Poliovirus type 3: molecular cloning of the genome and nucleotide sequence of the region encoding the protease and polymerase proteins

Alan J.Cann, Glyn Stanway, Rudi Hauptmann*, Philip D.Minor⁺, Geoffrey C.Schild⁺, Lindsay D.Clarke, Roger C.Mountford and Jeffrey W.Almond

Department of Microbiology, Leicester University Medical School, University Road, Leicester LE1 7RH, and ⁺National Institute for Biological Standards and Control, Holly Hill, London NW3 6RB UK

Received 14 January 1983; Revised and Accepted 16 February 1983

ABSTRACT

Overlapping cDNA clones representing the entire genome of poliovirus type 3 have been prepared in *E. coli* by two separate methods. Cloning of RNA.cDNA hybrids produced a more comprehensive set of clones with generally larger cDNA inserts than cloning of double - stranded cDNA. A restriction map of the entire genome and the nucleotide sequence of 2003 bases from the 3' terminus, comprising the region encoding the protease and polymerase proteins, are presented.

INTRODUCTION

Poliovirus can be classified by serological methods into three distinct antigenic types (1) each of which is capable of causing paralytic disease in humans. In common with other members of the Picornaviridae family, all three types possess a single-stranded positive-sense RNA genome approximately 7,500 nucleotides in length packaged into a 27nm diameter icosahedral particle. This RNA is analogous in structure and function to a eukaryotic messenger RNA and functions most probably as a single translational unit during virus replication (1).

The molecule contains a poly A tract at its 3' terminus (2,3) and a small covalently-linked protein, (VpG) at its 5' terminus (4,5).

Molecular cloning and nucleic acid sequencing of poliovirus genomes is of interest for several reasons. Although vaccines against all three strains of poliovirus are available and have been widely and successfully used throughout the developed world, immunisation is not without problems (6,7). Genetic instability, particularly of the poliovirus type 3 Sabin vaccine strain has been demonstrated *in vitro* (8), and evidence implicates the vaccine as a possible cause of paralytic disease

in certain rare instances (7,9,10). Knowledge of the structure of those elements of the virus involved in immunogenicity should provide the groundwork for the development of alternative vaccines (11,12,13). Nucleotide sequence information also contributes to a clearer understanding of some of the molecular events of virus replication, such as the mechanism of processing of the polyprotein precursor of viral polypeptides (14,15). Finally, comparison of nucleotide sequences of polio strains and other enteroviruses, should illustrate the extent of genetic diversity within the group, and may highlight those regions of the genome likely to be of importance in influencing their differing biological properties.

As part of an investigation with these and other objectives, we have begun molecular cloning and nucleotide sequencing of poliovirus type 3 Sabin vaccine strain. The construction and restriction mapping of cDNA clones together with the nucleotide sequence of 2003 bases from the 3' terminus, encoding the putative protease and polymerase proteins, are presented. Comparisons with the previously published poliovirus type 1 genome sequences are discussed (16,17,18).

MATERIALS AND METHODS

Chemicals and enzymes

Reverse transcriptase was a gift from Dr. J.W. Beard, NIH, Bethesda. E. coli DNA polymerase I and T4 DNA ligase were purchased from New England Nuclear; large fragment E. coli DNA polymerase I from Boehringer Mannheim; terminal deoxynucleotidyl transferase from P-L Biochemicals; Nuclease S₁ and proteinase K from Sigma; restriction endonucleases from Bethesda Research Laboratories and [α -³²P] dATP and dCTP (2000-3000 Ci/mmol) from Amersham International. E. coli JA221 (rk⁻, mk⁺, Trp ES del, Leu B6⁻, Rec A, lac y) was supplied by Dr. P. Meacock (University of Leicester / ICI Joint Laboratory).

