chu Cha Cha And Chu Cha Cha And Chu Cha Cha Cha And Chu Cha Chu
Leu Thr Ala Aan Glu Thr Gly Ala Thr MET Pro Val Leu Pro Ser Asp Ser Ile Glu Thr Arg Thr Thr Tyr MET His Phe Asn Gly Ser Glu Thr Asp Val Glu Cys Phe Leu Gly Arg Ala Ala Cys Val His Val Thr Glu Ile Gln 2633 2653 2653
AAC AAA CAT GET ATT GEA ATA CAT AGE GAA GAA GEA GAA TTG TTC AAT GAT TGG AAA ATC AAC CTG TGC AGE CTT GTC AAC TT AGA AAG AAA CTG GAA CTC TTC ACT TAT GTT AGG TTT GAT TGT GAG TAT ACC ATA CTG GCC ACT Asn Lys Asp Ala Thr Gly lle Asp Asn His Arg Glu Ala Lys Leu Phe Asn Asp Trp Lys lle Asn Leu Ser Ser Leu Val Glu Leu Arg Lys Lys Leu Clu Leu Phe Thr Tyr Val Arg Phe Asp Ser Glu Tyr Thr lle Leu Ala Thr 2743 2763
GCA TCT CAA CCT GAT TCA GCA AAC TAT TCA AGC AAT TTG GTG GTC CAA GCC ATG TAT GTT CCA CAT GGT GCC CCG AAA TCG AAA GGT GGC GAT TAC ACA GTG GCA AGT GCT TCA AAC CCC AGT GTA TTC TTC AAC GTG GGG GAT ACA Ala Ser Cln Pro Asp Ser Ala Asn Tyr Ser Ser Asn Leu Val Val Gln Ala MET Tyr Val Pro Nts Gly Ala Pro Lys Ser Lys Arg Val Gly Asp Tyr Thr Trp Gln Ser Ala Ser Asn Pro Ser Val Phe Phe Lys Val Gly Asp Thr 2893 2013
TCA AGG TTT ACT GTG CCT TAT GTA GGA TTG GCA TCA GCA TAT AAT TGT TTT TAT GAT GGT TAC TCA CAT GAT GAT GGA GAA ACT CAG TAT GGC ATA ACT GTT GTC AAAC CAT AAT GGT AGT GGC ATA GTA ATA GTA AAT GAA CAT GAT Ser Arg Phe Ser Val Pro Tyr Val Cly Leu Ala Ser Ala Tyr Aan Cys Phe Tyr Aap GLy Tyr Ser His Aap Aap Ala GLU Thr Gln Tyr GLy Lle Thr Val Leu Aan His MET GLy Ser MET Ala Phe Arg ILe Val Aan GLU His Aap 3043 3133 3163
GAA CAC AAA ACT CIT GTC AAG ATT TAT CAC AGG GCA AAG CTC GTT GAA GCA TGG ATT CAC AGG GCA CAG AGG CCC CAC ACT CAT CAC TA CGG GCC ACA AT TAT CCT AAG CAN CCA CTA ATT AAG CAN CCA CTA CTA CTA CAG CAN CCA CTA CTA CTA CTA CAG CAN CCA CTA CTA CTA CTA CAG CAN CCA CTA CTA CTA CTA CTA CTA CTA CTA CTA
3193 123 GOT GGC AT AM TCC TATAGCAT TAG GA CCT AGG TA GGC TG GATA GTC AAA TG TT AAA AT AT CA CA CTG ATG ACC GGA GAC GAC CAT AM TCC ATA GCA TG TG AGA GAC GAC CAT AM TCC ATA GCA TG TG AGA GATA GTC CATA GCA GATA GTC CA ATA GGC ATA GTC CA ATA GTC ATA GA ATA ATA ATA ATA ATA ATA ATA ATA
GGA GGA CAT GGT GCA GAA ACA ATA CCA CAC TGT AAC CGT ACA TCA GGT GTT TAC TAT TCC ACA TAT TAC AGA AAG TAT TAC CCC ATA ATT TGC GAA AAG CCC ACT ATT CG ATT GAA GGA AGC CCT TAT TAC CCA AGT AGA TTT CAA Cly Gly His Cly Ala Glu Thr lle Pro His Cys Asn Arg Thr Ser Gly Val Tyr Tyr Ser Thr Tyr Tyr Arg Lys Tyr Tyr Pro Ile Ile Cys Glu Lys Pro Thr Asn Ile Trp Ile Glu Gly Ser Pro Tyr Tyr Pro Ser Arg Phe Gln
