Volume 13 Number 6 1985 Nucleic Acids Research

Human rhinovirus 2: complete nucleotide sequence and proteolytic processing signals in the capsid protein region

Tim Skern, Wolfgang Sommergruber, Dieter Blaas, Peter Gruendler, Franz Fraundorfer, Christian Pieler+, Ingrid Fogy+ and Ernst Kuechler*

Institut für Biochemie der Universität Wien, Währingerstrasse 17, 1090 Vienna, Austria, and +Dr. Ernst Boehringer Institut für Arzneimittelforschung, Dr. Ernst Boehringer Gasse 5-11, 1120 Vienna, Austria

Received 11 January 1985; Revised and Accepted 22 February 1985

ABSTRACT

cDNA clones representing the entire genome of human rhinovirus 2 have been obtained and used to determine the complete nucleotide sequence. The genome consists of 7102 nucleotides and possesses a long open reading frame of 6450 nucleotides; this reading frame is initiated 611 nucleotides from the 5'end and stops 42 nucleotides from the polyA tract. The N-terminal sequences of three of the viral capsid proteins have been elucidated, thus defining the positions of three cleavage sites on the polyprotein. The extensive amino acid sequence homology with poliovirus and human rhinovirus 14 enabled the other cleavage sites to be predicted. Cleavages in the 3' half of the molecule appear to take place predominantly at Gln-Gly pairs, whereas those in the 5' half (including the capsid proteins) are more heterogeneous.

INTRODUCTION

The common cold is one of the most frequent viral infections of man. This disease, although in itself not serious, is of economic importance through working days lost and can lead to secondary infections (1). The main causative agents of this disease are human rhinoviruses (2). Acquisition of immunity to rhinoviruses is difficult because of the presence of over 100 antigenically distinct serotypes (3, 4).

Rhinoviruses are members of the family picornaviridae. Typically the RNA from these viruses is a single-stranded, positive sense molecule of approximately 7500 nucleotides in length. Unlike most eukaryotic mRNAs, this RNA is not capped but is joined at the 5'end to a small virus encoded protein VPg; it is however polyadenylated at the 3'end (5 - 8). The primary translational product of this RNA is assumed to be a single, large polyprotein which is subsequently processed by proteolytic cleavage to yield the mature virus proteins. Thus, only one large

reading frame will be present in the RNA (9). An icosahedral viral capsid containing 60 copies each of 4 virus polypeptides, VP1 - VP4, surrounds the RNA (10).

The molecular basis of serotypic diversity, inherent in rhinoviruses, is not understood. 50 of the serotypes have been put in 16 distinct groups, based on the low level of antigenic cross-reactivity which does exist (11, 12). However, it has been shown that despite this antigenic diversity, there are only two receptors on cells susceptible for human rhinoviruses (13). Therefore some common receptor binding sites may be present on the viruses. As part of a programme to investigate the problem of antigenic diversity at the molecular level, we have determined the complete nucleotide sequence of human rhinovirus 2 (HRV2) from the cloned cDNA. In this report, we present this sequence, as well as the N-terminal amino acid sequences of three of the viral capsid proteins, which enables these capsid proteins to be defined more closely on the precursor polypeptide. Limited comparison of the derived amino acid sequence from HRV2 with that of poliovirus type 1 (14, 15) and with that of the recently published sequence of human rhinovirus 14 (HRV14) are also presented (16). A more extensive analysis of the polymerase protein of HRV2 has already been presented (17).

MATERIALS AND METHODS

Materials

Nycodenz was obtained from Nyegaard and Company, Oslo, Norway; deoxynucleoside triphosphates, terminal transferase, and polynucleotide kinase were from P.L. Biochemicals; reverse transcriptase was a gift of Dr. J. Beard, NIH, Bethesda or was purchased from Anglian Biotechnology Co, Cambridge; oligodG tailed, PstI cleaved pBR322 and bacterial alkaline phosphatase were from Bethesda Research Labs; restriction enzymes from either New England Biolabs, Bethesda Research Labs or P.L. Biochemicals; [α-^{32}P]dCTP and [γ-^{32}P]ATP were from Amersham International.

Propagation and Isolation of Human Rhinovirus 2.

Propagation and purification of HRV2, as well as the isolation of the viral RNA were as previously described (17).

Synthesis and Cloning of HRV2 cDNA.

OligodT-primed synthesis of cDNA on HRV2 RNA by AMV reverse transcriptase and cloning of the cDNA-RNA hybrid produced was carried out as previously described (17), except that tailing of the cDNA-RNA hybrid was performed at a dCTP concentration of 80 μM dCTP in a 5 minute reaction.

In cloning experiments using oligodT as primer, clones derived from the 5'end of HRV2 were underrepresented. To overcome this problem, two restriction fragments were prepared from clones which had been previously mapped onto the HRV2 genome and were used to prime reverse transcriptase. Annealing of the restriction fragments to the HRV2 RNA was performed in a reaction mixture (10μl) containing 80 % formamide, 40 mM PIPES, 400 mM NaCl, 1 mM EDTA, pH 6.4. The mixture was held at 68° C for 10 min and allowed to cool overnight to 40° C. This material was then used as a template for reverse transcriptase and the resulting cDNA-RNA hybrid was cloned as described previously (17).

Analysis of Cloned cDNA.

Initially, cloned cDNA was analysed by making small-scale plasmid preparations, excising the inserts with PstI and examining them on 1.4 % agarose gels. Southern blots (18) of these gels were then probed with ^{32}P-labelled HRV2 cDNA; the strength of the signal gave an indication of the position of the insert on the HRV2 genome. Inserts were mapped relative to each other using restriction enzyme mapping or were sequenced directly and were ordered by computer analysis. Recombinants containing inserts representing specific regions of the human rhinovirus 2 genome were identified by colony hybridisation (19). Previously mapped fragments were nick-translated (20) and used as probes.

DNA Sequencing and Computer Analysis.

Inserts were either subcloned into the plasmid pUC9 (21) and sequenced chemically (22) or into M13mp9 and were sequenced by the chain termination method (23). Sequences so generated were analysed on a Cyber 170 computer using the programs developed by Staden (24) as modified by Isono (25).

N-terminal Protein Sequencing.