Isolation of virus RNA

A plaque-purified isolate of the Sabin vaccine strain (Leon 12a₁ b) was propagated in HEP2c cells and purified on sucrose gradients as previously described (9). RNA was extracted directly from the appropriate sucrose fractions by the addition of SDS to 0.2% and proteinase K to 20 µg/ml. After incubation at

37°C for 20 min, an equal volume of phenol/ chloroform/ isoamyl alcohol/ 8-hydroxyquinoline (100v/100v/4v/0.8w) was added and the mixture shaken for 10 min at room temperature. After a second extraction, RNA was recovered by ethanol precipitation, assayed by measuring OD₂₆₀ and checked for quality by electrophoresis on a 1% agarose gel.

Preparation of poliovirus-specific cDNA

Purified RNA (25 µg) was used as a template for cDNA synthesis by AMV reverse transcriptase (210u) using pdT₁₇ (2.5 µg) as primer. The reaction was carried out in a 400 µl incubation mixture containing 50 mM Tris-HCl (pH 8.3), 8 mM MgCl₂, 20 mM mercaptoethanol, 100 mM NaCl, 0.5 mM each of dATP, dCTP, dGTP, dTTP and 10 µCi of [α-³²P] dATP. After 60 min at 37°C, EDTA was added to 20 mM, the mixture heated to 65°C, phenol extracted and passed through a Sephadex G100 column. The yield of cDNA was estimated by monitoring the incorporation of [α-³²P] dATP into material eluting in the void volume.

Cloning of RNA.cDNA hybrids

Homopolymer tailing of RNA.cDNA hybrids was carried out using terminal deoxynucleotidyl transferase (19). A 250 µl reaction mixture contained 4 µg RNA.cDNA hybrid, 140 mM K-cacodylate (pH 6.9), 1 mM CaCl₂, 0.1 mM dithiothreitol, 5 nmol of dCTP, 10 µCi [α - ³²P] dCTP and 80 units of terminal transferase. The mixture was incubated for 15 min at 37°C and the reaction placed on ice until the number of incorporated dC residues per 3' end could be calculated. Upon addition of an estimated 15-20 residues per end, the reaction was stopped by the addition of EDTA to 20 mM. Following phenol extraction, the mixture was passed through a Sephadex G100 column and DNA was recovered from the excluded volume by ethanol precipitation.

Plasmid pAT 153 (20) was digested with Pst I and dG-tails were added (19,21) by incubation at 37°C in a 250 µl reaction mixture containing pAT 153 (24 µg), 140 mM K-cacodylate (pH 6.9), 4 mM MgCl₂, 0.1 mM dithiothreitol, 8 nmol [³H]-dGTP (specific activity 0.37 Ci/mmol) and 60 units of terminal transferase. The reaction was stopped after 30 min when an estimated 20 dG residues/end had been added. The dG-homopolymer tailed DNA was purified as described above. 400 ng of

homopolymer tailed plasmid was annealed with 1 μ g of RNA.cDNA hybrid in 250 μ l of 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.2 mM EDTA by incubating at 65°C for 5 min., 45°C for 2 h followed by slow cooling to room temperature. The annealed DNA was used to transform *E. coli* JA221 which was then plated out onto Luria agar containing 10 μ g/ml tetracycline. Recombinants were identified by their sensitivity to ampicillin at 100 μ g/ml.

Cloning of ds cDNA

2 μ g cDNA, prepared as described above, was dissolved in water (25 μ l) and dissociated from the RNA template by heating to 100°C for 2 min followed by rapid chilling. The cDNA was rendered double stranded by incubation at 13°C for 7 h with *E.coli* DNA polymerase I (60U) in 200 μ l of 200 mM HEPES (pH 6.9), 5 mM MgCl₂, 70 mM KCl, 20 mM dithiothreitol, 0.5 mM each of dCTP dGTP and dTTP, 0.1 mM dATP and 20 μ Ci [α -³²P] dATP (22). After heating to 65°C to stop the reaction, the mixture was phenol extracted and passed through a Sephadex G100 column. An aliquot of the recovered ds cDNA (0.3 μ g) was incubated with nuclease S₁ (10u) for 15 min at 37°C in 30 mM Na-acetate (pH 4.5), 300 mM NaCl and 3 mM ZnSO₄ to remove any single-stranded material. The reaction was stopped by the addition of EDTA to 20 mM, phenol extraction, and passage through a Sephadex G100 column. Addition of dC-tails and cloning were exactly as described above for RNA.cDNA hybrids.