3493 3513 3583 3613 3613 3613 3613 3613 3613 3613 36
128 3643 3703 3763 3763 3763 3763 3763 3763 1763 GAT AGE GAT AGE GAT AGE GAT CACA GAA GTC ACCA GAA CTA CAA GAT CTC CTC ACC ACA AAA GTT TTG CTC ACG CAA ATC GTT GT GTC GTC ACA GAA GTC ACCA GAA GTC ACCA GAA GTC GCC GAA GAT TTG CTC ACC ACA AAA GTC TTG GTC GTC AAA ATC GTT GL U GL GL U GL GL U GL AL BLAS ASP PHE LEU TH THT LAS LEVE SET LASP ST AND ATC GTT GTC GTC GTC ACT ACT ACT ACT ACT ACT ACT ACT ACA GAT GTC GTC ACT ACT ACT ACT ACT ACT ACT ACT ACT A
3933 383 383 383 383 383 383 383 383 383
4003 4003 4003 4003 4003 4003 4003 4003
Asn Asp Gly Trp Phe Arg Lys Phe Asn Asp Ala Cys Asn Ala Ala Lys Gly Leu Glu Trp Ile Ala Asn Lys 118 Ser Lys Leu Ile Glu Trp Ile Lys Asn Lys Val Leu Pro Gln Ala Lys Glu Lys Leu Glu Phe Cys Ser Lys Leu 403 413 413 413 413 413 413 413 413 413 41
Lys Gin Leu Asp lie Leu Giu Arg Gin lie Thr Thr MET His lie Ser Asn Pro Thr Gin Glu Lys Arg Giu Gin Leu Phe Asn Asn Val Leu Trp Leu Giu Gin MET Ser Gin Lys Phe Ala Pro Phe Tyr Ala Val Giu Ser Lys Arg 4030 4030 4030 4030 4030 4030 4030 403
ILE Arg Glu Leu Lys Asn Lys MET Val Asn Tyr MET Gln Phe Lys Ser Lys Gln Arg Thr Glu Pro Val Cys Val Leu ILe His Gly Thr Pro Gly Ser Gly Lys Ser Leu Thr Thr Ser ILe Val Gly Arg Ala ILe Ala Glu His Phe 4453 AAT TCA GCA GTA TAT CCA CTA CCA GAT CCC AAG CAC THT GAT GGT TAT CAG GAA GTT GTC ATT ATG GAT GTAT CTG AAA TCC GAC GGA CAT GAT ATA AGC ATG TTT GTC CAA ATG GTT TCT TCA GTG GAT TTT GT 4451 AAT TCA GCA GTA TAT CAC GTT CCA CCA GAT CCC AAG CAC THT GAT GGT TAT CAG GAA GTT GTC ATT ATG GAT GTT GTC GAT ATA AGC ATG GTT TGT CTAA ATG GTT TCT TCA 4451 AAT TCA GCA GTA TAT CAC TT CCA CCA GAT CCC AAG CAC THT GAT GGT TAT CAG GAA CAG GAA GTT GTC ATT ATG GAT GAT CTG AAC CCAA AAT CCA GAT GGA CAT GAT ATA AGC ATG TTT TGT CAA ATG GTT TCT TCA 4451
Asn Ser Ala Val Tyr Ser Leu Pro Pro Asp Pro Lys His Phe Asp Gly Tyr Gln Gln Gln Glu Val Val I le MET Asp Asp Leu Aan Gln Asn Pro Asp Gly Gln Asp The Leu Gan Rit Gli Gln Glu Val Ser Ser Val Asp Phe Leu 4603 4513 4623 4633 4632 4632 4633 4632 4633 463 46
Pro Pro HET Ala Ser Leu Asp Asn Lys Gly HET Leu Phe Thr Ser Asn Phe Val Leu Ala Ser Thr Asn Ser Asn Thr Leu Ser Pro Pro Thr Ile Leu Asn Pro Glu Ala Leu Val Arg Arg Phe Gly Phe Asp Leu Asp Ile Cys Leu 4693 4723 473 473 4813 4813
La nor nor no chan and nor and nor out and the set of the act and the last nor set of an internet and and the set of the