Pure preparations of each of the capsid proteins VP1, VP2

and VP3 were obtained as follows. 2 mg of HRV2 was heated in Laemmli sample buffer (26) and loaded onto a 12.5 % SDS-polyacrylamide gel. The gel was stained with a saturated solution of Coomassie blue in 50 mM Tris-HCl, pH 7.4 and the protein bands cut out with a scalpel. Elution was performed in an ISCO elution apparatus at 50 volts for 16 hrs using Laemmli running buffer, and the proteins were recovered by precipitation with TCA.

N-terminal sequencing was carried out using the method of Hunkapiller and Hood (27) on an AB-470A protein sequencing apparatus (Applied Biosystems, Inc., Foster City, CA, USA). 2 nmol VP1, 2 nmol VP2 and 1 nmol VP3 respectively were taken for each run. Derivatised amino acids obtained were analysed by HPLC.

RESULTS AND DISCUSSION

Molecular Cloning and Sequencing of the HRV2 Genome

The construction of clones using the cDNA-RNA hybrid method representing 1425 nucleotides at the 3' end of HRV2 has been reported previously (17). Examination of the remaining 350Amp^{s}Tet^r clones generated in this experiment provided clones representing about 60 % of the genome. Cloning experiments using cDNA hybrids generated by priming reverse transcriptase on HRV2 RNA with two restriction fragments (indicated as A and B in Fig. 1) enabled recombinants covering the rest of the genome to be obtained. Inserts representing sequences at the 5' end of the HRV2 genome were detected by Grunstein-Hogness screening (19) using the primers themselves as probes.

17 clones spanning the HRV2 genome are shown in Fig. 1. 3 clones (numbers 61, 100 and 109) obtained from the cloning experiment using primer A (Fig. 1) were found to contain a PstI site and to each possess a common fragment of the same length. Sequencing of these clones revealed that all 3 shared the sequence TTAAAAC directly adjacent to the dC-dG tails. As this sequence corresponds to the first seven nucleotides of each of the three poliovirus serotypes (14, 15, 28, 29) and of HRV14 (16), it was concluded that these clones contained the exact 5' terminus of the HRV2 genome.

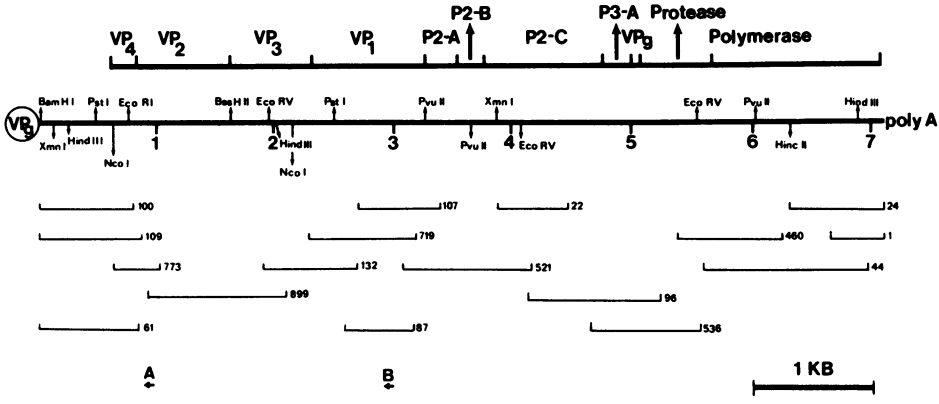


Fig.1: 17 overlapping cDNA clones spanning the HRV2 genome. Clones 61, 100 and 109 contain the presumed 5' terminus of HRV2; clones 1 and 24 contain part of the polyA tract. These 17 clones and a further 25 (not shown) were used to determine the nucleotide sequence. The arrows A and B depict two restriction fragments which were used to prime reverse transcriptase to obtain clones from the 5' end. The location of the individual mature virus proteins on the genome of HRV2 is shown at the top of the figure. The L434 nomenclature system of Rueckert and Wimmer (35) was used to name the proteins.

To obtain the sequence of the entire genome of HRV2, the 17 clones shown in Fig. 1 plus a further 25 (not shown) were used. Sequencing was performed from the ends of clones or from restriction sites within them. 85 % of the sequence was obtained on both strands and each nucleotide was determined at least twice. Analysis of the sequence data was aided by the extensive homology of HRV2 to poliovirus and by the assumption that the virus genome should contain only one large open reading frame. Regions at the 5' and 3' ends thought not to be translated were sequenced 100 % on both strands. It is thus felt that the sequence presented here is an accurate representation of the genome of HRV2.

Aspects of the Sequence of HRV2.

General remarks. The complete nucleotide sequence of HRV2, comprising of 7102 nucleotides (excluding the polyA tract) is shown in Fig. 2. An open reading frame coding for 2150 amino acids is present; this reading frame is initiated with the AUG at position 611 and ends with the stop codon UAA 42 nucleotides before