Analysis of cloned cDNA

Colony hybridization was performed on amp^S tet^R colonies using the method of Grunstein and Hogness (23,24). Plasmid DNA was isolated from those colonies displaying strong hybridization by a small-scale partial purification method (25). This DNA was further characterized by restriction enzyme mapping. Plasmids containing large cDNA inserts were purified on CsCl gradients and more detailed restriction maps were produced by the method of Smith & Birnstiel (26).

Sequencing and Data Handling

Cloned poliovirus cDNA was subcloned into the mp7, 8 and 9 derivatives of bacteriophage M13 (27) and sequenced by the dideoxynucleotide method (28). Sequence data thus generated were

analysed on a Digital PDP 11/44 computer using programs developed by Staden (29).

RESULTS

Synthesis and molecular cloning of cDNA

Poliovirus type 3 RNA, extracted from purified virus after proteinase K treatment, was analysed on 1% agarose gels and visualized by staining with ethidium bromide. Only batches of RNA showing a single discrete band with no evidence of degradation were considered suitable for cloning experiments. 25 µg of this material was reverse transcribed in a 400 µl reaction mixture and approximately 10 µg of cDNA was synthesized. After gel filtration and ethanol precipitation this material was analysed by electrophoresis on 1% alkaline agarose (30) and 2.6% acrylamide/urea gels. The cDNA synthesized migrated primarily in the region of the gel corresponding to full-length molecules, although a gradation of small material was also apparent. Digestion of the cDNA with the restriction endonuclease Hae III after alkaline hydrolysis of input RNA, produced several distinct bands with a relatively weak background smear (Fig. 1).

Aliquots of cDNA from a single reverse transcription reaction were cloned into E. coli plasmid vector pAT 153 using two different procedures as follows.

i) Cloning of double-stranded cDNA. Approximately 2 µg of cDNA was used in a second-strand synthesis reaction. After purification by gel filtration, the material was digested with S_1 nuclease to remove any single-stranded material present in the preparation. Approximately 80% of radioactivity incorporated in the reverse transcription reaction was found to be resistant. An estimated 0.3 µg of this ds cDNA was extended by homopolymer tailing with dC using terminal deoxynucleotidyl transferase (19). Incorporation of [α - 32 P] dCTP suggested that based on an average estimated length of 2000 bp, approximately 20 dC residues /3' end had been added. This material was annealed to 0.44 µg of pAT 153 tailed at the Pst I site with dG (19,21). One quarter of the annealed mixture was used to transform competent E. coli JA221, and produced 836 amp^S tet^r colonies. The vast

Figure 1. Single-strand product of reverse transcription of the genome of poliovirus type 3. Electrophoresis of cDNA is on a 2.6% acrylamide 8M urea TBE gel, (100 mM Tris-base, 50mM boric acid, 2.5mM EDTA), before and after cleavage with restriction endonuclease Hae III. O- signifies position of origin; V- indicates the migration of unlabelled polio vRNA (approximately 7500 nucleotides) on the same gel.

majority, (>90%), of these had polio cDNA inserts (see below). This corresponds to approximately 1×10^4 colonies / μ g ds cDNA and 8×10^3 colonies /pmol of RNA starting material.