Tyr Ser Val Asp Gin Leu Val Thr Ala Ile Ile Ser Asp Phe Lys Ser Lys MET Gin Ile Thr Asp Ser Leu Glu Thr Leu Phe Gln'Gly Pro Val Tyr Lys Asp Leu Glu Ile Asp Val Cys Asn Thr Pro Pro Ser Glu Cys Ile Asn 4993 5023
GAT TTA CTG AAA TCT GTA GAT TCA GAA CAG ATT AGG GAA TAT TCT AAG AAG AAA TGG ATT ATA CCT GAA ATT CCT ACA TAT GAA AGG GGT ATG AAT CAG CCAGC ATG ATT ATA ACT ATT ACT ATT CTG ATG TTT GTC AGT ACA TTA Asp Leu Ley Lys Ser Val Asp Ser Glu Glu Ile Arg Glu Tyr Cys Lys Lys Lys Trp lle Ile Pro Glu Ile Pro Th Asn Ile Glu Arg Ala MET Asn GCA ALS CAG ATG ATT ATA ACT ATT CTG ATG TTT GTC AGT ACA TTA Sp Leu Ley Lys Ser Val Asp Ser Glu Glu Ile Arg Glu Tyr Cys Lys Lys Lys Trp lle Ile Pro Glu Ile Pro Th Asn Ile Glu Arg Ala MET Asn GLn Ala Ser MET Ile Lle Asn Thr Ile Leu MET Phe Val Ser Thr Leu S143 S103 S103 S103 S103 S103 S103 S103 S10
5143 GGT ATT TAT GTC ATT TAT MA TTC TTT GCT CAA ACT CAAGEG CGT ATT CT GGT AAC CCG CCT CAC AAT AAA CTA AAA GCC CCA ACT TTA CCC CAC GTT GT CTC CCAGEGA ACC AACA GAA TTA GTA TAG GAAA Gly lie Val Tyr Val lie Tyr Lys Leu Phe Ala Gin Thr Gin Gly Pro Tyr Ser Gly Asn Pro Pro His Asn Lys Leu Lys Ala Pro Thr Leu Arg Pro Val Val Val Gin Gly Pro Asn Thr Glu Phe Ala Leu Ser Leu Leu Arg Lys 5293 5293
AAC ATA ATG ACT ATA ACA ACC TCA AAG GGA GAG TTC ACA GGG TTA GGC ATA CAT GAT CGT GTC TGT GTC ATA CCC ACA CAC GCA CAG CGT GGT GAT GTA GTA GTA GTG GAT GGT CAG AAA ATT AGA GTT AGA GTT AGA GAT AAG TAC AAA TTA GTA Asn lle HET Thr lle Thr Thr Ser Lys Gly Glu Phe Thr Gly Leu Gly Ile His Asp Arg Val Cys Val Ile Pro Thr His Ala Gln Pro Gly Asp Asp Val Leu Val Asn Gly Cln Lys Ile Arg Val Lys Asp Lys Tyr Lys Leu Val
GAT CCA GAC AAT AAT CTA GAG CTT ACA GTG TTG ACT TTA GAT AGA AAT GAA AAA TTC AGA GAT ATC AGG GGA TTT ATA TCA GAA GAT CTA GAA GGT GTG GAT GCC ACT TTG GTA GTA GTA CAT TCA AAT AAC TTT ACC AAC ACT ATC TTA ASP Pro Glu Asn Ile Asn Leu Glu Leu Thr Val Leu Thr Leu Asp Arg Asn Glu 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 Thr Asn Thr Ile Leu
GA OTT GGC CTT GTA ACA ATG GGA GGA CTT ATT ATT TTG AGT AGC AGC CCC AGT AGA AGA ATG TTT GTT ATT GAT AGA ATG GGA AGA TTG TTG
AT CCA AGA CAA CTA TH TCA CCT CAA CTT AAA AAA CAA TAT TT CTA GAG AAA CAA GGC CAA GTA ATA CCT AGA CAA MG CTT AGG GAG TT AGA CATA AAT CCA GGC GAA CTA AAA TTA CAT CCC AGG CAA TT AT AAA CGA AGA CAA CTA TH TCA CCT CAA CTT AAA AAA CAA TAT TT GTA GAG AAA CAA GGC CAA GTA ATA CCT AGA CTA AAA TCA CAT CCC AGT CAT TT TAT AAA CGI AGU Phe Ser Ala Gin Leu Lys Lys Cin Tyy Full Giu Lys Cin Giy Cin Vai lie Ala Arta His Lys Vai Ara Giu Phe Aan Ile Aan TCA VAI Ason Thr Ala Thr Lys Ser Lys Leu His Pro Ser Vai Phe Tyr Ser Vai Con Cin Phe Ser Ala Gin Leu Lys Lys Cin Tyy Full Giu Lys Cin Giy Cin Vai lie Ala Arta His Lys Vai Ara Giu Phe Aan Ile Aan Tro Vai Ason Thr Ala Thr Lys Ser Lys Leu His Pro Ser Vai Phe Tyr