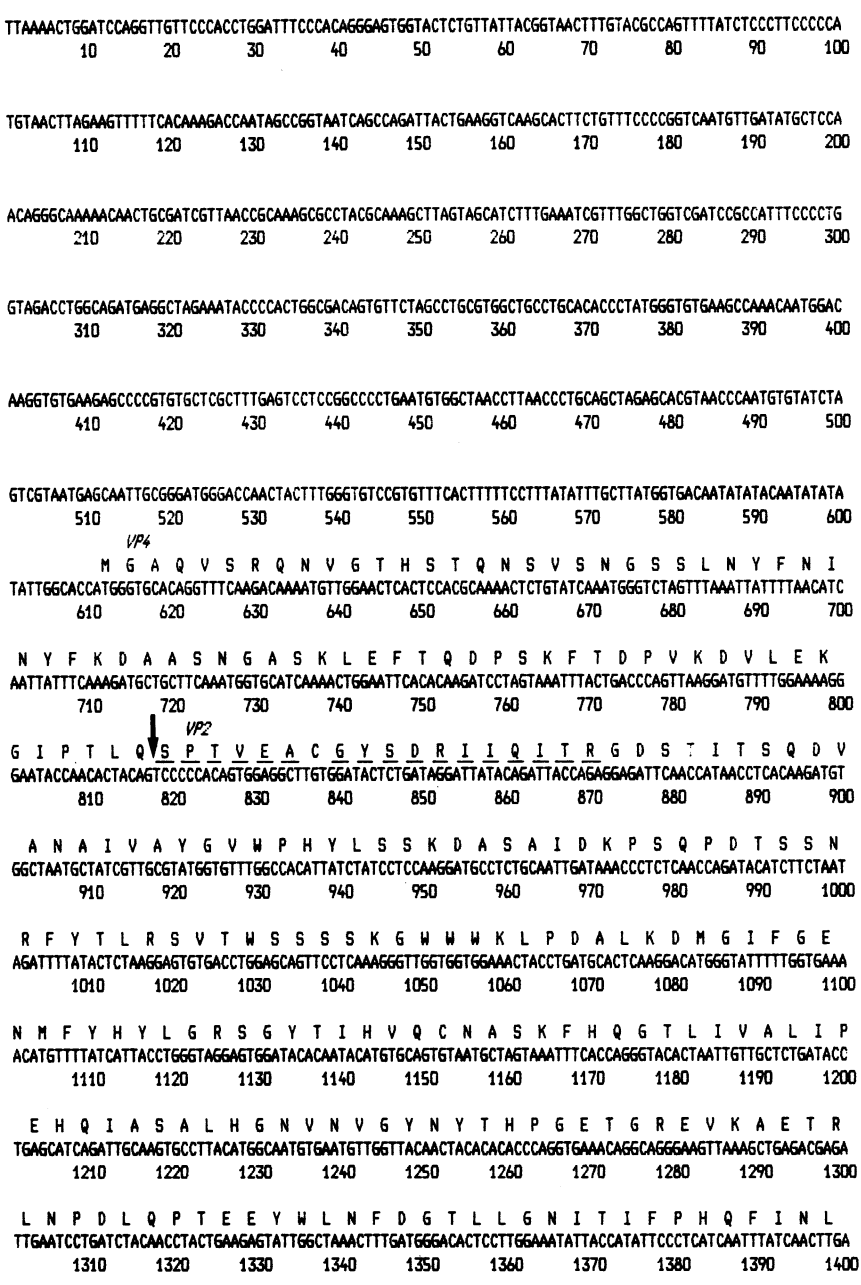
the start of the polyA tract. The presence of numerous stop codons in both the other possible reading frames and the fact that the N-terminal amino acid sequences which we have determined can be found in this derived amino acid sequence proves that this reading frame is used *in vivo*.

The 5'-proximal Region. At present no function has been ascribed to this region in picornaviruses; it is assumed to be non-coding (14). Analysis of the 610 nucleotides between the first nucleotide and the start of the long open reading frame in HRV2 does show the presence of small open reading frames. However, counterparts in the 5'proximal regions of polioviruses and HRV14 are not present,so that these reading frames are of unknown significance.

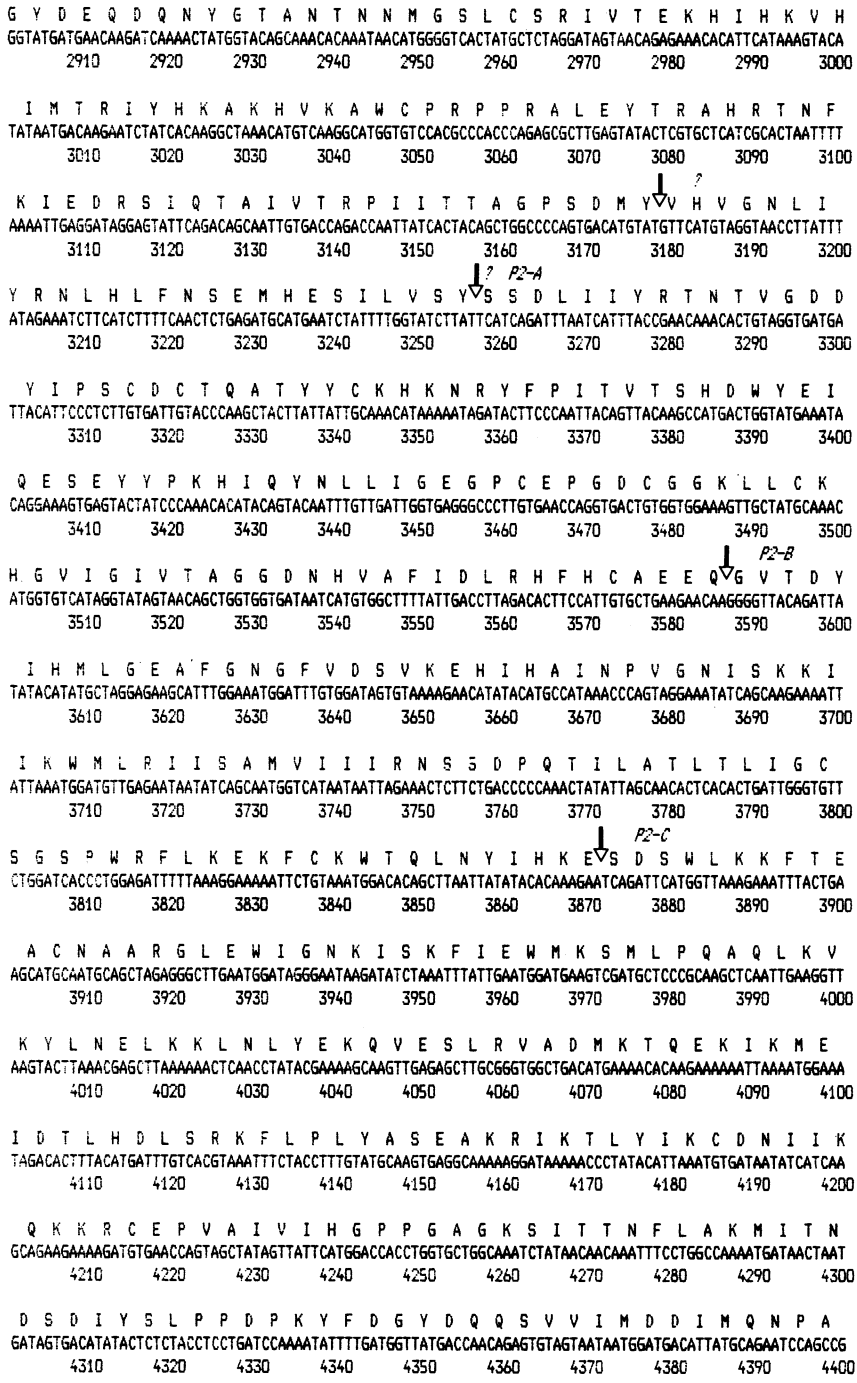
On the other hand, comparison of the nucleotide sequence of the 5' proximal region with those of HRV14 and poliovirus type 1 reveals an outstanding degree of homology. Alignment of the sequences to optimise homology results in values of about 65 % between HRV2 and HRV14 and about 55 % between HRV2 and poliovirus type 1. Furthermore, the homology is not distributed randomly throughout this region, but tends to be present as blocks, a fact already pointed out for HRV14 and poliovirus type 3 by Stanway *et al.* (16). 5 blocks containing 16 or more continuously identical nucleotides can be found between HRV2 and HRV14 and 5 between HRV2 and poliovirus type 1. However, only two stretches are conserved in HRV2, HRV14 and all three poliovirus types. These are a stretch of 16 bases (beginning at 436 in HRV2) and a stretch of 23 bases (beginning at 531 in HRV2).