ii) Cloning of RNA.cDNA hybrids. An aliquot of the RNA.cDNA material as recovered from the reverse transcription was homopolymer tailed directly with dCTP and terminal deoxynucleotidyl transferase under conditions identical to those used for tailing of ds cDNA. Making an assumption (based in part on gel analysis) that the average chain-length of cDNA was approximately 5000 nucleotides and that free, single- stranded RNA would not act as a substrate-acceptor in the reaction, dCTP was incorporated to give an approximate average tail length of 20 residues /end. These conditions for tailing of RNA.cDNA hybrid molecules are similar to those described by others except that trimming of hybrid molecules with ribonuclease A was omitted (31). The material obtained was annealed to an aliquot

of the same batch of dG tailed, Pst I cleaved pAT 153 vector as was used in the experiment described above, at a molar ratio of approximately 1:1, and used to transform competent E. coli JA221. Approximately 85% of the resulting colonies were of the $\text{amp}^S \text{tet}^R$ phenotype and 2×10^3 of these were obtained per μg cDNA hybrid, corresponding to a yield of 5×10^3 transformants / μmol of RNA used in the initial reverse transcription. The vast majority (>90%) of these colonies hybridized with a poliovirus cDNA probe (see below).

Analysis of cDNA clones

Colonies with an $\text{amp}^S \text{tet}^R$ phenotype from both cloning experiments were grown on nitrocellulose filters and screened for the presence of poliovirus cDNA by the method of Grunstein & Hogness (23,24). Poliovirus cDNA, prepared by reverse transcription of genome RNA, was used as a probe. Clones containing cDNA corresponding to the 3' region of the genome were identified using a cDNA probe prepared from a reverse transcription reaction containing 1/4000th normal amount of dATP (specific activity 2000-3000 Ci/mmol). cDNA synthesized under these conditions was of much lower molecular weight than that synthesized under standard conditions (data not shown), and is therefore probably enriched for sequences adjacent to the poly A tract. Clones containing cDNA corresponding to regions towards the 5' end of the genome were detected using a cDNA probe synthesized under standard conditions, or cDNA from a reaction in which DNase I treated, sheared, salmon sperm DNA was used as a random primer (32). Results of hybridization studies are shown in Fig. 2. Strongly positive colonies were isolated and grown initially in 1 ml cultures. Plasmid DNA was extracted and analysed on 1% agarose gels using supercoiled pAT 153, pBR 322 and M13 mp7 RF as markers. Those with an insert size of greater than 1000 bp were grown in 10 ml cultures and plasmid DNA was digested with restriction endonucleases. The vast majority, (>95%), of plasmids examined in this way were cleavable with Pst I to generate at least two fragments, one of which was the same size as pAT 153. Representatives of plasmids showing overlapping hexameric restriction sites were characterized further by the restriction mapping method of Smith & Birnstiel (26). A sample

Figure 2. Colony hybridization of amp^S tet^R transformants. Filters were probed with the ^{32}P -radiolabelled products of a reverse transcription reaction containing $10\mu Ci$ [α - ^{32}P] dATP at a specific activity of 2000-3000 Ci/mmol (A) and a normal reaction in which sheared salmon-sperm DNA was used as a primer (B).

of cDNA clones, which together contain material corresponding to the entire poliovirus type 3 genome, (see below), are illustrated in Fig. 3. The orientation of clones in relation to virus RNA was based on results obtained with the probe synthesized under conditions of reduced [dATP] and was later confirmed by sequencing. Clones labelled pSGA were obtained by the ds cDNA method and those labelled pSAG were from RNA.cDNA hybrid cloning. In all, 836 colonies were examined by hybridization and 205 by gel analysis to obtain the pSGA clones represented in Fig. 3, plus a number of others of similar or smaller size, overlapping but not extending those shown. By comparison, rather fewer hybrid clones were examined, 773 by hybridization and 114 by gel analysis, to obtain those shown plus many others of similar or smaller size. From these observations it seemed that large cDNA clones (i.e. $>20000bp$) and clones containing sequences from the 5' half of the genome were more easily obtained by the hybrid cloning method than by the cloning of ds cDNA. However, as observed by others (31) clones, (produced by either method), containing cDNA corresponding to the extreme 3' terminus of the genome were relatively less common.