GAT GTT TTT CCA GGT GAC AAG GAA CCT GCT GTA TTG AGT GAC AAT GAT CCC AGA CTC GAA GTT AAA TTG ACT GAA TCA TTA TTC TCT AAG TAC AAG GGG AAT GTA ATT ACG GAA CCC ACT GAA AAT ACG GTA GAT GCT GTA GGC GTT GTA GAC CAT TAT Asp Val Phe Pro Gly Asp Lys Glu Pro Ala Val Leu Ser Asp Asn Asp Pro Arr Leu Glu Val Lys Leu Thr Glu Ser Leu Phe Ser Lys Tyr Lys Glu Asn Thr Glu Pro Thr Glu Asn TTT Leu Val Asp Val Asn Tte GT
6043 613 613 613 613 613 613 613 613 613 61
613 623 623 623 623 623 623 623 623 623 62
AT GAT CT GT AGA ATG AAA ATG AGA ATG AAA CTA GGG ATG CT TAC AAA GA ATG CGA CAA GGT CTG CT GGA CCA GGG GGT CTG GAT CGA GTG CTG TG CTG CTG CTG CTG CTG CTG CTG
TTT GAT TAC TOT MAT TIT GAT GCC TOT TIG TCA CCA GTT TGC TGT GTG CTA GAG AAG GTT TTG ACT AGGC TTT GCA GGC TCT TCA TTA ATT CAA TCA ATT CAT AGC CAT CAT ATT CTT AGG GAT CAA ATA TAT CTT CTT
Phe Asp Tyr Ser Asm Phe Asp Ala Ser Leu Ser Pro Val Trp Phe Val Cys Leu Clu Lys Val Leu Thr Lys Leu Cly Phe Als Cly Ser Ser Leu Ile Cln Ser Tie Cys Asm Thr His His File Phe Val Cys Leu Clu Lys Val Val 6703 6643 6643 6673 6673 6673 6663 6673 667 GCC TCA GCG ACC ASC ASC TAT TIC ANT CC ATC ACC ATA ATC ATT AGG ACT TIG ATA TIA GAT GCA TAT GAT TIA GAT GAT ATA ATC CTT AGT ACT GAT GAT TIG ATA ATC TIA GCT TAT GAT GAT ATA ATC TIA GCT TIC GAT GAT TIG ATA ATC ATT AGG ACT TIG ATA ATC ATT AGG GAT ATG GAT TIA GAT GAT ATA GAT ATA GAT ATA ATC TIA GAT CAA ATC TIA GAT GAT GAT TIA GAT GAT GAT TIA GAT GAT GAT TIA GAT GAT GAT GAT GAT GAT GAT GAT GAT GA
Clu Cly Cly MET Pro Ser Cly Cys Ser Cly Thr Ser Ile Phe Asn Ser MET ILe Asn Asn Ile Ile Ile Arg Thr Leu Ile Leu Asp Ala Tyr Lys Cly Ile Asp Leu Asp Lys Leu Lys Ile Leu Ala Tyr Cly Asp Asp Leu Ile Val 6793 6793 679 679 679 679 679 679 679 679 679 679
Set by root by concerning the use of the concerning
Con the FTO THE LEW VAL HIS PTO VAL HET PTO HET LyS ASP ILE HIS GLU SET ILE ATG TTP THE LYS ASP PTO LYS ASP THE GIN ASP HIS VAL ATG SET LEW CAS MET LEW ALA TTP HIS SET GLY GLU LYS GLU TYT ASP GLU PHE 7093 7103 7103 7103 7103 7103 710 710 710 710 710 710 710 710 710 710
Ile Gin Lys Ile Arg Thr Thr Asp Ile Giy Lys Cys Leu Ile Leu Pro Giu Tyr Ser Val Leu Arg Arg Arg Trp Leu Asp Leu Phe