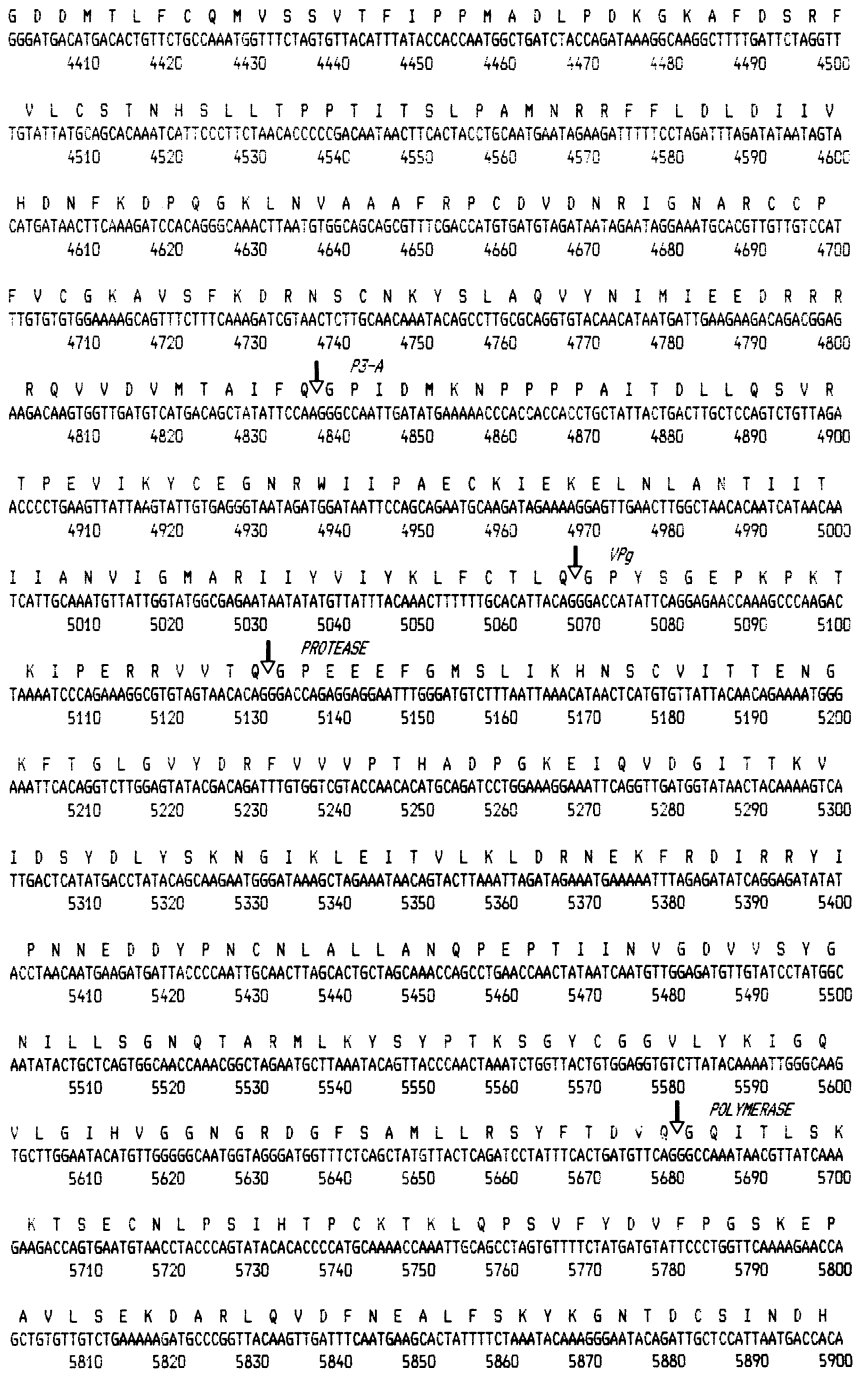
A comparison of the coding regions of HRV2, HRV14 and poliovirus 1 reveals only one stretch of more than 16 identical nucleotides (a block of 20 between HRV2 and HRV14, starting at position 5360 in HRV2). The conservation of these two blocks in two rhinoviruses and in all three poliovirus serotypes strongly suggests the involvement in an essential function during the life cycle of these picornaviruses. However, suggestions as to the nature of this function must remain speculative as no data on the role of this region are available.

Structure of the Polyprotein. Analysis of the polyprotein coded for by the open reading frame shows that the homology with HRV14