Figure 3. Restriction endonuclease map of poliovirus type 3 cDNA. The detailed restriction map of the 3' terminal 2003 bases was determined using the method of Smith and Birnstiel (26). pSAG clones were obtained by the hybrid method. pSGA clones were obtained by the double-stranded method.

Nucleotide sequencing

Plasmids pSGA 31 and pSAG 6 were used for nucleotide sequencing studies. Following digestion with Hae III, Alu I, or Sau 3A, the cDNA was subcloned into Sma I and Bam HI singly or doubly digested M13 mp7, mp8 or mp9 (27). Recombinant phage, ("whites"), generated by this procedure were subjected to nucleotide sequence analysis by the dideoxynucleotide method (28). The sequence of 2003 nucleotides adjacent to the 3' poly A tract of poliovirus type 3 Sabin vaccine strain, is shown in Fig. 4. 95% of this sequence was obtained from at least duplicate sequencing reactions. Wherever coding function differences from poliovirus type 1 occurred (16,17), these were checked carefully, allowing no possible ambiguity. The sequence shown corresponds to the 3' untranslated region of 71 nucleotides, together with the region of the genome encoding the putative protease and polymerase proteins, based on extrapolation from the published data for polio type 1 (14,15).

[illegible]

```

V D Y I D Y L H H S H H L Y K N K I Y C V K B G H P S G C S
BTG BAT TAC ATA GAC TAC CTT AAC CAT TCA CAC CAC TTG TAC AAA AAC AAG ATA TAT TGT GTT AAG GGC GGC ATG CCA TCT GGC TGC TTA
1340 1350 1360 1370 1380 1390 1400 1410 1420
C..T ..C ... ..C ... ..A ... ..C ... ..C ... ..G ..T ..A ..C ... ..C ... ..G ..T ... ..G ... ..A ... ..A ... ..
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

G T S I F N S M I N H L I I R T L L L K T Y K G I D L D H L
GBC BAT TCA ATT TTT GGT GAC GAT GTA ATA GCT TCC TAT CCC CAT GAG GTT GAC GCT AGT CTC CTA GCC CAA TCA GGA AAA GAC TAT GGA
1430 1440 1450 1460 1470 1480 1490 1500 1510
..G ... ..C ... ..C ... ..C ... ..C ... ..T ..C ... ..C ... ..C ... ..G ... ..G ... ..A ... ..C ... ..C ... ..
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

K M I A Y G D D V I A S Y P H E V D A S L A Q S G K D Y G
AAA ATG ATT GCC TAT GGT GAC GAT GTA ATA GCT TCC TAT CCC CAT GAG GTT GAC GCT AGT CTC CTA GCC CAA TCA GGA AAA GAC TAT GGA
1520 1530 1540 1550 1560 1570 1580 1590 1600
... ..C ... ..C ... ..T ... ..T ... ..C ... ..C ... ..A ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

L T N T P A D K S A T F E T V T M E N V T F L K R F F R A D
CTA ACC ATG ACT CCG GCA GAT AAA TCT GCC ACT TTT GAG ACA GTC ACA TGG GAG AAT GTA ACT TTC TTG AAA AGA TTC TTC AGA CCA GAT
1610 1620 1630 1640 1650 1660 1670 1680 1690
... ..T ... ..A ... ..C ... ..A ... ..T ..A ... ..A ... ..C ... ..C ... ..C ... ..A ... ..C ... ..G ... ..C ... ..C ... ..C ... ..C ... ..
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

E K Y P F L I H P V M P M K E I H E S I R W T K D P R N T Q
GAG AAA TAC CCC TTC CTC ATA CAT CCA GTA ATG CCA ATG AAG GAA ATT CAT GAA TCA ATC AGA TGG ACA AAA GAT CCT CGG AAT ACG CAG
1700 1710 1720 1730 1740 1750 1760 1770 1780
... ..C ... ..A ... ..T ... ..T ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

D H V R S L C L L A W H N G E E E Y N K F L A K I R S V P I
GAC CAT GTA CBC TCC TTG TGT CTA TTG GCT TGG CAC AAC GGG GAA GAA GAA TAC AAC AAA TTT TTA GCT AAA ATT AGG AGT GTG CCA ATC
1790 1800 1810 1820 1830 1840 1850 1860
..T ..C ... ..T ... ..C ... ..C ... ..T ..A ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

G R A L L L P E Y S T L Y R R W L D S F R R
GGA AGA GCT TTG TTG CTC CCA GAG TAC TCA ACA TTG TAC CCG CGT TGG CTT GAC TCA TTT TAG TAA CCT ACC TCA GTC GAA TTG GAT TGG
1880 1890 1900 1910 1920 1930 1940 1950 1960
... ..C ... ..A ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

GTC ATA CTG TTG TAG GGG TAA ATT TTT CTT TAA TTC GGA GG-POLY A
1970 1980 1990 2000
... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..C ... ..
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