FIG. 2. Nucleotide and derived protein sequence of HRV-14. The complete nucleotide sequence of the HRV-14 RNA genome is presented. Putative cleavage sites (arrows) and assignment of viral proteins to specific locations are indicated.

Biochemistry: Callahan et al.

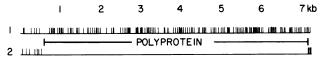


FIG. 3. Translation of HRV-14 genomic RNA. Computer translation of the sequence in Fig. 2 was performed in all three reading frames, numbered 1–3 on the left. The vertical lines designate the occurrence of stop codons. kb, Kilobases.

and replication. In addition, determination of the genome sequence allows definitive comparisons between the four families of picornaviruses to determine sequence conservation in viral proteins and evolutionary relationships between viruses.

We have cloned and sequenced the entire RNA genome of HRV-14. Four cDNA clones, designated 198 (866 bp, nucleotides 13-878), 7 (3231 bp, nucleotides 443-3673), 57 (1911 bp, nucleotides 2624-4534), and 186 (3224 bp, nucleotides 4004-7227), were isolated which overlapped and represented 99.8% of the entire genome sequence. Random deletion of the 3' and 5' ends of cDNA clones 7 and 186, respectively, yielded a subset of clones used to facilitate rapid sequencing of the cDNA. The map positions of all these clones are shown in Fig. 1.

Translation of the cDNA in all three reading frames shows a single large open reading frame 6637 nucleotides long with initiation 34 codons downstream at position 629. This would predict a 5' 628-nucleotide nontranslated region, which contains four unused AUG codons upstream and in the same reading frame. The four unused AUG codons would initiate peptides with amino acid lengths of only 2, 8, 10, and 24. Analysis of the other two reading frames in the 5' noncoding region shows nine additional initiation codons. Six of these would result in proteins <29 amino acids long. The remaining three (nucleotide locations 496, 526, and 595) fall within the first reading frame (Fig. 3) and could initiate an amino acid sequence of up to 77 amino acids, depending on which AUG was utilized as a start codon. Although these starts cannot be ruled out completely, they seem highly unlikely for the following reasons: First, they would result in relatively short peptides; second, they lack efficient flanking sequences for ribosome recognition as described by Kozak (19), which are present around the AUG at position 629; third, comparative studies among the three serotypes of poliovirus demonstrated that the AUG codons found in similar positions in the 5' noncoding region are not conserved, in contrast to the strong homology found within the true coding sequence (20). In addition, comparison of the 1A (first capsid protein to be translated) sequences of HRV-14 and PV-1 shows that initiation at position 629 of HRV-14 gives a 1A

Table 1. Size of predicted HRV-14 proteins

Protein	Nucleotides	Number of amino acids	Predicted M
Polyprotein	629-7165	2178	242,593
1A (VP4)	629-835	69	7,310
1B (VP2)	836-1621	262	28,506
1C (VP3)	1622-2329	236	26,198
1D (VP1)	2330-3196	289	32,384
2A	3197-3634	146	16,090
2B	3635-3925	97	10,796
2C	3926-4915	330	37,400
3A	4916-5170	85	9,751
3B (VPg)	5171-5239	23	2,470
3C	5240-5785	182	19,999
3D	57867165	460	52,217