VP4
VP2

```
R S N N S A T I I A P Y V N A V P M D S M R S H N N W S L V I I P I
GGAGTAATAATTCTGCCACAATAATTGCCCCTTATGTCAATGCAGTTCCTATGGATTCAATGCGGAGCCACAATAATTGGAGTTTGGTAATAATACCAAT
      1410      1420      1430      1440      1450      1460      1470      1480      1490      1500

 C P L E T S S A I N T I P I T I S I S P M C A E F S G A R A K R Q
ATGTCCCCTTGAGACATCAAGTGCAATTAACACAATACCTATTACAATATCTATAAGCCCCATGTGTGCAGAGTTTTCCGGCGCGCGTGCCAAGCGTCAA
      1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
          VP3
G L P V F I T P G S G Q F L T T D D F Q S P C A L P W Y H P T K E
GGATTACCAGTTTTCATCACACCAGGTTCAGGACAGTTTTTGACAACAGATGATTTCCAATCCCCATGTGCACTTCCCTGGTATCACCCAACTAAGGAAA
      1610      1620      1630      1640      1650      1660      1670      1680      1690      1700

I S I P G E V K N L V E I C Q V D S L V P I N N T D T Y I N S E N M
TTTCTATTCCAGGTGAGGTTAAAAATTTGGTTGAAATTTGTCAAGTAGACAGCCTAGTACCAATAAATAACACTGACACCTACATCAATAGTGAAAATAT
      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800

 Y S V V L Q S S I N A P D K I F S I R T D V A S Q P L A T T L I G
GTATTCTGTTGTATTGCAATCATCAATTAATGCACCAGATAAGATCTTCTCTATTCGAACAGATGTTGCTTCCCAACCTTTAGCTACTACTTTGATTGGT
      1810      1820      1830      1840      1850      1860      1870      1880      1890      1900

 E I S S Y F T H W T G S L R F S F M F C G T A N T T V K L L L A Y
GAGATATCTAGCTATTTCACCCACTGGACAGGGAGTCTCCGTTTCAGCTTCATGTTTTGTGGTACTGCCAACACTACTGTTAAGCTTTTGTTGGCATACA
      1910      1920      1930      1940      1950      1960      1970      1980      1990      2000

T P P G I A E P T T R K D A M L G T H V I W D V G L Q S T I S M V V
CACCACCTGGTATCGCAGAACCCACCACAAGAAAGGATGCAATGCTAGGCACTCATGTTATATGGGATGTGGGGTTGCAGTCTACAATATCAATGGTAGT
      2010      2020      2030      2040      2050      2060      2070      2080      2090      2100

 P W I S A S H Y R N T S P G R S T S G Y I T C W Y Q T R L V I P P
GCCATGGATTAGCGCTAGTCATTATAGAAACACATCACCAGGTAGATCTACATCTGGGTACATAACATGCTGGTATCAGACTAGATTAGTCATTCCACCT
      2110      2120      2130      2140      2150      2160      2170      2180      2190      2200

 Q T P P T A R L L C F V S G C K D F C L R M A R D T N L H L Q S G
CAGACCCCACCAACAGCTAGATTGTTATGTTTTGTATCTGGGTGCAAAGACTTTTGCTTGCGCATGGCACGAGATACTAACCTACACCTGCAAAGTGGTG
      2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
          VP1
A I A Q N P V E N Y I D E V L N E V L V V P N I N S S N P T T S N S
CAATAGCACAGAACCCTGTTGAGAATTATATAGATGAAGTTCTTAATGAAGTTTTAGTTGTCCCAAATATTAATAGTAGTAACCCCACAACATCAAATTC
      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400

 A P A L D A A E T G H T S S V Q P E D V I E T R Y V Q T S Q T R D
TGCCCCAGCATTAGATGCTGCAGAAACAGGGCACACTAGTAGTGTTCAACCAGAGGATGTCATTGAAACTAGGTATGTGCAGACATCACAAACAAGAGAT
      2410      2420      2430      2440      2450      2460      2470      2480      2490      2500

 E M S L E S F L G R S G C I H E S K L E V T L A N Y N K E N F T V
GAAATGAGTTTAGAGAGTTTTCTTGGCAGATCAGGATGCATACATGAATCTAAATTAGAGGTTACACTTGCAAATTATAACAAGGAGAATTTTACAGTGT
      2510      2520      2530      2540      2550      2560      2570      2580      2590      2600

W A I N L Q E M A Q I R R K F E L F T Y T R F D S E I T L V P C I S
GGGCTATTAATCTACAAGAAATGGCTCAAATTAGAAGGAAATTTGAATTGTTCACCTATACTAGGTTTGATTCTGAAATAACCCTAGTTCCATGCATTTC
      2610      2620      2630      2640      2650      2660      2670      2680      2690      2700

 A L S Q D I G H I T M Q Y M Y V P P G A P V P N S R D D Y A W Q S
CGCCCTTAGTCAGGACATTGGACACATCACAATGCAATACATGTATGTTCCACCAGGTGCACCGGTGCCCAATAGTAGGGACGATTATGCATGGCAGTCT
      2710      2720      2730      2740      2750      2760      2770      2780      2790      2800

 G T N A S V F W Q H G Q A Y P R F S L P F L S V A S A Y Y M F Y D
GGCACTAATGCCTCTGTTTTCTGGCAACATGGACAGGCTTATCCAAGATTTTCCTTACCTTTCCTAAGTGTGGCATCTGCTTATTACATGTTTTATGATG
      2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
```


G Y D E Q D Q N Y G T A N T N N M G S L C S R I V T E K H I H K V H
GGTATGATGAACAAGATCAAAACTATGGTACAGCAAACACAAATAACATGGGGTCACTATGCTCTAGGATAGTAACAGAGAAACACATTCATAAAGTACA
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
I M T R I Y H K A K H V K A W C P R P P R A L E Y T R A H R T N F
TATAATGACAAGAATCTATCACAAGGCTAAACATGTCAAGGCATGGTGTCCACGCCCACCCAGAGCGCTTGAGTATACTCGTGCTCATCGCACTAATTTT
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
? ?
K I E D R S I Q T A I V T R P I I T T A G P S D M Y V H V G N L I
AAAATTGAGGATAGGAGTATTCAGACAGCAATTGTGACCAGACCAATTATCACTACAGCTGGCCCCAGTGACATGTATGTTCATGTAGGTAACCTTATTT
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
? P2-A
Y R N L H L F N S E M H E S I L V S Y S S D L I I Y R T N T V G D D
ATAGAAATCTTCATCTTTTCAACTCTGAGATGCATGAATCTATTTTGGTATCTTATTCATCAGATTTAATCATTTACCGAACAAACACTGTAGGTGATGA
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
Y I P S C D C T Q A T Y Y C K H K N R Y F P I T V T S H D W Y E I
TTACATTCCCTCTTGTGATTGTACCCAAGCTACTTATTATTGCAAACATAAAAATAGATACTTCCCAATTACAGTTACAAGCCATGACTGGTATGAAATA
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
Q E S E Y Y P K H I Q Y N L L I G E G P C E P G D C G G K L L C K
CAGGAAAGTGAGTACTATCCCAAACACATACAGTACAATTTGTTGATTGGTGAGGGCCCTTGTGAACCAGGTGACTGTGGTGGAAAGTTGCTATGCAAAC
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
P2-B
H G V I G I V T A G G D N H V A F I D L R H F H C A E E Q G V T D Y
ATGGTGTCATAGGTATAGTAACAGCTGGTGGTGATAATCATGTGGCTTTTATTGACCTTAGACACTTCCATTGTGCTGAAGAACAAGGGGTTACAGATTA
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
I H M L G E A F G N G F V D S V K E H I H A I N P V G N I S K K I
TATACATATGCTAGGAGAAGCATTTGGAAATGGATTTGTGGATAGTGTAAAAGAACATATACATGCCATAAACCCAGTAGGAAATATCAGCAAGAAAATT
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
I K W M L R I I S A M V I I I R N S S D P Q T I L A T L T L I G C
ATTAAATGGATGTTGAGAATAATATCAGCAATGGTCATAATAATTAGAAACTCTTCTGACCCCCAAACTATATTAGCAACACTCACACTGATTGGGTGTT
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
P2-C
S G S P W R F L K E K F C K W T Q L N Y I H K E S D S W L K K F T E
CTGGATCACCCTGGAGATTTTTAAAGGAAAAATTCTGTAAATGGACACAGCTTAATTATATACACAAAGAATCAGATTCATGGTTAAAGAAATTTACTGA
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
A C N A A R G L E W I G N K I S K F I E W M K S M L P Q A Q L K V
AGCATGCAATGCAGCTAGAGGGCTTGAATGGATAGGGAATAAGATATCTAAATTTATTGAATGGATGAAGTCGATGCTCCCGCAAGCTCAATTGAAGGTT
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
K Y L N E L K K L N L Y E K Q V E S L R V A D M K T Q E K I K M E
AAGTACTTAAACGAGCTTAAAAAACTCAACCTATACGAAAAGCAAGTTGAGAGCTTGCGGGTGGCTGACATGAAAACACAAGAAAAAATTAAAATGGAAA
4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
I D T L H D L S R K F L P L Y A S E A K R I K T L Y I K C D N I I K
TAGACACTTTACATGATTTGTCACGTAAATTTCTACCTTTGTATGCAAGTGAGGCAAAAAGGATAAAAACCCTATACATTAAATGTGATAATATCATCAA
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
Q K K R C E P V A I V I H G P P G A G K S I T T N F L A K M I T N
GCAGAAGAAAAGATGTGAACCAGTAGCTATAGTTATTCATGGACCACCTGGTGCTGGCAAATCTATAACAACAAATTTCCTGGCCAAAATGATAACTAAT
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
D S D I Y S L P P D P K Y F D G Y D Q Q S V V I M D D I M Q N P A
GATAGTGACATATACTCTCTACCTCCTGATCCAAAATATTTTGATGGTTATGACCAACAGAGTGTAGTAATAATGGATGACATTATGCAGAATCCAGCCG
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400