```

Figure 4. Nucleotide sequence of 2003 bases of cloned poliovirus cDNA corresponding to the 3' terminal region of the genome. Based on comparison with poliovirus type 1 (15,16,17) the putative protease and polymerase are separated by a single gln-gly cleavage at position 547-552. Differences in amino acid and nucleotide sequences from type 1 are indicated.

These are separated from each other by a single gln-gly cleavage junction at nucleotide positions 547-552 (15,16).

DISCUSSION

Overlapping cDNA clones which together comprise the entire genome of the Sabin vaccine strain of poliovirus type 3 have been obtained. Material from a single reverse transcription reaction was cloned in *E. coli* by two different procedures, using aliquots of the same batch of dG tailed pAT 153 vector DNA. A direct comparison of the classical ds cDNA method with an RNA.cDNA hybrid cloning method was therefore possible.

Quantitatively the results obtained from the two methods

were similar, 8×10^3 amp^S tet^R colonies from the ds cDNA method and 5×10^3 from the hybrid method / μmol of RNA starting material were obtained. This is similar to other published results using these techniques (31,33). Qualitatively, both methods produced clones with a range of cDNA inserts, many of which were <1 kb in size. Clones with larger cDNA inserts, however, were obtained, and these were characterized further by restriction mapping. The largest clones obtained, pSAG 16 and pSAG 20, had cDNA inserts of 3.8 and 3.7 kb respectively. Both of these, and several others of similar sizes, e.g. pSAG 28 and pSAG 19, were obtained by the hybrid method. The largest clones obtained from the ds cDNA method, pSGA 21 and pSGA 19, had insert sizes of 2.05 and 1.65 kb respectively.

In agreement with the results of others (31) clones containing cDNA corresponding to the 5' terminus of viral RNA, (sequence verification to be published), were obtained by the hybrid method, whereas such clones cannot be obtained from the ds cDNA method because of the S_1 nuclease step. Alternative cDNA cloning methods, designed to obtain the 5' terminal regions, have been described (34,35,36,37). In one of these (37) cDNA synthesis requires many more steps, including the use of two separate primers, and two separate homopolymer tailing reactions. The most recent method described (38) gives higher efficiencies than those achieved here and produces full length cDNA clones. However, special vectors and a larger number of synthetic reactions are required.

On consideration of the relative merits of both methods, our conclusions are that although efficiencies are comparable, the speed and simplicity of the hybrid method, together with the possibility of producing clones spanning the entire genome in a single experiment, makes this the method of choice for cloning large RNA molecules of this type, e.g. whole viral genomes. Amongst the clones presented here, (fig. 3), and others made in our laboratory from other polio strains (to be published) we have not observed any artefacts that might be due to insertions, deletions or other re-arrangements of cDNA (34).

One unexpected result was that the 3' terminus seemed underrepresented in clones. This observation is in agreement with

other published data (31) and is the same for both double stranded cDNA and RNA.cDNA cloning methods.