Table 2. Predicted HRV-14 polyprotein cleavage sites

	leteu III v -14 polyprotein eleavage sites
Proteins	Cleavage site
1A/1B	-Pro-Ala-Leu-Asn/Ser-Pro-Asn-Val-
1 B /1C	-Ile-Val-Pro-Gln/Gly-Leu-Gly-Asp-
1C/1D	-Ala-Leu-Thr-Glu/Gly-Leu-Gly-Asp-
1D/2A	-Ile-Lys-Ser-Tyr/Gly-Leu-Gly-Pro-
2A/2B	-Ala-Glu-Glu-Gln/Gly-Leu-Ser-Asp_
2B/2C	-Ile-Glu-Arg-Gln/Ala-Asn-Asp-Gly-
2C/3A	-Thr-Leu-Phe-Gln/Gly-Pro-Val-Tyr-
3A/3B	-Ala-Gln-Thr-Gln/Gly-Pro-Tyr-Ser-
3B/3C	-Val-Val-Val-Gln/Gly-Pro-Asn-Thr-
3C/3D	-Val-Glu-Lys-Gln/Gly-Gln-Val-Ile-

sequence identical to that of PV-1 1A at positions 1-6 and at 13 of the first 17 amino acids.

Alignment of the HRV-14 5' noncoding sequence to the first 600 bases of the PV-1 sequence shows an RNA homology in the range of 75%, depending on how the alignment is done. Interestingly, the major difference between them occurs just 5' of the proposed initiation codon where HRV-14 is missing 114 bases that are present in the PV-1 5' noncoding region. This is the precise location where each of the different poliovirus serotypes differs dramatically (20). The function of this region in picornaviruses remains unknown. In addition, computer programs predict a stem and loop structure at the 5' end of the HRV-14 genome sequence that is very similar to that found in PV-1 (4).

Protein-to-protein comparisons between HRV-14, PV-1, EMCV, and FMDV sequences show that PV-1 and HRV-14 are closely related, whereas sequence homology to EMCV or FMDV is not apparent. The map positions of the viral proteins were determined solely by analogy to PV-1, since no amino acid sequence was determined directly from isolated proteins. However, based on the strong homology with PV-1 proteins and similarity in the predicted size of the HRV-14 viral proteins, the assignment of amino acid sequence to proteins appears to be correct. The general genomic organization is similar to those of other picornaviruses. Protein nomenclature used is the 4-3-4 system recently described by Rueckert and Wimmer (17). Comparisons of homologies for non-structural proteins between HRV-14 and PV-1 ranged from 44% in the protease gene 3C to 65% in the replicase gene 3D. Recent studies have postulated that the protease of picornaviruses is a cysteine protease with the cysteine at position 146 of the PV-1 3C within the active site (21). The protease gene of HRV-14 encodes the same conserved amino acid sequence (Gly-Gln-Cys-Gly-Gly-Val at residues 154-159) that is found in PV-1. A similar sequence (Gly-Trp-Cys-Gly-Ser-Ala) is found in EMCV protease (6).

Comparison of the four structural proteins gave interesting results. 1A (VP4), which is an internal structural protein, and 1B (VP2) show a 60% amino acid homology to 1A and 1B of PV-1. Although 1A shows conserved sequences throughout its length, 1B contains a very different region (amino acids 134–184), which corresponds to a PV-1 neutralization epitope (22) and is probably the same in HRV-14. Comparisons of 1D (VP1) and 1C (VP3), which contain most of the immunodominant neutralizing epitopes of PV-1, show only 47% and 44% homology, respectively. Comparisons of these sequences have also allowed us to predict possible neutralizing sites. For example, the amino acid sequences 62-95 of 1C (VP3) and 91-95 of 1D (VP1) are unique to HRV-14 and map to known neutralizing sites in PV-1 (ref. 23; E. Emini, personal communication). No open reading frame large enough to code for the leader proteins found in EMCV and FMDV is found 5' of the amino-terminal methionine codon of 1A (VP4) in HRV-14.

The viral-encoded protease of HRV-14, like those of other picornaviruses, is presumably crucial for proper proteolytic cleavage of the viral polyprotein and for effective replication of the virus. The predicted proteolytic cleavage sites are shown in Table 2. Like other picornaviruses, there is a predominance of Gln-Gly sites (6 of 10) in addition to a Glu-Gly, a Tyr-Gly, an Asn-Ser, and a Gln-Ala. Eight of the cleavage dipeptides are identical in both sequence and location to those in PV-1. The Glu-Gly cleavage site between 1C and 1D is not found in either PV-1 (4) or EMCV, but five are present in FMDV (7). Only one uncleaved Gln-Gly sequence exists in the HRV-14 sequence (within the protease region), but multiple uncleaved dipeptides identical to the other dipeptide cleavage sites are found. The Gln-Ala cleavage site, if utilized, is unique to HRV and not found in PV-1, EMCV, or FMDV. As in other picornaviruses, the flanking amino acid sequences around the cleavage sites do not fall into a specific pattern but do contain a proline either amino- or carboxylterminal to the cleaved amino acid pair in half the sites, which is similar to the situation in EMCV (6). This lack of uniformity of sequence surrounding the cleavage sites in all four picornaviruses clearly indicates involvement of secondary or tertiary structure in the precursor molecule.