P3-A
VPg
PROTEASE
POLYMERASE

```
I R I A S S H Y A A Q L I T L D I D P K P I T L E D S V F G T D G L
TAAGAATTGCATCATCACATTATGCAGCACAACTCATTACCTTAGATATTGACCCAAAACCTATTACACTTGAGGACAGTGTCTTTGGCACTGATGGATT
     5910      5920      5930      5940      5950      5960      5970      5980      5990      6000

 E A L D L N T S A G F P Y I A M G V K K R D L I N N K T K D I S K
AGAGGCTCTTGATTTGAACACTAGCGCAGGATTTCCATATATTGCAATGGGAGTTAAAAAGAGAGATTTAATAAACAACAAGACCAAGGATATAAGCAAA
     6010      6020      6030      6040      6050      6060      6070      6080      6090      6100

 L K E A I D K Y G V D L P M V T F L K D E L R K H E K V I K G K T
CTTAAAGAAGCAATTGACAAATACGGAGTTGACTTACCTATGGTCACCTTCTTGAAAGATGAACTCAGAAAGCATGAAAAGGTAATTAAAGGTAAAACTA
     6110      6120      6130      6140      6150      6160      6170      6180      6190      6200

R V I E A S S V N D T L L F R T T F G N L F S K F H L N P G I V T G
GAGTTATTGAAGCTAGTAGTGTGAATGATACCCTATTATTTAGAACAACTTTTGGCAACCTCTTTTCAAAGTTCCACTTGAATCCTGGAATTGTTACTGG
     6210      6220      6230      6240      6250      6260      6270      6280      6290      6300

 S A V G C D P E V F W S K I P A M L D D K C I M A F D Y T N Y D G
ATCAGCAGTTGGATGTGATCCAGAGGTGTTTTGGTCAAAAATACCAGCAATGTTGGATGATAAATGTATTATGGCTTTTGATTATACAAATTATGATGGT
     6310      6320      6330      6340      6350      6360      6370      6380      6390      6400

 S I H P I W F E A L K Q V L V D L S F N P T L I D R L C K S K H I
AGTATACACCCTATTTGGTTTGAAGCTCTTAAACAGGTACTGGTAGATCTATCATTTAATCCAACATTAATAGATAGCTATGCAAGTCTAAACACATCT
     6410      6420      6430      6440      6450      6460      6470      6480      6490      6500

F K N T Y Y E V E G G V P S G C S G T S I F N T M I N N I I I R T L
TCAAAAATACATACTATGAAGTGGAGGGAGGTGTACCATCTGGGTGTTCAGGTACTAGTATTTTTAACACTATGATCAATAATATTATCATAAGGACCTT
     6510      6520      6530      6540      6550      6560      6570      6580      6590      6600

 V L D A Y K N I D L D K L K I I A Y G D D V I F S Y I H E L D M E
AGTGTTAGATGCATACAAGAATATAGATCTAGATAAGCTTAAGATAATTGCCTATGGTGATGATGTCATATTCTCATACATACATGAACTGGACATGGAG
     6610      6620      6630      6640      6650      6660      6670      6680      6690      6700

 A I A I E G V K Y G L T I T P A D K S N T F V K L D Y S N V T F L
GCTATAGCAATAGAGGGTGTTAAATATGGTTTGACTATAACTCCTGCTGATAAATCTAACACATTTGTAAAATTAGACTATAGCAATGTTACTTTTTTAA
     6710      6720      6730      6740      6750      6760      6770      6780      6790      6800

K R G F K Q D E K Y N F L I H P T F P E D E I F E S I R W T K K P S
AAAGAGGGTTTAAGCAAGATGAGAAGTATAACTTTCTAATACATCCAACTTTCCCTGAAGATGAAATATTTGAATCCATCAGATGGACAAAGAAACCATC
     6810      6820      6830      6840      6850      6860      6870      6880      6890      6900

 Q M H E H V L S L C H L M W H N G R D A Y K K F V E K I R S V S A
ACAAATGCATGAACATGTGTTGTCTCTGTGTCACTTAATGTGGCACAATGGACGTGACGCATACAAAAAATTTGTGGAGAAGATACGCAGTGTAAGCGCT
     6910      6920      6930      6940      6950      6960      6970      6980      6990      7000

 G R A L Y I P P Y D L L L H E W Y E K F
GGTCGTGCACTGTACATCCCTCCGTATGATTTGCTTTTGCATGAGTGGTATGAAAAATTTTAAAGATATAGAAATAGTAAACTGATAGTTTATTAGTTTT
     7010      7020      7030      7040      7050      7060      7070      7080      7090      7100


AT poly(A)
     7110      7120      7130      7140      7150      7160      7170      7180      7190      7200
```

Fig.2: The complete nucleotide sequence (as cDNA) of HRV2 and the encoded amino acid sequence from the long reading frame. Amino acids confirmed by protein sequencing are underlined. Cleavage sites so defined are indicated with a closed arrow, those predicted from the homology with poliovirus with an open arrow.