The sequence of 2003 nucleotides adjacent to the 3' poly A tract has been determined. This comprises the region encoding, (based on a direct comparison with poliovirus type 1), the putative protease and polymerase proteins, separated by a single gln-gly cleavage site at position 547-552, together with the 3' untranslated region. Within the coding region there is 85% homology in nucleotide sequence and 97% homology in predicted amino-acid sequence with the polio type 1 (Mahoney strain) published by others (16,17). This level of homology is in agreement with previous estimates of homology between polio types 1 and 3 based on hybridisation (39), and underlines the close evolutionary relationship between the two types. All coding function differences observed are the result of point mutations; we have not detected any insertions or deletions in this region. These changes are evenly distributed throughout the two proteins, except that there are three changes clustered within six amino-acids coded at 1147-1164 in the polymerase protein. Also, there are 182 conserved amino-acids at the carboxyl-terminus of this protein, suggesting a functional importance of this region. Amino-acid changes between the two types are largely conservative, and the great similarity in structure suggests close functional homology between the two types. In the case of the protease, both homology in sequence and the conservation of the gln-gly cleavage site between the two proteins, suggest that processing in both types occurs by an identical mechanism. However, an alternative cleavage occurs in type 1 at a tyr-gly cleavage site encoded by bases 6420-6425 of this strain (15) generating proteins 6a and 6b. These proteins, which have been identified by one-dimensional peptide maps, are not found in cells infected with Sabin type 3 virus or many strains related to it (9), although the tyr-gly amino acids are conserved at this site, (position 991-996 in our sequence).

The 3' untranslated region of 71 nucleotides (together with the final 48 nucleotides of the coding region) is totally conserved between types 1 (Mahoney) and 3 (Sabin). This remarkable conservation strongly implies a functional role for

this region. The type 3 sequence contains GG as the two bases preceding the poly A tract, and in this respect resembles the type 1 sequences previously published (16,18,40). There is one report of only a single G at this position (17).

ACKNOWLEDGEMENT

We are grateful to the Medical Research Council of Great Britain for their support of this work (Project Grant No. G80/0815/2CA).

*Present address: Ernst-Boehringer-Institute for Arzneimittelforschung, Bender & Co. Ges mbH, Dr. Boehringer-Gasse 5-11, A-1121 Vienna, Austria

REFERENCES

1. Rueckert, R.R. (1975) In *Comprehensive Virology* 6. Ed. Fraenkel-Conrat and Wagner, 131-213.
2. Armstrong, J.A., Edmonds, M., Nakazato, H., Phillips, B.A., and Vaughn, M. (1972) *Science* 176, 526-528.
3. Yogo, Y. and Wimmer, E. (1972) *P.N.A.S. U.S.A.* 68, 1877-1882.
4. Lee, Y., Nomoto, A., Detjen, M. and Wimmer, E. (1977) *P.N.A.S. U.S.A.* 74, 59-63.
5. Pettersson, R.F., Flanagan, J.B., Rose, J.K. and Baltimore, D. (1977) *Nature* 268, 270-272.
6. Bazin, M. (1980) *Nature* 285, 7-8.
7. Cossart, Y.E. (1977) *British Medical Journal* 1, 1621-1623.
8. Yoshii, K., Matsukura, T., Nakano, M. and Tagaya, I. (1979) *Journal of Biological Standardisation* 7, 169-178.
9. Minor, P. (1980) *Journal of Virology* 34, 73-84.
10. Minor, P. (1982) *Journal of General Virology* 59, 307-317.
11. Kleid, D.G., Mc Kercher, P.D., Morgan, D.O., Robertson, B.H. and Bachrach, H.L. (1981) *Science* 214, 1125-1128.
12. Bittle, J.L., Houghton, R.A., Alexander, H., Shinnick, T.M., Sutcliffe, S.G., Lerner, R.A., Rowlands, D.J. and Brown, F. (1982) *Nature* 298, 30-33.
13. Minor, P.D., Schild, G.C., Bootman, J., Evans, D.M.A., Ferguson, M., Reeve, P., Spitz, M., Stanway, G., Cann, A.J., Hauptmann, R., Clarke, L.D., Mountford, R.C., and Almond, J.W. *Nature*, (In Press).
14. Rueckert, R.R., Matthews, T.J., Kew, O.M., Pallansch, M., McClean, C. and Omilanski, S. (1979) In *The Molecular Biology of Picornaviruses*, Ed. Perez-Bercoff, R., 113-125.
15. Hanecak, R., Semler, B.L., Anderson, C.W. and Wimmer, E. (1982) *P.N.A.S. U.S.A.* 79, 3973-3977.
16. Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.J., Dorner, A.J., Emmini, E.A., Hanecak, R., Lee, J.L., van der Werf, S., Anderson, C.W. and Wimmer, E. (1981) *Nature* 291, 547-553.
17. Racaniello, V.R. and Baltimore, D., (1981) *P.N.A.S. U.S.A.* 78, 4887-4899.
18. Nomoto, A., Omata, T., Toyoda, H., Kuge, S., Horie, H., Katoda, Y., Genba, Y., Nakano, Y. and Imura, N. (1982) *P.N.A.S. U.S.A.* 79, 5793-5797.