The 3' noncoding region is 47 nucleotides long in HRV-14. This compares to 72 in PV-1, 96 in FMDV, and 126 in EMCV (4-7). Both HRV-14 and FMDV have single stop codons, whereas PV-1 and EMCV have double stops. All four picornaviruses lack the polyadenylylation signal sequence A-A-U-A-A-A, which supports the finding that the poly(A) tail is transcribed from a minus-strand template.

Further sequence comparisons to other HRV serotypes will help to generate a genomic map of common and unique regions. This information will aid in mapping neutralization sites, active sites of the protease and the replicase, and viral receptor binding sites. Recent competition binding studies have shown clearly that HRV-14 competes with the vast majority of other HRV serotypes for attachment to HeLa cell receptors (8). Since this receptor is distinct from other picornaviruses, determination of this highly conserved region in HRV serotypes will be of major importance. We thank Nancy Dunn and Peter Kniskern for synthesis of DNA primers and thank Ann Palmenberg (University of Wisconsin) for helpful discussions.

- Cooper, P. D., Agol, V. I., Bachrach, H. L., Brown, F., Ghendon, Y., Gibbs, A. J., Gillespie, J. H., Lonberg-Holm, K., Mandel, B., Melnick, J. L., Mohanty, S. B., Povey, R. C., Rueckert, R. R., Schaffer, F. L. & Tyrrell, D. A. J. (1978) Intervirology 10, 165-180.
- 2. Melnick, J. L. (1980) Prog. Med. Virol. 26, 214-232.
- 3. Agol, V. I. (1980) Prog. Med. Virol. 26, 119-157.
- Kitamura, N., Semler, B. L., Rothberg, P. G., Larsen, G. R., Adler, C. J., Dorner, A. J., Emini, E. A., Hanecak, R., Lee, J. J., van der Werf, S., Anderson, C. W. & Wimmer, E. (1981) Nature (London) 291, 547-553.
- Racaniello, V. R. & Baltimore, D. (1981) Proc. Natl. Acad. Sci. USA 78, 4887–4891.
- Palmenberg, A. C., Kirby, E. M., Janda, M. R., Drake, N. L., Duke, G. M., Potratz, K. F. & Collett, M. S. (1984) Nucleic Acids Res. 12, 2969-2985.
- Carroll, A. R., Rowlands, D. T. & Clarke, B. E. (1984) Nucleic Acids Res. 12, 2461–2472.
- 8. Abraham, G. & Colonno, R. J. (1984) J. Virol. 51, 340-345.
- 9. Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
- 10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 11. Holmes, D. S. & Quigley, M. (1981) Anal. Biochem. 114, 193-197.
- 12. Grunstein, M. & Hogness, D. S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3965.
- 13. Hong, G. F. (1982) J. Mol. Biol. 158, 539-549.
- 14. Hansen, J. N., Pheiffer, B. H. & Boehnert, J. A. (1980) Anal. Biochem. 105, 192-201.
- 15. Maxam, A. M. & Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
- 16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 17. Rueckert, R. R. & Wimmer, E. (1984) J. Virol. 50, 957-959.
- 18. Gwaltney, J. M. (1975) Yale J. Biol. Med. 48, 17-45.
- 19. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
- Toyoda, H., Kohara, M., Kataoka, Y., Suganuma, T., Omata, T., Imura, N. & Nomoto, A. (1984) J. Mol. Biol. 174, 561-585.
- 21. Argos, P., Kamer, G., Nicklin, M. J. H. & Wimmer, E. (1984) Nucleic Acids Res. 12, 7251-7267.
- 22. Emini, E. A., Jameson, B. A. & Wimmer, E. (1984) J. Virol. 52, 719-721.
- 23. Emini, E. A., Jameson, B. A. & Wimmer, E. (1983) Nature (London) 304, 699-703.