Table 1. Homologies (expressed as percentages) between the derived amino acid sequences of proteins considered to be equivalent in HRV2, HRV14 and poliovirus type 1.

	HRV2/HRV14	HRV2/Poliovirus	HRV14/Poliovirus
VP4	51	58	58
VP2	60	45	49
VP3	49	49	41
VP1	32	17	39
P2-A	34	31	45
P2-B	42	41	50
P2-C	44	43	61
P3-A	38	36	49
VPg	58	30	48
Protease	50	35	48
Polymerase	55	56	65

and poliovirus extends into the coding region as well. The presence of this high degree of homology has proved useful, as it has enabled the mature virus proteins to be mapped onto the polyprotein and predictions concerning the proteolytic cleavage sites within the polyprotein to be made, despite the lack of biochemical and genetic data.

Homologies between the different proteins of HRV2, HRV14 and poliovirus type 1 are given in Tab. 1. Comparison of the derived protein sequences from HRV2 and HRV14 indicates that the least homology (32 %) was found in VP1 and the greatest in VP2 (60 %). Interestingly, therefore, the coat protein VP2 is more conserved between these two rhinoviruses than the polymerases (55 %) or the proteases (50 %).A plausible explanation may be that VP2 plays an important role in determining the structure of the virus and does not contribute greatly to the antigenicity. Similar analysis of the homology between HRV2 and poliovirus type 1 reveals that the homology varies between 17 % in VP1 and 56 % in the polymerase protein. It is perhaps worth noting that although the small genome linked protein VPg of HRV2 is 2 amino acids shorter than that of HRV14, the first five amino acids of the two VPgs are identical (this includes the tyrosine through which the VPg is linked to the RNA). Moreover, the two VPg sequences can be aligned to give 58 % homology; in contrast, the homologies between the two rhinovirus VPgs and that of polio-

virus type 1 are significantly lower (Table 1). Generally, the degree of homology between HRV2 and poliovirus is less than that between HRV14 and poliovirus, suggesting that HRV14 may be more closely related to poliovirus than HRV2. HRV2 and HRV14 have recently been placed into different receptor groups (13). Furthermore, HRV14 shares a common cellular receptor with coxsackie A21, another enterovirus (30). It seems therefore possible that HRV2 and HRV14 represent two members of the rhinovirus genera that have diverged considerably from one another. Additional support for this conclusion comes from a very recent study (36) which showed the presence of strong antigenic relationships between the P-2C proteins of poliovirus type 1 and several picornaviruses (including HRV2 and HRV14). Thus, although the P-2C proteins of HRV2 and HRV14 would appear to contain at least one conserved antigenic site, the relatively low level of amino acid sequence homology between these two proteins (44 %) implies that the sequences around this site have diverged to a marked degree.

Proteolytic Cleavage Sites in the Polyprotein. The mature virus proteins are derived from the polyprotein by proteolytic cleavage carried out by the viral protease, although some steps are also believed to be effected by a host encoded protease (31 - 33). The extensive amino acid homology of HRV2 with poliovirus type 1 (for which the cleavage sites have been determined by amino acid sequencing analysis (14)) and HRV14 enabled most of the cleavage sites in HRV2 to be predicted (Table 2). From this homology, it was however not possible to predict with certainty the cleavage site which would define exactly the position of the capsid protein VP1. The problem was in part resolved by the determination of the first 18 amino acids at the N-terminus of VP1. N-terminal sequencing of VP2 and VP3 was also performed enabling these proteins to be positioned. Amino acids determined by protein sequencing are underlined in Fig. 2; the cleavage sites are shown in Table 2.

The cleavage at the carboxy-terminus of VP1, i.e. between VP1 and P2-A, remains to be determined. This cleavage takes place in poliovirus and presumably also in HRV14 at a tyrosine-glycine amino acid pair; such a pair is not present in HRV2 at

Table 2. Proteolytic cleavage sites within the HRV2 polyprotein. Those amino acids identified by protein sequencing are underlined, the other cleavage sites were predicted from the homology with poliovirus and are considered putative. For comparison, the cleavage sites of HRV14 as predicted from the homology with poliovirus (16) and the cleavage sites of poliovirus type 1 as determined by protein sequencing (14) are also shown.

	Cleavage sites		
Protein Junction	HRV2	HRV14	Poliovirus
VP4/VP2	LQ SP	LN SP	LN SP
VP2/VP3	RQ GL	PQ GL	LQ GL
VP3/VP1	AQ NP	TQ TI ?	AQ GL
		or SQ TV ?	
		or TE GL ?	
VP1/P2-A	MY VH ?	SY GL ?	TY GF
	or SY SS ?	or RY GG ?	
P2-A/P2-B	EQ GV	EQ GL	GQ GI
P2-B/P2-C	KE SD	RQ AN	KQ GD
P2-C/P3-A	FQ GP	FQ GP	FQ GP
P3-A/VPg	LQ GP	TQ GP	HQ GA
VPg/Protease	TQ GP	VQ GP	VQ GP
Protease/Polymerase	VQ GQ	KQ GQ	SQ GE

the required position. However, a tyrosine-serine pair is present between nucleotides 2644 - 2649. Cleavage at this site would result in a VP1 protein having a molecular weight of 36K, sufficiently close to the experimentally determined value as obtained from the mobility of VP1 on SDS-polyacrylamide gels (2, 10). Nevertheless, in the absence of confirmatory sequence data, this cleavage site must be regarded as tentative.