19. Deng, G. and Wu, R. (1981) *Nucleic Acids Research* 9, 4173-4188.
20. Twigg, A.J. and Sherratt, D. (1980) *Nature* 283, 216-218.
21. Otsuka, A. (1981) *Gene* 13, 339-346.
22. Bothwell, A.L., Paskind, M., Reth, M., Imanshi-Kari, T., Rajewsky, K. and Baltimore, D. (1981) *Cell* 24, 625-637.
23. Grunstein, M. and Hogness, D.S. (1975) *P.N.A.S. U.S.A.* 72, 3961-3965.
24. Grunstein, M. and Wallis, J. (1979) *In Methods in Enzymology* 68, Ed. Wu, R., 379-389.
25. Holmes, D.S. and Quigley, M. (1981) *Analytical Biochemistry* 114, 193-197.
26. Smith, H.O. and Birnstiel, M.L. (1976) *Nucleic Acids Research* 3, 2387-2398.
27. Messing, J., Crea, R. and Seeburg, P. (1981) *Nucleic Acids Research* 9, 309-321.
28. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *P.N.A.S. U.S.A.* 74, 5463-5467.
29. Staden, R. (1980) *Nucleic Acids Research* 8, 3673-3694.
30. McDonnell, M.W., Simon, M.N. and Studier, F.W. (1977) *Journal of Molecular Biology* 110, 119-146.
31. van der Werf, S., Brègègere, F., Kopecka, H., Kitamura, N., Rothberg, P.G., Kourilski, P., Wimmer, E. and Girard, M. (1981) *P.N.A.S. U.S.A.* 78, 5983-5987.
32. Taylor, J.M., Illmensee, R. and Summers, J. (1976) *Biochim. Biophys. Acta* 442, 324-330.
33. Kurtz, D.T. and Nicodemus, C.F. (1981) *Gene* 13, 145-152.
34. Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979) *Cell* 16, 851-861.
35. Rowekamp, W. and Firtel, R.A. (1980) *Developmental Biology* 79, 409-418.
36. Vitek, M.P., Kreissman, S.G. and Gross, R.H. (1981) *Nucleic Acids Research* 9, 1191-1202.
37. Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schultz, G. (1981) *Nucleic Acids Research* 9, 2251-2266.
38. Okayama, H. and Berg, P. (1982) *Molecular and Cellular Biology* 2, 161-170.
39. Tracey, S. and Smith, R.A. (1981) *Journal of General Virology* 55, 193-199.
40. Porter, A.G., Fellner, P., Black, D.N., Rowlands, D.J., Harris, T.J.R. and Brown, F. (1978) *Nature* 276, 298-301.