Examination of Table 2 shows that more than half the cleavages in the HRV2 polyprotein appear to take place at glutamine-glycine pairs. Interestingly, 3 of the 4 cleavage sites which do not follow this pattern lie in the capsid protein region, an observation which may not be without functional significance. The viral protease may recognise certain exposed regions of the polyprotein rather than the cleavage site *per se*. Mutations which lead to the exposure of other regions of the polyprotein could make a new cleavage site accessible and result in mature proteins of different sizes. The use of different cleavage sites may lead to an altered antigenicity and thus contribute to the generation of new serotypes. Pertinent to this hypothesis is

the observation that the capsid proteins of two other rhinoviruses have markedly different sizes (13).

In order to compare further the three VP1s, hydrophobicity plots (34) of the VP1 proteins from HRV2, HRV14 and poliovirus type 1 were carried out (data not shown). Given the relatively low degree of homology between these three proteins (32 % between HRV2 and HRV14, 17 % between HRV2 and poliovirus type 1 (Tab. 1)), the hydrophobicity plots are remarkably similar. A major antigenic site in poliovirus type 1 lies around amino acid 95; in both HRV2 and HRV14, a hydrophilic region of different sequence lies in the same area, suggesting thatthis region could represent a major antigenic site in HRV2.

In summary, the complete nucleotide sequence of HRV2 has been determined, and the derived amino acid sequence compared to that of HRV14 and poliovirus type 1. In contrast to the high degree of homology existing between the 3 poliovirus serotypes, HRV2 and HRV14 appear to be more distantly related. Surprisingly though, HRV2 and HRV14 show about as much homology to poliovirus type 1 as to each other. This supports the proposal of Stanway et al. (16) that rhinoviruses and enteroviruses should be considered as one genera of the picornaviridae family.

ACKNOWLEDGEMENTS

We thank A. Gupta, L. Frasel and M. Balasko for expert technical assistance, A. Barta and Z. Rattler for helpful discussions, and B. Gamperl for typing the manuscript. This work was supported by the "Österreichischer Fonds zur Förderung der wissenschaftlichen Forschung" and by Boehringer Ingelheim. We are grateful to Dr. J. Almond for communicating sequence data prior to publication.

*To whom correspondence should be addressed

REFERENCES

1. Gwaltney, J.M.jr (1975) Yale J. of Biol. and Med. 48, 17 - 45.
2. Stott, E.J. and Killington, R.A. (1972) Ann.Rev.Microbiol. 26, 503 - 524.
3. Fox, J.P. (1976) American J. Epidemiol.103, 345 - 354.
4. Melnick, J.L., (1980) Prog. Med. Virol. 26, 214 - 232.
5. Cooper,P.D., Agol, H.L., 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. and Tyrrell,D.A.J. (1978) Intervirology 10, 165 - 180.
6. Golini, F., Nomoto, A. and Wimmer, E. (1978) Virology 89, 112 - 118.
7. Ahlquist, P. and Kaesberg, P. (1979) Nucleic Acids Res. 7, 1195 - 1204.
8. Gaunt, C.J. (1980) Intervirology 13, 7 - 15.
9. Butterworth, B.E. and Korant, B.D. (1974) J. Virol. 14, 282 - 291.
10. Medappa, K.C., McLean, C. and Rueckert, R.R. (1971) Virology 44, 259 - 270.
11. Cooney, M.K. Fox, J.P. and Kenny, G.E. (1982) Infect. Immun. 37, 642 - 647.
12. Halfpap, L.M. and Cooney, M.K. (1983) Infect. Immun. 40, 213 - 218.
13. Abraham, G. and Colonno,R.J. (1984) J. Virol. 51, 340 - 345.
14. Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.J., Dorner, A.J., Emini, E.A., Hanecak, R., Lee, J.L., van der Werf, S., Anderson, C.W. and Wimmer, E. (1981) Nature 291, 547 - 553.
15. Racaniello, V. and Baltimore, D. (1981) Proc.Natl. Acad.Sci. U.S.A. 78, 4887 - 4891.
16. Stanway, G., Hughes, P.J., Mountford, R.C, Minor, P.D. and Almond, J.W. (1984) Nucleic Acids Res. 12, 7859 - 7877.
17. Skern, T., Sommergruber, W., Blaas, D., Pieler, Ch. and Kuechler, E. (1984) Virology 136, 125 - 132.
18. Southern, E.M. (1975) J.Mol.Biol. 98, 503 - 517.
19. Grunstein, M. and Hogness, D.S. (1975) Proc.Natl.Acad.Sci. U.S.A. 72, 3961 - 3965.
20. Rigby, P.W.J., Beckman, M., Rhodes, C. and Berg, P. (1977) J.Mol.Biol. 119, 237 - 251.
21. Vieisa, J. and Messing, J.G. (1982) Gene 19, 259 - 268.
22. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499 - 560.
23. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463 - 5467.
24. Staden, R. (1980) Nucleic Acids Res. 8, 3673 - 3694.
25. Isono, K. (1982) Nucleic Acids Res. 10, 85 - 89.
26. Laemmli, U.K. (1970) 227, 680 - 685.
27. Hunkapiller, M.W. and Hood, L.E. (1983) Science 219, 650 - 659.
28. Toyoda, H., Kohara, M., Kataoka, Y., Suganuma, T., Omata, T., Imura, N. and Nomoto, A., (1984) J.Mol.Biol. 174, 561 - 586.
29. Stanway, G., Cann, A.J., Hauptmann, R., Hughes, P., Clarke, L.D., Mountford, R.C., Minor, P.D., Schild, G.C. and Almond, J.W. (1983) Nucleic Acids Res. 11, 5629 - 5643.
30. Lonberg-Holm, K., Crowell, R.L. and Philipson, L. (1976) Nature 259, 679 - 681.
31. Pelham, H.R. (1978) Eur. J. Biochem. 85, 457 - 462.
32. Svitkin, Y.V., Gorbalenya, A.E., Kazachkov, Y.A. and Agol, V.I. (1979) FEBS Lett. 87, 7 - 11.
33. Hanecak, R., Semler, B.L., Anderson, C.W. and Wimmer, E. (1982) Proc.Natl.Acad.Sci. U.S.A. 79, 3973 - 3977.
34. Hopp, T.H. and Woods, K.R. (1981) Proc.Natl.Acad.Sci. U.S.A. 78, 3824 - 3828.
35. Rueckert, R.R. and Wimmer, E. (1984) J. Virol. 50, 957 - 959.
36. Emini, E.A., Schleif, W.A., Colonno, R.J. and Wimmer, E. (1985) Virology 140, 13 - 20.